Banking: A Mechanism Design Approach[*]

Fabrizio Mattesini   Cyril Monnet

University of Rome Tor-Vergata   Federal Reserve Bank of Philadelphia

Randall Wright

University of Wisconsin - Madison and

Federal Reserve Bank of Philadelphia

February 10, 2010

Abstract

We study banking using mechanism design, without prior assumptions about
what banks are, who they are, or what they do. Given preferences, technologies,
and frictions – including imperfect commitment, monitoring and collateral – we
characterize incentive feasible and efficient allocations, and interpret the out-
comes in terms of institutions that resemble banks. Our bankers accept deposits
and make investments, and their liabilities help others in making transactions
(like bank notes, checks or debit cards). This activity is essential: without
it, the set of feasible allocations would be inferior. We discuss how many and
which agents should be bankers. Agents who are more patient, more visible,
have a bigger stake in the system, or have a lower abiltity to liquidate collateral
for strategic reasons make better bankers, because they are less inclined to re-
nege on obligations. Other things equal, bankers should have good investment
opportunities, but it can be efficient to sacrifice return by using a bank that is
more trustworthy – less inclined to renege – since this can aid in making other
transactions. We compare these predictions with banking history.

1 Introduction

The goal of this project is to study banking, without making prior assumptions about what banks are, who they are, or what they do. To this end, we adopt the approach of mechanism design. This method, in general, begins by describing an economic environment, including preferences, technologies, and certain frictions – by which we mean spatial or temporal separation, information problems, commitment issues, and so on. One then studies the set of allocations that are attainable, respecting both resource and incentive feasibility constraints. Sometimes one also describes allocations that are optimal according to particular criteria. One then looks at these allocations and tries to interpret the outcomes in terms of institutions that can be observed in actual economies. We want to see if something that looks like banking emerges out of such an exercise. To reiterate, we do not take a bank as a primitive concept. Our primitives are preferences, technologies and frictions, and we want banking to arise endogenously.

Much has been written about the virtues of mechanism design in general.[1] Our particular approach is close to that advocated by Townsend (1987, 1990). He describes the method as asking if institutions that we see in the world, such as observed credit or insurance arrangements, can be derived from simple but internally consistent economic models, whereby internal consistency we mean that one cannot simply assume a priori that some markets are missing, contracts are incomplete, prices are sticky, etc. Of course, something that looks like missing markets or incomplete contracts may emerge, but the idea is to specify an environment explicitly and derive this as an outcome.[2] Simple models, with minimal frictions, often do not generate

---

[1] For an informal description of mechanism design, following the Nobel prize going to some of its pioneers, see http://nobelprize.org/nobel_prizes/economics/laureates/2007/ecoadv07.pdf.

[2] As Townsend (1988) puts it: "The competitive markets hypothesis has been viewed primarily as a postulate to help make the mapping from environments to outcomes more precise ... In the end though it should be emphasised that market structure should be endogenous to the class of general equilibrium models at hand. That is, 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." Relatedly, speaking more directly about banking, Williamson (1987) says "what makes financial intermediation potentially worthy of study are its special functions (such as diversification,

arrangements that resemble those in actual economies; for example, they typically predict that credit and insurance work better than the institutions we observe. So, one asks, what additional complications can be introduced to bring the theory more in line with experience? We want to apply this method to banking.

Obviously some frictions are needed, since models like Arrow-Debreu have no role for banks. As has been discussed often, frictionless models have no role for any institution whose purpose is to facilitate the process of exchange. The simplest example is the institution of money, and a classic challenge in monetary economics is to ask what frictions make money essential, in the following sense: Money is said to be essential when the set of allocations that satisfy incentive and other feasibility conditions is bigger or better with money than without it.[3] We study the essentiality of banks in the same sense. Just like monetary economists ought not take the role of money as given, for this issue, we cannot take banks as primitive. In our environment, the planner – or the mechanism – may choose to have some agents perform certain functions resembling salient elements of banking: they accept deposits and make investments, and their liabilities (claims on deposits) are used by others to facilitate exchange. This activity is essential, in the sense that if it were ruled out the set of feasible allocations would be inferior.

The vast literature on banks and financial intermediation is surveyed by e.g. Gorton and Winton (2002) and Freixas and Rochet (2008). Much of this research is based on informational frictions, including adverse selection, moral hazard, and costly state verification, that hinder the channeling of funds from investors to entrepreneurs. One can distinguish broadly three main strands. One approach originating with Diamond and Dybvig (1983) interprets banks as coalitions of agents providing insurance against liquidity shocks. Another approach pioneered by Leland and Pyle (1977) and developed by Boyd and Prescott (1986) interprets bank as coalitions sharing infor-

---

information processing, and asset transformation). We cannot expect to generate these special activities or derive many useful implications if our approach does not build on the economic features that cause financial intermediaries to arise in the first place."

