

Banking: A New Monetarist Approach

CHAO GU

University of Missouri-Columbia

FABRIZIO MATTESINI

University of Rome Tor-Vergata

CYRIL MONNET

University of Bern and Study Center Gerzensee

and

RANDALL WRIGHT

University of Wisconsin - Madison, FRB Minneapolis, and FRB Chicago

First version received August 2010; final version accepted May 2012 (Eds.)

We develop a model where: (i) banks take deposits and make investments; (ii) their liabilities facilitate third-party transactions. Other models have (i) or (ii), not both, although we argue they are intimately connected: we show that they both emerge from limited commitment. We describe an environment, characterize desirable allocations, and interpret the outcomes as banking arrangements. Banks are essential: without them, the set of feasible allocations is inferior. As a technical contribution, we characterize dynamically optimal credit allocations with frictions, show they involve backloading, and analyse how this interacts with banking. We also confront the theory with economic history.

Key words: Banking, Limited Commitment

JEL Codes: D02, E02, E44, G21

1. INTRODUCTION

We develop a theory designed to capture two salient features of banking: (i) banks accept deposits and make investments on behalf of their depositors; (ii) their liabilities, claims on said deposits, facilitate exchange with other parties. While many models of banking consider either the first or second feature, it is desirable to have a framework incorporating both, because the two activities are connected in a fundamental way: as we show, both originate from limited commitment. Of course, banks may do more—*e.g.* provide liquidity insurance and information processing. We downplay these functions, as they have been studied extensively elsewhere, and focus instead on banks arising endogenously as a response to commitment problems. Commitment issues are central because banking concerns the intertemporal allocations of resources, which hinges on the incentive to make good on one's obligations. In our model, banks emerge as agents that are relatively trustworthy, in the sense that they have stronger incentives to honour promises, and this allows claims on deposits to serve as a means of payment—*i.e.* as inside money.

The formal model incorporates the following ingredients. There are two types of infinitely lived agents. Each period is divided into two subperiods, where type 1 wants to consume in the first and type 2 in the second. Type 1 can produce and invest in the first subperiod, thus generating second-subperiod output valued by type 2. In a first-best world, it would be efficient to have type 2 lend to type 1, enabling the latter to consume and invest in the first subperiod, with type 2 consuming the product of the investment in the second. In the second subperiod, however, type 1 is tempted to abscond with the proceeds, as in the cash-diversion models of DeMarzo and Fishman (2007) or Biais *et al.* (2007). This is more of a problem to the extent type 1 is impatient, has better opportunities to divert the proceeds of the investments, and has a smaller probability or a higher cost of getting caught. In general, we need to impose a repayment constraint guaranteeing type 1 does not behave opportunistically, hindering the ability to exploit intertemporal gains from trade.

We then introduce another agent, who is like the first type, but may have less incentive to behave opportunistically, or a higher expected cost of getting caught. Even if this third agent is less efficient than type 1 at producing second-subperiod output, when the incentive problem is severe the following scheme is efficient: type 1 works in the first subperiod and deposits the output with the third agent, who invests on his behalf. Since the third agent is more inclined to deliver the goods, type 2 is willing to produce more in the first subperiod for type 1. This resembles banking: type 1 deposits resources with, and delegates investment to, his banker; and claims on these deposits facilitate transactions between 1 and 2. These liabilities constitute inside money, a role played historically by banknotes, and later checks and debit cards, backed by demand deposits. This arrangement allows type 1 to get more from type 2, compared to pure credit, because the banker is more trustworthy than the type 1 agent. Again, this can be true even if the banker does not have access to the highest-return investment opportunities. The function of formal theory is to make these ideas precise, and hopefully derive new insights. We also put the model to work in several substantive applications, and compare its predictions with some historical observations.¹

Before getting more into results we briefly discuss the methodological approach. Our aim is to proceed with minimal prior assumptions about who banks are or what they do. The agents that become bankers here are not fundamentally different from depositors—*e.g.* in terms of preferences—although they may have less of a problem dealing with certain frictions, like imperfect monitoring and commitment. Obviously some frictions are needed, since models like Arrow-Debreu have no role for banks, or, for that matter, any other institution whose *raison d'être* is to facilitate the process of exchange. The simplest such institution is money, and a classic challenge in monetary economics is to ask what frictions make money essential in the following sense (see, *e.g.* Wallace 2001, 2010): money is said to be essential when the set of feasible allocations is bigger, or better, with it than without it. We similarly want to know when banking is essential.

Here the planner, or mechanism, instructs certain agents to perform functions resembling elements of banking. This activity is essential: if it were ruled out the set of incentive-feasible allocations would be inferior. We call this a *New Monetarist* approach because it is in the tradition of research labelled this way in recent surveys by Williamson and Wright (2010a, b) and Nosal and Rocheteau (2011). Given the existence of these sources, this is not the place to go into detail on the methods or models in that literature, except to say that it focuses on studying the process of exchange explicitly, and on deriving endogenously institutions designed to facilitate that process.

1. The applications are previewed below, but we mention here that we concentrate on issues other than ones that have been the focus of previous research, *e.g.* banks' tendency to borrow short and lend long, or to make deposits available on demand except in unusual circumstances like suspensions. These can be analysed in our framework, too, but we prefer to discuss more novel results, like the role of claims on deposits in the exchange process.

To pick just one example, Kiyotaki and Wright (1989) set out to show that money can emerge endogenously, and is essential, as a medium of exchange that helps deal with frictions making direct barter difficult. Our goal is to think about banking in a similar way.²

Our analysis formalizes the idea that agents are better suited to banking (accepting and investing deposits) when they have a good combination of the following characteristics that make them more trustworthy (less inclined to renege on obligations):

- they are relatively patient;
- they are more visible, by which we mean more easily monitored;
- they have a greater stake in, or connection to, the economic system;
- they have access to better investment opportunities; and
- they derive lower payoffs from opportunistically diverting resources.

Some of these (like the first) points may be obvious, but others seem less so—*e.g.* the idea that it may be good to delegate investments to those with a greater stake in the system, even if they have less lucrative investment opportunities, since this facilitates transactions.

In terms of the banking literature, Gorton and Winton (2002) and Freixas and Rochet (2008) provide surveys. Much of this work is based on information frictions, including adverse selection, moral hazard, and costly state verification. One strand, originating with Diamond and Dybvig (1983), interprets banks as coalitions providing liquidity insurance.³ Another approach, pioneered by Leland and Pyle (1977) and developed by Boyd and Prescott (1986), interprets banks as information-sharing coalitions. A related approach, following Diamond (1984) and Williamson (1986, 1987), interprets banks as delegated monitors taking advantage of returns to scale. Relative to these theories, although monitoring is also part of the story, we focus more on commitment problems. Rajan (1998) previously criticized banking theory for not concentrating more on incomplete contracts, or markets, based on limited enforcement.⁴ We agree that commitment/enforcement issues are central, but we think this needs to be endogenous. In this regard, we build on Kehoe and Levine (1993) and Alvarez and Jermann (2000). We also highlight Cavalcanti and Wallace (1999a, b), where inside money also facilitates trade.⁵ However, that model does not have deposits, delegated investments, loans, or endogenous monitoring. Our approach captures these features as well as the role of bank liabilities in the exchange process.

The rest of the article is organized as follows. Section 2 describes the basic environment. Section 3 discusses feasible and efficient allocations when there is a single group, consisting of two types. This provides a simple model of credit with imperfect commitment, monitoring and collateral, but no banks. Section 4 considers two groups, and shows it can be efficient for some agents in one to act as bankers for the other group. Section 5 shows how to implement efficient

2. Mattesini *et al.* (2009) go into more detail concerning the way this approach relates to the one advocated by Townsend (1987, 1988). This method first lays out an environment, including frictions (*e.g.* information or commitment problems), and then tries to interpret outcomes (*e.g.* incentive-feasible or efficient allocations) in terms of institutions one observes in actual economies. We want to know which frictions lead to banking. For this question, one cannot *assume* missing markets, incomplete contracts, etc., although something like that may *emerge*: “the theory should explain why markets sometimes exist and sometimes do not, so that economic organisation falls out in the solution to the mechanism design problem” (Townsend, 1988). Relatedly, we adhere to a generalization of Wallace’s (1998) dictum: “money should not be a primitive in monetary theory—in the same way that a firm should not be a primitive in industrial organization theory or a bond a primitive in finance.” By extension, banks should not be a primitive in banking theory; they should arise endogenously. See Araujo and Minetti (2011) for recent work in a similar spirit.

3. See Ennis and Keister (2009) for a recent contribution with references.

4. In addition, see Calomiris and Kahn (1991), Myers and Rajan (1998), and Diamond and Rajan (2001).

5. In addition, see Wallace (2005), Koepl *et al.* (2008), Andolfatto and Nosal (2009), Huangfu and Sun (2011), He *et al.* (2005, 2008), Mills (2008), Sanches and Williamson (2010), and Monnet and Sanches (2012).

allocations using inside (bank) money. Section 6 takes up various extensions, to study which agents should be bankers, how many we should have, and how big they should be, as well as asking how we should monitor them when it is costly, what is the tradeoff between trust and return, and why might intermediated lending be useful.

For simplicity, most of the analysis focuses on stationary allocations. Section 7 relaxes this, by considering efficient dynamic allocations, and shows the main economic conclusions survive. These results probably constitute the biggest technical contribution of the article—which is for the most part meant to be as simple as we could make it—since it was not trivial (at least to us) to derive the properties of efficient dynamic allocations, even if the general approach goes back a long way (at least to Thomas and Worall, 1988). In terms of substantive findings, we show that efficient credit allocations involve backloading rewards for borrowers—*e.g.* offering small loans at relatively high interest at the beginning, followed by larger loans at better terms in the longer run. This provides the greatest incentives for borrowers to repay. Given all of this theory, Section 8 compares the predictions of the model to some facts from banking history. Section 9 concludes.

2. THE ENVIRONMENT

Time is discrete and continues forever. There are two *groups*, $i = a, b$, each with a $[0, 1]$ measure of agents. In a group, agents can be one of two *types*, $j = 1, 2$.⁶ We refer to agents of type j in group i as j^i (*e.g.* 1^a is a type 1 agent in group a). Each period, each agent can be active or inactive. Inactive agents do not consume or produce, and get utility normalized to 0, in that period. Agents in group i are active with probability γ^i and inactive with probability $1 - \gamma^i$, where γ^i can differ across groups, so that they can have different degrees of connection to the economic system. In each period there are two group-specific goods, 1^i and 2^i , $i = a, b$. What defines a group is that agents have utility functions with only goods produced in their group as arguments. Active agents 1^i consume good 1^i and produce good 2^i , while active agents 2^i consume good 2^i and produce good 1^i . Letting x_j and y_j denote consumption and production by agents j , utility $U^j(x_j, y_j)$ is increasing in x_j , decreasing in y_j and satisfies the usual differentiability and curvature conditions. We assume $U^j(0, 0) = 0$, normal goods, and a discount factor across periods $\beta \in (0, 1)$.

Each period is divided into two subperiods, and good j must be consumed in subperiod j . Thus, type 1^i agents must consume before 2^i , making credit necessary. To have a notion of collateral, good 2^i is produced in the first subperiod, and invested by either type 1^a or 1^b , with fixed gross return ρ^a or ρ^b in terms of second-subperiod goods (there is no investment across periods, only across subperiods). This may be as simple as pure storage, perhaps for safekeeping, or any other investment; merely for ease of presentation do we impose a fixed return. To generate gains from trade in a simple way, type 2^i agents cannot invest for themselves; more generally, we could let them invest, just not as efficiently. Also, type 1^i agent can invest goods produced in either group. An important friction is this: when type 1^i agents are supposed to deliver the goods, in the second subperiod, they can renege to obtain a payoff λ^i per unit of diverted resources, over and above $U^1(x_1, y_1)$. If $\lambda^i = 0$, investment constitutes perfect collateral, since type 1^i has no gain from renegeing when the production cost is sunk. But if $\lambda^i > 0$, there is an opportunity cost to delivering the goods.

Formally, diversion can be interpreted as type 1 consuming the investment returns, but this is meant to stand in for the general idea that investors can extract resources opportunistically. Also,

6. Types are permanent. The main results go through when agents are randomly assigned types each period, but the analysis is messier (Mattesini *et al.*, 2009).

when we say utility for agents is defined only over goods produced in their own group, we mean U^i is only a function of these goods. Type 1^i also gets a payoff $\lambda^i y$, over and above U^i , from absconding with y units of the proceeds of investments of goods from either group. This is the key incentive issue in the model.⁷ We assume $U^1(x_1, y_1) + \lambda^1 \rho^1 y_1 \leq U^1(x_1, 0)$ for all x_1 and y_1 , so that *ex ante* it is never efficient for type 1^i to produce and invest for their own consumption; they only consider consuming the proceeds opportunistically *ex post*. In this setup, by design, any trade or other interaction across groups is only interesting for its incentive effects, not for more standard mercantile reasons. Also, we point out that there is no outside money here, so that we can concentrate more clearly on inside money in what follows.

