A Microfoundation of Monetary Economics*

Shouyong Shi
Department of Economics
University of Toronto
150 St. George Street, Toronto,
Ontario, Canada, M5S 3G7
(email: shouyong@chass.utoronto.ca)
(phone: 416-978-4978; fax: 416-978-6713)

2006

Abstract

In this lecture, I explain what the microfoundations of money are about and why they are necessary for monetary economics. Then, I review recent developments of a particular microfoundation of money, commonly known as the search theory of money. Finally, I outline some unresolved issues.

JEL classification: E40, E50, E31
Keywords: Money; Search; Microfoundation.

* This paper is based on a lecture given at the Canadian Economic Association Meetings (Hamilton, 2005). I am grateful to Angelo Melino, Steve Williamson and an anonymous referee for valuable comments, and to Hilary Shi for editorial assistance. I gratefully acknowledge the financial support from the Bank of Canada Fellowship and from the Social Sciences and Humanities Research Council of Canada (SSHRCC). The opinion expressed here is my own and it does not represent the view of the Bank of Canada.

1. Monetary Issues

In this lecture, I review the recent development of a microfoundation of monetary economics, known as the search theory of money. To explain what a microfoundation of monetary economics is about and why it is necessary, let me list the main issues in monetary economics as follows.

> (I1) Existence and essentiality of fiat money. Why would intrinsically worthless money have value? How can fiat money improve the efficiency of resource allocations?
>
> (I2) Return dominance. Why is money dominated in the rate of return by other assets and, in particular, by government issued nominal bonds?
>
> (I3) Optimal monetary policy. What is the optimal inflation rate?
>
> (I4) Money, banking and payments systems. What is the role of fiat money in banking and payments systems?
>
> (I5) Monetary propagation. How are monetary shocks propagated to affect the price level and real activities in business cycles?
>
> (I6) International monetary arrangements. What determines the nominal exchange rate between currencies? Should a country maintain its national currency or join a currency union?

Three terms need to be defined – fiat money, the essentiality of money and a microfoundation of money. Fiat money is an object that circulates in the market and that has two characteristics emphasized by Wallace (1980). One is intrinsically uselessness in the sense that fiat money does not yield direct utility or facilitate production; the other is that fiat money is not backed or expected to be backed by any government policy such as taxes. According to this definition, fiat monies have been rare. Commodity monies, such as gold, silver or even seashells, have intrinsic value. Paper monies have very little intrinsic value, but they are occasionally defended by the authorities that issue them, especially in episodes of currency crises. Despite this reality, I would argue that monetary theory should not be built upon the intrinsic value of money or government intervention. Any such theory would fail to account for the additional value that money has in the market over and beyond its intrinsic value and government intervention, and hence the theory would fail to uncover the critical differences between money and other assets. On the other hand, a theory that focuses on fiat money can be extended easily to incorporate the intrinsic value that money might have, or the government intervention to which money might be exposed. Thus, fiat money should be the primary object to be studied in monetary economics.

Money is essential if it improves the efficiency of resource allocations relative to an economy without money. In this lecture, I will often use a social welfare function to measure efficiency, but sometimes Pareto efficiency will also be used. Essentiality is not a vacuous concept. In many models, money can have a positive value in a particular equilibrium, while the equilibrium is inferior to a non-monetary equilibrium. Essentiality of money is a property to be sought for two reasons. One is that we want to know how much better money can do relative to non-monetary methods of exchange. The other is that a monetary equilibrium which is inferior to a non-monetary equilibrium could not have survived the test of time.

A microfoundation of monetary economics is a theoretical framework that endogenizes the value and essentiality of fiat money by explicitly specifying the frictions that impede the functioning of markets. Such a framework must be able to trace all potential changes in the role of money to changes in the underlying environment and policy. A large class of models in monetary economics, some of which are popular for policy analysis, fail to qualify as the microfoundations of monetary economics. When a model cannot resolve fundamental issues such as existence and essentiality, one cannot give much confidence to the answers that the model gives to other questions listed above.

In the next section, I will explain why the issue of existence is difficult and review some earlier efforts that tackled the issue. Then, I will report the recent progress in the literature under the label "a value theory of money", which is commonly known as the "search theory of money". Finally, I will list the challenges to monetary economics that still lie ahead.[1]

2. The Need for a Microfoundation of Monetary Economics

"To anyone who comes over from the theory of value to the theory of money, there are a number of things which are rather startling." This was how John Hicks (1935, p.1) described the state of monetary theory seventy years ago. The main problem Hicks confronted was that monetary theory in the classical dichotomy was inconsistent with the theory of value. On the real side, value theory described how the value of goods arose from their intrinsic properties, such as the marginal utility of consuming the goods and the marginal productivity of the goods. On the monetary side, the quantity equation was imposed separately to link the price level to the quantity of money. Hicks tried to integrate monetary theory into value theory by arguing that intertemporal choices of consumption and production could induce the demand for liquidity, and hence for money.

Value theory since Hick's time has developed into the celebrated framework of general equilibrium. The progress on the monetary side pales in comparison. Most monetary models have imposed ad hoc assumptions to graft money onto the general equilibrium framework. Such shortcuts include specifications with money in the utility function (Sidrauski, 1967, and Brock, 1974) or in the production function (Levhari and Patinkin, 1968), and the specifications with transactions costs (Heller and Starr, 1976, Kimbrough, 1986). An extreme version of the transactions cost approach is the cash-in-advance constraint (Clower, 1967).

[1] Other critical reviews of monetary economics are provided by include Kareken and Wallace (1980), Wallace (2001) and, to some extent, by Wright (2005).

These shortcuts are convenient because they do not have to modify radically the analytical and computational techniques developed for the general equilibrium framework. But analytical convenience often comes at a price. With the shortcuts, the functional forms in which money appears in the models are fixed by assumption. In reality, however, these functional forms are likely to change as agents modify their money demand to respond to changes in policy, the information technology, and other aspects of the trading environment. Thus, the shortcuts are not good microfoundations of money. In particular, they are subject to the well-known Lucas (1976) Critique, which renders the policy recommendations of these models questionable.

To be more specific, consider the shortcut that introduces real money balances into preferences or production technology. Since money appears in the primitives of the model, it is not fiat money. In Hicks' (1935) words, "merely to call that marginal utility [of money] X, and then proceed to draw curves, would not be very helpful." A typical justification for putting money in the primitives is that it approximates for the role of money in saving resources, such as time, that must be used if transactions are non-monetary. There is no denying that money serves this role in reality. The problem is that the approximation does not specify explicitly the link between money and the trading environment. As a result, it leaves many questions unanswered, such as the following. Does money appear in the utility function in the same way, regardless of whether agents want to trade a million different types of goods or only two types of goods? How should the substitution between money and other resources in these functions change when agents improve the communication between trades? How does public or private debt change the role of money? What if there are alternative means of exchange? How should the functional form change when there are two currencies as opposed to one? Should a country join a currency union? These questions are not just intellectual curiosities – they are clearly important for policy. Any attempt to answer these questions with the shortcut must add more ad hoc assumptions on the functional forms in which money appears.

Exogenous cash-in-advance constraints (e.g., Clower, 1967) or exogenous transaction functions (e.g., Kimbrough, 1986) are not really better shortcuts – they are just other ways of fixing the role of money exogenously. However, the literature has attempted to generate cash-in-advance constraints endogenously from explicit descriptions of the trading environment. These attempts should be distinguished from the shortcuts, because they do derive the role of money from the trading environment. As such, they are part of the microfoundations of monetary theory. I now review this literature briefly.

The quest for the microfoundations of money started formally when Kareken and Wallace (1980) edited their volume. There, Kareken and Wallace identified fiat money as the main object to be studied in monetary economics. The volume contained three specifications of monetary models that have been widely used since then.

The first specification is the overlapping generations model of money originated from Samuelson (1958). Lucas' (1972) pioneering work on the neutrality of money stirred interest in this model for macroeconomists. Although Lucas focused on the positive implications of the model, he did specify the physical environment explicitly to support a role of money. The main friction in the overlapping generations model is that markets between currently alive agents and future generations do not exist. Money is a device used to trade between generations. A few papers in the Kareken-Wallace volume formally explored when this friction could give rise to a monetary equilibrium and how money affected efficiency. One drawback of the overlapping generations model of money is that the role of money is so tied to the generational structure that it can be superseded by other means of intergenerational transfers, such as social security. Another drawback is that money has the same rate of return as other assets, unless additional ad hoc assumptions are imposed to give money a special role.

The second specification is Townsend's (1980) models of spatial separation. In these models, agents differ in the timing of receiving endowments and in the (exogenous) itineraries by which they travel across spatially separated markets. Spatial separation serves as a metaphor for the difficulties of communicating between markets and enforcing contracts. Townsend constructed three models of this kind. One model endogenously delivers a cash-in-advance constraint; the second model has a structure of overlapping generations, with agents being born at different dates but living forever; the third model allows privately issued debt to exist in the equilibrium. Because agents are differently situated, these models are suitable for studying the redistributive role of monetary policy. The main drawback of these models is that the itineraries assumed for agents are very rigid.[2]

The third specification is the model by Bewley (1980). In that model, each agent faces idiosyncratic shocks to preferences and/or endowments. Although agents can barter perfectly within each period, they cannot enforce intertemporal contracts and, in particular, cannot enforce insurance contracts. Money serves as a device to self-insure against unfavourable shocks. However, Bewley did not model why it is difficult to enforce insurance contracts. Later developments have explained this difficulty by introducing private information regarding tastes or actions (see Green, 1987, for a non-monetary example). These models, under the label "dynamic contracts", have been widely used for examining the distribution of wealth and the redistributive consequences of policy (see Lucas, 1992). However, the efficient contracts in these models can often induce better allocations than a monetary equilibrium; that is, money is not essential.

Given these earlier attempts, why do we need another microfoundation of money? There are two reasons. First, it is difficult to communicate the results between these earlier models, because each of them uses a particular structure to emphasize a particular trading friction. To improve the communication, we need to construct a framework to unify the language in the field. Not only

[2]The trading-post models have some resemblance to Townsend's models but are closer to the general equilibrium framework. See Ostroy and Starr (1974) and Shapely and Shubik (1977).

can a unified language help us to describe the important frictions that are common in all previous models, but also it may lead us to uncover other important frictions that are implicit but not emphasized in previous models. Second, previous models are not convenient for examining the competition between different media or methods of exchange. Often, money is the only asset in these models or money is not dominated in the rate of return by other assets. Some of Townsend's models give a hope of how to address the competition between money and private debt, but their rigid structures are difficult to modify.

In 1989, Nobuhiro Kiyotaki and Randall Wright constructed a model of money that appeared very different from previous models of money.[3] They dispensed with the notion of Walrasian markets completely and instead described the exchange process as bilateral matches. This change in the model environment simplified the description of the frictions and the task of illustrating why money is needed to overcome those frictions. Kiyotaki and Wright (1989) examined commodity monies. One of their main results is that good intrinsic qualities, such as a low storage cost, are not a pre-requisite for an object to serve as a medium of exchange – there is an equilibrium where the object with the highest storage cost serves as one of the monies. This result suggests the possibility that fiat money, being intrinsically useless, may also be valued as a medium of exchange. This possibility was explored in detail in a later paper by Kiyotaki and Wright (1993). From these two papers, a literature has grown into what is now called the *search theory of money.* I will call it a "value theory of money", instead, for reasons to be given in the next section.

3. A Value Theory of Money

I will focus on fiat money instead of commodity money. It then makes sense to start with Kiyotaki and Wright (1993). Also, I will describe the model in discrete time rather than continuous time.

3.1. An Environment of Decentralized Exchange

There is a continuum of types of goods and a continuum of types of agents. All agents have the common discount rate $r > 0$. The types are indexed by points along a unit circle. Agents are specialized in production and consumption, so that there is need for exchange. A type i agent consumes only the goods in the interval $[i^* - \alpha/2, i^* + \alpha/2]$, which are called the agent's consumption goods, and produces only good i which is called the agent's production good. The parameter $\alpha \in (0, 1)$ is an important parameter, to be explained later. The index i^* is drawn randomly from the unit circle in each period. All goods are perfectly storable at no cost.

In addition to the goods, there is another object called fiat money. This object is intrinsically useless in the sense that it does not generate utility to any agent and it does not facilitate production. The total supply of this object is fixed at M per capita, where $M \in (0, 1)$. At the

[3] A precursor to Kiyotaki and Wright (1989) is Jones (1976). Jones formulates the model with adaptive expectations, which prevents him from getting very far with the problem.

beginning of time, this object is distributed randomly in the population so that a fraction, M, of the agents hold money, each holding one unit.

I make two restrictive assumptions. First, money and goods are all indivisible. Second, an agent cannot hold both money and goods at any given time; that is, there is a unit upper bound on an agent's inventory. With these assumptions, a trade is always a swap of two indivisible objects. As such, the model cannot say anything meaningful about the determination of prices.

Let the utility of consuming a good be $u > 0$ if it is a consumption good and 0 otherwise.[4] Let the disutility of production be $c \in (0, u)$ and denote $U = u - c$. To keep the focus on fiat money, it is useful to exclude the possibility that an agent accepts a non-consumption good in a match. This can be achieved by assuming that an agent who accepts a good incurs disutility (transactions cost) $\varepsilon \in (0, U)$, where ε can be made arbitrarily small.[5]

The exchange process is described by the following assumptions:

(a1) Random matches: agents meet each other randomly;

(a2) Separation of trades: in each period an agent meets only one other agent;

(a3) No public record-keeping: each agent's trading history is private information.

As I will discuss later, these extreme assumptions are not necessary for the main results. At this point, however, let me discuss the trading patterns implied by these assumptions.

Imagine that two agents, say i and j, meet in the market. Agent i has desire for good j if and only if good j lies in the interval $[i^* - \alpha/2, i^* + \alpha/2]$. This occurs with probability α, because agent j is a random draw from the population and the index i^* is from the unit circle. Similarly, agent j has desire for good i with probability α. Thus, the two agents have a double coincidence of wants with probability α^2. There are two cases of a single coincidence of wants, each occurring with probability $\alpha(1 - \alpha)$. Finally, with probability $(1 - \alpha)^2$, the meeting has no coincidence of wants. Therefore, the smaller the value of α, the more difficult it is to have a double coincidence of wants. Although this difficulty of exchange is important for monetary exchange (see Jevons, 1875), it is not sufficient.

To support monetary exchange, there must also be restrictions on agents' abilities to communicate and to monitor actions. Assumptions (a2) and (a3) posit that the costs of communicating between matches and of monitoring agents' actions are infinite. Although these costs need not be infinite in order to induce monetary exchange, the absence of such costs would make monetary exchange unnecessary. To see this, suppose first that agents can communicate costlessly between

[4] The assumption of a uniform utility level over all consumption goods is used here for simplicity. See Kiyotaki and Wright (1991) for a general specification where agent i's utility decreases in the distance between the type of a good and the agent's ideal consumption good i^*.