[3]This notion of essentiality in monetary economics is usually attributed to Frank Hahn; for recent discussions, see Wallace (2001, 2008).

mation in ways that induce agents to truthfully reveal the quality of investments. A third approach based on Diamond (1984) interprets banks as delegated monitors taking advantage of returns to scale. These papers provide many useful insights. We think we have something different to offer, which complements existing models, but also helps shed new light on several issues, perhaps especially when we study which agents should become bankers, and when we highlight the role of their liabilities in facilitating payments.[4]

Relative to information-based theories, we focus on limited commitment (although imperfect monitoring is also part of the story). We are of course not the first to highlight commitment issues. Rajan (1998) has criticized standard banking theory on the grounds that is typically assumes agents have a perfect ability to contract, and argues instead for model that rely on incomplete contracting, or incomplete markets, based on limited enforcement (see also Calomiris and Kahn 1991, Myers and Rajan 1998, and Diamond and Rajan 2001). We agree that limited enforcement or commitment should be central, but rather than taking the degree of market incompleteness as given, we want to delve into this a little further, using the tools of mechanism design. While we think our approach to banking is novel, there is related work on money and credit, including Sanches and Williamson (2009), Nosal and Rocheteau (2009), Koeppl, Monnet and Temzelides (2008), and Andolfatto (2008), and there is much work in general equilibrium with macro applications building on the models of limited commitment in Kehoe and Levine (1993, 2001). These papers all study environments that are similar to ours, although the applications are different.

---

[4]Work on the Diamond-Dybvig model is a large branch of the literature; see e.g. Jacklin (1987), Wallace (1988, 1990), Peck and Shell (2003), Green and Lin (2003), Andolfatto et al. (2007), and Ennis and Keister (2008). Usually these models do not interpret the bank as a self-interested agent, but as a contract or a mechanism, nor do they derive which agents should be bankers. In the papers that emphasize information sharing or delegated monitoring, banks are agents, but their role is restricted to solving information problems, and again they typically do not derive which agents will play this role. The fact that bank liabilities are useful in transactions is usually not discussed at all; for exceptions see Andolfatto and Nosal (2008), Huangfu and Sun (2008), Kiyotaki and Moore (2005), He et al. (2005, 2008), Cavalcanti and Wallace (1999a, 1999b), Wallace (2005), and Mills (2008), but arguably these are better characterized as papers in monetary theory trying to get something that looks like a bank into the model, rather than mainstream papers on banking that study the role of their obligations in the exchange process.

Commitment issues are central because banking concerns intertemporal reallocation, and we want to take seriously dynamic incentives to make good on one's obligations. Agents in our model have investment opportunities, which they can in principle use as collateral to ameliorate commitment problems. But this does not work well if investments are easily liquidated – if e.g. a debtor can simply consume the proceeds, an investment cannot credibly be used as collateral.[5] An implication is that delegated investment may be useful. If you deposit resources with a third party, who has less incentive or ability to liquidate for strategic reasons, others will be more willing to extend you credit. Thus, claims on deposits can facilitate other transactions, and this resembles banking. Other things being equal, it is better if a bank has good investment opportunities, but it may be efficient to sacrifice rate of return by depositing with one that is more trustworthy – less inclined to renege – since this helps facilitate other transactions. This puts in new perspective Hicks' (1935) rate of return dominance puzzle: our agents hold assets with lower rates of return, because these aid in exchange, because they constitute the liabilities of more trustworthy parties that we interpret as banks.

The idea is obviously correct that sellers often accept the obligations of third parties, which throughout history took the form of notes, checks, credit/debit cards, or other instruments issued by commercial banks, when they would not accept one's personal IOU. Of course this begs the question, why is a bank less inclined to renege on obligations? In the model, future rewards and punishments mitigate strategic behavior, so patience is relevant, but monitoring is imperfect (opportunistic behavior is detected only probabilistically). Agents with a higher likelihood of being monitored, or greater visibility, have more incentive to make good on obligations, and so they are better suited for the responsibility of accepting and investing deposits.[6] However,

---

[5]This is related to Kiyotaki and Moore (2008), Mills and Reed (2008), and references therein.