Although Section 7 generalizes this, for now we focus on stationary symmetric allocations. These are given by vectors $(x_1^i, y_1^i, x_2^i, y_2^i)$ for each group i , plus descriptions of cross-group transfers, investment, and diversion. We sometimes proceed as if a planner collects production and allocates it to investors and consumers, but this is only to ease the presentation—all the planner really does is make suggestions concerning actions or allocations. When there are no transfers across groups or diversion, allocations are resource feasible if $x_1^i = y_2^i$ and $x_2^i = \rho^i y_1^i$, so we can summarize these allocations by (x_1^i, y_1^i) . When there is no ambiguity we drop the subscript and write (x^i, y^i) . Finally, the planner/mechanism has an imperfect monitoring technology: a deviation from suggestions in group $i = a, b$ is detected with probability π^i , and goes undetected with probability $1 - \pi^i$. This random monitoring technology differs across groups to capture the idea that some are more visible than others, and thus, presumably, less inclined to misbehave.⁸

3. ONE GROUP

With a single group we drop the superscript i . Now all a planner/mechanism can do is recommend a resource-feasible allocation (x, y) for agents in the group. This recommendation is incentive feasible, or IF, as long as no one wants to deviate. Although we focus on the case where agents cannot commit to future actions, suppose as a benchmark they can commit to some degree. One notion is full commitment, at the beginning of time, even before agents know their type, chosen at random before production, exchange, and consumption commence. Then (x, y) is IF as long as the total surplus is positive,

$$S(x, y) \equiv U^1(x, y) + U^2(\rho y, x) \geq 0. \quad (1)$$

Another notion is partial commitment, where agents can commit at the beginning but only after knowing types. Then IF allocations entail two participation constraints:

$$U^1(x, y) \geq 0 \quad (2)$$

$$U^2(\rho y, x) \geq 0. \quad (3)$$

The situation in which we are actually interested involves no commitment. This means that, at the start of *every* period there are two participation conditions:

$$U^1(x, y) + \beta V^1(x, y) \geq (1 - \pi) \beta V^1(x, y) \quad (4)$$

$$U^2(\rho y, x) + \beta V^2(x, y) \geq (1 - \pi) \beta V^2(x, y), \quad (5)$$

7. The introduction of $\lambda^i > 0$ is motivated by the idea that, although investment acts as collateral, as Ferguson (2008) puts it “Collateral is, after all, only good if a creditor can get his hands on it.”

8. There are several ways to interpret random monitoring, but a straightforward one is to assume imperfect record keeping: information concerning deviations simply “gets lost” with probability $1 - \pi^i$ across periods. Later we make monitoring endogenous.

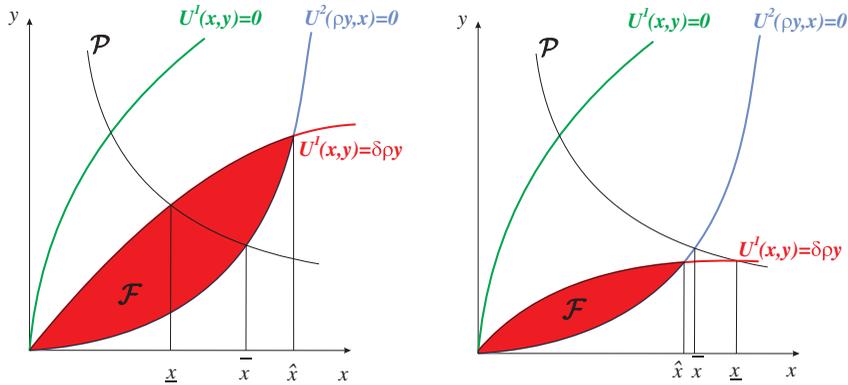


FIGURE 1
Incentive feasible allocations

where $V^j(x, y)$ is the continuation value of agent j . In (4) and (5) the LHS is the payoff from following the recommendation, while the RHS is the deviation payoff.⁹ A deviation is detected with probability π , which results in a punishment to future autarky with payoff 0 (one could consider weaker punishments but this is obviously the most effective). With probability $1 - \pi$ it goes undetected and hence unpunished. Since agents are active with probability γ each period, $V^1(x, y) = \gamma U^1(x, y) / (1 - \beta)$ and $V^2(x, y) = \gamma U^2(\rho y, x) / (1 - \beta)$. From this it is immediate that (4) and (5) hold if and only if (2) and (3) hold, so dynamic considerations for now only involve happenings across subperiods within a period.

When agent 1 invests y , he promises to deliver ρy units of good 2 in the next subperiod, but can always renege for a short-term gain $\lambda \rho y$. If caught, he is punished with future autarky, and so he delivers the goods only if

$$\beta V^1(x, y) \geq \lambda \rho y + (1 - \pi) \beta V^1(x, y),$$

where the RHS is the payoff to behaving opportunistically, again detected with probability $1 - \pi$. Inserting $V^1(x, y)$ and letting $\delta \equiv \lambda(1 - \beta) / \pi \gamma \beta$, this reduces to

$$U^1(x, y) \geq \delta \rho y. \tag{6}$$

As shown in Figure 1, the *repayment constraint* (6) is a clockwise rotation of 1's participation constraint. This plays a prominent role in the sequel. A low β , low monitoring probability π , low stake in the system γ , or high diversion value λ all increase δ and the temptation to default. We say an agent is more *trustworthy* when he has smaller δ , which means he can credibly promise more, and therefore would have a bigger credit limit.

The IF set with no commitment is denoted

$$\mathcal{F} = \{(x, y) \mid (3) \text{ and } (6) \text{ hold}\}.$$

Clearly, \mathcal{F} is convex, compact, and contains points other than $(0, 0)$, so there are gains from trade, under the usual Inada conditions. For comparison, the IF set with partial commitment \mathcal{F}_P satisfies

9. At the suggestion of a referee, we mention that although we do not explicitly define a formal game, we can still use methods from game theory, including the one-shot deviation principle—which, for our purposes, is nothing more than the unimprovability principle of dynamic programming.

(2) and (3), and with full commitment \mathcal{F}_F only requires (1). Notice $\mathcal{F} \subset \mathcal{F}_P \subset \mathcal{F}_F$. In Figure 1 (\hat{x}, \hat{y}) is the unique point other than $(0, 0)$ where (3) and (6) intersect, and $\delta^b < \delta^a$ implies (\hat{x}^b, \hat{y}^b) is northeast of (\hat{x}^a, \hat{y}^a) . A more general result, also easy to verify, is:

Lemma 1. *If $\delta^a > \delta^b$ and $\rho^a = \rho^b$ or $\delta^a = \delta^b$ and $\rho^a < \rho^b$ then $\mathcal{F}^a \subset \mathcal{F}^b$.*

One can define various notions of allocations that are Pareto optimal, or PO. The *ex ante* PO allocation is the (x^o, y^o) that maximizes the surplus $S(x, y)$. A natural criterion for *ex post* (conditional on type) welfare, which we use below, is

$$W(x, y) = \omega_1 U^1(x, y) + \omega_2 U^2(\rho y, x). \quad (7)$$

As we vary the weights ω_i in (7) we get the contract curve, or Pareto set,

$$\mathcal{P} = \left\{ (x, y) \mid \rho \frac{\partial U^1(x, y)}{\partial x} \frac{\partial U^2(\rho y, x)}{\partial y} = \frac{\partial U^2(\rho y, x)}{\partial x} \frac{\partial U^1(x, y)}{\partial y} \right\}. \quad (8)$$

It is possible that $\mathcal{P} \cap \mathcal{F} = \emptyset$ or $\mathcal{P} \cap \mathcal{F} \neq \emptyset$. Other useful results, also simple enough to state without proofs, are given by:

Lemma 2. *Given normal goods, \mathcal{P} defines a downward-sloping curve in (x, y) space.*

Lemma 3. *Let (x', y') solve $\max W(x, y)$ subject to $(x, y) \in \mathcal{F}$. Then $(x', y') \in \mathcal{P}$ iff (6) is not binding.*

4. MULTIPLE GROUPS

Consider two groups a and b , where $\rho^a = \rho^b = \rho$, but $\delta^a > \delta^b$ so that 1^a have more of a commitment problem than 1^b . The IF set for the economy as a whole is given by allocations (x^i, y^i) for each group, plus descriptions of interactions across groups, as we now discuss. Consider first a pure transfer t : all 1^b agents produce an extra $t > 0$ units of good 2^b and give it to agents 1^a , who invest it and use the proceeds for their own benefit.¹⁰ Since there are γ^b/γ^a active 1^b agents for each active 1^a , payoffs are

$$\hat{U}^1(x^a, y^a, t) \equiv U^1(x^a, y^a) + \lambda^a \rho t \gamma^b / \gamma^a \quad (9)$$

$$\hat{U}^1(x^b, y^b, t) \equiv U^1(x^b, y^b + t). \quad (10)$$

We need to analyse transfers for the following reason. We are ultimately interested in a different scheme, where output from one group is transferred to the other group to invest, with the proceeds transferred back. This delegated investment activity can change the IF set, but so can pure transfers. To make the point that delegated investment can do more we must first analyse transfers.

With $t > 0$, the participation conditions for 2^i are as before,

$$U^2(\rho y^i, x^i) \geq 0, \quad i = a, b, \quad (11)$$

10. Transfers in the other direction, from 1^a to 1^b , are given by $t < 0$, and it is never useful to have simultaneous transfers in both directions. Note that t is like a lump sum tax on 1^b , with the proceeds going to 1^a , except it is not compulsory: 1^b can choose to not pay t , at the risk of punishment to future autarky, which happens with probability π^b .

but the repayment constraints for 1^i change to¹¹

$$\hat{U}^1(x^i, y^i, t) \geq \delta^i \rho y^i, i = a, b. \quad (12)$$

The IF set with $t > 0$ satisfies (11) and (12). Notice t only enters these conditions through $\hat{U}^1(x^i, y^i, t)$. Thus, when it comes time to settle obligations, t affects the continuation values for 1^a and 1^b , but not short-run temptations. Since agents 1^a are better off and agents 1^b worse off with $t > 0$, this relaxes the repayment constraints in group a and tightens them in group b . Therefore, if these constraints are binding in group a but not b , this expands the IF set.

To see just how much we can accomplish with transfers, consider the biggest t satisfying (11) and (12). This maximization problem has a unique solution \tilde{t} , and implies allocations $(\tilde{x}^i, \tilde{y}^i)$. Clearly, \tilde{t} rises as δ^b falls. Suppose, e.g. $U^1(x, y) = x - y$, $U^2(\rho y, x) = u(\rho y) - x$, and to make the case stark $\lambda^b = 0$. Then IF allocations in group b solve

$$u(\rho y^b) - x^b \geq 0 \quad (13)$$

$$x^b - y^b - t \geq 0. \quad (14)$$

The maximum t and the implied allocation for group b are given by $\tilde{y}^b = y^*$, $\tilde{x}^b = u(\rho y^*)$ and $\tilde{t} = u(\rho y^*) - y^*$, where $\rho u'(\rho y^*) = 1$. In this example, production by 1^b is efficient, 2^b give all their surplus to 1^b , and \tilde{t} taxes away the entire surplus of group b . Giving \tilde{t} to 1^a relaxes their repayment constraint as much as is feasible (any $t > \tilde{t}$ leads to defection in group b).

We now introduce deposits, $d > 0$, production of good 2^a by 1^a transferred to 1^b for investment, then transferred back for consumption by 2^a . Since 1^a is now only obliged to pay out $\rho(y^a - d)$ in the second subperiod, his repayment constraint becomes¹²

$$\hat{U}^1(x^a, y^a, t) \geq \delta^a \rho (y^a - d). \quad (15)$$

Similarly, since 1^b is now obliged to pay out $\rho(y^b + d\gamma^a/\gamma^b)$,

$$\hat{U}^1(x^b, y^b, t) \geq \delta^b \rho (y^b + \gamma^a d / \gamma^b). \quad (16)$$

We also face a resource constraint

$$0 \leq d \leq y^a. \quad (17)$$

The IF set with deposits \mathcal{F}_d is given by an allocation (x^i, y^i) for each group i , together with t and d , satisfying (11) and (15) and (17).