[5] Note that accepting money does not incur the transactions cost ε. This assumption is innocuous, since all the qualitative results here will continue to hold even after such a cost is added.

matches, a polar assumption to (a2). Then, agents will arrange trades between matches to circumvent the problem of a double coincidence of wants. Suppose next that it is costless to keep a public record of all transactions, a polar assumption to (a3). Then, the economy can function efficiently without the use of money. Every agent is required to produce for the trading partner (as a gift) whenever the partner desires the good. If the agent refuses to produce for the partner in such a match, the agent will be excluded from the trading system forever. Because trades can be costlessly monitored, the punishment can be enforced. If agents are sufficiently patient, then the punishment supports the gift-giving scheme.[6] Assumption (a3) precludes not only this gift-giving mechanism of exchange, but also all forms of credit trade.

Assumptions (a1) - (a3) are intended to simplify the analysis and to contrast sharply with the Walrasian paradigm. However, no one would believe seriously that they are good characterizations of realistic markets. I can relax these assumptions as follows:

(A1) Sufficiently asymmetric trade: in a large fraction of trade, an agent desires for the trading partner's goods more than the partner desires for the agent's goods.

(A2) Costly communication between matches.

(A3) Anonymity: it is costly to update the public record of transactions.

These assumptions should not be strange to monetary economists. All microfoundations of monetary theory have imposed these assumptions, explicitly or implicitly. This is done in Townsend (1980) by spatial separation, in Samuelson (1958) by overlapping generations, and in Lucas (1980) by the communication cost. In comparison with previous models, the Kiyotaki-Wright model seems more natural for spelling out these frictions explicitly. In particular, the assumption of anonymity was not emphasized in previous models, although it was implicit there.[7]

Putting the assumptions in the general form should also help us to correct two misunderstandings of the theory that one might conceive from the restrictive assumptions (a1) – (a3). One misunderstanding is the view that the theory requires some matches to have no double coincidence of wants. This view is not warranted, because one can construct a model in which money is valued and essential despite the fact that every match has a double coincidence of wants (see Engineer and Shi, 1998, and Berentsen and Rocheteau, 2003). What is needed is that a large fraction of matches are asymmetric in the sense described in (A1) so that money is a better way than barter to transfer the gains from trade between agents. Moreover, the asymmetry of trade can arise in many forms. One example is private information about the quality of the goods, as Williamson and Wright (1994) demonstrate.

[6] This result is an implication of the well-known "Folk Theorem" in repeated games (e.g., Fudenberg and Maskin, 1986). In addition to the gift-giving scheme, there are many other outcomes that can be supported by punishments.

[7] Kiyotaki and Wright (1989, 1993) did not specify the general assumptions (A1) - (A3). Kocherlakota (1998) seems the first to emphasize the role of money as a record keeping device.

Another misunderstanding is that the theory requires buyers not to know where to get their consumption goods. Contrary to this impression, monetary exchange can still arise in an environment where matching is directed (see Corbae et al., 2003). In this regard, it is unfortunate that the theory spawned from the Kiyotaki-Wright model is labelled the search theory of money. For a better description, I will sometimes refer to a market which satisfies Assumptions (A1) - (A3) as decentralized exchange, and a monetary model with these assumptions as a value theory of money.

3.2. Individuals' Decisions and Equilibria

To illustrate how fiat money can be valued in the environment described in the previous subsection, I now analyze the agents' decisions. Call an agent who holds money a *money holder* and an agent who does not have money a *producer*. A money holder's decision is simple. Because money is intrinsically useless, a money holder is always willing to exchange money for a good he desires. But a money holder will not exchange money for a non-consumption good, because doing so incurs a transaction cost $\varepsilon > 0$. Thus, commodity money does not arise in this economy. Similarly, a producer will not trade a good for a non-consumption good. In a meeting with a double coincidence of wants, the two producers barter by swapping their goods.

The only non-trivial decision is a producer's decision as to whether to accept money in a match where the money holder desires the producer's good. Let me pick an arbitrary producer in such a match and denote the probability with which the producer accepts money as $\pi \in [0, 1]$. The decisions of other producers in similar matches are denoted as Π.

To analyze agents' decisions, it is convenient to use the Bellman equation. Let V_m be the value function of holding a unit of money and V_p the value for a producer. Both are measured at the *end of a period*. Consider only the stationary economy in which these value functions are constant over time. Then, the Bellman equation for V_m is

$$V_m = \frac{1}{1+r} \left\{ (1-M)\alpha\Pi(U - \varepsilon + V_p) + [1 - (1-M)\alpha\Pi]\, V_m \right\}.$$

Because the value is measured at the end of a period, the terms on the right-hand side of the equation are the expected values of future payoffs. To explain these terms, notice that a money holder trades the money away only when all of the following events occur: he meets a producer, the producer has a consumption good he wants, and the producer accepts money. These three events occur with probability $(1-M)$, α and Π, respectively. Thus, the probability of a successful trade for a money holder is $(1-M)\alpha\Pi$. In this case, the money holder exchanges money for the good, which yields net utility $(u - \varepsilon)$. He then immediately incurs the cost c to produce a good, the net value of which is $(V_p - c)$. Thus, the value of an exchange to the money holder is $(U - \varepsilon + V_p)$, where $U = u - c$. When a trade fails to occur, the money holder holds onto his money, which has the value V_m.

An alternative (and perhaps easier) interpretation for the above equation is that the "permanent income" from holding money, rV_m, is equal to the present value of expected gains or surplus from trade. A trade yields a gain, $(U - \varepsilon + V_p - V_m)$, to a money holder. Incorporating the probability of a successful trade, the alternative interpretation leads to:

$$rV_m = (1 - M)\alpha\Pi\left(U - \varepsilon + V_p - V_m\right). \tag{3.1}$$

This equation can be verified by re-arranging the original equation.

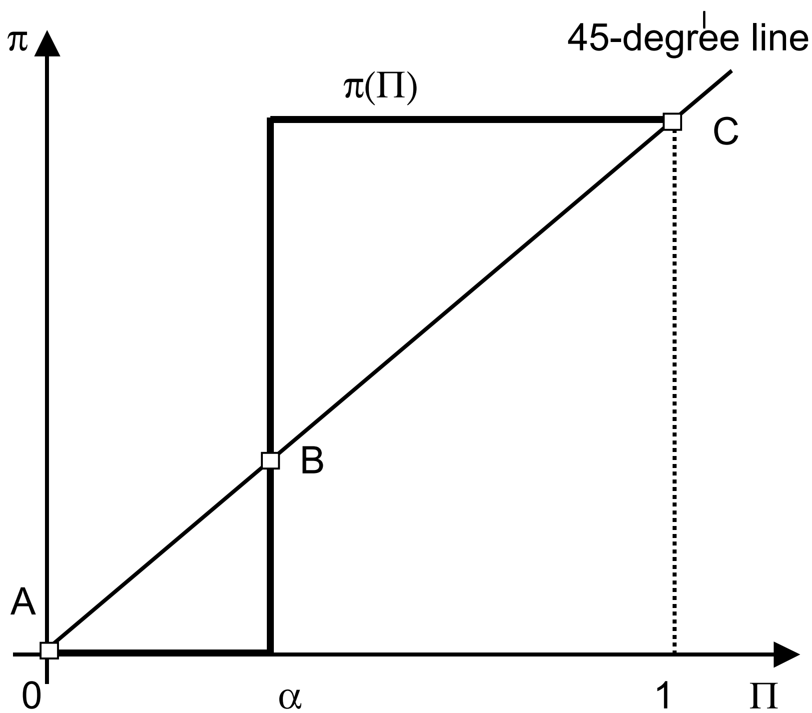

Figure 1. An individual producer's best response and equilibria

For a producer, there can be two possible trades. One is barter, which occurs with probability $(1 - M)\alpha^2$. The gain to a producer from a barter trade is $(U - \varepsilon)$. The other possible trade is a monetary trade. For a monetary trade to occur, the producer must meet a money holder who desires the producer's good, which occurs with probability $M\alpha$. In this match, the producer chooses the probability π to accept money, and this choice maximizes the expected gain to the producer: $\pi(V_m - V_p)$. Therefore, the Bellman equation for V_p is:

$$rV_p = (1 - M)\alpha^2\left(U - \varepsilon\right) + M\alpha \max_{\pi}\left[\pi(V_m - V_p)\right]. \tag{3.2}$$

A *symmetric equilibrium* in this economy can be defined as (π, Π, V_p, V_m) such that: (i) given (Π, V_p, V_m), each individual producer's decision π is optimal; (ii) the value functions satisfy (3.1) and (3.2); and (iii) producers' decisions are symmetric, i.e., $\pi = \Pi$.

If $\pi = \Pi = 0$, the equilibrium is a barter equilibrium. A monetary equilibrium exists if $\pi = \Pi > 0$. When $\pi = \Pi = 1$, money is accepted by every producer and the equilibrium is called a pure monetary equilibrium. If $0 < \pi = \Pi < 1$, the monetary equilibrium is a mixed one – producers sometimes accept money and sometimes do not.

Define $D = V_m - V_p$. An individual producer's decision on π is given as follows: $\pi = 1$ if $D > 0$; $\pi = 0$ if $D < 0$; and $\pi \in [0, 1]$ if $D = 0$. In all equilibria, $\max_\pi(\pi D) = \Pi D$. Using (3.1) and (3.2), I can solve for D and verify that $D > 0$ if and only if $\Pi > \alpha$. Then, an individual producer's decision on π is a best response to other producers' decisions, $\pi(\Pi)$, where $\pi(\Pi) = 1$ if $\Pi > \alpha$, $\pi(\Pi) = 0$ if $\Pi < \alpha$, and $\pi(\Pi) \in [0, 1]$ if $\Pi = \alpha$. An equilibrium is a solution to the equation $\pi(\Pi) = \Pi$.

The best response is depicted in Figure 1. Points A, B and C in the figure are the symmetric equilibria. Thus, I have the following proposition.

Proposition 3.1. *For any $M \in (0, 1)$ and $\alpha \in (0, 1)$, there exist three equilibria. One is the barter equilibrium, where $\pi = 0$. The second is a pure monetary equilibrium, where $\pi = 1$. The third is a mixed monetary equilibrium, where $\pi = \alpha$.*

3.3. Money and Efficiency

What does this simple model say about the value of money? To find out, let me solve for V_m from (3.1) and (3.2):

$$V_m = (1 - M)\alpha\Pi \left[1 - \frac{(1 - M)\alpha}{r + \alpha\Pi}(\Pi - \alpha)\right] \left(\frac{U - \varepsilon}{r}\right). \tag{3.3}$$

First, the value of money depends on agents' expectations. More precisely, multiple equilibria are self-fulfilling and money is not always valued. Multiplicity arises from a strategic complementarity among producers' decisions; that is, an individual producer's acceptance of money increases with other producers' acceptance. If all agents believe that money will have no value, then it is indeed optimal for each producer to reject money in trade. On the other hand, if all agents believe that money will be valued highly, then it is optimal for each producer to always accept money. If all agents believe that money has the same value as goods, then it is optimal to be indifferent between the two objects. Because such complementarity is common in coordination games, it is then useful to view fiat money as a social contrivance and the above model as a game to coordinate on the contrivance. Viewed this way, the non-monetary equilibrium is a result of coordination failure and should be an equilibrium of every model where money is fiat.[8]

Second, money can be essential in the pure monetary equilibrium. To see this, let me compute agents' lifetime utilities. Change the notation of the value functions to $V_m(M)$ and $V_p(M)$ to

[8] Wallace (1980) refers to the existence of the non-monetary equilibrium as an indication that a monetary equilibrium with fiat money is always tenuous. This feature cannot be appreciated if money is introduced into preferences, the production function, or cash-in-advance constraints.

emphasize the dependence on the money stock. In the barter economy, an agent's maximized lifetime utility is $V_p(0)$. In the pure monetary equilibrium, it can be verified that $V_p'(0) > 0$ if and only if $\alpha < (1-r)/2$. Thus, when $r < 1$ and $\alpha < (1-r)/2$, we have $V_p(M) > V_p(0)$ for sufficiently small M. Because $V_M(M) > V_p(M)$ in all monetary equilibria, the pure monetary equilibrium Pareto improves upon the barter equilibrium if agents are patient, if the coincidence of wants is difficult and if the money stock is low.

The efficiency gain does not arise from improved terms of trade – the quantities of goods and money in each trade are fixed by the assumption of indivisible goods and money. Rather, the efficiency gain comes from the increased probability of exchange. In the pure monetary equilibrium, a money holder exchanges with probability $(1-M)\alpha$ and a producer with probability $(1-M)\alpha^2 + M\alpha$. Weighted by the population of each type of agents, the average probability of exchange for a randomly selected agent in the pure monetary equilibrium is $(1-M)\alpha\,[2M + (1-M)\alpha]$, where the weights are the sizes of the two types of agents. This expression is increasing in M if and only if $M < (1-\alpha)/(2-\alpha)$. Because the probability of exchange in a barter economy is given by the same expression with $M = 0$, then money increases the probability if only a small fraction of agents are endowed with money.

However, if the fraction of agents endowed with money is high, then the probability of exchange in the pure monetary equilibrium is lower than in the barter economy. In this case, too much money is chasing too few goods, which makes the monetary equilibrium less efficient than the barter economy. For example, in the extreme case where everyone holds money, the probability of exchange is zero in a monetary economy; so is every agent's utility. Also, the probability of exchange is lower in the mixed monetary equilibrium than in the barter economy.

Finally, the value of M that maximizes social welfare can be interior. To see this, let me measure social welfare by the weighted sum of individual agents' utilities, where the weights are the same for all agents. I define the steady state welfare per period as:

$$W(M) = MrV_M(M) + (1-M)rV_p(M). \tag{3.4}$$

Again, focus on the pure monetary equilibrium. The welfare level is:

$$W(M) = \alpha(1-M)[(1-M)\alpha + M\pi](U-\varepsilon).$$

Because goods are indivisible, aggregate output is equal to $W(M)/(U-\varepsilon)$. Thus, social welfare and output are both maximized at the same particular level of M. If $\alpha \geq 1/2$, the optimal level of M is zero. If $\alpha < 1/2$, the optimal level of M is $M^o = (1-2\alpha)/[2(1-\alpha)]$, which is interior. Setting $M = 1$ is never optimal because, when everyone holds money, production and consumption are zero. In fact, it is never optimal to endow a half or more of the population with money.

The optimal level of money, M^o, decreases in α. Because a lower value of α corresponds to more severe frictions, the optimal level of money increases with the trading frictions.

The appropriate interpretation of M is the distribution of money, rather than the money stock. It is difficult to associate the above result on optimal M to the optimal quantity of money in standard monetary economics (e.g., Friedman, 1969). The problem is two-fold. First, because money is indivisible, the model cannot examine money growth in a meaningful way – positive money growth in the model will destroy the value of money by leading to $M = 1$ while negative money growth will drive the economy to a non-monetary economy. Second, because goods are also indivisible in the model, the price level is fixed at one. It is difficult to see whether the real effects of M discussed above are just artifacts of the fixed price level. In the next section, I will examine flexible prices and discuss other methods of exchange that compete with money.

4. Prices, Distribution and Competing Methods of Exchange

4.1. Determination of Prices

As discussed above, endogenizing the price level is necessary for distinguishing real money balances from nominal balances. Also, the price level itself is important because it is the reciprocal of the purchasing power of money. What determines the purchasing power of money? Is the price level uniquely determined? These questions are ones that a monetary model is supposed to answer.