[6]In terms of the literature, imperfect monitoring has been studied by many people, but in theories of money and banking it is worth mentioning Kocherlakota (1998), Kocherlakota and Wallace (1998), and Cavalcanti and Wallace (1999a, 1999b). Our version, where agents are monitored probabilistically, differs from theirs, and also from imperfect public monitoring in game theory where players cannot observe only signals about each other's actions (see Mailath and Samuelson, 2006).

we go beyond simply assuming that some can be monitored while others cannot, by allowing agents to have different probabilities of gaining from economic activity, or different stakes in the system. Even with equal visibility, those with higher stakes are less inclined to deviate from proscribed behavior because they have more to lose. This allows us to endogenize monitoring when we analyze which agents, and how many, should be bankers.

Summarizing, we show that agents are better suited to perform the activities of banking (accepting and investing deposits) to the extent that they have a good combination of the following characteristics making them more trustworthy (less inclined to renege on obligations):

- they are relatively patient;

- they are more visible, or more easily monitored;

- they have a greater stake in, or connection to, the economic system;

- they have access to relatively good investment opportunities;

- they derive lower payoffs from liquidating investments for strategic reasons.

Some of these findings, like patience relaxing incentive constraints, may be obvious; we think that others are more subtle, like the idea that it can be better to delegate your investments to parties with a greater stake in the system, even if they have relatively poor investment opportunities, because their trustworthiness facilitates other transactions. And the results are consistent with at least our reading of economic history, as discussed below. Even when the conclusions are straightforward, we think the formal analysis is useful because it makes the effects and the nature of the trade-offs precise. This is the case when we study the trade-off between rate of return and facilitating transactions. And when we choose which agents to monitored and make bankers, we analyze how to select those with the right combination of the above-mentioned characteristics. Similarly, when we discuss the efficient number of

6

banks, we can lay out the trade-off between having fewer, which reduces monitoring costs, and having more deposits per bank, which increases the incentive to misbehave. All of this comes directly out of a mechanism design approach, without making assumptions about what is a bank, who is a bank, or what banks do.

The rest of paper is organized as follows. Section 2 describes the basic environment, emphasizing the roles of temporal separation, limited commitment, collateral and monitoring. Section 3 characterizes incentive feasible and optimal allocations in a baseline version of the model. This version has a single group of agents that are heterogeneous with respect to type – so that at various points in time some want to borrow while others want to lend – but are homogenous within a given type. Section 4 considers multiple groups, in the sense that a given type can differ across groups with respect to visibility, connection to the system, etc. This section contains the main results on essentiality. Section 5 generalizes the analysis to discuss which individuals are best suited for banking, how to monitor when it is costly, and rate-of-return dominance. Section 6 reviews some banking history. Section 7 concludes.

2 The Environment

Time is discrete and continues forever. Agents belong to one of $N \geq 1$ groups, and in each group they can be one of 2 types. Within a group, agents of a given type are homogeneous, while across groups types can be heterogeneous. The role of heterogeneous groups will be clear later; for now we focus on a representative group with a set of agents $\mathcal{A}$.[7] Each period, all agents of type j in the group can be active or inactive, and we partition $\mathcal{A}$ into three subsets: inactive agents $\mathcal{A}_0$; active type 1 agents $\mathcal{A}_1$; and active type 2 agents $\mathcal{A}_2$. These sets have measure γ_0, γ_1, and γ_2, respectively, and type j agents take as given that they belong each period to $\mathcal{A}_j$ or $\mathcal{A}_0$ with probabilities γ_j and $1 - \gamma_j$. To ease ease the presentation, without affecting too many results, we set $\gamma_1 = \gamma_2 = \gamma$. Active agents can produce, consume, and

---

[7]No special restirctions on $\mathcal{A}$ are needed – it could be a continuum, countably infinite, or finite with as few as two agents.

derive utility each period as described below. Inactive agents get utility normalized to 0, say, because they have no desire to consume or ability to produce, that period. Letting γ differ across groups captures the idea that they can have different degrees of connection to the economic system: a bigger γ means one has more frequent gains from trade, and hence more at stake.