We relax the repayment constraint in group a while tightening it in group b with $d > 0$, as we did with $t > 0$, but deposits and transfers are different in the way they impact incentives: t only affects continuation values, while d affects directly within-period temptations to renege. This implies that deposits are essential in the technical sense: if we start with $d = 0$, and then allow $d > 0$, for some parameters, the IF set expands.¹³ The following states this formally:

11. In case it is not clear, (12) is the condition for 1^i to pay off 2^i (i.e. agents in their own group). For 1^a who is meant to divert the returns from t , this can be written

$$\lambda^a \rho t \gamma^b / \gamma^a + \beta \gamma^a \hat{U}^1(x^a, y^a, t) / (1 - \beta) \geq \lambda^a \rho (t \gamma^b / \gamma^a + y^a) + (1 - \pi) \beta \gamma^a \hat{U}^1(x^a, y^a, t) / (1 - \beta),$$

which simplifies to (12).

12. Notice that these conditions allow transfers as well as deposits, since they use the payoffs defined in (9) and (10).

13. We are not claiming $\mathcal{F} \subset \mathcal{F}_d$ for all parameters, or for any $d > 0$; the claim is that deposits can be essential for some parameters when we get to choose d .

Proposition 1. For all parameters $\exists d$ such that $\mathcal{F} \subset \mathcal{F}_d$; for some parameters $\exists d$ such that $\mathcal{F}_d \setminus \mathcal{F} \neq \emptyset$.

Proof Since any allocation in \mathcal{F} can be supported once deposits are allowed by setting $d=0$, it is trivial that $\mathcal{F} \subset \mathcal{F}_d$. To show more allocations may be feasible with deposits, it suffices to give an example. To make it easy, suppose $\lambda^b=0$, so that holding deposits does not affect the repayment constraint in group b . Then there are some allocations for group a that are only feasible with $d > 0$. To see this, set $t = \tilde{t}$ to maximize the transfer from b to a , as discussed above. Given $(x^b, y^b, t) = (\tilde{x}^b, \tilde{y}^b, \tilde{t})$, all incentive constraints are satisfied in group b . In group a , the relevant conditions (11) and (15) are

$$U^2(\rho y^a, x^a) \geq 0$$

$$\hat{U}^1(x^a, y^a, \tilde{t}) \geq \delta^a \rho (y^a - d).$$

For any allocation such that $\delta^a \rho y^a \geq \hat{U}^1(x^a, y^a, \tilde{t})$, $d > 0$ relaxes the repayment constraint for 1^a , and hence expands the IF set. \parallel

While the example in the proof uses $\lambda^b=0$, this is not necessary. Suppose for both groups $U^1(x, y) = u(x) - y$, $U^2(y, x) = y - x$, $\rho = 1$, $\delta^i = \delta$ and $\lambda^i = \lambda \in (0, 1)$, and consider maximizing welfare with weights $\omega_1^i = 1$ and $\omega_2^i = 0$. We can summarize \mathcal{F}_d when b makes transfers t and accepts deposits d by pairs (x^a, x^b) satisfying

$$u(x^a) - x^a + \lambda t \geq \delta(x^a - d) \quad (18)$$

$$u(x^b) - x^b - t \geq \delta(x^b + d); \quad (19)$$

and symmetrically for group a . To show deposits expand \mathcal{F} beyond what can be achieved with transfers, notice $t > 0$ relaxes (18) by λt and tightens (19) by t , while $d = \lambda t / \delta$ relaxes (18) by the same amount and only tightens (19) by $\lambda t < t$. To obtain the same level of slack in one group, deposits require less tightening in the other. Figure 2 shows feasible outcomes in (x^a, x^b) space for three cases. If $t = d = 0$, \mathcal{F} is the square. Using transfers but no deposits, it expands by the dark-shaded areas. Using deposits as well further expands \mathcal{F} to also include the light-shaded areas.

The economic intuition is simple. Suppose that you are 1^a , and want goods from 2^a in exchange for a pledge to pay him back later. When the time comes to pay up, if $\lambda^a > 0$ you are tempted to renege, opportunistically diverting the resources that were earmarked for settlement. This limits your credit. Your temptation is relaxed by depositing $d > 0$ with 1^b to invest on your behalf. Of course, we must also consider 1^b 's temptation, but generally, whenever $\delta^b < \delta^a$, deposits allow you to get more from 2^a than a personal pledge. As a special case, $\lambda^b = 0$ means 1^b is totally credible—or, his investments constitute perfect collateral—so you may as well deposit all your resources with him. Another case of interest is $\pi^a = 0$, where your personal promise is worth nothing, and therefore absent deposits you cannot trade at all. In any event, we think it is accurate to call 1^b your banker, and we expand on this idea in the next two sections.¹⁴

14. Although we find the applications in Sections 5–6 interesting, this article is written so that readers can skip any or all of this material, and proceed directly to the analysis of optimal dynamic allocations in Section 7, if so inclined, with no loss of conceptual continuity.

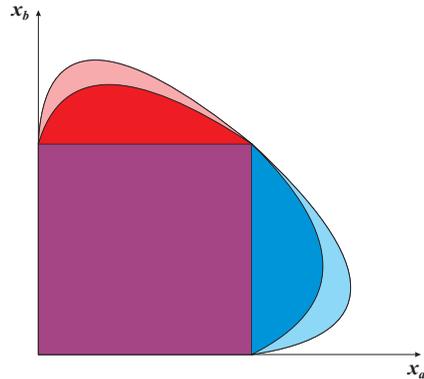


FIGURE 2

Incentive feasible allocations with transfers and deposits

5. INSIDE MONEY

Having deposits used in payments is imperative for a complete model of banking, as over time various bank-issued instruments have played this role, from notes to checks to debit cards. This is one of the most commonly understood functions of banking, as evidenced by Selgin's (2007) entry on "Banks" for *Encyclopedia Britannica*:

Genuine banks are distinguished from other kinds of financial intermediaries by the readily transferable or "spendable" nature of their IOUs, which allows those IOUs to serve as a means of exchange, that is, money. Commercial bank money today consists mainly of deposit balances that can be transferred either by means of paper orders known as checks or electronically using plastic "debit" cards.

Most formal banking models fail to speak to this issue. In Diamond–Dybvig, *e.g.* agents with a desire to consume withdraw deposits and eat them. Presumably this is not meant to be taken literally, but stands in for the idea that they want to buy something. But why cannot they buy it using claims on deposits as a means of payment? We understand that the model is not intended to address this question. But then, as useful as it may be for some purposes, the model is incomplete. Inside money does aid in transactions in Calvo–Wallace, but that has nothing that resembles deposits or investment. A complete model ought to have both.

We proceed by first presenting a relatively heuristic discussion, then give the equations to make the ideas precise. The question with which we begin is, how can a mechanism keep track of agents' actions in the arrangement discussed above? One way that seems especially appealing when record keeping is imperfect or costly is the following: when an agent 1^a wants to consume in the first subperiod, he produces and deposits output y^a with an agent 1^b in exchange for a receipt. Think of this receipt as a *bearer note* for ρy^a . He gives this note to 2^a in exchange for his consumption good x^a . Naturally, 2^a accepts it, since the note is backed by the promise of the trustworthy 1^b . Agent 2^a carries this note to the second subperiod, when he wants to consume (here he wants to consume with probability 1, but it is not hard to make it random, if one wants the setup to look more like Diamond–Dybvig). At that point 2^a redeems the note for his consumption good. Banker 1^b pays 2^a out of deposits—principle plus return on investments—to clear, or settle, the obligation. See Figure 3.¹⁵

15. Another scheme, suggested by Chris Phelan, that one might consider is this: suppose 1^a gives his output directly to 2^a who then gives it to 1^b to invest. This has deposits and delegated investment, but not inside money. One can rule

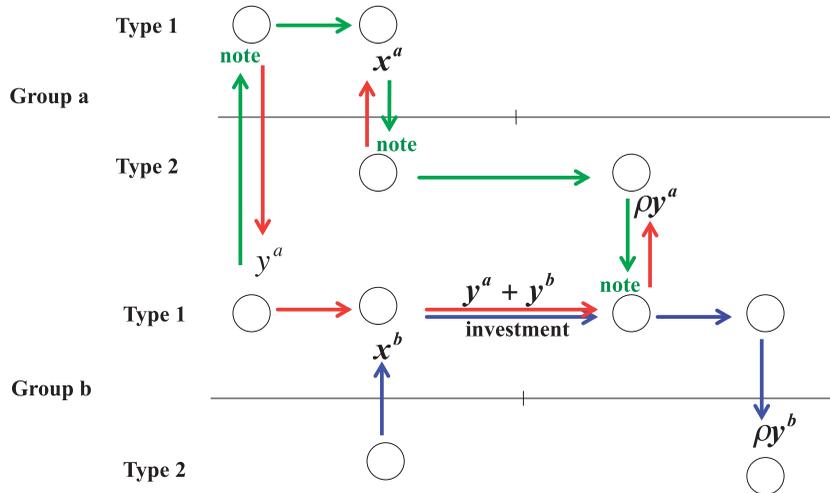


FIGURE 3
Inside money

To make this more precise, we need to be specific about how agents meet and what is observed. Therefore, we now explicitly interpret groups a and b as inhabiting different locations, or islands, to use a metaphor common in the search literature. To ease the exposition, let $\gamma^i = \rho^i = 1$, $U^1 = u(x) - y$ and $U^2 = u(y) - x$. Also, let $\pi^a = 0 < \pi^b$. To discuss circulating paper, we assume that any agent can costlessly produce indivisible, durable, intrinsically worthless objects that could, in principle, function as bearer notes. To avoid technical details, we assume agents can store at most one note. This does not affect the interesting results, but it means we do not have to rule out potential deviations where agents accumulate multiple notes over time, and cash them in bundles—which they do not want to do, anyway, but it clutters the presentation to have to prove it. In passing, we mention that this is actually an assumption shared with Cavalcanti–Wallace, as well as many of the earlier search-based models of outside money.

Meetings occur as follows. Within each group j , each agent 1^j is randomly matched with one 2^j for the entire period. We know from standard arguments (*e.g.* Wallace, 2010) that some medium of exchange is necessary for trade on island a , given $\pi^a = 0$, but notes issued by 1^a have no value because 1^a has no incentive to redeem them. And these notes cannot have value as fiat objects, since no one would give up anything to get one when he can print his own for free. This is not the case for notes issued by 1^b . In addition to the above matching structure, before 1^j and 2^j pair off, 1^a agents travel to island b , and meet some 1^b at random. Then, in subperiod 2, agent 2^a can travel to island b and meet anyone they like—*i.e.* search by 2^a is directed. This means that 2^a takes the note back to the *same* 1^b agent that issued it. Alternatively, one could imagine setting up the model so that type 1^b agents redeem notes issued by *any* bank, as in Cavalcanti *et al.* (1999); this may have interesting consequences, but for now we take the route of directed search, along the lines of Corbae *et al.* (2003). In any case, for completeness we have to specify what happens if $n > 1$ agents of type 2^a try to match with the same 1^b agent. In this case, we assume that all n have the same probability $1/n$, but only one actually meets (can trade with) him.

that out, however, by assuming 2^a cannot transport first-subperiod goods, just like they cannot invest them. This makes the receipts or bearer notes essential.

Now consider a simple mechanism that suggests a particular set of actions—basically, meetings and trades—but agents can either accept or reject suggestions. If in a meeting both agents accept, they implement a suggested trade; otherwise, there is no trade in the meeting. If someone rejects a suggested trade, as above, with probability π^i they are punished with future autarky. There are four types of trades we need to consider: (1) when 1^a meets 1^b , the former should produce and deposit $d = y^a$ in exchange for the latter's note; (2) when 1^a meets 2^a , the latter should produce x^a in exchange for a note if the former has one, and otherwise there is no trade; (3) when 2^a meets the 1^b who issued the note, the latter redeems it for y^a ; and (4) within group b , x^b is produced by 2^b for 1^b in the first subperiod and y^b is delivered to 2^b in the second, as in the previous sections, without using notes.