To determine prices, I relax the assumption of indivisibility. Ideally, I would like to allow both money and goods to be divisible, but that would complicate the problem significantly (see section 5.1). For price determination, it suffices to allow the object on one side of a monetary trade to be divisible but keep the object on the other side of the trade indivisible. This subsection presents a model in which goods are divisible and money is indivisible. The next subsection will discuss the opposite environment in which money is divisible and goods are indivisible.

The presentation here closely follows the papers by Shi (1995) and Trejos and Wright (1995), which will be referred to as the Shi-Trejos-Wright model. To allow for divisible goods, let me assume that each producer can produce any non-negative quantity of goods. The cost of producing q units of goods is $\phi(q)$ in terms of utility. For simplicity, I normalize $\phi(q) = q$. The utility of consuming q units of consumption goods is $u(q)$, where u is strictly increasing and concave, with $u(0) = 0$, $u'(0) = \infty$ and $u'(\infty) = 0$. Also, assume that there exists a positive and finite level $\bar{q}$ such that $u(\bar{q}) = \phi(\bar{q})$. Then, there exists a unique $q^* \in (0, \bar{q})$ such that $u'(q^*) = \phi'(q^*)$ $(= 1)$. Call q^* the efficient quantity of production since it equates the marginal utility of consumption and the marginal cost of production. Denote $u^* = u(q^*)$.

In contrast to the Kiyotaki-Wright (1993) model, I assume that goods are perishable. This assumption rules out commodity money, and so I can set the transaction cost of accepting goods to zero. However, with perishable goods, a producer may decide to trade away the goods regardless of the match, because the goods will perish anyway. To rule out such spurious trading, I change the assumption so that a producer can produce on the spot after finding a trade.

Money is still indivisible, with a unit upper bound on each money holder's holdings. A money holder cannot produce goods. The taste patterns and the matching process are exactly the same as in the previous model. In particular, Assumptions (a1) – (a3) are still imposed.

In the described environment, each trade involves only a small number (two) of agents. Because matches are separated from each other, the two agents in a particular match do not have immediate access to the agents in other matches. Thus, each agent in a match has local monopoly power, despite the fact that there are many agents in the economy. The Walrasian mechanism does not seem reasonable for determining the terms of trade in each match. On that, there seems to be an agreement. However, there is no agreement on which non-Walrasian pricing mechanism should be selected. Many non-Walrasian mechanisms can give both sides of the match local monopoly power. A large bulk of the literature has used Nash bargaining to determine the terms of trade. I will use this formulation and will discuss other pricing mechanisms later. A desirable feature of Nash bargaining is that its outcome coincides with that of some non-cooperative sequential bargaining games (see Binmore, 1987).

To formulate the Nash bargaining problem, I specify an agent's threat point as the value that an agent can get by quitting the match and continuing to search. This value is V_p for a producer and V_m for a money holder. The agents in each particular match take these values as given, because they are determined by future matches rather than the current match. As before, denote $D = V_m - V_p$.

Consider first a barter trade between, say, producers i and j. The two producers bargain over two quantities, q_i and q_j. The level q_i is the quantity of goods agent i gives to agent j, and q_j the quantity agent j gives to agent i. Such a trade yields a surplus $[u(q_j) - q_i]$ to agent i and $[u(q_i) - q_j]$ to agent j. The threat point V_p does not appear in these surpluses, because a producer remains a producer after a barter trade. The Nash bargaining problem is:

$$\max_{(q_i,q_j)} [u(q_j) - q_i]^{1/2} [u(q_i) - q_j]^{1/2} \text{ s.t. } u(q_j) \geq q_i \geq 0 \text{ and } u(q_i) \geq q_j \geq 0.$$

The two constraints are necessary for the two agents to participate in the trade. The two agents have the same bargaining weights, because they are symmetric. The unique solution to the above problem is $q_i = q_j = q^*$, where q^* is defined above. Since the solution maximizes the joint surplus in the match, the quantities of goods traded in barter are (ex post) efficient.

Now consider a monetary trade. The money holder and the producer bargain over the quantity of goods, q, that the producer produces for the unit of money. The nominal price of goods in this trade is $p = 1/q$. The joint surplus in the trade is $[u(q) - q]$. The surplus to the money holder is $[u(q) - D]$; to the producer, it is $(D - q)$. For the two agents to participate in the trade, q must satisfy the constraints $u(q) \geq D$ and $D \geq q \geq 0$. Let $\theta \in [0, 1]$ be the bargaining weight of the money holder. Then, the two agents take D as given and choose q to solve:

$$\max_{q} [u(q) - D]^{\theta} [D - q]^{1-\theta} \text{ s.t. } u(q) \geq D \geq q \geq 0.$$

I focus on stationary monetary equilibria, although a non-monetary equilibrium exists here as in the Kiyotaki-Wright (1993) model. For a monetary equilibrium to exist, two conditions are necessary. First, $q > 0$; otherwise a money holder would rather discard the money and become a producer. Second, $u(q) > D$; that is, the money holder's participation constraint in the Nash bargaining problem does not bind. If $u(q) = D$, then the money holder's surplus would be zero, and again the money holder would be better off discarding the money. Since $D \geq q$, these two conditions imply $D > 0$ and $u(D) > D$. That is, the value of being a money holder is strictly greater than the value of being a producer, and the joint surplus in a monetary trade is strictly positive. Using the earlier notation $\bar{q}$, I write these implications as $D \in (0, \bar{q})$.

The producer's participation constraint does not bind in monetary equilibria. To see this, one can verify that the objective function in the Nash bargaining problem is strictly decreasing in q at $q = D$. By reducing q slightly below D, the objective function is increased. Thus, $q < D$.

Therefore, the quantity q solves the first-order condition of the above bargaining problem. Let the solution be $q = q(D)$. This is the best response of an individual producer to the aggregate variable D. For all $D \in (0, \bar{q})$ and all $\theta \in (0, 1)$, $q(D)$ lies in the interior of $(0, D)$. Furthermore, $q'(D) > 0$, $q(0) = 0$, and $q(\bar{q}) = \bar{q}$.

The producer also chooses the probability, π, with which to accept money. This choice maximizes the producer's expected gain from the trade, $[\pi(D - q)]$. Because $q < D$ in all monetary equilibria, as argued above, then $\pi = 1$. That is, all monetary equilibria are pure equilibria, in contrast to the Kiyotaki-Wright (1993) model with indivisible goods. With this result, I set $\pi = 1$ in the following analysis.

Now, I modify the Bellman equations for V_m and V_p. Let Q be the quantity of goods that are traded in monetary matches other than the particular one analyzed above. Although $Q = q$ in all symmetric equilibria, it is useful to distinguish the two variables. Then,

$$rV_m = (1 - M)\alpha[u(Q) - D], \tag{4.1}$$

$$rV_p = (1 - M)\alpha^2(u^* - q^*) + M\alpha(D - Q). \tag{4.2}$$

Note that V_m and V_p are functions of the aggregate variables Q and D only. This feature supports the earlier treatment that individual agents take V_m and V_p as given when choosing q. Use the above conditions and the definition of D to solve for $D = D(Q)$, where $D(0) < 0$, $D(q^*) > 0$, $D'(Q) > 0$ and $D''(Q) < 0$. The relationship $D = D(Q)$ is a requirement for aggregate consistency – it determines the aggregate variable D when all trading producers trade Q units of goods for money.

In the symmetric equilibrium, all producers trade the same quantity of goods for money, i.e., $q = Q$. Thus, an equilibrium is a joint solution to the following two equations: $D = D(q)$ and $q = q(D)$. The following proposition characterizes the solutions.

Proposition 4.1. *Monetary equilibrium does not exist if either r is very large or θ is close to zero. Suppose $r \to 0$. Then, a sufficient condition for existence is $\theta \geq M + (1 - M)\alpha$. In this case, there exist two monetary equilibria, one with $q < q^*$ and another with $q \geq q^*$. The two monetary equilibria coincide only when $\theta = M + (1 - M)\alpha$, in which case $q = q^*$. Generically, there are an even number of monetary equilibria. In all monetary equilibria, money is accepted with probability one and the quantity of trade satisfies: $u(q) - q > \alpha(u^* - q^*)$.*

The proposition states several intuitive results. First, monetary equilibria exist only if the bargaining weight of a money holder is not too low and if agents are patient. When money holders have a very low bargaining weight, the gain from a trade to a money holder is small, in which case an agent would rather discard money and become a producer. To explain why heavy discounting destroys the existence of a monetary equilibrium, notice the following difference between a barter trade and a monetary trade. When a producer trades in barter, he can consume immediately. But if he trades goods for money, he must wait until he finds a producer in a future match who can produce his consumption goods. The cost of this delay in consumption increases with time discounting. If agents are sufficiently impatient, barter gives the producer a higher present value than monetary exchange.

Second, because $q \neq q^*$ in general, the quantity of goods traded in a monetary match is inefficient ex post. Time discounting is one reason for the inefficiency and asymmetric bargaining weights another. But even when $r \to 0$ and $\theta = 1/2$, the inefficiency still exists except for a particular quantity of money. Despite this ex post inefficiency, a monetary trade is more ex ante efficient than barter in the sense that the *expected* joint surplus is higher in a monetary trade. More precisely, $u(q) - q > \alpha(u^* - q^*)$, as stated in the above proposition.

Figure 2 depicts an individual producer's best response $q = q(D)$ and the aggregate consistency condition $D = D(q)$. The intersections between these two curves are monetary equilibria, as depicted by points H and L. In general, the number of monetary equilibria is even except for the non-generic case where the two curves are tangent to each other at some point. However, the case with exactly two monetary equilibria reveals almost all essential properties of monetary equilibria in general. Denote the level of q in these two equilibria as q_H and q_L, respectively, where $q_H > q_L$. Money has a higher purchasing power in the equilibrium with q_H than in the equilibrium with q_L. For this reason, I call the equilibrium with q_H the *strong equilibrium* and the equilibrium with q_L the *weak equilibrium*. The strong equilibrium has a lower price level, higher real money balances and higher output than the weak equilibrium.

As in the simple model, multiple monetary equilibria arise from the strategic complementarity among producers' decisions. However, the crucial decision here is on how much to trade for money rather than how likely to accept money – a producer accepts money with probability one in all monetary equilibria. An individual producer is willing to exchange a high quantity of goods for money if the excess value for a money holder relative to a producer is high, as shown by the feature $q'(D) > 0$. In turn, if all producers are willing to exchange a high quantity of goods for money, then the excess value of money is indeed high, as shown by the feature $D'(q) > 0$. This interaction leads to the strong equilibrium. On the other hand, if all other producers are only willing to exchange a low quantity of goods for money, then the value of money will be low and an individual producer also will exchange a low quantity of goods for money. This interaction leads to the weak equilibrium.

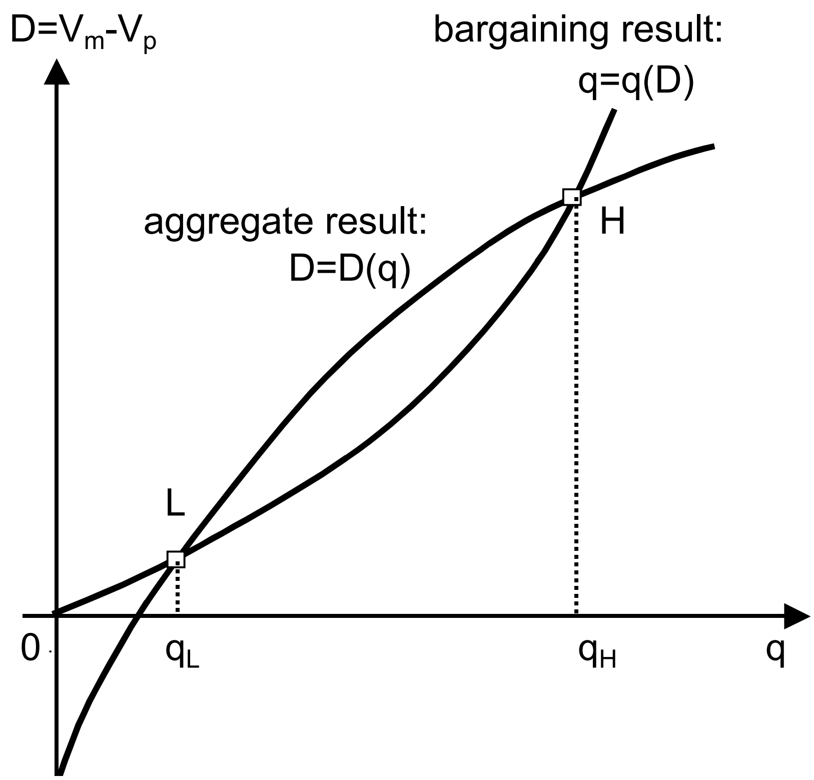

Figure 2. Monetary equilibria with Nash bargaining

I can rank the two equilibria according to the Pareto criterion. To do so, let me focus on the case where r is small. Note that V_m and V_p approach infinity when $r \to 0$, but rV_m, rV_p and D are all positive and finite. In fact,

$$rV_m \to rV_p = (1-M)\alpha \left\{ M \left[u(q) - q \right] + (1-M)\alpha(u^* - q^*) \right\}. \tag{4.3}$$

The values for the agents are maximized when the joint surplus in a monetary trade is maximized. The joint surplus is larger in the strong equilibrium than in the weak equilibrium. Thus, the strong equilibrium Pareto dominates the weak equilibrium.

What does this model say about the price level? First, the price level is not uniquely determined; rather, it differs in different equilibria. The multiplicity is generic – it exists except for a measure zero subset of parameter values. This indeterminacy is quite different from the one in standard monetary models. For example, in models with money in the utility function, the price level can be indeterminate only when real money balances and consumption are strongly complementary with each other, e.g., Grandmont (1985). Also, Diamond (1984) constructs a search model with cash-in-advance constraints and shows that multiple monetary equilibria can arise when the matching technology has increasing returns to scale. These features of preferences and technology are not needed here for the indeterminacy. Note that all monetary equilibria in the current model are stationary. In contrast, in overlapping generations models, a continuum of non-stationary monetary equilibria exist and they all converge to the non-monetary equilibrium (see Kareken and Wallace, 1980).

Second, the price level is sensitive to events that do not affect preferences and technology, such as sunspots. More precisely, there exists an equilibrium in which such non-fundamental events affect the price level (see Shi, 1995). Because changes in the price level affect real activities here, there may be a theoretical justification for a monetary authority to stabilize prices by providing a nominal anchor to expectations. Moreover, if the economy is extended to include two fiat monies, then there will be multiple equilibria with differing exchange rates between the two monies.[9]

Third, the price level and real money balances respond to changes in the distribution of money, M, differently in the two equilibria. In Figure 2, an increase in M shifts down the curve $D(q)$ but leaves the curve $q(D)$ intact. Thus, q_H decreases and q_L increases. That is, an increase in the fraction of agents who are endowed with money increases the price level in the strong equilibrium but decreases the price level in the weak equilibrium. Aggregate real money balances can be measured as Mq. Then, an increase in M increases aggregate real balances in the weak equilibrium, but may either increase or decrease real balances in the strong equilibrium.

Finally, this model extends the welfare properties of the simple Kiyotaki-Wright (1993) model to an environment with divisible goods and endogenous prices. Let me measure social welfare according to $W(M)$ defined in (3.4). In the limit $r \to 0$, W approaches rV_p (see (4.3)). The optimal distribution of money, in terms of M, is always less than 1. The optimal M is positive iff $W'(0) > 0$. This requirement is equivalent to $\theta u'(q) < 1 - \theta$ at $M = 0$. Because $q_H \geq q^*$, the requirement is satisfied in the strong equilibrium if $\theta < 1/2$. Recall that a sufficient condition for the existence of a monetary equilibrium is $\theta \geq M + (1-M)\alpha$. At $M = 0$, this condition becomes $\theta \geq \alpha$. Thus, if $\alpha \leq \theta < 1/2$ and if r is sufficiently small, it is optimal to endow an interior fraction of the population with money in the strong equilibrium.

4.2. Other Pricing Mechanisms and the Distribution of Money Holdings

The Shi-Trejos-Wright model is specific, but the main predictions of the model are more general. The literature has examined several variations of the model.

[9] The set of equilibrium values of the exchange rate depends on the supply of each money. This feature distinguishes the indeterminacy from that in Kareken and Wallace (1981), where there is a continuum of equilibria and the exchange rate in each equilibrium is endogenously fixed.

One variation is to change the bargaining game. The Nash bargaining formulation used above is equivalent to a non-cooperative sequential bargaining game where agents can search during bargaining; that is, agents in a match can search after rejecting an offer and can come back to the current match if they do not find a new match. Let me now consider a different bargaining game where agents can search only after breaking up the current match. This change in the bargaining game changes the equilibrium values for being a money holder and a producer. Because these values are important for monetary equilibria, it is interesting to see whether the main results will change radically.

When agents cannot search while bargaining, the bargaining outcome is equivalent to that of Nash bargaining where each agent's threat point is set to zero. This amounts to replacing the buyer's surplus with the value $[u(q)+V_p]$ and the producer's surplus with the value $[V_m-q]$ in the Nash bargaining formulation. In comparison with the previous formulation, the main difference here is that the quantity of goods traded in a monetary match is less than the efficient quantity q^* in all monetary equilibria. However, other qualitative predictions are still valid. Chief among these are the multiplicity of monetary equilibria, the ex post inefficiency of a monetary trade, and the comparative statics of the price level and welfare.

Another variation is to explore alternative pricing mechanisms. For example, price posting by sellers seems an obvious alternative to bargaining. With price posting, the model will encounter the well-known paradox illustrated by Diamond (1971). That is, if a buyer always encounters only one price quote in each period, then the seller will set the price at the buyer's reservation value, even though there are many sellers in the market. In the monetary economy, Diamond's paradox implies that a money holder gets zero surplus from trade, in which case money has no value in the equilibrium. There are at least two ways to resolve this puzzle. One is to allow each buyer to receive a second price quote with positive probability, as Burdett and Judd (1983) did in a non-monetary model and Head and Kumar (2005) in a monetary model. Another way is to introduce private information about buyers' tastes (see Faig and Jerez, 2005). In either way, the main new feature of the equilibrium relative to the Shi-Trejos-Wright model is that there is a non-degenerate distribution of prices.

Finally, one can allow agents to accumulate money while keeping goods indivisible. Green and Zhou (1998) examine such an environment. Their model generates two different results of particular interest. First, there is a continuum of monetary equilibria instead of just two. These equilibria differ from each other in prices and aggregate output but, in each equilibrium, all monetary trades occur at a single price level. Second, the distribution of money holdings across agents is non-degenerate. This distribution arises from the facts that matching is random and that agents can accumulate money. Even if two agents have the same initial money holdings, they may experience different matching sequences and, hence, they may end up with different amounts of money holdings. For example, an agent who has been a producer in every match in the past will accumulate more money than another agent who has been a producer in only a half of the matches. When the distribution of money holdings is non-degenerate, monetary policy can have distributive effects among the buyers that do not exist in the Shi-Trejos-Wright model. See Molico (1998), Deviatov and Wallace (2001), Green and Zhou (2005), Camera and Corbae (1999), Berentsen et al. (2004) and Zhu (2003).

Despite these differences, the model by Green and Zhou shares three important properties with the model described in section 4.1. First, the multiple equilibria are self-fulfilling; thus, the value of money depends on expectations. Second, different equilibria can be Pareto ranked. The lower the average price level, the higher the level of welfare. Third, the optimal fraction of agents who hold money can be positive and interior.

4.3. Competing Methods of Exchange and Record-Keeping

So far the analysis has maintained Assumption (a3) to exclude all forms of public record-keeping. To examine the competition between money and other ways of exchange, I need to relax (a3) to allow some form of record-keeping. The literature has done so with two different approaches.

One approach is to use mechanism design. In this approach, a social planner designs the mechanism to allocate resources and the mechanism must be compatible with agents' incentives. The set of incentive compatible allocations depends on what the planner knows about agents' actions and, hence, on what public record-keeping device is available. Among the incentive compatible allocations, some are efficient in the sense that they maximize social welfare. If these efficient allocations rely on the use of money, then money is essential. Using this approach, Kocherlakota (1998) establishes the result that money is an imperfect record-keeping device. That is, all incentive compatible allocations that can be implemented with the use of money can be implemented alternatively with some forms of record-keeping, and money is essential only if the record-keeping device is sufficiently imperfect.

Let me first show how a record-keeping device can help to implement the allocations in the monetary equilibria in the Shi-Trejos-Wright model. Assume that the society can keep a record on every agent. To produce the allocation in the monetary equilibria, suppose that the record has three possible values, 0, 1 and -1. At the beginning of the economy, the record is set to 1 for a fraction $M \in (0, 1)$ of the agents and 0 for the remaining agents. The exchange takes place and the records are updated according to the following rules (mechanism). In a meeting with a double coincidence of wants, the two agents exchange only when both agents' records are 0, in which case the two swap a quantity q^* of goods and their records remain at 0 after the trade. In a meeting with a single coincidence of wants, the two agents exchange only when the agent who wants the partner's good has record 1 and the agent who does not want the partner's good has record 0, in which case the latter agent produces a quantity q for the former agent and their records are swapped after the trade. In all other cases, agents do not trade. If an agent deviates from these rules, then the agent's record is set to -1 and the agent is banished from the exchange system forever. When agents are sufficiently patient, the punishment to autarky is severe, and so the above trading mechanism is incentive compatible. Setting $q = q_H$ and q_L, respectively, the mechanism generates the same allocation as the one in the two monetary equilibria in the Shi-Trejos-Wright model.

The social planner can do better than the monetary equilibria. For example, the planner can improve the efficiency of trade by setting $q = q^*$ in all trades. Moreover, the planner can achieve the efficient allocation by giving the following instructions: an agent produces q^* for the partner whenever the partner wants the goods and he has not deviated, and produces zero for the partner if he deviated in the past. Thus, money is only an imperfect record-keeping device. To relax Assumption (a3) without destroying the essential role of money, other competing record-keeping devices must also be imperfect.

Kocherlakota and Wallace (1998) examine one such imperfect record-keeping device by assuming that a transaction is recorded with probability $\rho \in (0, 1)$. They characterize the optimal allocations which maximize agents' ex ante welfare. An optimal allocation is said to have no role for money if production according to the described rule does not depend on money holdings, to have some role for money if production depends on money holdings, and to be entirely monetary if every consumer has money and every producer does not. Kocherlakota and Wallace establish the following results when agents are sufficiently patient. First, if the record-keeping device is sufficiently good in the sense that ρ is close to 1, then the optimal allocation has no role for money. Second, if the record-keeping device is sufficiently imperfect in the sense that ρ is close to 0, then the optimal allocation has some role for money. The optimal allocation is entirely monetary if and only if $\rho = 0$. These results seem to accord well with the evolution of payments systems. As a society improves its capacity of record-keeping, the role of money gradually declines. Wallace, with others, modifies this framework to examine the societal benefits of using other forms of exchange, such as banks, e.g., Cavalcanti and Wallace (1999).

Another approach to examining the competition between money and other methods of exchange is to characterize the equilibria with these methods directly. Bilateral credit and middlemen have been examined in this way.

To introduce bilateral credit, I assume that two agents in a match can keep a record of the transactions between themselves, but cannot share records with agents outside the match (see Shi, 1996). That is, actions can be monitored within a match but not outside a match. To illustrate how this limited record-keeping ability enlarges the trading possibilities, recall that two producers in a match with a single coincidence of wants cannot trade with each other in the Shi-Trejos-Wright model. Now they can. The producer who likes the trading partner's good can issue an IOU to get the good. To make the IOU credible, the issuer (the debtor) gives the creditor a collateral that must be regained to assist future consumption. To prevent the

IOU from circulating, let me assume that each agent's collateral is specific to himself and hence not transferable. To ensure that the debtor repays the IOU, the creditor monitors the debtor by following the debtor into future matches until the IOU is repaid. Such monitoring is costly, because the creditor cannot trade with other agents until the IOU is repaid. The repayment can be either in terms of the creditor's consumption goods or money. Once the IOU is repaid, the debtor regains the collateral and the creditor ceases to be able to track the debtor.[10]

In this environment, bilateral credit can co-exist with money and can dominate money in the rate of return. Money still has value, because a monetary exchange allows agents to trade and consume more quickly. A monetary trade is instantaneous, after which the two agents can match with other agents. In contrast, a credit trade ties the creditor and the debtor for some time until the debt is repaid. This cost of monitoring the debtor makes a bilateral IOU an imperfect substitute for money. Hence, bilateral credits must generate a higher explicit rate of return than money in order to be valued. A debtor is willing to pay such a higher return to the creditor, because he does not always have money. Moreover, bilateral credits improve efficiency. The monetary equilibrium with credit Pareto dominates the monetary equilibrium without credits.

The model also has implications for the forms of repayments. First, the model admits monetary equilibria where IOUs are repaid only in terms of goods, only in money, or in both. Second, these equilibria with different repayment methods can co-exist. Third, when repayments are real, there are multiple equilibria that resemble those in the Shi-Trejos-Wright model. However, insisting on nominal repayments eliminates the weak monetary equilibrium by requiring the purchasing power of money to be bounded below by that of bilateral credits. In some parameter regions, the equilibrium with only nominal repayments can improve efficiency relative to the equilibrium with real repayments.

To introduce middlemen into the Shi-Trejos-Wright model, Li (1999), Johri and Leach (2002) and Shevchenko (2004) assume that middlemen can improve the quality of matches. However, the details are different in these two models. Li models the quality of goods in a match as private information. Agents can become middlemen by investing in an information technology that gives them better abilities to discern the quality of goods than other agents. In contrast, Johri and Leach, and Shevchenko assume that the qualities of goods are publicly observed. In their models, a middleman is an agent who has invested to acquire a larger capacity of storage than other agents. By storing more types of goods, a middleman will be more likely to have the good for the agent whom the middleman will meet in the future. In both models, a middleman must be compensated for the investment by charging a higher price. An equilibrium with middlemen exists only when the cost of investment by a middleman is not too high. Money is still valued in the equilibrium with middlemen. The reason is simple: when exchanges are decentralized, even

[10] Corbae and Ritter (2004) extend the model to allow the two agents to form a long-term relationship and to decide when to terminate that relationship. Li (2001) allows the credit to circulate in a limited way.

the middlemen need money to speed up their trades!

5. Bringing Monetary Theory Closer to Policy

The theory presented so far still contains assumptions that severely limit its ability to study policy. One is the indivisibility of money or goods. With indivisible money, the theory is unable to study money growth or optimal inflation. With divisible money and indivisible goods, there is a continuum of monetary equilibria (see section 4.2), which makes it ambiguous whether money is neutral. In both cases, one cannot easily compare the results with those in traditional macro models. Another limit of the theory is the lack of other assets, such as capital, which is important for studying business cycles. I now describe the effort in the literature to eliminate these limitations.

5.1. Divisible Money and Inflation

When money and goods are divisible, there will be a non-degenerate distribution of money holdings in the equilibrium, as I explained in section 4.2. This distribution is intractable analytically and cumbersome numerically (see Molico, 1998).[11] Although the distribution is useful for the distributive effects of monetary policy, there are many other issues for which we may want to abstract from the distribution. We can achieve this abstraction by allowing agents to insure against the matching risks. One way is provided by Shi (1997) and another by Lagos and Wright (2005). The two approaches are similar, so I will present a variation of Shi (1997) here.

For agents to insure against the matching risks, I assume that each agent experiences a large number of trades in each period. To illustrate this assumption, suppose that each agent has one unit of time in each period, which can be divided into many small pieces. Let $\Delta > 0$ be the length of each piece of time and $1/\Delta$ be the number of such pieces in each period. The agent acts as a buyer in the market in a fraction b of the time and a seller in a fraction $(1-b)$ of the time. During each piece of time, the agent is randomly matched with another agent. At the beginning of each period, the agent must allocate the amount of money to each piece of time in which he will be a buyer and decide how much to trade. Furthermore, the agent is not able to move money between two pieces of time in a period. At the end of the period, the agent collects the goods and money received during the period, and then consumes. As Δ becomes sufficiently small, the number of trades during each period becomes sufficiently large over which the agent's matching risks are smoothed. Thus, the agent's money holdings at the end of each period are deterministic. If all agents have the same amount of money holdings initially, then the distribution of money holdings across agents will remain degenerate in all periods.

[11] This technical difficulty is not unique here, but rather common to all models in which agents face uninsurable idiosyncratic risks. See Green (1987), Lucas (1980, 1992), Bewley (1980), and Imrohoroglu (1992).

The above modelling device is isomorphic to the following economy with large households. Imagine that the economy consists of many households. Each household consists of a large number of members, a fraction b of whom are buyers and a fraction $(1-b)$ are sellers. At the beginning of a period, the household makes all decisions for the members. The household allocates money to the buyers and instructs the members on how much to trade. The members do not make decisions – they simply carry out the instructions. After matching and trading, the members bring back the goods and money received, and all members share the same amount of consumption. Because the household experiences a large number of matches, the matching risks are smoothed in each household in each period.

I introduce money growth in the usual way. Let $\bar{M}_t$ be the average money holdings per household in period t. At the beginning of each period, a lump-sum monetary transfer is distributed to the households, in the amount $\tau_t = (\gamma - 1)\bar{M}_{t-1}$, where $\gamma > 0$ is a constant. Thus, the aggregate stock of money holdings grows over time at the gross rate γ. All nominal prices and quantities are normalized by the aggregate stock $\bar{M}$, and so they are stationary in the steady state despite the fact that the money stock is not constant over time.

Let me examine an individual household's decisions. A household has the preferences and productive abilities of an individual agent in the Shi-Trejos-Wright model. Since these households are symmetric, except for the types of goods they want to consume and the goods they can produce, I can pick an individual household as the representative household. Use lower-case letters for the representative household's decisions and upper-case letters for other households' decisions. The representative household's intertemporal utility function is:

$$\sum_{t=0}^{\infty} \beta^t \left[u(c_t) - \Phi_t \right], \quad \beta \in (0,1).$$

Here, c is the household's consumption, and Φ is the sum of the sellers' disutilities of production. Recall that a seller's cost of producing q units of goods is $\phi(q)$. Instead of the linear function used earlier, this function is now assumed to be strictly convex (and increasing), with $\phi(0) = 0$ and $\phi(\infty) = \infty$. The efficient quantity of trade, q^*, is still defined by $u'(q^*) = \phi'(q^*)$.