To describe payoffs, let $\tilde{v}_1^a(m)$ be the expected utility of 1^a when he meets 1^b and $v_1^a(m)$ his expected utility when he meets 2^a , given he has $m \in \{0, 1\}$ notes. Then

$$\begin{aligned}\tilde{v}_1^a(0) &= v_1^a(1) - y^a \\ \tilde{v}_1^a(1) &= v_1^a(1) \\ v_1^a(1) &= u(x^a) + \beta \tilde{v}_1^a(0) \\ v_1^a(0) &= \beta \tilde{v}_1^a(0).\end{aligned}$$

Thus, if 1^a has $m=0$ when he meets 1^b , he produces/deposits y^a in exchange for a note, while if $m=1$ they do not trade. Then, when 1^a meets 2^a , if $m=1$ he swaps the note for x^a , while if $m=0$ he leaves without consuming, and in either case starts next period with $m=0$. Similarly, for 2^a

$$\begin{aligned}v_2^a(0) &= \tilde{v}_2^a(1) - x^a \\ v_2^a(1) &= \tilde{v}_2^a(1) \\ \tilde{v}_2^a(1) &= u(y^a) + \beta v_2^a(0) \\ \tilde{v}_2^a(0) &= \beta v_2^a(0),\end{aligned}$$

where $v_2^a(m)$ is the payoff when 2^a has m notes and meets 1^a , while $\tilde{v}_2^a(m)$ is the payoff when he meets 1^b . Since $\pi^a=0$, on island a the relevant incentive conditions are $v_1^a(1) - y^a \geq v_1^a(0) \geq 0$ and $\tilde{v}_2^a(1) - x^a \geq \tilde{v}_2^a(0) \geq 0$, which reduce to

$$u(x^a) \geq y^a \tag{20}$$

$$u(y^a) \geq x^a. \tag{21}$$

Let \tilde{v}_1^b be the payoff of 1^b , a representative banker, when he meets 1^a , and v_1^b his payoff when he meets 2^b in the first subperiod. Let $\hat{v}_1^b(a)$ be his payoff when he meets 2^a and $\hat{v}_1^b(b)$ his payoff when he meets 2^b in the second subperiod. Then

$$\begin{aligned}\tilde{v}_1^b &= v_1^b = u(x^b) - y^b + \hat{v}_1^b(a) \\ \hat{v}_1^b(a) &= \hat{v}_1^b(b) = \beta \tilde{v}_1^b.\end{aligned}$$

The important decision for 1^b is repayment. If he reneges on either 2^a or 2^b , he is detected and punished with probability π^b . But 2^a only gets y^a if he gives 1^b one of his notes; otherwise, the mechanism says 1^b can use the resources for his own benefit $\lambda^b y^a$. It is this part of the

implementation scheme that gives 2^a the incentive to produce for a note in the first place. The payoff for 2^b is $v_2^b = u(y^b) - x^b + \beta v_2^b$, which we need to consider, of course, since we need 2^b to produce for 1^b —this is how we give 1^b an equilibrium payoff that provides him with the incentive to honour his obligations. Therefore, the incentive conditions for group b reduce to

$$u(y^b) \geq x^b \quad (22)$$

$$u(x^b) \geq y^b \quad (23)$$

$$u(x^b) - y^b \geq \delta^b (y^b + y^a). \quad (24)$$

Given the utility functions used here, (22) and (23) are the participation constraints, and (24) is the repayment constraint, for group b described earlier. Similarly, for group a (20) and (21) are the participation constraints, and there is no repayment constraint given $\pi^a = 0$. Summarizing, we have:

Proposition 2. *Any (x^a, y^a) and (x^b, y^b) satisfying (20)–(24) can be decentralized using banknotes. Since these same constraints define the IF set, any IF allocation can be decentralized in this way.*

In terms of economics, deposit-backed paper issued by b is used as a payment instrument by a , which is essential since $\pi^a = 0$ implies there is no trade on island a when $d = 0$. In equilibrium, all transactions on island a are now spot trades of goods for notes—*i.e.* the economy has been fully *monetized*. This is related to the use of currency in models like Kiyotaki and Wright (1993), except we use bank liabilities rather than fiat objects. The difference from most banking theory is that our bank liabilities facilitate transactions. The difference from Calvo–Wallace banks is that our banks do more than issue notes, they also take deposits and make investments.

6. EXTENSIONS AND APPLICATIONS

Having shown that banking is essential, and that IF allocations can be decentralized using deposit-backed notes, we take up other questions. For this, we use welfare criterion (7), with the same weights in each group: $\omega_j^a = \omega_j^b$.

6.1. Who should hold deposits?

With two groups and $\delta^a > \delta^b$, we claim it may be desirable in a Pareto sense to have group a deposit resources with group b , but not the other way around. Let

$$(x_0^i, y_0^i) = \arg \max W^i(x^i, y^i) \text{ s.t. } (x^i, y^i) \in \mathcal{F}^i \quad (25)$$

be the best IF allocation in group i with no transfers or deposits. At (x_0^i, y_0^i) , we obviously cannot make type 1^i better off without making type 2^i worse off, or vice-versa, given $d = 0$. So, we ask, can we make agents in one group better off without hurting the other group with $d \neq 0$? If deposits can help, in this regard, we say they are *Pareto essential*, or PE.¹⁶

16. Essentiality means the IF set becomes bigger or better. By PE, we mean better, according to (25). Also note that transfers cannot help here, since they make one group worse off, so we ignore them.

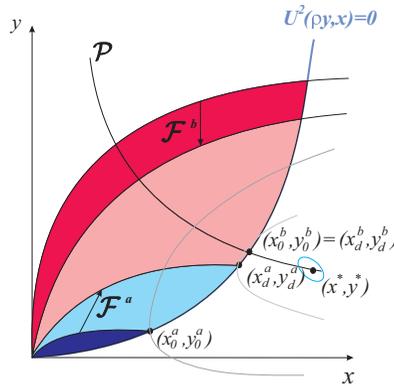


FIGURE 4
Pareto essential deposits

Consider the allocation that, for some d , solves

$$(x_d^i, y_d^i) = \arg \max W^i(x^i, y^i) \text{ s.t. } (x^i, y^i) \in \mathcal{F}_d^i \tag{26}$$

where, compared to (25), the constraint set in (26) is now \mathcal{F}_d^i . Deposits are PE if there is d such that $W^i(x_d^i, y_d^i) \geq W^i(x_0^i, y_0^i)$ for both i with one strict inequality.

Proposition 3. *Deposits from a to b are PE iff the repayment constraint binds for a and not b .*

We omit a formal proof (see Mattesini *et al.*, 2009), but the idea is simple: if the repayment constraint binds in one group, but not the other, bankers should be selected from the latter. Suppose, e.g. $\mathcal{F}^a \subset \mathcal{F}^b$. Then, since the welfare weights are the same, if the repayment constraint does not bind in group a at (x_0^a, y_0^a) then it cannot bind in b . This is shown in Figure 4 for a case in which the best unconstrained allocation (x^*, y^*) is not feasible in either group. When $d=0$, $(x_0^b, y_0^b) \in \mathcal{P}$ solves (25) for group b , but the commitment problem is so severe in group a that $(x_0^a, y_0^a) \notin \mathcal{P}$. Introducing $d > 0$ as a fraction of y rotates in the repayment constraint for group b and rotates out the one for group a . This has no effect on group b , since (x_0^b, y_0^b) is still feasible, but makes group a better off.

6.2. How should we monitor?

We now choose monitoring intensity, and thus endogenize δ^i . Assume monitoring in group i with probability π^i implies a utility cost $\pi^i k^i$. Define a new benchmark with $d=0$ as the solution (x^i, y^i, π^i) to

$$\max_{(x, y, \pi)} W^i(x, y) - \pi k^i \text{ s.t. } (x, y) \in \mathcal{F}^i \text{ and } \pi \in [0, 1]. \tag{27}$$

Notice the repayment constraint must bind, $U^1(x^i, y^i) = \delta^i \rho y^i$, since otherwise we could reduce monitoring costs. Also notice that (x^*, y^*) is generally not efficient when monitoring is endogenous, since reducing π implies a first-order gain while moving away from (x^*, y^*) entails only a second-order loss.

Suppose for the sake of illustration that we want to minimize total monitoring costs. For now, assume there is exactly one active type 1 agent, which means there is a single candidate banker, in

each group at each date. Obviously, if agents in one group deposit with the other, we can reduce monitoring costs in the former only by increasing it in the latter. Still, this may be desirable. In the Appendix we prove that if $\gamma^b \geq \gamma^a$, $\lambda^b \leq \lambda^a$ and $k^b \leq k^a$, then $d > 0$ may be desirable but $d < 0$ can never be. Also, we show that when 1^b has a big enough stake in the economy, he should hold all the deposits, so that we can give up monitoring agents 1^a entirely.

Proposition 4. *Fix (x^a, y^a) and (x^b, y^b) . If $\gamma^b \geq \gamma^a$, $\lambda^b \leq \lambda^a$, and $k^b \leq k^a$, then efficient monitoring implies $\delta^b < \delta^a$. Also, if $\gamma^b > \bar{\gamma}$, defined in the proof, then $\pi^a = 0$.*

Other applications of endogenous monitoring are available (see Mattesini *et al.*, 2009). One can show, *e.g.* that $d > 0$ may be desirable even if 1^a must compensate 1^b for increased monitoring costs. One can also consider the efficient number of bankers, more generally. Fewer bankers reduce total monitoring cost, but this means more deposits per bank, so that we might need to monitor them more rigorously. In fact, even if there is only one group, if one considers asymmetric allocations, it can be desirable to designate some subset of the type 1^b agents as bankers, and concentrate all monitoring on them. We leave further exploration of these ideas to future work.

6.3. Rate of return dominance

To this point we assumed $\rho^a = \rho^b$, but, by continuity, of course deposits in group b can be PE even if $\rho^a > \rho^b$. Heuristically, this explains why individuals keep wealth in demand deposits, despite the existence of alternatives with higher yields: deposits are better payment instruments. In other words, deposits are more liquid. Still, there is an interaction between liquidity and return worth making precise. The next result is also proved in the Appendix.

Proposition 5. *If $d > 0$ is PE when $\rho^b = \rho^a$, there exists $\varepsilon > 0$ such that $d > 0$ is PE when $\rho^b = \rho^a - \varepsilon$. Deposits in group b are PE if $\delta^a > \bar{\delta}^a$, and either: (a) $\delta^b \leq \bar{\delta}^b$ and $\delta^a \rho > (\rho - 1)u'(\rho \underline{y}^a)$; or (b) $\delta^b > \bar{\delta}^b$ and $\delta^a \rho > \delta^b + (\rho - 1)u'(\rho \underline{y}^a)$, with the thresholds $\bar{\delta}^a$, $\bar{\delta}^b$ and \underline{y}^a defined in the proof.*

6.4. Intermediated lending

To this point banks undertake investments directly. In reality, although banks do invest some deposits directly, they also lend to borrowers who then make investments. The reason this is worth mentioning is that once we introduce borrowers explicitly, one may wonder how they can credibly commit to repay the bank but not depositors. What is the use of banks as intermediaries if depositors can lend directly to investors? To address this, we now suppose there is a third group c , as well as a and b . For illustration, all parameters are the same across groups, except $\pi^b > \pi^c > \pi^a = 0$. Also, $\rho^a = \rho^b = 1$, and in all groups, $U^1(x, y) = u(x) - y$ and $U^2(y, x) = u(y) - x$. To incorporate lending, agents in group c have a special technology $f(I)$ that requires at least \bar{I} units of good y^a . Precisely, for $I < \bar{I}$ we have $f(I) = 0$, and for $I \geq \bar{I}$, we have $f(I) = \alpha I$ with $\alpha > 1$.

All is well if the minimum investment \bar{I} is small. But when it is large, \bar{I} may be too expensive for a single group a agent to lend to a group c investor. Absent other frictions, the solution is to have many 1^a agents lend to a given group c investor. But suppose we add an additional friction, that agents in group c can meet at most n other agents each period.¹⁷ Now direct lending may fail,

17. We do not regard this as particularly deep; it is simply an example to show how other frictions can interact with limited commitment.

as 1^a would have to produce enough y^a to meet the minimum investment threshold, which may not be worthwhile. And since $\pi^a = 0$, it is impossible for group a lenders to pool their resources, and have one agent 1^a lend it all to some agent 1^c , since the designated 1^a agent would certainly run off with the proceeds. Here is where intermediated lending can help: a trustworthy agent 1^b collects resources from many agents 1^a , and lends them to a 1^c investor. Delegating lending through a bank allows us to meet the minimum investment level, despite commitment problems within group 1.

To fill in the details, first, note that $\pi^a = 0$ implies group a cannot consume at all without deposits. In principle, 1^a could deposit resources with 1^b for direct investment, but it may be desirable for 1^b to lend the deposits to 1^c , if $f(\cdot)$ constitutes a better opportunity. When group a lends $d = y^a$ directly to group c , the relevant incentive constraints for group a are

$$u(x^a) - y^a \geq 0 \text{ and } u(\alpha y^a) - x^a \geq 0.$$

Similarly, for b and c ,

$$\begin{aligned} u(y^b) - x^b &\geq 0 \text{ and } u(x^b) - y^b \geq \delta^b y^b \\ u(y^c) - x^c &\geq 0 \text{ and } u(x^c) - y^c \geq \delta^c (\alpha \tilde{n} y^a + y^c), \end{aligned}$$

where $\tilde{n} \leq n$ is the number of agents 1 in group a pooling their resources for lending, with $\tilde{n} y^a \geq \bar{I}$. Since an agent 1^c can meet at most n agents, assuming symmetry, the minimum resources that each 1^a must commit is \bar{I}/n . If $u(\alpha \bar{I}/n) < \bar{I}/n$, this is too large for direct lending by group a to be viable.