Add two assumptions to simplify the analysis. First, there is no match in which the two agents have a double coincidence of wants, and so barter does not occur. If the producer in a match can produce the consumption goods for the buyer's household, the match is called a trade match. Second, the buyer in every trade match makes a take-it-or-leave-it offer.

Let $\widehat{\alpha}$ replace the earlier notation α, the probability with which an agent likes another agent's good. Denote $\alpha = \widehat{\alpha}(1-b)$, which is the probability with which a buyer encounters a trade match. A seller encounters a trade match with probability $\alpha b/(1-b)$. Then, αb is the total number of trade matches that involve the representative household's buyers. It is also the total number of trade matches that involve the representative household's sellers. In each trade match involving a buyer of the representative household, let $x\bar{M}$ be the amount of money and q the amount of goods traded, where $\bar{M}$ is the average money holdings per household. In each trade match involving a seller of the representative household, let $X\bar{M}$ and Q be the corresponding quantities of trade. Then, $c_t = b\alpha q_t$ and $\Phi_t = b\alpha\phi(Q_t)$. The representative household takes Q as given, because it is other households' offer.

For each trade involving the representative household's buyer, the household chooses the quantities of money and goods to be traded, (x, q).[12] To characterize these decisions, let m_t be the household's money holdings in period t before the exchange occurs, normalized by M. Let $v(m_t)$ be the household's value function. Define $\omega_t = (\beta/\gamma)v'(m_{t+1})$ as the household's future marginal value of money, discounted by both β and money growth. Define Ω_t similarly for other households. Then, x and q must satisfy two constraints. First, the buyer cannot offer more money than he has; therefore, $m/b \geq x \geq 0$. This constraint reflects the separation between the buyers in different matches. As such, it is a trading restriction, rather than a usual cash-in-advance constraint.[13] Second, the surplus of trade to the seller in the match must be non-negative, i.e., $\Omega x \geq \phi(q)$. This constraint binds for the buyer. Accordingly, $q_t = q(x_t)$, where $q(x) = \phi^{-1}(\Omega x)$.

The representative household's decision problem can be formulated as a dynamic programming problem. For this purpose, suppress the subscript t and use the subscript $\pm j$ to denote the variables in period $t \pm j$, where $j > 0$. Substituting $q = q(x)$ into the expression for consumption, I express the representative household's decision problem as follows:

$$v(m) = \max_{0 \leq x \leq m/b} \left[u\left(b\alpha q(x)\right) - b\alpha\phi(Q) + \beta v(m_{+1})\right],$$

where $m_{+1} = \left[m + b\alpha X - b\alpha x + \tau_{+1}\right]/\gamma$. The amount $b\alpha X$ is the total amount of money received by the household from selling goods, and the amount $b\alpha x$ is the total amount of money spent by the household. The factor $1/\gamma$ appears in the formula, because m is normalized by $\bar{M}$ and m_{+1} by $\bar{M}_{+1}$. In this optimization problem, the representative household takes all upper-case variables as given.

In a symmetric equilibrium, the decisions are the same for all households, and so the lower-case variables are equal to the corresponding upper-case variables. As a result, $m = M = 1$. A monetary equilibrium requires $\omega > 0$. Also, I impose $\omega < \infty$ to satisfy the transversality condition on money holdings. This condition is necessary for the first order conditions of the maximization problem to indeed characterize the household's optimal decisions. I focus on the steady state of the monetary equilibrium. In particular, $\omega_{+1} = \omega$.

Let $b\alpha\lambda$ be the shadow cost of the constraint $m/b \geq x$ in the above maximization problem. The optimal choice of x leads to the condition: $\lambda/\omega = u'(c)/\phi'(q) - 1$. Then, the envelope

[12]The household chooses the quantities (x, q) for each of its members, and so these quantities should also be indexed by the member's identity. I suppress this index to shorten the notation, but it is useful to keep this fact in mind when computing the surplus from an individual trade.

[13]The easiest way to distinguish this constraint from a cash in advance constraint is to allow the possibility of barter, as in the Shi-Trejos-Wright model. In such an extension, although a household can trade through barter in some matches, each buyer in a monetary trade is still constrained by the above money constraint.

condition of money holdings yields:

$$\frac{\gamma}{\beta} - 1 = \alpha \left[\frac{u'(c)}{\phi'(q)} - 1 \right]. \tag{5.1}$$

It is clear that when $\gamma > \beta$, the marginal utility of consumption exceeds the marginal cost of production, in which case the money constraint on a buyer binds. I will assume $\gamma > \beta$ throughout the analysis. Whenever I discuss the equilibrium with $\gamma = \beta$, I mean the allocation in the limit $\gamma \downarrow \beta$. The limit case $\gamma = \beta$ is the so-called Friedman (1969) rule.[14]

The equation (5.1) and the accounting equation $c = b\alpha q$ together determine a unique level of q in the steady state. Other variables in the steady state are determined accordingly. In particular, $x = 1/b$, $c = b\alpha q$, and $\omega = b\phi(q)$. Therefore, a unique monetary steady state exists.

Note that the solutions for the real variables are independent of the money stock. Thus, money is neutral in the model. However, money is not superneutral. Since the money growth rate appears in (5.1), money growth and hence long-run inflation affects the real allocation.

The steady state is inefficient for all $\gamma > \beta$. To see the inefficiency, consider a social planner who chooses (q, Q) to maximize the representative household's utility function, under the constraint of the matching technology and the symmetry constraint. The socially optimal allocation satisfies $u'(c) = \phi'(q)$. In contrast, the equilibrium steady state generates $u'(c) > \phi'(q)$ for all $\gamma > \beta$ (see (5.1)). This inefficiency is reminiscent of the inefficiency of the monetary equilibria in the Shi-Trejos-Wright model. It implies that the Friedman rule is optimal. That is, steady state utility is maximized by setting γ to β.

The optimality of the Friedman rule is not surprising. In the economy described here, the number of trades is assumed to be fixed. The only inefficiency here is in the intensive margin of trade, i.e., the quantity of goods traded in each match. The Friedman rule eliminates this inefficiency by reducing the opportunity cost of holding money to zero. However, the number of trades in an economy with decentralized exchange is likely to vary with inflation. This creates a meaningful trade-off between the inefficiency in the intensive margin and the extensive margin. The next subsection will explore this trade-off.

5.2. Friedman Meets Hosios

I endogenize the extensive margin of trades by allowing each household to choose the search effort (or intensity) for its buyers.[15] This choice generates externalities, because each household ignores the effects of its own choice on the other households' matching rates. For example, suppose

[14] The Friedman rule requires the nominal interest rate to be zero. However, the nominal interest rate is not an operational concept here because nominal bonds are not present in the setup. Instead, I use the nonpecuniary return to money to serve the role of interest. The rate of such a return is $\gamma\omega_{-1}/(\beta\omega) - 1$. Setting this rate to zero in the steady state leads to $\gamma = \beta$.

[15] Another way to endogenize the extensive margin is to allow the household to choose b, the fraction of buyers or the fraction of time spent in buying (see Shi, 1997). However, the algebra is more cumbersome.

that each household increases the buyers' search effort. This creates a positive externality by increasing the matching rate for the sellers. It also creates a negative externality by reducing the matching rate for other buyers. Unless the equilibrium compensates the search effort efficiently, it cannot internalize the externalities. Monetary policy can affect the extent of this inefficiency.

Let the representative household's choice of search effort for each of its buyers be $s > 0$. Let the disutility of such effort be $h(s)$, where $h' > 0$, $h'' > 0$ and $h(0) = 0$. Let S be the search effort of other households' buyers. For simplicity, assume that a seller's search effort is fixed at 1.

In the presence of the search effort, I need to modify the matching rates. The "effective" number of buyers in the market is bS, and so the effective number of participants in the market is $(bS + 1 - b)$. The total number of trade matches in each period is assumed to be given by a matching function with constant returns to scale: $\widehat{\alpha}(bS)(1-b)/[bS+1-b]$. Let α denote the probability with which one unit of a buyer's search effort generates a trade match. Then,

$$\alpha = \alpha(S) \equiv \frac{\widehat{\alpha}(1-b)}{bS+1-b}. \tag{5.2}$$

A buyer who searches with effort s gets a trade match with probability αs, and a seller gets a trade match with probability $bS\alpha/(1-b)$. The total number of trade matches involving the representative household's buyers is $b\alpha s$, and the total number of trade matches involving its sellers is $b\alpha S$. The household's consumption in a period is $c = b\alpha sq$ and disutility of production is $\Phi = b\alpha S\phi(Q)$. Net utility in a period is $[u(c) - \Phi - bh(s)]$.

For the two externalities discussed above to be interesting, the choice s must be interior. To achieve this, we need to alter the earlier assumption on bargaining to give the seller a positive surplus. The easiest way to do so is to assume that the buyer in a trade match still makes all the offers but must give the seller a minimum surplus $R > 0$. This constraint on the seller's minimum surplus is: $\Omega x - \phi(q) \geq R$. I assume that the two agents in the particular match takes R as given, although R will be determined in the equilibrium. This assumption simplifies the analysis.[16] The constraint binds for the buyer, which solves for $q = q(x) \equiv \phi^{-1}(\Omega x - R)$.

The representative household's decisions on x and s solve the following problem:

$$v(m) = \max_{s \geq 0,\ 0 \leq x \leq m/b} \left[u\left(b\alpha s q(x)\right) - b\alpha S\phi(Q) - bh(s) + \beta v(m_{+1}) \right],$$

where $m_{+1} = \left[m + b\alpha SX - b\alpha sx + \tau_{+1}\right]/\gamma$. The representative household takes α and all upper-case variables as given.

For the equilibrium, I need to determine the minimum surplus for the seller, R. To do so, imagine that R is determined by Nash bargaining between a fictitious buyer and a fictitious seller from two average households. Assume that the quantities in such an average trade are

[16] It may be interesting to use a more elaborated bargaining protocol to deliver this assumed outcome, but I do not do so here. At the end of subsection 5.3 I will discuss the additional effects – the holdup effects – that will arise if the two agents in a match can directly affect R.

(Q, X). The seller's surplus is $[\Omega X - \phi(Q)]$ and the buyer's surplus is $[u'(C)Q - \Omega X]$. Note that the quantity of goods is evaluated with the marginal utility of consumption, because each trade contributes only marginally to the household's consumption. Let the seller's bargaining weight be $\theta \in (0, 1)$. The Nash bargaining formulation is:

$$\max_{(Q,X)} [u'(C)Q - \Omega X]^{\theta} [\Omega X - \phi(Q)]^{1-\theta} \quad \text{s.t. } M/b \geq X.$$

It is easy to confirm that the money constraint in this problem binds iff $u'(C) > \phi'(Q)$. Under this condition, $X = M/b$. Solving the problem yields:

$$R = (1 - \Theta) \left[u'(C)Q - \phi(Q) \right], \tag{5.3}$$

where Θ is the following share of the match surplus to the buyer:

$$\Theta \equiv \frac{\theta u'(C)}{\theta u'(C) + (1 - \theta)\phi'(Q)}. \tag{5.4}$$

This share is endogenous, which is an important feature for optimal monetary policy.

Now, I turn to the steady state of the equilibrium. In the steady state, the buyer's optimal search effort satisfies the following first-order condition:

$$h'(s) = \alpha(s)\Theta \left[u'(c)q - \phi(q) \right]. \tag{5.5}$$

The right-hand side of this equation is the expected marginal gain to a buyer's search effort. It is equal to a buyer's matching rate per search effort, α, multiplied by the buyer's surplus from a trade. In addition, the envelope condition for money holdings can be derived as follows:

$$\frac{\gamma}{\beta} - 1 = \alpha(s)s \left[\frac{u'(c)}{\phi'(q)} - 1 \right]. \tag{5.6}$$

For all $\gamma > \beta$, I have $u' > \phi'$ (provided $s > 0$); therefore, the money constraint in a trade binds.

Before examining optimal money growth, let me examine the socially efficient quantity of trade and search effort. Suppose that a social planner chooses (q, Q) and (s, S) to maximize the representative household's steady state utility, under the constraints of the matching function and symmetry. Then, the efficient choices are characterized by:

$$h'(s) = \left[\frac{d\,(s\alpha(s))}{ds} \right] (u'q - \phi), \tag{5.7}$$

$$u'(c) = \phi'(q). \tag{5.8}$$

Compare the equilibrium equations, (5.5) and (5.6), with the above equations. An equilibrium must satisfy two requirements in order to be socially efficient. One is the Friedman rule, i.e., $\gamma = \beta$, which makes the equilibrium condition (5.6) coincide with the efficiency condition (5.8). I call this margin of efficiency the intensive margin, because it has to do with the quantity of goods traded in each match. The other requirement is $\alpha(s)\Theta(c) = d(s\alpha(s))/ds$. I call this margin of efficiency the extensive margin, because it has to do with the search effort and the number of trades. I can express this requirement as $\Theta = \eta(s)$, where $\eta(s)$ denotes the share of the aggregate number of matches contributed by the buyers' search units. Thus, to internalize the matching externalities, the buyers' share of surplus must be equal to the buyers' share of contribution to matches. This condition is known as the Hosios rule (see Hosios, 1990).

Efficiency in the extensive margin requires the Hosios rule for the following reason. The social benefit of a buyer's search effort is the increase in the aggregate number of matches. For a household, however, the private benefit of increasing a buyer's search effort is the share of surplus the buyer will obtain in trade. By equating the buyers' share of surplus and the buyers' contribution to the aggregate number of matches, the Hosios rule internalizes the externalities that buyers' search effort generates in the matching process.

Generically, there does not exist a money growth rate that allows the equilibrium to implement both the Friedman rule and the Hosios rule. That is, the monetary equilibrium is not the first-best. If monetary policy could not affect the search externalities, then the Friedman rule would be the second-best policy. However, monetary policy does affect the search externalities in this economy, and so there is an interesting trade-off between the efficiency in the two margins.

To see the trade-off, consider an example where u' is constant and $\phi(q) = \phi_0 q^{\xi}$, where $\phi_0 > 0$ and $\xi > 1$. Let the equilibrium allocation under the Friedman rule be $q = q^*$ and $s = s^*$. Let γ_h be the money growth rate at which the equilibrium allocation implements the Hosios rule. Define optimal monetary policy as the money growth rate that maximizes the household's steady state utility. Then, optimal monetary policy achieves the first-best only if $\eta(s^*) = \theta$. When $\eta(s^*) \neq \theta$, optimal monetary policy can achieve only the second-best. The following proposition describes the second-best policy (see Berentsen et al., 2003, for a proof).

Proposition 5.1. *If $\theta < \eta(s^*)$, then the optimal money growth rate is above the Friedman rule and below the rate implied by the Hosios rule. If $\theta > \eta(s^*)$, then the rate implied by the Hosios rule is inconsistent with a monetary equilibrium, in which case the Friedman rule is optimal.*

Loosely speaking, increasing money growth can improve welfare in some cases, because it can induce the buyers to search more intensively. To understand how money growth generates this effect, I need to describe the feature of the buyers' share of surplus in trade, which is denoted as Θ. In this model, the buyers' share is a function of the quantity of goods traded in a match, i.e., $\Theta = \Theta(q)$ (see (5.4)). This function is decreasing and, when $u' = \phi'$, $\Theta = \theta$. These features of the surplus share arise from the money constraint in trade. To explain, suppose $u' > \phi'$, so that the money constraint binds. The binding constraint serves as a credible device for the buyer to limit the offer in bargaining and, hence, to extract a larger fraction of the surplus. Thus, the buyer's share exceeds θ. The more severely binding is the money constraint, the higher is the fraction of the surplus the buyer can extract. Since the money constraint is more binding when the difference $(u' - \phi')$ is larger, and since this difference decreases in q (for all $q < q^*$), then the buyer's share decreases with the quantity of goods traded in the match.

The quantity of goods traded in a match is a decreasing function of the money growth rate when the money growth rate is low. Thus, the buyer's share of surplus increases with the money growth rate. It is equal to θ when the money growth rate follows the Friedman rule.

Now, I explain the results in the proposition. First, consider the case where $\theta < \eta(s^*)$. In this case, the buyers' share of surplus under the Friedman rule is below their share of contribution to the aggregate number of matches. Buyers are under-compensated for their search effort, and so the aggregate number of matches is inefficiently low. Increasing money growth above the Friedman rule reduces q and, as explained above, increases the buyers' share of surplus. As a result, buyers' search effort increases, which improves efficiency in the extensive margin. This efficiency gain comes at a loss of efficiency in the intensive margin, because the decrease in q widens the gap between the marginal utility of consumption and the marginal cost of production. The best trade-off between the two margins is achieved by a money growth rate between the Friedman rule and the level implied by the Hosios rule.[17]

The case with $\theta > \eta(s^*)$ is different. In this case, the buyers are over-compensated for their search effort, even when money growth is at the Friedman rule. To improve efficiency in the extensive margin, money growth must be reduced below the Friedman rule, which is inconsistent with the existence of a monetary equilibrium. The Friedman rule remains optimal in this case.

The above analysis highlights the tension between the efficiency in the extensive margin and the efficiency in the intensive margin. This analysis brings up the following two questions. First, is the extensive effect important for aggregate economic activities relative to the intensive effect? There is no direct evidence on how the aggregate number of trades fluctuates over business cycles. However, it is reasonable to believe that the volume of sales is high during the Christmas season, for example, because people make more purchases as well as a larger quantity in each purchase. The distinction between the two margins of trade is natural in economies where the exchange is decentralized. In such economies, the suboptimality of the Friedman rule is a general possibility.

The second question is: should fiscal policy, instead of monetary policy, be used to correct the inefficiency in the extensive margin? The answer to this question depends on how well fiscal policy can be implemented. In the current model, inflation affects the extensive margin by increasing the bargaining weight of the buyers. A lump-sum tax on sellers, together with a lump-sum transfer to buyers, can also achieve this redistribution of the match surplus. If the tax were costless to collect, then the fiscal policy would be able to restore efficiency without monetary policy deviating

[17]The extensive effect has some resemblance to the effect of inflation on search, which Benabou (1992) analyzes. His model has sticky prices and inflation can increase buyers' search effort by reducing sellers' monopoly power. The differences are that there is an essential role for money in the current model and that prices are fully flexible.

from the Friedman rule. But the tax is likely to be costly to collect, because the exchange occurs in separated matches and record-keeping is difficult. One way to model the cost is to assume that the fiscal authority must send tax collectors out to the market to meet private agents, as modelled by Li and Wright (1998). Then, there can be a trade-off between the tax and inflation, which may cause optimal inflation to be above the Friedman rule.

5.3. Inflation and Capital Accumulation

Now, I introduce capital and analyze the long-run relationship between capital and inflation.[18] To put the analysis into context, let me briefly review the literature on the relationship between inflation and capital. Tobin (1965) argues that inflation increases the capital stock by reducing the rate of return on money relative to capital. This conclusion depends critically on his assumption of an exogenous savings function. Sidrauski (1967) replaces this exogenous function with intertemporal utility maximization where money appears in the utility function. The result is that, if labour supply is inelastic, the long-run capital stock is independent of inflation. Instead, the long-run capital stock is determined by equating the marginal productivity of capital to the rate of time preference. When labour supply is elastic, inflation reduces the long-run capital stock by inducing agents to increase leisure. Thus, the standard conclusion is that inflation either reduces or has no effect on the long-run capital stock.

Similar conclusions hold in models with cash-in-advance constraints. Moreover, if purchases of investment goods are subject to the cash-in-advance constraint, such as in Stockman (1981), then increasing inflation reduces the long-run capital stock by reducing the real value of wealth available for the purchases. Given the role that the extensive margin of trade played in the previous section, it is interesting to see whether it also has important implications for the relationship between money and capital.

To allow for capital, I assume that goods are storable and add two assumptions: (i) capital has the same attributes as consumption goods and the transformation between the two is costless; (ii) only the households that want to consume the goods can store the goods as capital. Assumption (i) is common in the neoclassical growth theory and it makes the relative price between consumption and capital accumulation unity. Assumption (ii) distinguishes one's output from one's capital, which is not common in the neoclassical theory. Together, the two assumptions rule out commodity monies. In this environment, a household can accumulate capital by storing consumption goods, but not by storing the household's own output. A household will not trade capital for other goods, either.

To the extent that assumption (ii) exogenously restricts the liquidity of capital, it does not appear to have any advantage over a cash-in-advance constraint that excludes capital from circulating as a medium of exchange. The defence against this potential criticism is that assumption

[18] The analysis here is based heavily on the paper by Shi (1999).

(ii) can be relaxed so that all households can store all types of goods. Because capital has different types (by assumption (i)), an argument similar to that in Engineer and Shi (1998) and Berentsen and Rocheteau (2003) implies that a trader accepts money more frequently than capital. That is, the difference in liquidity between money and capital can be endogenous. However, the resulting circulation of capital as commodity money would complicate the analysis significantly.

Let me continue the description of the technology. Capital and labour are the inputs in production. In addition to selling goods, sellers also perform the role of producers. Let k be the total stock of capital in the representative household at the beginning of a period. Let the number of buyers in the household be denoted as $b \in (0,1)$ and the number of sellers be $n = 1 - b$. The household divides capital evenly to the producers before the agents go to the market; thus, each producer has k/n units of capital. A producer uses the capital and the labour input, l, to produce the following quantity of output: $Q = l^{\varepsilon}(k/n)^{1-\varepsilon}$, where $\varepsilon \in (0,1)$. Inverting this function, I can solve for the amount of labour input as $l = L(Q,k)$, where

$$L(Q,k) \equiv Q^{\frac{1}{\varepsilon}} (k/n)^{-\frac{1-\varepsilon}{\varepsilon}} . \tag{5.9}$$

The disutility of the labour input is $\widehat{\phi}(l) = \phi_0 l^{\sigma}$, where $\phi_0 > 0$ and $\sigma > 1$. Define

$$\phi(Q,k) \equiv \widehat{\phi}(L(Q,k)) = \phi_0 Q^{\frac{\sigma}{\varepsilon}} (k/n)^{-\frac{\sigma(1-\varepsilon)}{\varepsilon}} .$$

Note that $L_2 < 0$ and $\phi_2 < 0$. That is, for any given amount of output, a higher capital input allows the producer to reduce the labour input and the disutility.

The representative household's net utility in a period is $[u(c) - b\alpha S\phi(Q,k) - bh(s)]$, where c is consumption and s the buyer's search effort. For simplicity, I assume that the marginal utility of consumption is a constant $u' > 0$.

To guarantee a minimum surplus R to the seller, the buyer faces the constraint: $\Omega x - \phi(q,K) = R$. As before, R is assumed to be exogenous to the two agents in a specific match. Denote the solution for q to the above constraint as $q = q(x,K)$. The derivatives of the function q with respect to the two arguments are denoted as q_1 and q_2.

Let $v(k,m)$ be the representative household's value function. The household's choices of (x, s, c, k_{+1}) solve the following problem:

$$v(k,m) = \max \left[u\left(c\right) - b\alpha S\phi(Q,k) - bh(s) + \beta v(k_{+1}, m_{+1}) \right],$$

subject to the law of motion of money, the constraint $0 \le x \le m/b$, and

$$c + k_{+1} - k \le b\alpha s q(x,K). \tag{5.10}$$

Here, capital is assumed not to depreciate. The representative household takes as given the matching rate α and all upper-case variables. In particular, by taking Q as given, the seller can derive some benefit from his capital input.[19]

The constraint (5.10) resembles the budget constraint in a standard neoclassical model. The main difference is the right-hand side of the constraint, i.e., the amount of resource that is available for consumption and capital accumulation. In a standard model, this resource is equal to output. In the current model, the resource is equal to the receipts of the household's consumption goods. Note that (5.10) effectively requires that capital accumulation and consumption purchases both be constrained by the household's money holdings.

In the equilibrium, the household's optimal decision on capital accumulation satisfies

$$b\alpha s \frac{[-\phi_2(q,K)]}{u'} = \frac{1}{\beta} - 1. \tag{5.11}$$

This condition requires that the expected rate of return to capital be equal to the real interest rate. Since ϕ_2 (< 0) is the reduction in disutility of labour generated by an additional unit of capital input, the left-hand side of the equation is the expected rate of return to capital from the labour-saving role. The right-hand side of the equation is the discount rate. It is also the real interest rate, because the marginal utility of consumption is constant.

The new feature in this condition is that the expected return to capital depends on the extensive margin of trade, i.e., on the expected number of matches ($b\alpha s$). This is because capital is utilized only when the seller has a trade match. The rate of utilization is equal to the probability with which the seller gets a trade match. This extensive margin of trade provides a novel channel through which inflation can affect capital. The following analysis shows that this additional channel can be the dominating force in some cases.

Other equilibrium conditions are similar to those in subsection 5.2. Replace the function $\phi(q)$ with $\phi(q,K)$ and $\phi'(q,K)$ with $\phi_1(q,K)$. Then, the buyers' share of surplus is $\Theta(Q,K)$, which is given by (5.4). The buyers' search effort obeys (5.5) and R obeys (5.3). The dynamic version of (5.6), which is the envelope condition of money holdings, is as follows:

$$\frac{\gamma\omega_{-1}}{\beta\omega} - 1 = \alpha(s)s\left[\frac{u'}{\phi_1(q,k)} - 1\right]. \tag{5.12}$$

Consider the special case where $\theta = 1$ and $\sigma = 1$, in which case $\Theta = 1$ and $R = 0$. To determine the equilibrium in this case, use (5.5) and (5.11) to solve s and q as functions of k. Denote these functions as $s = s(k)$ and $q = g(k)$. Then, I can express ϕ as a function of k, say, $\phi = f(k)$. Since $R = 0$ and $x = 1/n$, then $\omega = b\phi$. Substituting these results, I rewrite (5.12) as

[19] An alternative assumption is that the seller takes the function $Q = q(X,k)$, rather than the number Q, as given. In this case, the seller's labour input is $L(q(X,k),k)$, which is independent of k. The seller does not obtain any benefit from his higher capital, since the buyer extracts all of the benefit. That is, the buyer holds up on the seller's capital. Since we have abstracted from the seller's ability to hold up on the buyer's money, it seems reasonable to abstract from the buyer's ability to hold up on the seller's capital as well. At the end of this section, I will examine the two holdup effects together by allowing R to be determined inside the match.

$k_{-1} = G(k, \gamma)$, where the function G is implicitly defined as follows:

$$f(G(k,\gamma)) = \frac{\beta}{\gamma} f(k) \left\{ 1 + \alpha(s(k))s(k) \left[\frac{u'}{\phi_1(g(k),k)} - 1 \right] \right\}. \tag{5.13}$$

The steady state level of capital, denoted k^*, solves the equation $k^* = G(k^*, \gamma)$. The steady state is locally stable if $G_1(k^*, \gamma) > 1$. The following proposition can be established:[20]

Proposition 5.2. *Suppose $\theta = 1$, $\sigma = 1$ and that the money growth rate is low. Then, the steady state exists. Under very mild conditions, the steady state is dynamically stable. Moreover, the capital stock, the buyers' search effort, and aggregate output in the steady state are all increasing functions of the money growth rate.*

The positive effect of inflation on the long-run capital stock contrasts sharply with the negative effect in the literature, especially when one considers the nature of the constraint, (5.10). As said earlier, this constraint requires that purchases of both goods and capital be constrained by the resource acquired with cash. In traditional models with a similar constraint, e.g., Stockman (1981), an increase in inflation reduces the long-run stock of capital.

The extensive margin of trade is the key to the positive effect of inflation on capital. When inflation increases, real money balances fall. Buyers try to spend their money more quickly by increasing the search effort. Since the buyers' search effort creates a positive externality for the sellers by increasing the sellers' matching rate, the sellers are able to utilize their capital more often. Hence, the expected rate of return to capital increases. Responding to this higher return, households accumulate more capital. With a higher capital stock and more trades, aggregate output increases.

The main point of the above proposition is to illustrate that the effect of inflation on the extensive margin of trade can be the dominating force in some cases. However, for several reasons, one should not conclude from the above proposition that inflation will always increase the long-run capital stock and output. First, the proposition is restricted to special functional forms. Second, even in the special case, too high a money growth rate can destroy the monetary steady state. Third, the model has abstracted from other effects of inflation. One is that inflation

[20] The existence of the steady state can be verified directly. To establish other features of the steady state, differentiate (5.5) and (5.11) to find $s'(k)$ and $g'(k)$. It can be shown that $s'(k) > 0$. Then, $f'(k)$ has the same sign as the following expression:

$$\frac{sh''}{h'} - \frac{\alpha' s}{\alpha} - \left(1 + \frac{\alpha' s}{\alpha}\right)(1-\varepsilon)\frac{u'q}{u'q - \phi}.$$

Because $u'q > q\phi_1 = \phi/\varepsilon$ for all $\gamma > \beta$, the above expression is greater than $\left[\frac{sh''}{h'} - 1 - \frac{2\alpha' s}{\alpha}\right]$. Because $\alpha' < 0$, then a sufficient condition for $f'(k) > 0$ is $sh''/h' \geq 1$, which is easily satisfied by a wide class of functional forms of h. With $f' > 0$, differentiating (5.13) with respect to k yields $G_1(k^*, \gamma) > 1$. Thus, the steady state is dynamically stable. From (5.13), it is then easy to see that $G_2(k, \gamma) < 0$. Thus, $dk^*/d\gamma > 0$. Because $s'(k) > 0$, then $ds/d\gamma > 0$.

induces consumers to substitute low-quality goods for high-quality goods (see Peterson and Shi, 2004). The other is the holdup effects described below.

The holdup effects arise if the two agents in a trade can affect R directly. To see this, let me change the assumption so that R is determined inside each match by Nash bargaining. Then, the surplus R is given by a similar formulation to (5.3), except that the average variables in that formula are replaced with variables specific to the match. That is,

$$R(q, K) = [1 - \Theta(q, K)] \left[u'q - \phi(q, K) \right],$$

where $\Theta(q, K)$ is given by (5.4), with $\phi'(Q)$ being replaced with $\phi_1(q, K)$ and $u'(C)$ with u'. Now, the constraint on the seller's minimum surplus is $\Omega x - \phi(q, K) = R(q, K)$, which solves for $q = q(x, K)$.