With intermediated lending, the relevant constraints are the same for group a , but now for group b they are

$$u(y^b) - x^b \geq 0 \text{ and } u(x^b) - y^b \geq \delta^b (y^b + \hat{n} y^a - d + \alpha d).$$

The total amount received by banker 1^b from \hat{n} agents in group a is $\hat{n} y^a$, of which he lends $d \leq \hat{n} y^a$ to agents in group c . In the second subperiod, these loans return αd . Given he also invests y^b for agents in his own group, his repayment constraint is as given above. For group c the relevant constraints are

$$u(y^c) - x^c \geq 0 \text{ and } u(x^c) - y^c \geq \delta^c (\alpha \tilde{n} d + y^c),$$

where $\tilde{n} d \geq \bar{I}$, and $\tilde{n} \leq n$ is the number of bankers lending d to an investor. If $\tilde{n} > 1$, the minimum investment is $\bar{I}/\tilde{n} n < \bar{I}/n$.

For \tilde{n} large, $u(\alpha \bar{I}/\tilde{n} n) > \bar{I}/\tilde{n} n$, and the IF set for group a contains points other than autarky. The smaller is δ^b , the larger we can set \tilde{n} . We conclude that intermediated lending can be PE, if we add additional frictions, and in particular if there is a large fixed investment \bar{I} . This is reasonable, since firms often need funds beyond what a single lender can provide. Moreover, a single lender may not want the risk exposure implied by single large investment, although we do not model this explicitly. The point is that we can extend the framework to explain how banks usefully intermediate between depositors and investors, based in part on limited commitment, and in part on other frictions.

7. NON-STATIONARY ALLOCATIONS

So far, we have restricted attention to stationary allocations. One might suspect that relaxing stationarity could be good for incentives, and if this works too well, deposits may no longer be

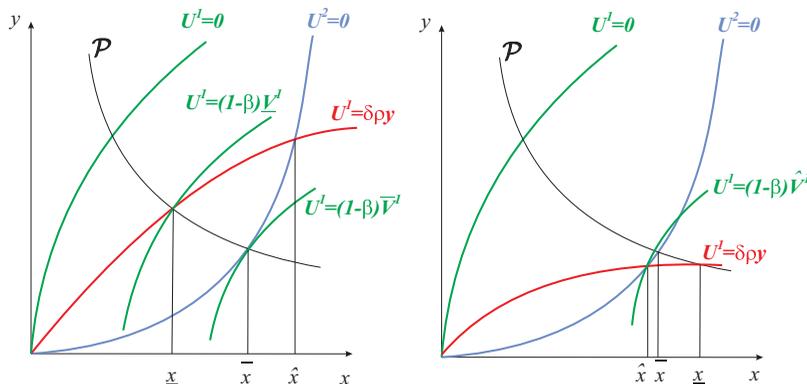


FIGURE 5

Stationary participation and repayment constraints

PE. Clearly, this is not true in general—*e.g.* if $\pi^a = 0$ then the only IF allocation in group a with $d = 0$ is autarky, and so it seems obvious that $d > 0$ can be PE. But it would be good to know how the results are affected, by relaxing stationarity, more generally. This is the issue studied here. We assume $\rho = \lambda = \gamma = 1$, for this exercise, mainly to ease the presentation, and start with a single group.

To begin, we identify three points of reference. Let $(\underline{x}, \underline{y})$ be the stationary allocation given by the intersection of the stationary Pareto set \mathcal{P} with 1's stationary repayment constraint, and let $\underline{V}^1 = U^1(\underline{x}, \underline{y}) / (1 - \beta)$ be the associated payoff for 1. Let (\bar{x}, \bar{y}) be the intersection of \mathcal{P} with 2's stationary participation constraint and $\bar{V}^1 = U^1(\bar{x}, \bar{y}) / (1 - \beta)$. Let $(\hat{x}, \hat{y}) \neq (0, 0)$ be the intersection of 1's stationary repayment constraint with 2's participation constraint and $\hat{V}^1 = U^1(\hat{x}, \hat{y}) / (1 - \beta)$. Figure 5 shows these reference points for two cases, with the same preferences, but in the left panel the stationary repayment constraint is loose, which implies $\bar{x} < \hat{x}$, and in the right it is tight, which implies $\bar{x} > \hat{x}$. On the left, we show 1's indifference curves associated with \underline{V}^1 and \bar{V}^1 , but not \hat{V}^1 , since the latter does not play a role when the repayment constraint is loose; on the right, we show 1's indifference curves associated with \hat{V}^1 , which is relevant when repayment is tight.

For a single group, consider the recursive Pareto problem

$$V^2(V^1) = \max_{x, y, V_{+1}^1} U^2(y, x) + \beta V^2(V_{+1}^1) \quad (28)$$

$$\text{s.t. } V^1 = U^1(x, y) + \beta V_{+1}^1 \quad (29)$$

$$\beta \pi V_{+1}^1 \geq y \quad (30)$$

$$V^2(V_{+1}^1) \geq 0, \quad (31)$$

where the subscript on V_{+1}^1 indicates next period. The objective is to maximize V^2 taking as given V^1 , where of course we only consider $V^1 \geq 0$, to satisfy 1's participation constraint at the initial date. Constraint (29), often called “promise keeping” in the literature, is the law of motion for V^1 ; (30) is the dynamic repayment constraint for 1; and (31) is the dynamic participation constraint for 2, guaranteeing he does not defect next period. This problem is complicated by the fact that the value function V^2 appears in the constraint (31), rendering it a non-standard dynamic

program. But, using standard methods,¹⁸ we can replace (31) with $V_{+1}^1 \leq \tilde{V}^1$, where \tilde{V}^1 is the largest payoff we can give to type 1 such that the problem has a solution with $V^2 \geq 0$. Note that $V^2(\tilde{V}^1) = 0$.

Given this, one can determine the upper bound \tilde{V}^1 explicitly, and characterize the dynamic outcome as a function of the initial condition V_0^1 , for two cases depicted by the two situations in Figure 5. Proofs of the next two Propositions, which constitute the main technical contribution of the article, are somewhat lengthy and are available in the Appendix.

Proposition 6. *Suppose $\bar{x} < \hat{x}$ (repayment loose). Then the upper bound is $\tilde{V}^1 = \bar{V}^1$. If $V_0^1 \in [\underline{V}^1, \bar{V}^1]$, the efficient allocation is stationary at the intersection of \mathcal{P} and the indifference curve $U^1(x, y) = (1 - \beta)V_0^1$. If $V_0^1 \in [0, \underline{V}^1)$ the efficient allocation is non-stationary, V_t^1 is strictly increasing in t until it converges to some $V_\infty^1 \in [\underline{V}^1, \bar{V}^1]$, and (30) is binding during the transition.*

Proposition 7. *Suppose $\bar{x} > \hat{x}$ (repayment tight). Then the upper bound is $\tilde{V}^1 = \hat{V}^1$. If $V_0^1 = \hat{V}^1$ the efficient allocation is stationary at (\hat{x}, \hat{y}) . If $V_0^1 \in [0, \hat{V}^1)$ the efficient allocation is non-stationary, V_t^1 is strictly increasing in t and converges to $V_\infty^1 = \hat{V}^1$. In either case, $V_0^1 = \hat{V}^1$ or $V_0^1 < \hat{V}^1$, (30) is always binding.*

In terms of economics, suppose the stationary repayment constraint is loose. Then, if we want to treat 1 well at the initial date, in the sense that $V_0^1 \geq \underline{V}^1$, the efficient outcome is stationary; and if we want to treat him less well, in the sense that $V_0^1 < \underline{V}^1$, then V_t^1 increases with t . In the latter case, the efficient way to encourage repayment is to backload 1's rewards. In practice, this can take the form of offering small loans at high interest at the beginning, followed by larger loans at better terms. Intuitively, it is wasteful to give 1 a stationary payoff high enough to discourage misbehaviour, since over time past rewards become sunk, and no longer affect incentives. This does not matter if we want to treat 1 well, but if we want to treat him less well, it is better to give him an increasing sequence of utilities. But V_t^1 cannot increase indefinitely, since we have to satisfy 2's participation, so $V_t^1 \rightarrow V_\infty^1$. In the other case, when the stationary repayment constraint is tight, except for the extreme case $V_0^1 = \hat{V}^1$ the outcome is always nonstationary, backloading 1's rewards for the same reason. Also note that, in either case, the dynamic repayment constraint always binds along the transition path.

The above results fully characterize the solution with one group. With two groups, there are several possibilities. First, suppose repayment is loose in both, and in group a the initial condition is $V_0^{1a} < \underline{V}^{1a}$ while in b it is $V_0^{1b} > \underline{V}^{1b}$. Ignoring deposits, from Proposition 6, the efficient allocation is stationary and unconstrained by repayment in b , while it is non-stationary and constrained in a . It is obvious that in this case $d > 0$ is PE, since deposits relax a binding constraint for a without affecting b . Indeed, we can set $d = \bar{d}$, so that the repayment constraint just binds for b , and get the maximum slack for a . The resulting allocation for a may now be stationary. To see this, notice that $d > 0$ rotates the repayment constraint in Figure 5, moving $(\underline{x}^a, \underline{y}^a)$ to the northwest along \mathcal{P}^a and lowering $\underline{V}^{1a} = U^1(\underline{x}^a, \underline{y}^a)/(1 - \beta)$. This may reverse the inequality $V_0^{1a} < \underline{V}^{1a}$, making the efficient allocation for a stationary. Hence, if we have enough slack in group b , by letting type 1^b be the banker for 1^a we can relax the latter's constraints so that we no longer need backloading. Even if \bar{d} is not big enough to reverse $V_0^{1a} < \underline{V}^{1a}$, so

18. These techniques go back to Thomas and Worrall (1988); see Ljungqvist and Sargent (2004) for a textbook treatment.

we still need backloading, deposits are still PE because they slacken repayment for a along the transition.¹⁹

So far we assumed $V_0^{1b} > \underline{V}^{1b}$. We can also have $V_0^{1i} < \underline{V}^{1i}$ for $i=a$ and $i=b$, with a non-trivial transition for both groups. Now the problem is more intricate, but it seems clear that $d_t > 0$ cannot be PE for all t since this only slackens repayment in one group at the expense of the other. However, $d_t > 0$ may be PE at some t in the future. Thus, over time $V_t^{1b} \rightarrow V_\infty^{1b}$, and suppose we get there in finite time, before V_t^{1a} converges. If repayment no longer binds in group b , but is still binding in a , it is efficient to set $d > 0$ since it will not hurt b . There is also the case where the repayment constraint is tight for both groups, or tight for one and loose for the other. In general, there are non-trivial transitions, and dynamically efficient allocations can be complicated. But the goal here is to show that $d > 0$ can be PE, not that $d > 0$ is always PE. This has been established. While there is much more one can do with this model, in terms of both theory and quantitative work, we leave that to future research.

To summarize the main point, consider what we think of as the leading case: in group b , repayment is loose and $V_0^{1b} > \underline{V}^{1b}$, so the efficient outcome is stationary; in group a , repayment is either loose and $V_0^{1a} < \underline{V}^{1a}$, or repayment is tight, both of which imply backloading for 1^a . Then we can set $d = \bar{d} > 0$, so repayment just binds in group b , and relax the constraint for a . In the case where in group a repayment is loose and $V_0^{1a} < \underline{V}^{1a}$, this may reverse the inequality and render the outcome stationary. This means that banking may replace backloading. In any case, $d = \bar{d}$ relaxes constraints for group a . Also, notice that we can do no better if we use a non-stationary banking scheme: setting $d_t = \bar{d}$ for all t maximizes slack in group a . We are not saying the optimal d_t must be constant, and if it were, *e.g.* costly to use deposits, one might want d_t to decline over time. But absent *ad hoc* reasons, there is nothing to gain from non-stationary d_t . At least, this is so in the leading case where repayment is loose and the efficient outcome is stationary in group b .