There are two holdup effects. One is the feature $R_1 > 0$. The other is the feature $R_2 < 0$, which arises when the bargaining weight of buyers (θ) is sufficiently high. The feature $R_1 > 0$ is a holdup effect that a seller can exert on the buyer's money, which Lagos and Wright (2005) emphasize. That is, a seller in a match does not take into account the cost that the buyer incurred in the past to acquire money and holds down the relative "price" of money to goods below their relative value. More precisely, $R_1 > 0$ implies $q_1 < \Omega/\phi_1$. The feature $R_2 < 0$ is a holdup effect that a buyer can exert on the seller's capital, which is absent in Lagos and Wright (2005). The quantities of trade tend to under-compensate for the seller's capital, because the buyer does not take into account the cost that the seller incurred in the past to acquire capital. More precisely, for any given amount of money traded in the match, the amount of disutility of production that a unit of capital saves at the margin is smaller when $R_2 < 0$ than when $R_2 = 0$.

Money growth at low levels affects the two holdup effects in opposite directions. By reducing the value of holding onto money, money growth increases the holdup effect on the buyer's money. On the other hand, because money growth can increase the effective rate of return to capital, as explained before, it can reduce the holdup effect on the seller's capital. It is not clear whether the overall effect of money growth on the holdup is significant. A comprehensive study of the welfare effect of inflation should include these effects, in addition to the extensive effect. Such a study is likely to be quantitative.

6. Addressed Issues and Unfinished Tasks

I have described a value theory of money and reported its recent progress. It is now time to check what issues the theory has addressed and what others the theory still needs to address.

6.1. Taking Stock

Let me use the list of issues at the beginning of this lecture as a checklist.

On (I1): existence and essentiality of fiat money. The theory successfully addresses this issue. It provides a list of assumptions, (A1) - (A3), under which money is valued and essential. By doing so, the theory also gives a common language to describe a monetary economy.

On (I2): return dominance. The theory is partially successful. There are two types of assets that dominate fiat money in rates of return: one is real assets and the other is government-issued nominal bonds. The theory can explain reasonably well why real assets have higher rates of return than fiat money. The explanation can be built on the one in Engineer and Shi (1998), which I used earlier to motivate Assumption (A1). In particular, a real asset yields a stream of dividends in terms of goods whose values are different to different types of agents. When matches are asymmetric, the agents whose assets yield a high value of dividends are less willing to part with their own assets for other goods or assets. In contrast, agents are more willing to depart with money, because it is devoid of any intrinsic value. As a result, money is more liquid than other assets and the liquidity compensates for its low rate of return. Wallace (2000) captures a particular version of this argument using the difference in the size of dividends to generate different degrees of asset liquidity.

On return dominance of money by bonds, the theory can explain the positive aspect but still needs to do more work on the normative aspect. The positive aspect is the question why bonds dominate money in return in reality. A typical explanation is that legal restrictions make bonds illiquid and the illiquidity must be compensated by a higher return (e.g., Wallace, 1983). For this explanation to have force in traditional models, the legal restrictions must be non-negligible. With decentralized exchange, in contrast, even arbitrarily small legal restrictions are sufficient to make bonds illiquid (see Shi, 2005). In this sense, the value theory of money provides a robust explanation for the illiquidity of bonds. The normative aspect of the issue is whether illiquid bonds are welfare-improving. On this question, Kocherlakota (2003) uses a variation of Townsend's models to show that illiquid bonds can increase welfare by improving intertemporal consumption-smoothing among heterogeneous agents. However, the illiquidity of bonds in his model is a physical feature and the welfare-improving role of illiquid bonds lasts for only one period. The task remains to combine the two sides of the research so that the illiquidity of bonds is an equilibrium feature and that such bonds improve welfare persistently.

On (I3): optimal monetary policy. The theory has renewed the debate on whether the Friedman rule is optimal. It has uncovered additional effects of inflation, such as the effects on the extensive margin of trade and the holdup effects. However, successful resolution of this issue is likely to be quantitative, rather than qualitative. Although some authors have calculated the welfare cost of inflation (e.g., Lagos and Wright, 2005), their calculations are biased toward emphasizing the negative holdup effect of inflation on buyer's money holdings. It is important to incorporate the extensive effect of inflation and the holdup effect on sellers. Also, the quantitative assessment of optimal inflation is empirically relevant only if it incorporates fiscal policy as well as monetary policy, as Phelps (1973) argued.

Issues (I4), (I5) and (I6) are largely unresolved. I will explain why each of these issues deserves attention in future research.

6.2. Banking and Payments Systems

There are a few reasons why banking and payments systems are important for monetary economics. First, the history of money has always been interwoven with the history of banking and payments systems. Inside money created by banks and payment systems facilitates exchange just as fiat (outside) money does. Some economists, like Hicks (1989), have even gone so far to argue that the development of banking and inside money preceded the prevalent use of outside money. Given the importance of banking, we should ask the following questions. Is outside money essential in the presence of inside money? Is outside money a complement with or a substitute for inside money? These questions are important for the role of a central bank. If outside money is just a substitute for inside money, then its role will diminish as the private banking system improves. In this case, the central bank will gradually retreat from the arena of monetary policymaking, although it may still remain as a regulatory body. On the other hand, if inside money and outside money are complementary with each other, then the central bank's role in monetary policy will enlarge as the private banking system improves.

Second, there is a long-standing debate on whether policy should separate money from credit. On one side, Friedman (1960) argues that the separation is desirable, because inside money generates instability in the money supply and the price level. He advocates a 100% reserve requirement on the part of banks' liabilities that serve as a medium of exchange. Sargent and Wallace (1982) make the opposite argument using an overlapping generations model. They show that the separation of money from credit is Pareto inferior to a system without the separation, because the separation restricts the ability of economy to counter credit demand shocks.

Third, monetary policy is often conducted through payments systems. For example, the Bank of Canada uses the Large Value Transfer System to control the overnight rate. To understand how monetary policy works, it seems important to incorporate a payment system into the theory. For example, if monetary policy is to target inflation, one needs to know how changes in narrowly defined interest rates such as the overnight rate affect inflation. But there are more fundamental questions about a payments system and its relationship to fiat money. Why should payments be settled in outside money? How frequently should payments be settled? Should a central bank restrict credit to a payments system, say, by either imposing limits on the amount of credit or charging interests on daily credit?

The difficulty of examining banking and payments systems is that there is tension between the assumptions needed to support fiat money and those to support banking and payments systems. Money is a decentralized medium of exchange and it has value only when record-keeping is limited.

In contrast, banking and payments systems are forms of centralized exchange which rely on the use of credit. To support credit, the society must have a good record-keeping ability. It is a challenge to construct a model where both money and credit are essential. In section 4.3, I reviewed some attempts to model money and credit in a value theory of money. There is also an effort to study competitive banking and note issuing, e.g., Cavalcanti et al. (1999). However, the indivisibility assumption in those models makes their results difficult to interpret. Berentsen et al. (2005) try to eliminate this shortcoming. In addition, banks in those models lack some key features of banks in reality and nominal bonds are absent, which makes nominal interest rates unconnected to the price of nominal bonds.

Models of payments systems have used variations of Townsend's model of spatial separation, e.g., Freeman (1996), Kahn and Roberds (2001), Temzelides and Williamson (2001), and Williamson (2003). These models provide insights into the role of money as a means of settlement and the relative merits of different settlement arrangements. There are gains to be had from integrating this literature with the value theory of money presented here. First, the literature on payments systems assumes that a particular fiat money is the only means to settle payments. Second, the literature on payments systems assumes, implicitly or explicitly, that the cost of record-keeping between periods is so large that payments must be settled each period. These assumptions leave the following questions open. Can there be alternative means of settlements? Is a fiat money essential if its only role is to settle payments? What is the optimal frequency of settlement and how does it depend on the record-keeping ability? The value theory of money is suitable for addressing these questions. Because current monetary policy has so much to do with payments systems, thinking deeply about these questions may radically change the way in which monetary policy should be conducted.

6.3. Monetary Propagation and Quantitative Studies

The literature of monetary business cycles has tried to answer how the effects of monetary shocks are propagated on to affect the price level, interest rates and real activities in business cycles. This literature uses traditional monetary models where money appears either in the utility function or cash-in-advance constraints. Although there is a disagreement on how to identify monetary shocks, the common mechanism which allows monetary shocks to affect real activities in these models is the liquidity effect (Lucas, 1990, and Fuerst, 1992). This effect is a redistributive effect that arises from the assumption of limited participation. That is, only a fraction of agents in the economy can participate in the market where the monetary injection takes place; thus, the nominal interest rate is initially determined by the funds (liquidity) available to these agents. It is well known that the liquidity effect is too short-lived to account for the persistent effects of monetary shocks on interest rates and output (Christiano and Eichenbaum, 1992). To prolong the liquidity effect, the literature has added the assumption of rigid nominal prices and wages. However, the amount of nominal rigidities required to account for the persistent effects of monetary shocks is too high to be realistic (see Chari et al., 2000).

Can the value theory of money contribute to the understanding of monetary propagation? There are three reasons to believe that it can. First, the value theory of money is not subject to the Lucas Critique on policy analysis that applies to the traditional models. Second, the separation of trades is a building block of the value theory of money, which makes the theory a natural environment for generating the liquidity effect and prolonging the propagation. As an illustration, Katzman et al. (2003) show that, if agents must learn about the size of monetary shocks, then temporary shocks can have persistent effects on the price level and output. Third, the value theory of money introduces an additional dimension – the extensive margin of trade – along which monetary shocks can affect output. As I have shown (see Shi, 1998), this dimension can interact with buyers' search intensity and sellers' inventory to prolong the effects of monetary shocks even when the shocks are serially uncorrelated.[21]

Perhaps the most urgent work in the value theory of money is to conduct quantitative analyses. Issues like monetary propagation are quantitative. Not long ago, the theory was too primitive to conduct sensible quantitative analyses, but the theory has progressed so much that it is now time to take quantitative analysis seriously. The work has already begun. Shi (1998) computed impulse responses of output, employment and search intensity to monetary shocks. Wang and Shi (2004) calibrated a model with monetary and productivity shocks to quantify the variability of the velocity of money. Menner (2005) computed the dynamic responses of the variables in a model with capital accumulation to those shocks. These exercises have shown that the quantitative predictions of the value theory of money are different from those of traditional models. While this quantitative difference is encouraging, the exercises also introduce additional parameters through search and bargaining. The literature must find an agreement on how to identify these parameters. There should be no doubt that more quantitative analyses will be forthcoming.

6.4. International Monetary Issues

One of the main limitations of traditional models is that they are inadequate for studying international monetary issues. In particular, they could not explain why two monies or currencies could co-exist, nor evaluate the efficiency gain from currency unification. Now let us see what the theory presented in this lecture can say about these issues.

On the co-existence of multiple currencies, Matsuyama et al. (1993) provide the following answer: two currencies can both have positive values in the equilibrium if agents expect so and if the supplies of the two currencies are not very high. The two currencies can circulate nationally or internationally, again depending on agents' expectations. There is no need to require that

[21] Other papers that study monetary propagation with decentralized exchange include Williamson (2005), Shi (2004), and Head et al. (2005).

a country's goods be bought with a particular currency or that only a particular type of real money balances yields utility to a country's residents. Matsuyama et al. established these results with the assumptions of indivisible money and goods, but their main results continue to hold when goods are divisible (see Shi, 1995). With endogenous prices, the exchange rate between two currencies is endogenous and depends on expectations.

On currency unification, the theory indicates that unification can generate potentially large efficiency gains by reducing the deviations from the law of one price. When the exchange is decentralized, an arbitrage between different trades is costly, and so identical goods can have different prices. In fact, the price of a good depends on the currency used, the country in which it is sold, and the country from which the buyer comes (see Head and Shi, 2003, and Liu and Shi, 2005). This is true even when an agent can use any of the currencies to buy a particular country's goods. Currency unification eliminates the dependence of prices on the type of currency used in the trade, although prices may still vary across locations.

Given these results, international monetary economics seems a fertile ground for future research. First, one can compute numerically the welfare gain from currency unification. Second, one can investigate international business cycles and see how well the theory can explain the stylized facts and puzzles documented by Backus et al. (1995). Third, by introducing country-specific shocks, one can evaluate different exchange regimes and international monetary cooperation.

7. Concluding Remarks

Mainstream monetary economics has used shortcuts to generate a value for money. Because the shortcuts are not derived from fundamental economic principles, they are inappropriate for assessing monetary policy and inadequate for addressing issues regarding competing means of exchange. Starting with Kareken and Wallace (1980), or even further back with Lucas (1972), monetary theorists have been trying to replace these shortcuts with the microfoundations of money. In this lecture I reported the development of a particular theory originated from Kiyotaki and Wright (1989, 1993), which I labeled the value theory of money. By reviewing this development, I hope to have made clear two points about constructing and using a microfoundation of money: it can be done and it makes a difference. The models reviewed here are not more difficult to work with than any of the shortcuts, they uncover new effects of monetary policy, and they suggest quantitatively different results from traditional models. This development should encourage us to take monetary theory seriously.

There might be resistance to this call for theorizing monetary economics, particularly among New Keynesian Macroeconomists. Using the shortcuts to model money (Clarida et al., 1999) or sidestepping money entirely (Woodford, 2003), New Keynesian Macroeconomists have created a tool kit for analyzing monetary policy. Having witnessed the rising popularity of this tool kit, one might think that monetary economics can be agnostic about the microfoundations of money. This thought would be an error for the following reason. An essential ingredient of all New Keynesian models is the Phillips Curve, which links output and unanticipated inflation. To obtain the Phillips Curve, the models must have a role for money. Being agnostic about the microfoundation of money would mean that the models could have no escape from the Lucas (1976) Critique on policy analysis and that their welfare analysis on monetary policy could not be reliable. In contrast, a microfoundation of money that produces the Phillips Curve can instruct us when and how the Phillips Curve shifts around. For example, Lucas' (1972) model tells us that, if the lack of communication between markets on the aggregate money supply is the genesis of the Phillips Curve, then improved communication will increase the slope of the Phillips Curve and reduce the real effect of monetary policy.

If a good theory is important for making good policy, how can we then explain the success of monetary policy in recent years? One task which monetary policy seems to have succeeded in many countries is to bring inflation under control. This success has more to do with the commitment by central banks to simple and transparent policy rules than with the use of New Keynesian Economics. Another claimed success for monetary policy is that the policy has helped to smooth business cycles in the last fifteen years or so. It is too early to judge whether the experience in the last fifteen years is a success or simple luck, because the economy has not had major shocks in that period. Even if the policy did succeed in smoothing cycles, it is important to know why. One cannot get a deep answer by just looking at the Phillips Curve.

Much work lies ahead for the microfoundations of monetary theory. The value theory of money reviewed here is unlikely to be the last word on such microfoundations – just look at how much monetary theory has changed in the last twenty-five years! Wherever the future development may lead, we should resist the temptation to fall back on the shortcuts. If we want monetary economics to be treated as a scientific discipline, we ought to make it so. I believe we can.

References

[1] Backus, David, Kehoe, Patrick and Finn Kydland, 1995, "International business cycles: theory and evidence." In Thomas F. Cooley (ed.): Frontiers of Business Cycle Research, pp. 331-356, Princeton University Press, Princeton, New Jersey.

[2] Benabou, Roland, 1992, "Inflation and efficiency in search markets," Review of Economic Studies 59, 299-329.

[3] Berentsen, Alesander, Camera, Gabriele and Christopher Waller, 2004, "The distribution of money and prices in an equilibrium with lotteries," Economic Theory 24, 887-906.

[4] Berentsen, Alesander, Camera, Gabriele and Christopher Waller, 2005, "Money, credit and banking," manuscript, University of Notre Dame.

[5] Berentsen, Alesander and Guillaume Rocheteau, 2003, "Money and the gains from trade," International Economic Review 44, 263-297.

[6] Berentsen, Alesander, Rocheteau, Guillaume and Shouyong Shi, 2003, "Friedman meets Hosios: efficiency in search models of money," Economic Journal, forthcoming.

[7] Bewley, Truman, 1980, "The optimum quantity of money." In: J.H. Kareken and N. Wallace (eds.), Models of Monetary Economies, Federal Reserve Bank of Minneapolis, Minneapolis, pp 169-210.

[8] Binmore, Kenneth, 1987, "Perfect equilibria in bargaining models." In: Binmore, Kenneth and P. Dasgupta (eds.), The Economics of Bargaining, Basil Blackwell, Oxford.

[9] Brock, William A., 1974, "Money and growth: the case of long-run perfect foresight," International Economic Review 15, 750-777.

[10] Burdett, Kenneth and Kenneth Judd, 1983, "Equilibrium price dispersion," Econometrica 51, 955-970.

[11] Camera, Gabriele and Dean Corbae, 1999, "Money and price dispersion," International Economic Review 40, 985-1008.

[12] Cavalcanti, Ricardo, Erosa, Andres and Ted Temzelides, 1999, "Private money and reserve management in a random matching model," Journal of Political Economy 107, 929-945.

[13] Cavalcanti, Ricardo and Neil Wallace, 1999, "Inside and outside money as alternative media of exchange," Journal of Money, Credit, and Banking 31, 443-457.

[14] Chari, V.V., Kehoe, Patrick J. and Ellen R. McGratten, 2000, "Sticky prices models of the business cycle: can the contract multiplier solve the persistence problem?" Econometrica 68, 1151-1179.

[15] Christiano, Lawrence and Martin Eichenbaum, 1992, "Liquidity effects and the monetary transmission mechanism," American Economic Review 82, 346-353.

[16] Clarida, Richard, Gali, Jordi and Mark Gertler, 1999, "The science of monetary policy: a new Keynesian perspective," Journal of Economic Literature 37, 1661-1707.

[17] Clower, Robert, 1967, "A reconsideration of the microfoundations of monetary theory," Western Economic Journal 6, 1-8.

[18] Cooley, Thomas F. and Gary D. Hansen, 1995, "Money and the business cycle." In: Thomas F. Cooley (ed.), Frontiers of Business Cycle Research, Princeton University Press, Princeton, New Jersey.

[19] Corbae, Dean and Joseph Ritter, 2004, "Decentralized credit and monetary exchange without public record keeping," Economic Theory 24, 933-951.

[20] Corbae, Dean, Temzelides, Ted and Randall Wright, 2003, "Directed matching and monetary exchange," Econometrica 71, 731-756.

[21] Deviatov, A. and Neil Wallace, 2001, "Another example in which lump-sum money creation is beneficial," Advances in Macroeconomics 1.

[22] Diamond, Peter, 1971, "A model of price adjustment," Journal of Economic Theory 2, 156-168.

[23] Diamond, Peter, 1984, "Money in a search equilibrium," Econometrica 52, 1-20.

[24] Engineer, Merwan and Shouyong Shi, 1998, "Asymmetry, imperfectly transferable utility, and fiat money in improving the terms of trade," Journal of Monetary Economics 41, 153-183.

[25] Faig, Miquel and Belen Jerez, 2005, "A theory of commerce," Journal of Economic Theory 122, 60-99.

[26] Freeman, Scott, 1996, "The payments system, liquidity, and rediscounting," American Economic Review 86, 1126-1138.

[27] Friedman, Milton, A Program for Monetary Stability. Fordham University Press, New York, 1960.

[28] Friedman, Milton, 1969, "The optimum quantity of money." In: Friedman, Milton (ed.): The Optimum Quantity of Money and Other Essays. Aldine Publishing Company, Hawthorne, New York, pp. 1-50.

[29] Fudenberg, D. and Eric Maskin, 1986, "The folk theorem in repeated games with discounting or with incomplete information," Econometrica 54, 533-554.

[30] Fuerst, Timonthy, 1992, "Liquidity, loanable funds, and real activity," Journal of Monetary Economics 29, 3-24.

[31] Grandmont, J.M., 1985, "On endogenous competitive business cycles," Econometrica 53, 995-1045.

[32] Green, Edward, "Lending and the smoothing of uninsurable income." In: Edward Prescott and Neil Wallace (eds.), Contractual Arrangements for Intertemporal Trade, University of Minnesota Press, Minneapolis, 1987.

[33] Green, Edward and Ruilin Zhou, 1998, "A rudimentary model of search with divisible money and prices," Journal of Economic Theory 81, 252-271.

[34] Green, Edward and Ruilin Zhou, 2005, "Money as a mechanism in a Bewley economy," International Economic Review 46, 351-371.

[35] Head, Allen and Alok Kumar, 2005, "Price dispersion, inflation, and welfare," International Economic Review 46, 533-572.

[36] Head, Allen, Kumar, Alok and Beverly Lapham, 2005, "Limited participation, market power and inflation dynamics," Manuscript, Queen's University.

[37] Head, Allen and Shouyong Shi, 2003, "A fundamental theory of exchange rates and direct currency trades," Journal of Monetary Economics 50, 1555-1591.

[38] Heller, Walter P. and Ross M. Starr, 1976, "Equilibrium with non-convex transactions costs: monetary and non-monetary economies," Review of Economic Studies 43, 195-215.

[39] Hosios, Arthur, 1990, "On the efficiency of matching and related models of search and unemployment," Review of Economic Studies 57, 279-298.

[40] Hicks, John, 1935, "A suggestion for simplifying the theory of money," Economica 2, 1-19.

[41] Hicks, John, A Market Theory of Money, Oxford University Press, Oxford, 1989, pp.55-63.

[42] Imrohoroglu, Ayse, 1992, "The welfare cost of inflation under imperfect insurance," Journal of Economic Dynamics and Control 16, 79-92.

[43] Jevons, William Stanley, Money and the Mechanism of Exchange, Appleton, London, 1875.

[44] Johri, Alok and John Leach, 2002, "Middlemen and the allocation of heterogeneous goods," International Economic Review 43, 347-362.

[45] Jones, Robert A., 1976, "The origin and development of media of exchange," Journal of Political Economy 84, 757-775.

[46] Kahn, Charles and William Roberds, 2001, "Real-time gross settlement and the costs of immediacy," Journal of Monetary Economics 47, 299-319.

[47] Kareken, John and Neil Wallace (eds.), Models of Monetary Economies, Federal Reserve Bank of Minneapolis, Minneapolis, 1980.

[48] Kareken, John and Neil Wallace, 1981, "On the indeterminacy of equilibrium exchange rates," Quarterly Journal of Economics 96, 2207-222.

[49] Katzman, Brett, Kennan, John and Neil Wallace, 2003, "Output and price level effects of monetary uncertainty in a matching model," Journal of Economic Theory 108, 217-255.

[50] Kimbrough, Kent P., 1986, "Inflation, employment, and welfare in the presence of transaction costs," Journal of Money, Credit, and Banking 18, 127-140.

[51] Kiyotaki, Nobuhiro and Randall Wright, 1989, "On money as a medium of exchange," Journal of Political Economy 97, 927-954.

[52] Kiyotaki, Nobuhiro and Randall Wright, 1991, "A contribution to the pure theory of money," Journal of Economic Theory 53, 215-235.

[53] Kiyotaki, Nobuhiro and Randall Wright, 1993, "A search-theoretic approach to monetary economics," American Economic Review 83, 63-77.

[54] Kocherlakota, Narayana, 1998, "Money is memory," Journal of Economic Theory 81, 232-251.

[55] Kocherlakota, Narayana, 2003, "Societal benefits of illiquid bonds," Journal of Economic Theory 108, 179-193.

[56] Kocherlakota, Narayana and Neil Wallace, 1998, "Incomplete record-keeping and optimal payment arrangements," Journal of Economic Theory 81, 272-289.

[57] Lagos, Ricardo and Randall Wright, 2005, "A unified framework for monetary theory and policy analysis," Journal of Political Economy 113, forthcoming.

[58] Levhari, David and Don Patinkin, 1968, "The role of money in a simple growth model," American Economic Review 58, 713-753.

[59] Li, Yiting, 1999, "Money and middlemen in an economy with private information," Economic Inquiry 37, 1-12.

[60] Li, Yiting, 2001, "A search model of money and circulating private debt with applications to monetary policy," International Economic Review 42, 925-946.

[61] Li, Yiting and Randall Wright, 1998, "Government transaction policy, media of exchange, and prices," Journal of Economic Theory 81, 290-313.

[62] Liu, Qing and Shouyong Shi, 2005, "Currency area and monetary coordination," Manuscript, University of Toronto.

[63] Lucas, Robert, Jr., 1972, "Expectations and the neutrality of money," Journal of Economic Theory 4, 103-124.

[64] Lucas, Robert, Jr., 1976, "Econometric policy evaluation: a critique." In: Brunner, Karl and Allan H. Meltzer (eds.), The Phillips curve and labor markets, Carnegie-Rochester Conference Series on Public Policy 1, 19-46, North-Holland, Amsterdam.

[65] Lucas, Robert E., Jr., 1980, "Equilibrium in a pure currency economy." In: Kareken, John H. and Neil Wallace (eds.), Models of Monetary Economies, Federal Reserve Bank of Minneapolis, Minneapolis, pp. 131-145.

[66] Lucas, Robert E., Jr., 1990, "Liquidity and interest rates," Journal of Economic Theory 50, 237-264.

[67] Lucas, Robert E., Jr., 1992, "On efficiency and distribution," Economic Journal 102, 233-247.

[68] Matsuyama, Kiminori, Kiyotaki, Nobuhiro, and Akihiro Matsui, 1993, "Towards a theory of international currency," Review of Economic Studies 60, 283-307.

[69] Menner, Martin, 2005, "A search-theoretic monetary business cycle model with capital accumulation," manuscript, Universidad Carlos III de Madrid, Spain.

[70] Molico, Miguel, 1998, "The distribution of money and prices in search equilibrium," Ph.D. dissertation, University of Pennsylvania.

[71] Ostroy, Joseph and Ross M. Starr, 1974, "Money and the decentralization of exchange," Econometrica 42, 597-610.

[72] Peterson, Brian and Shouyong Shi, 2004, "Money, price dispersion and welfare," Economic Theory 24, 907-932.

[73] Phelps, Edmund, 1973, "Inflation in the theory of public finance," Swedish Journal of Economics 75, 67-82.

[74] Samuelson, Paul A., 1958, "An exact consumption-loan model of interest with or without the social contrivance of money," Journal of Political Economy 66, 467-482.

[75] Sargent, Thomas and Neil Wallace, 1982, "The real bills doctrine vs. the quantity theory: a reconsideration," Journal of Political Economy 90, 1212-1236.

[76] Shapely, Lloyd S. and Martin Shubik, 1977, "Trade using one commodity as a means of payment," Journal of Political Economy 5, 937-968.

[77] Shevchenko, Andrei, 2004, "Middlemen," International Economic Review 45, 1-24.

[78] Shi, Shouyong, 1995, "Money and prices: a model of search and bargaining," Journal of Economic Theory 67, 467-496.

[79] Shi, Shouyong, 1996, "Credit and money in a search model with divisible commodities," Review of Economic Studies 67, 627-652.

[80] Shi, Shouyong, 1997, "A divisible search model of fiat money," Econometrica 65, 75-102.

[81] Shi, Shouyong, 1998, "Search for a monetary propagation mechanism," Journal of Economic Theory 81, 314-352.

[82] Shi, Shouyong, 1999, "Search, inflation and capital accumulation," Journal of Monetary Economics 44, 81-103.

[83] Shi, Shouyong, 2004, "Liquidity, interest rates and output," manuscript, University of Toronto.

[84] Shi, Shouyong, 2005, "Nominal bonds and interest rates," International Economic Review 46, 579-612.

[85] Sidrauski, Miquel, 1967, "Rational choice and patterns of growth in a monetary economy," American Economic Review, Papers and Proceedings 57, 537-544.

[86] Stockman, Alan C., 1981, "Anticipated inflation and the capital stock in a cash-in-advance economy," Journal of Monetary Economics 8, 387-393.

[87] Temzelides, Ted and Stephen Williamson, 2001, “Payments systems design in deterministic and private information environments,” Journal of Economic Theory 99, 297-326.

[88] Tobin, James, 1965, “Money and economic growth,” Econometrica 33, 671-684.

[89] Townsend, Robert M., 1980, “Models of money with spatially separated agents.” In: Kareken, John H. and Neil Wallace (eds.), Models of Monetary Economies, Federal Reserve Bank of Minneapolis, Minneapolis, pp. 265-303.

[90] Trejos, Alberto and Randall Wright, 1995, “Search, bargaining, money and prices,” Journal of Political Economy 103, 118-141.

[91] Wallace, Neil, 1980, “The overlapping generations model of fiat money.” In: Kareken, John H. and Neil Wallace (eds.), Models of Monetary Economies, Federal Reserve Bank of Minneapolis, Minneapolis, pp. 49-82.

[92] Wallace, Neil, 1983, “A legal restrictions theory of the demand for ‘money’ and the role of monetary policy,” Federal Reserve Bank of Minneapolis Quarterly Review 7, 1-7.

[93] Wallace, Neil, 2000, “A model of the liquidity structure based on asset indivisibility,” Journal of Monetary Economics 45, 55-68.

[94] Wallace, Neil, 2001, “Whither monetary economics,” International Economic Review 42, 847-869.

[95] Wang, Weimin and Shouyong Shi, 2004, “The variability of velocity of money in a search model,” forthcoming in Journal of Monetary Economics.

[96] Williamson, Stephen, 2003, “Payments systems and monetary policy,” Journal of Monetary Economics 50, 475-495.

[97] Williamson, Stephen, 2005, “Monetary policy and distribution,” Manuscript, University of Iowa.

[98] Williamson, Stephen and Randall Wright, 1994, “Barter and monetary exchange under private information,” American Economic Review 84, 104-123.

[99] Woodford, Michael, Interest Rates and Prices: Foundations of a Theory of Monetary Policy, Princeton University Press, Princeton, 2003, pp. 61-138.

[100] Wright, Randall (ed.), Models of Monetary Economies II: the Next Generation, special issue in International Economic Review 46, 2005.

[101] Zhu, Tao, 2003, “Existence of a monetary steady state: indivisible money,” Journal of Economic Theory 112, 307-324.