We close this section by mentioning that this model, while bearing some similarity to other dynamic models with incentive problems, *e.g.* Atkeson and Lucas (1992), also has some interesting differences. In particular, our model does not lead to immiseration, where a small set of agents eventually end up with the entire surplus. In any case, we collect some of the main results here as follows:

Proposition 8. *Consider optimal allocations with $d_t = 0$, and suppose that at $t = 0$ the repayment constraint does (does not) bind for group a (group b). Then $d_t = \bar{d} > 0$ for all t is PE, and we can do no better by having d_t vary with t . Given $d_t = \bar{d}$, the efficient dynamic allocation is stationary for b , and can either be stationary or non-stationary for a , depending on whether \bar{d} lowers \underline{V}^{1a} enough to reverse the inequality $V_0^{1a} > \underline{V}^{1a}$.*

8. A BRIEF HISTORY OF BANKING

Above we have presented some theory and a variety of applications of the abstract model. Now we compare the results with some facts from banking history. First, as regards abstracting from outside money, this seems reasonable from the historical perspective, since institutions that accepted commodity deposits were operating long before the invention of coinage, let alone fiat currency.

19. One has to be slightly careful with this case. Proposition 6 says the transition in group a takes us to $V_\infty^{1a} \in [\underline{V}^{1a}, \bar{V}^{1a}]$, and it is important to note that we can get there in finite time, and can end up in the interior of the interval (since in discrete time we jump a finite distance at each step). Once we get to V_∞^{1a} , if it is interior, then from Proposition 6 one might think repayment is no longer binding and hence $d > 0$ is no longer PE. But this is not right, since repayment is no longer binding only because we have $d > 0$.

As Davies (2002) describes the situation, in ancient Mesopotamia and Egypt, goods were often deposited in palaces or temples, and later, private houses.

Grain was the main form of deposits at first, but in the process of time other deposits were commonly taken: other crops, fruit, cattle and agricultural implements, leading eventually and more importantly to deposits of the precious metals. Receipts testifying to these deposits gradually led to transfers to the order not only of depositors but also to a third party. In the course of time private houses also began to carry on such deposit business ... The banking operations of the temple and palace based banks preceded coinage by well over a thousand years and so did private banking houses by some hundreds of years.

Importantly, deposit receipts were transferrable, and hence facilitated transactions and payments, as in the model. In ancient Babylon, also, as Ferguson (2008) says: “Debts were transferable, hence ‘pay to the bearer’ rather than a named creditor. Clay receipts or drafts were issued to those who deposited grain or other commodities at royal palaces or temples.” And, also as in the model, “the foundation on which all of this rested was the underlying credibility of a borrower’s promise to repay.” This is, of course, exactly what the model was designed to capture.

In his discussion of medieval Venetian bankers, Mueller (1997) describes two types of deposits: *regular*, which were actual goods that bankers had to deliver on demand; and *irregular*, involving specie or coins that only had to be repaid with the same value, but not the same objects. The former were like modern-day safety-deposit boxes; the latter were more like demand deposits, and involved a tacit agreement that the banker would invest the resources. When one puts one’s money in a modern bank account, one usually does not expect to withdraw the same money, only something of appropriate value. This is true in the model, too: a bank’s liability is not the deposit *per se*, but the returns on investments. Because they are making investments, banks are more than mere storage facilities, although safekeeping may have something to do with it. Consider the English goldsmiths, who many regard as the first modern bankers (Joslin 1954; Quinn 1997; Selgin 2010). Originally they offered depositors mainly security, but early in the 17th century their deposit receipts began circulating in place of cash for payments, the first incarnation of banknotes; shortly after, deposits could be transferred by drawn note (checks).²⁰

Institutions of the type modelled here—*i.e.* acceptors of commodities on deposit that end up facilitating transactions—were common well after the emergence of modern banking. In colonial Virginia, *e.g.* tobacco was commonly used in transactions because of the scarcity of precious metals, and the practice of depositing tobacco in public warehouses and then exchanging certificates, attesting to its quality and quantity, survived for over 200 years (Galbraith, 1975). Similarly, in the 19th century, to facilitate transactions and credit arrangements between cocoon producers and silk weavers, warehouses commonly stored dried cocoons or silk and issued warrants that could be used to pledge for credit. The first of these warehouses was funded by a group of entrepreneurs in Lyons in 1859, and later imitated by a series of Italian banks (Federico, 1997). What we take away from these examples is that a very important feature of much early, and some more modern, deposit banking is that claims on deposits were used to facilitate exchange.

In Venice, Mueller (1997) says deposit banking served “a function comparable to that of checking accounts today ... not intended primarily for safekeeping or for earning interest but rather as a means of payment which facilitated the clearance of debts incurred in the process

20. Although many say goldsmiths were the first modern bankers, others mention the Templars (Weatherford, 1997), who during the crusades specialized in moving and protecting money and other valuables. But it is not clear if their liabilities *circulated* as a means of payment, the way goldsmiths’ receipts did. Some say that, before the goldsmiths, transferring funds from one account to another “generally required the presence at the bank of both payer and payee” (Kohn, 1999; see also Quinn, 2004). Even so, deposits facilitated payments. On checks, Spufford (1988) says the Florentines were using these as early as 1368.

of doing business. In short, the current account constituted ‘bank money,’ money based on the banker’s promise to pay.” Of course, this only works if bankers are trustworthy. The medieval Rialto banks offer evidence consistent with this: “Little capital was needed to institute a bank, perhaps only enough to convince the guarantors to pledge their limited backing and clients to deposit their money, for it was deposits rather than funds invested by partners which provided bankers with investable capital. In the final analysis, it was the visible patrimony of the banker—alone or as part of a fraternal compagnia—and *his reputation as an operator on the market place in general* which were placed on the balance to offset risk and win trust” (Mueller, p. 97, emphasis added).

Many bankers historically started as merchants, who almost by definition have a big connection to the market. The great banking families in Renaissance Italy and Southern Germany in the 16th century were originally merchants, who began lending their own capital, and then started collecting deposits from other merchants, nobles, clerics, and small investors. They were not the richest individuals: wealth was then concentrated in the hands of landowners, who controlled agriculture, forests and mineral rights. But merchants arguably had the most to lose in terms of reputation from renegeing on obligations: “because commerce involved the constant giving and receiving of credit, much of a merchant’s effort was devoted to ensuring that he could fulfill his own obligations and that others would fulfill theirs.” (Kohn, 1999). Further evidence on the bankers having a big connection to the market is given by Pressnell (1956) in its study of early English country banks during the Industrial Revolution: almost all of these emerged as a by-product of some other economic activity, often some kind of manufacturing.

Bankers also were subject to some kind of monitoring, like we emphasized in the model. For instance, “to maintain ‘public faith,’ the Senate in 1467 reminded bankers of their obligation to show their account books to depositors upon request.” (Mueller, p. 45). If caught cheating, the punishment was indeed lifetime banishment from Venetian banking, but this apparently happened rarely in history (like in the model). Going back to Roman times, Orsingher (1967) observes: “One of the most important techniques used by Roman bankers was the use of account books analogous to those which all citizens kept with scrupulous care. This account-book was called a Codex and was indispensable in drawing up contracts A procedure peculiar to bankers deserves to be noted: the ‘editio rationum’ or production of accounts. Anyone running a bank could be compelled at a moment’s notice to produce his accounts for his clients’, or even for a third party’s, inspection.”

Once again returning to Venice:

In the period from about 1330 to 1370, eight to ten bankers operated on the Rialto at a given time. They seem to have been relatively small operators on average... Around 1370, however, the situation changed [and] Venetian noble families began to dominate the marketplace. After the banking crisis of the 1370s and the War of Chioggia, the number of *banchi di scritta* operating at any given time on the Rialto dropped to about four, sometimes as few as three. These banks tended, therefore, to be larger and more important than before. (Mueller, p. 82)

We also tried to formalize issues concerning the efficient number of bankers, revolving around credibility vs. size, since obviously a bank can be “too big not to fail.” To summarize the general point, while we are not experts on history, this evidence on the whole seems to illustrate how the model is broadly consistent with the facts.

Finally, in terms of more recent history, what about banking panics and the recent financial crisis? Gorton (2010) argues the crisis was a wholesale panic, where some financial firms ran on others by not renewing repurchase agreements, similar to commercial bank customers withdrawing deposits. The location of subprime risk was unknown, depositors were confused, and consequently ran. Our approach is too stylized to capture all the intricacies, but we can use

it to think about the big picture. Suppose the probability of being active each period γ is subject to shocks, and uncertainty surrounding these shocks can induce agents to reduce deposits to re-establish incentives. Such shocks depend on the nature of business, *e.g.* γ could be affected by housing markets if one's business is originating mortgage loans. As γ goes down, banking works less well, but *this is as it should be*: when the frictions increase it can be efficient for credit to dry up. We do not claim recent events were not problematic; only that it is interesting to look at them through the lens of this model.

9. CONCLUSION

We began by specifying preferences, technologies, and frictions, including imperfect commitment and monitoring. We then examined incentive-feasible and efficient outcomes, and tried to interpret these in terms of arrangements observed in actual economies. The model illustrates how it can be desirable for some agents to perform certain functions resembling banking: they accept deposits, they invest or make loans, and their liabilities facilitate transactions among other parties. Banking is essential: without it, the set of feasible allocations is inferior. We showed how to implement good outcomes using circulating bank liabilities. Also, it helps if banks offer high interest rates, but it can also be efficient to sacrifice return for trust. We explored various other issues, such as endogenous monitoring and intermediated lending. We also solved for efficient dynamic credit allocations, showed how they generally can entail backloading, and discussed how this interacts with banking. Finally, we discussed some history related to the theory. This is all in the spirit of what is called the New Monetarist approach.

The framework can be generalized to analyse other commonly discussed phenomena, *e.g.* the tendency of banks to borrow short and lend long. We concentrated on different issues, like deriving a set of characteristics that make for good bankers. Other extensions and applications are available. In a companion paper (Gu *et al.* 2012) using exactly the same environment, instead of studying incentive-feasible and efficient outcomes, we impose various alternative solution concepts—*e.g.* Walrasian pricing or Nash bargaining—and study decentralized equilibria. We show that it is possible, indeed easy, for these economies to exhibit complicated endogenous dynamics, including cyclic, chaotic and stochastic (sunspot) equilibria. In all of these cases, market activity fluctuates over time, either deterministically or randomly, based on self-fulfilling prophecies concerning credit conditions. Based on this, we think the framework provides an interesting setting in which to analyze credit markets, generally, and not just banking.

We close with a question Ken Burdett asked about the article: Should we be surprised by a theory that predicts it is efficient to put agents into occupations they are good at? If the occupation is, say, singing, it is obvious that people with good range, pitch, timbre, etc. are right for the job. That is different, however, since we can all agree that music and hence musicians give people direct utility. *But no one likes bankers, just like no one likes dollars, for their own sake.* People like money for what it does, and we think it is useful to explain what makes a particular object an efficient or an equilibrium medium of exchange, not just assume it enters in a particular *ad hoc* way into preferences or constraints. Similarly, if we want to study banking, we think it is useful to have a theory that is explicit about the frictions that give rise to a role for banks in the first place. For us, this project was a step in that direction.

APPENDIX

Proof of Proposition 4. Since $\gamma^b > \gamma^a$, it must be that $U^1(x^b, y^b) \geq U^1(x^a, y^a)$. With deposits d , and since there is one candidate banker in each group, the repayment constraint in group b becomes $-\lambda^b \rho (y^b + d) + \pi^b \gamma^b U^1(x^b, y^b) / r = 0$,

where $r = (1 - \beta) / \beta$. Therefore, we obtain

$$\frac{\partial \pi^b}{\partial d} = \frac{r \lambda^b \rho}{\gamma^b U^1(x^b, y^b)}.$$

The repayment constraint in group a is $-\lambda^a \rho (y^a - d) + \pi^a \gamma^a U^2(x^a, y^a) / r = 0$, so that

$$\frac{\partial \pi^a}{\partial d} = -\frac{r \lambda^a \rho}{\gamma^a U^1(x^a, y^a)}.$$

Increasing deposits from group a to b reduces the overall monitoring cost $\pi^a k^a + \pi^b k^b$ since

$$\frac{\partial \pi^a}{\partial d} k^a + \frac{\partial \pi^b}{\partial d} k^b = r \left[\frac{\lambda^b k^b \rho}{\gamma^b U^1(x^b, y^b)} - \frac{\lambda^a k^a \rho}{\gamma^a U^1(x^a, y^a)} \right] < 0,$$

where the inequality follows from $U^1(x^a, y^a) \leq U^1(x^b, y^b)$, $\gamma^a \leq \gamma^b$, $\lambda^b \leq \lambda^a$ and $k^b \leq k^a$. Hence, starting from $d = 0$, only $d > 0$ can reduce total monitoring cost.

To prove the second part let $(\check{x}^a, \check{y}^a)$ solve $\max W^a(x, y)$, s.t. the participation constraint for 2^a only. If

$$\bar{\pi} \equiv \frac{r \lambda^b (y^b + \check{y}^a)}{\gamma^b U^1(x^b, y^b)} \leq 1$$

then it is optimal to set $\pi^b = \bar{\pi}$, $d = \check{y}^a$, and $\pi^a = 0$. Then \bar{y} is defined as

$$\bar{y} \equiv \frac{r \lambda^b (y^b + \check{y}^a)}{U^1(x^b, y^b)}.$$

This completes the proof. \parallel

Proof of Proposition 5. Consider an example with $U^1 = x - y$, $U^2 = u(\rho y) - x$, $\gamma^a = \gamma^b$, $\lambda^a = \lambda^b = 1$, $\omega_1 = \omega_2 = \omega$, and $\rho^a = \rho > 1 = \rho^b$. Given $\rho > 1$, we can have $\bar{\delta}^a < \bar{\delta}^b$, so that $(x^{*,a}, y^{*,a})$ is feasible in group a but $(x^{*,b}, y^{*,b})$ is not feasible in group b . Here, we focus on the case where deposits in group b are PE. The condition $\delta^a > \bar{\delta}^a$ implies that $y^{*,a}$ is not IF in group a , so that deposits potentially have a role. Consider the situation in group b . In the first Case (a), agents in group b do not have a commitment problem because $\delta^b \leq \bar{\delta}^b$, although they do have inferior storage technology. Therefore, making deposits in group b requires agents in group a produce more to make up for a lower return to sustain a given level of consumption. The condition $\delta^a \rho > (\rho - 1) u'(\rho y^a)$ insures that δ^a is high enough so that $d > 0$ is PE. Case (b) is similar, except agents in group b have a binding repayment constraint when $\delta^b > \bar{\delta}^b$. Therefore they need to be compensated for taking deposits to prevent default. A transfer from group a does just that, but it comes on top of the additional production required from group a to cover for the loss in return. Hence, $d > 0$ is PE if $\delta^a \rho > \delta^b + (\rho - 1) u'(\rho y^a)$, which is stricter than Case (a). Finally, if the commitment problem in group a is very severe, $u'(\rho y^a)$ will be large. In this case, if the investment technology in group a improves, their commitment problem must be worse for $d > 0$ to be PE.

The planner's problem with no interaction between groups is given by (25). The first best is $y^{*,i}$ solving $\rho^i u'(\rho^i y^{*,i}) = 1$. Denote by \underline{y}^i , the level of y^i that satisfies the repayment constraint at equality given δ^i . Also, Define $\bar{\delta}^i$ by $[u(\rho^i y^{*,i}) - y^{*,i}] / (\rho^i y^{*,i}) = \bar{\delta}^i$ as the level above which the repayment constraint binds in group i at $y^{*,i}$. The next two claims establish the result.

Claim 1. *Deposits in group b are PE if*

$$\delta^a > \bar{\delta}^a, \delta^b \leq \bar{\delta}^b \text{ and } \delta^a \rho > (\rho - 1) u'(\rho y^a).$$

Proof Note that given x_2^a and d , agents 1^a has to produce y^a such that $x_2^a = (y^a - d) \rho + d$. The repayment constraint is

$$u[(y^a - d) \rho + d] - \frac{y^a}{\rho} \geq \delta^a \rho (y^a - d).$$

To show deposits in group b are PE, we show increasing d relaxes the repayment constraint in group a . Hence it must be that at the allocation y^a

$$(1 - \rho) u'[(y^a - d) \rho + d] + \delta^a \rho > 0$$

$$\delta^a \rho > (\rho - 1) u'[(y^a - d) \rho + d]$$

So $d > 0$ is PE at \underline{y}^a iff $\delta^a \rho > (\rho - 1) u'(\rho y^a)$, establishing the claim. \parallel

Claim 2. Deposits in group b are PE if

$$\delta^a > \bar{\delta}^a, \delta^b > \bar{\delta}^b \text{ and } \delta^a \rho \geq \delta^b + (\rho - 1)u'(\rho y_2^a).$$

Proof When $\delta^b > \bar{\delta}^b$, the solution to (25) in group b is \underline{y}^b . Deposits are incentive compatible only if agents 1^a make a transfer τ to agents 1^b . The repayment constraint in group b with τ and d , evaluated at \underline{y}^b , is $u(\underline{y}^b) - \underline{y}^b + \tau \geq \delta^b(\underline{y}^b + d)$. By definition, $u(\underline{y}^b) - \underline{y}^b = \delta^b \underline{y}^b$ and the minimum transfer τ that satisfies the constraint is $\tau = \delta^b d$. The repayment constraint in group a is

$$u[(y^a - d)\rho + d] - \frac{y^a}{\rho} - \tau \geq \delta^a \rho (y^a - d).$$

Substituting $\tau = \delta^b d$, we get

$$u[(y^a - d)\rho + d] - \frac{y^a}{\rho} - \delta^a \rho y^a + (\delta^a \rho - \delta^b)d \geq 0,$$

so the repayment constraint is relaxed whenever $\delta^a \rho - \delta^b \geq (\rho - 1)u'[(y^a - d)\rho + d]$. Evaluating at \underline{y}_2^a , $\delta^a \rho - \delta^b \geq (\rho - 1)u'(\rho y_2^a)$. This establishes the claim, and therefore completes the proof. \parallel

Proof of Propositions 6-7. Problem (28) is a dynamic contracting problem with two-sided lack of commitment, and is solved using results of Thomas and Worrall (1988). These deliver the following: $\exists \tilde{V}^1$ such that $V^2(\tilde{V}^1) = 0$; $V^2(V^1) \geq 0$ iff $V^1 \leq \tilde{V}^1$; and $V^2(V^1)$ is decreasing, strictly concave and continuously differentiable. Hence, we can rewrite (28) as

$$V^2(V^1) = \max_{x, y, V_{+1}^1} U^2(y, x) + \beta V^2(V_{+1}^1) \quad (\text{A.1})$$

$$\text{s.t. (29), (30) and } \tilde{V}^1 \geq V_{+1}^1,$$

where we replace (31) by $\tilde{V}^1 \geq V_{+1}^1$ (We also need $V_{+1}^1 \geq 0$, but that is redundant given the repayment constraint $\beta\pi V_{+1}^1 \geq y$.) We eventually determine \tilde{V}^1 , but first we solve (A.1) taking it as given. Letting μ_1 , μ_2 and μ_3 be multipliers on (29), (30), and $\tilde{V}^1 \geq V_{+1}^1$, necessary and sufficient conditions for a solution are

$$0 = U_x^2(y, x) + \mu_1 U_x^1(x, y) \quad (\text{A.2})$$

$$0 = U_y^2(y, x) + \mu_1 U_y^1(x, y) - \mu_2 \quad (\text{A.3})$$

$$0 = \beta V^{2'}(V_{+1}^1) + \beta\mu_1 + \beta\pi\mu_2 - \mu_3. \quad (\text{A.4})$$

The envelope condition is $V^{2'}(V^1) = -\mu_1$, which combines with (A.4) to yield

$$0 = \beta V^{2'}(V_{+1}^1) - \beta V^{2'}(V^1) + \beta\pi\mu_2 - \mu_3. \quad (\text{A.5})$$

Given \tilde{V}^1 , we consider separately: (1) $\tilde{V}^1 > \underline{V}^1$; and (2) $\tilde{V}^1 \leq \underline{V}^1$.

Case 1: $\tilde{V}^1 > \underline{V}^1$. We describe the solution to (A.1) as a function of $V^1 \geq 0$. There are three possibilities: (1a) $V^1 \in [0, \underline{V}^1)$; (1b) $V^1 \in [\underline{V}^1, \tilde{V}^1]$; and (1c) $V^1 > \tilde{V}^1$. We first need the following:

Claim 3. In Case 1, the solution to (A.1) is stationary if and only if $V^1 \in [\underline{V}^1, \tilde{V}^1]$.

Proof First assume the solution is stationary, which means $V_{+1}^1 = V^1$. By (A.5), $V_{+1}^1 = V^1$ implies either: $\mu_2 = \mu_3 = 0$ or $\mu_2, \mu_3 > 0$. Suppose $\mu_2, \mu_3 > 0$. Since $\mu_2 > 0$ and the solution is stationary, (x, y) is on the stationary repayment constraint, and it is to the left of \mathcal{P} by virtue of (A.2) and (A.3). Type 1's payoff associated with this allocation is $V^1 = V_{+1}^1 < \underline{V}^1$. However, $\mu_3 > 0$ implies $V^1 = V_{+1}^1 = \tilde{V}^1$. This is a contradiction, since we are in Case 1, where $\tilde{V}^1 > \underline{V}^1$.

So we cannot have $\mu_2, \mu_3 > 0$, and we must have $\mu_2 = \mu_3 = 0$. Then (A.2) and (A.3) imply $(x, y) \in \mathcal{P}$, and given stationarity, we can combine (29) and (30) to yield the stationary repayment constraint. Type 1's payoff thus satisfies $V^1 \geq \underline{V}^1$, since by definition \underline{V}^1 is the payoff associated with the point where the stationary repayment constraint intersects \mathcal{P} . Also, given stationarity, $U^1(x, y)/(1 - \beta) = V^1 = V_{+1}^1 \leq \tilde{V}^1$. This establishes that stationarity implies $V^1 \in [\underline{V}^1, \tilde{V}^1]$.

To show the converse, start by assuming $V^1 \in [\underline{V}^1, \tilde{V}^1]$, and recall that (29), (30), $V_{+1}^1 \leq \tilde{V}^1$, (A.2), (A.3), and (A.5) are necessary and sufficient for (A.1). Set $\mu_2 = \mu_3 = 0$ and $\mu_1 = -U_x^2/U_x^1$. Set $V_{+1}^1 = V^1$. Let (y, x) be the point on \mathcal{P} where $U^1(y, x) = (1 - \beta)V^1$. Then all of the relevant conditions are satisfied. Hence, $V^1 \in [\underline{V}^1, \tilde{V}^1]$ implies stationarity. This establishes Claim 3. \parallel

Return to the analysis of (A.1) in Case 1, consider first (1a).

Claim 4. *In Case (1a) $V_{+1}^1 > V^1$.*

Proof Now $V^1 \in [0, \underline{V}^1]$. Suppose by way of contradiction that $V_{+1}^1 \leq V^1$. First suppose $V_{+1}^1 < V^1$. This implies $\mu_3 > 0$, by the strict concavity of V^2 and (A.5). But $\mu_3 > 0$ implies $V_{+1}^1 = \tilde{V}^1$, and we have $\tilde{V}^1 > \underline{V}^1$ since we are in Case 1. Thus $V_{+1}^1 = \tilde{V}^1 > \underline{V}^1$, which contradicts $V_{+1}^1 < V^1$, since $V^1 < \underline{V}^1$ in Case (1a). Now suppose $V_{+1}^1 = V^1$. This contradicts Claim 3, which tells us that the solution is stationary if and only if $V^1 \in [\underline{V}^1, \tilde{V}^1]$, and in Case (1a) $V^1 \in [0, \underline{V}^1]$. So cannot have $V_{+1}^1 \leq V^1$. \parallel

Now consider Case (1b), $V^1 \in [\underline{V}^1, \tilde{V}^1]$. By Claim 3 we know the solution is stationary and (x, y) is given by the intersection of \mathcal{P} and $U^1(x, y) = (1 - \beta)V^1$. Finally, consider Case (1c), where $V^1 > \tilde{V}^1$. This implies $\mu_3 > 0$ by (A.5). Therefore $V_{+1}^1 = \tilde{V}^1$, and after one period we are back to Case (1b), although as we show below, this cannot happen once we endogenize \tilde{V}^1 . In any event, we are done with Case 1, and proceed to:

Case 2: $\tilde{V}^1 \leq \underline{V}^1$. *The following possibilities are now relevant: (2a) $V^1 < \tilde{V}^1$; (2b) $V^1 = \tilde{V}^1$; and (2c) $V^1 > \tilde{V}^1$. We need the following:*

Claim 5. *In Case 2, the solution to (A.1) is stationary if and only if $V^1 = \tilde{V}^1$.*

Proof To verify this, note that one can show that if the allocation is stationary then $V^1 = \tilde{V}^1$ following the proof of Claim 3. To show the converse, assume $V^1 = \tilde{V}^1$. Again, the first-order conditions are necessary and sufficient. It is easy to check that these are satisfied by $V_{+1}^1 = V^1 = \tilde{V}^1$, an allocation (x, y) solving $U^1(x, y) = (1 - \beta)\tilde{V}^1$ and $y = \beta\pi\tilde{V}^1$, and multipliers μ_1, μ_2 , and μ_3 given by the first-order conditions. This establishes Claim 5. \parallel

We now return to (A.1). In Case (2a), where $V^1 < \tilde{V}^1$, $V_{+1}^1 > V^1$ by an argument is similar to Case (1a). In Case (2b), where $V^1 = \tilde{V}^1$, Claim 5 establishes $V_{+1}^1 = \tilde{V}^1$ and the allocation is stationary. In Case (2c), where $V^1 > \tilde{V}^1$, $V_{+1}^1 = \tilde{V}^1$ by an argument is similar to Case (1c), but we also show below that this cannot happen once we endogenize \tilde{V}^1 . This completes the analysis of the solution to (A.1) for an arbitrary $\tilde{V}^1 \geq 0$.

We now need to find the value of \tilde{V}^1 that makes (A.1) equivalent to (28)—*i.e.* we seek a solution to $V^2(\tilde{V}^1) = 0$. We consider separately the two possibilities shown in Figure 5: a loose stationary repayment constraint, where $\underline{V}^1 < \bar{V}^1$; and a tight one, where $\underline{V}^1 > \bar{V}^1$.

Loose: $\bar{x} < \hat{x}$. We claim $\tilde{V}^1 = \bar{V}^1$. To verify this, from the above results we know that $V^1 = \tilde{V}^1$ implies the optimal allocation is stationary. Given the allocation is stationary, $V^2(V^1) = U^2(y, x)/(1 - \beta)$, and we only need to find the stationary solution to $U^2(y, x) = 0$. This is the (y, x) that occurs at the (non-zero) intersection of the indifference curves $U^2(y, x) = 0$ and $U^1(y, x) = \bar{V}^1$ in Figure 5. Hence, with a loose repayment constraint $\tilde{V}^1 = \bar{V}^1$. It is obvious from Figure 5 that $\underline{V}^1 < \bar{V}^1$ is equivalent to $\bar{x} < \hat{x}$. This establishes Proposition 6.

Tight: $\bar{x} > \hat{x}$. We now claim $\tilde{V}^1 = \hat{V}^1$, using a similar argument. If we set $V^1 = \tilde{V}^1$ then the optimal allocation is stationary. Thus, we need to find the best point (x, y) for 1 consistent with $U^2(y, x) = 0$ and the stationary repayment constraint. From the right panel of Figure 5, this is given by the intersection of participation constraint and repayment constraint, which is $V^1 = \hat{V}^1$. Again, it is clear from the figure that $\underline{V}^1 > \bar{V}^1$ is equivalent to $\bar{x} > \hat{x}$. This establishes Proposition 7. \parallel

Acknowledgments. We thank many colleagues for comments and discussions, especially Steve Williamson, Erwan Quintin, Chris Phelan, V.V. Chari, Ken Burdett and Daniel Sanches. We also thank several referees and especially the Editor, Bruno Biais, for his generous input, including an outline for the Introduction that we followed closely (although, of course, we remain responsible for the execution). Wright thanks the NSF and the Ray Zemon Chair in Liquid Assets at the Wisconsin School of Business for research support. The usual disclaimers apply.

REFERENCES

- ALVAREZ, F. and JERMANN, U. (2000), "Efficiency, Equilibrium, and Asset Pricing with Risk of Default", *Econometrica*, **68**, 775–798.
- ANDOLFATTO, D. and NOSAL, E. (2009), "Money, Intermediation and Banking", *Journal of Monetary Economics*, **56**, 289–294.
- ATKESON, A. and LUCAS, R.E. Jr. (1992), "On Efficient Distribution with Private Information", *Review of Economic Studies*, **59**, 427–453.
- ARAÚJO, L. and MINETTI, R. (2011), "On the Essentiality of Banks", *International Economic Review*, **52**, 679–691.

- BLAIS, B., MARIOTTI, T., PLANTIN, G. and ROCHET, J.-C. (2007), "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications", *Review of Economic Studies*, **74**, 345–390.
- BOYD, J. and PRESCOTT, E. (1986), "Financial Intermediary Coalitions", *Journal of Economic Theory*, **38**, 211–232.
- CALOMIRIS, C. and KAHN, C. (1991), "The Role of Demandable Debt in Structuring Optimal Banking Arrangements", *American Economic Review*, **81**, 497–513.
- CAVALCANTI, R., EROSA, A. and TEMZILIDES, T. (1999), "Private Money and Reserve Management in a Random-Matching Model", *Journal of Political Economy*, **107**, 929–945.
- CAVALCANTI, R. and WALLACE, N. (1999a), "A Model of Private Bank Note Issue", *Review of Economic Dynamics*, **2**, 104–136.
- CAVALCANTI, R. and WALLACE, N. (1999b), "Inside and Outside Money as Alternative Media of Exchange", *Journal of Money, Credit, and Banking*, **31**, 443–457.
- CORBAE, D., TEMZELIDES, T. and WRIGHT, R. (2003), "Directed Matching and Monetary Exchange", *Econometrica*, **71**, 731–756.
- DAVIES, G. (2002), *A History of Money From Ancient Times to the Present Day*, 3rd edn (Cardiff: University of Wales Press).
- DEMARZO, P. and FISHMAN, M. (2007), "Agency and Optimal Investment Dynamics", *Review of Financial Studies*, **20**, 151–188.
- DIAMOND, D. and DYBIVIG, P. (1983), "Bank Runs, Deposit Insurance, and Liquidity", *Journal of Political Economy*, **91**, 401–419.
- DIAMOND, D. (1984), "Financial Intermediation and Delegated Monitoring", *Review of Economic Studies*, **51**, 393–414.
- DIAMOND, D. and RAJAN, R. (2001), "Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking", *Journal of Political Economy*, **109**, 287–327.
- ENNIS, H. and KEISTER, T. (2009), "Bank Runs and Institutions: The Perils of Intervention", *American Economic Review*, **99**, 1588–1607.
- FEDERICO, G. (1997), *An Economic History of the Silk Industry, 1830-1930* (Cambridge/New York: Cambridge University Press).
- FREIXAS, X. and ROCHET, J.-C. (2008), *Microeconomics of Banking* (Cambridge: MIT Press).
- FERGUSON, N. (2008), *The Ascent of Money* (New York: The Penguin Press).
- GALBRAITH, K. (1975), *Money: Whence It Came, Where It Went* (Boston: Houghton Mifflin).
- GORTON, G. (2010), *Slapped in the Face by the Invisible Hand: Banking and the Panic of 2007* (Oxford/New York: Oxford University Press).
- GORTON, G. and WINTON, A. (2002), "Financial Intermediation", in Constantinides, G., Harris, M. and Stulz, R. (eds) *Handbook of the Economics of Finance* (Amsterdam/Boston: Elsevier/North Holland).
- GU, C., MATTESINI, F., MONNET, C. and WRIGHT, R. (2012), "Endogenous Credit Cycles" (Mimeo).
- HE, P., HUANG, L. and WRIGHT, R. (2005), "Money and Banking in Search Equilibrium", *International Economic Review*, **46**, 637–670.
- HE, P., HUANG, L. and WRIGHT, R. (2008), "Money, Banking and Monetary Policy", *Journal of Monetary Economics*, **55**, 1013–1024.
- HUANGFU, X. and SUN, H. (2011), "Private Money and Bank Runs", *Canadian Journal of Economics*, **44**, 859–879.
- JOSLIN, D.M. (1954), "London Private Bankers, 1720-1785", *Economic History Review New Series*, **7**, 167–186.
- KEHOE, T.J. and LEVINE, D.K. (1993), "Debt-Constrained Asset Markets", *Review of Economic Studies*, **60**, 865–888.
- KIYOTAKI, N. and WRIGHT, R. (1989), "On Money as a Medium of Exchange", *Journal of Political Economy*, **97**, 927–954.
- KIYOTAKI, N. and WRIGHT, R. (1993), "A Search-Theoretic Approach to Monetary Economics", *American Economic Review*, **83**, 63–77.
- KOEPPL, T., MONNET, C. and TEMZELIDES, T. (2008), "A Dynamic Model of Settlement", *Journal of Economic Theory*, **142**, 233–246.
- KOHN, M. (1999), "Early Deposit Banking" (Dartmouth Economics Working Paper 99-03).
- LELAND, H.E. and PYLE, D.H. (1977), "Informational Asymmetries, Financial Structure and Financial Intermediation", *Journal of Finance*, **32**, 371–387.
- LJUNGQVIST, L. and SARGENT, T. (2004), *Recursive Macroeconomic Theory* (Cambridge: MIT Press).
- MONNET, C. and SANCHES, D. (2012), "Private Liquidity and Banking Regulation" (Mimeo, Federal Reserve Bank of Philadelphia).
- MATTESINI, F., MONNET, C. and WRIGHT, R. (2009), "Banking: A Mechanism Design Approach" (Federal Reserve Bank of Philadelphia Working Paper No. 09-26).
- MILLS, D. (2008), "Imperfect Monitoring and the Discounting of Inside Money", *International Economic Review*, **49**, 737–754.
- MUELLER, R.C. (1997), *The Venetian Money Market, Banks, Panics, and the Public Debt, 1200-1500* (Baltimore: Johns Hopkins University Press).
- MYERS, S.C. and RAJAN, R. (1998), "The Paradox of Liquidity", *Quarterly Journal of Economics*, **113**, 733–771.
- NORSAL, E. and ROCHETEAU, G. (2011), *Money, Payments, and Liquidity* (Cambridge: MIT Press).
- ORSINGHER, R. (1967), *Banks of the World* (New York: Walker and Company).
- PRESSNELL, L. S. (1956), *Country Banking in the Industrial Revolution* (Oxford: Clarendon Press).

- QUINN, S. (1997), "Goldsmith-Banking: Mutual Acceptance and Interbanker Clearing in Restoration London", *Explorations in Economic History*, **34**, 411–432.
- QUINN, S. (2004), "Money, Finance and Capital Markets", in Floud, R. and McCloskey, D. (eds) *The Cambridge Economic History of Britain since 1700*, Vol. I, 1700–1860 (Cambridge/New York: Cambridge University Press).
- RAJAN, R. (1998), "The Past and Future of Commercial Banking Viewed through an Incomplete Contract Lens", *Journal of Money, Credit and Banking*, **30**, 524–550.
- SANCHES, D. and WILLIAMSON, S. (2010), "Money and Credit with Limited Commitment and Theft", *Journal of Economic Theory*, **145**, 1525–1549.
- SELGIN, G. (2007), "Banking", entry written for *Encyclopedia Britannica*.
- SELGIN, G. (2010), "Those Dishonest Goldsmiths", (Mimeo).
- SPUFFORD, P. (1988), *Money and Its Uses in Medieval Europe* (Cambridge/New York: Cambridge University Press).
- THOMAS, J. and WORRALL, T. (1988), "Self-Enforcing Wage Contracts", *Review of Economic Studies*, **55**, 541–554.
- TOWNSEND, R. (1987), "Arrow-Debreu Programs as Microfoundations of Macroeconomics", in Bewley, T. (ed) *Advances in Economic Theory: Fifth World Congress* (Cambridge/New York: Cambridge University Press).
- TOWNSEND, R. (1988), "Models as Economies", *Economic Journal*, **98**, 1–24.
- WALLACE, N. (1998), "A Dictum for Monetary Theory", *Federal Reserve Bank of Minneapolis Quarterly Review*, **Winter**, 20–26.
- WALLACE, N. (2001), "Whither Monetary Economics?" *International Economic Review*, **42**, 847–869.
- WALLACE, N. (2005), "From Private Banking to Central Banking: Ingredients of a Welfare Analysis", *International Economic Review*, **46**, 619–636.
- WALLACE, N. (2010), "The Mechanism Design Approach to Monetary Theory", in Friedman, B. and Woodford, M. (eds) *Handbook of Monetary Economics*, 2nd edn (Amsterdam: Elsevier).
- WEATHERFORD, J. (1997), *The History of Money* (New York: Three Rivers Press).
- WILLIAMSON, S. (1986), "Costly Monitoring, Financial Intermediation and Equilibrium Credit Rationing", *Journal of Monetary Economics*, **18**, 159–179.
- WILLIAMSON, S. (1987), "Financial Intermediation, Business Failures, and Real Business Cycles", *Journal of Political Economy*, **95**, 1196–1216.
- WILLIAMSON, S. and WRIGHT, R. (2010a), "New Monetarist Economics: Methods", *Federal Reserve Bank of St. Louis Review*, **92**, 265–302.
- WILLIAMSON, S. and WRIGHT, R. (2010b), "New Monetarist Economics: Models", in Friedman, B. and Woodford, M. (eds) *Handbook of Monetary Economics*, 2nd edn (Amsterdam: Elsevier).