In each period there are two goods, 1 and 2. Agents in $\mathcal{A}_1$ consume good 1 and produce good 2, while agents in $\mathcal{A}_2$ consume good 2 and produce good 1. Letting x_j and y_j denote consumption and production by type j, we assume utility $U^j(x_j, y_j)$ is increasing in x_j, decreasing in y_j, satisfies the usual differentiability and curvature conditions, and $U^j(0,0) = 0$. When convenient, we also assume normal goods. A key friction is temporal separation: each period is divided in in two, and good j must be consumed in subperiod j. This generates a role for credit, since type 1 must consume before type 2. To have a notion of collateral, we assume good 2 is produced in the first subperiod and invested in some way that delivers goods for consumption in the second subperiod, with a fixed gross return $\rho \geq 1$. Investment here may be as simple as pure storage (perhaps for safe keeping). Or it could involve physical capital, or any other project; it is purely for simplicity that we assume a fixed return ρ.[8]

We do not allow type 2 agents to invest for themselves – or at least not as efficiently as type 1 – since if they could there would be no gain from intertemporal trade. Thus, only a producer of good 2, a type agent 1, can invest it. We sometimes (but only for the discussion of implementation at the end of Section 4) make the assumption that only type 1 can transport good 2 across groups, just like only type 1 can invest good 2 over time. Also, in the formal model agents all discount across periods at the same rate β, but in the economic discussion we sometimes proceed as if patience differed across groups, since it is apparent what would happen if it did. This is mainly to reduce notation, but also to avoid some technical issues that can arise with heterogeneous discount rates. Our treatment of differences in patience is therefore

---

[8]There are no investment opportunities across periods, only across subperiods, again purely for simplicity.

relatively heuristic, but we still think it is useful. We are more rigorous in modeling differences in visibility, connection to the system, and the opportunity to liquidate collateral.

Suppose we offer type 1 good 1, in exchange for good 2 that will be produced in the first subperiod, invested, and delivered in the second subperiod. In a real sense, they are getting a loan to consume good 1, with a promise to deliver good 2 later, backed by their investment. Such a collateralized loan works very well if type 1 agents get no payoff from consuming or otherwise liquidating the returns on the investment, since when it comes time to deliver the goods, the production cost has been sunk. To make it work less well, we let type 1 derive payoff λ per unit liquidated out of investments, over and above the payoff $U^1(x_1, y_1)$. If $\lambda = 0$, as we said, collateral works well, but if $\lambda > 0$ there is an opportunity cost to delivering goods even if the production cost is sunk. We assume $U^1(x_1, y_1) + \lambda \rho y_1 \leq U^1(x_1, 0)$ for all x_1, so that it is never efficient for type 1 to produce and invest for their own consumption. Also, a type 1 agent derives the same liquidation payoff from any good 2, even if it was produced by another type 1 agent, including one from a different group. But for type $j = 1, 2$ agents in any group i, only goods produced within the same group i enter their utility functions (this is what defines a group).

We focus on symmetric and stationary allocations, given by vectors $(x_1^i, y_1^i, x_2^i, y_2^i)$ for each group i, and when there is more than one group, descriptions of cross-group transfers, investment, and liquidation. We sometimes proceed with the discussion as if the planner, or mechanism, collects all production and allocates it to consumers; this is merely for convenience. The mechanism does not produce or invest goods – the planner's only job is to suggest ways to organize exchange. Assuming for the moment that there are no transfers across groups or liquidation, since $\gamma_1 = \gamma_2 = \gamma$, allocations are resource feasible if $x_1 = y_2$ and $x_2 = \rho y_1$, which means they can be summarized by (x_1, y_1). To reduce notation, we drop the subscript, and write (x, y). This completes the specification of the basic environment.

9

3   A Single Group

3.1   Baseline Results

For now, $N = 1$, so that all a planner or mechanism can do is recommend a resource-feasible allocation (x, y) for the group. This recommendation is incentive feasible, or IF, as long as no one wants to deviate. Although we focus below on the case in which agents cannot commit to future actions, and hence may deviate whenever they like, we begin with benchmarks where they can commit to some degree. One notion is full commitment, by which we mean they can commit at the beginning of time before they even know their type. In this case, (x, y) is IF as long as the total surplus is positive,

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

Another notion is partial commitment, where agents can commit at the beginning but only after knowing their type. In this case, IF allocations are constrained by two participation constraints

$$U^1(x, y) \ \geq \ 0 \tag{2}$$
$$U^2(\rho y, x) \ \geq \ 0. \tag{3}$$

   With no commitment, at the start of each period, we face the participation conditions

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

where $V^j(x, y)$ is the continuation value of type j. In these conditions, the LHS is type j's payoff from following the recommendation, while the RHS is the payoff from deviating, where as usual, without loss in generality we can restrict attention to one-shot deviations. 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 obviously is the most effective). But with probability $1 - \pi$, deviations go

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 the dynamic participation conditions hold iff (**??**)-(**??**) hold.

Moreover, type 1 consumes, produces and invests all in the first subperiod, in exchange for a promise to deliver good 2 in the second, but he can always renege and liquidate the investment for a short-term gain $\lambda \rho y$. If he is caught, he is punished with autarky, but again, he is only caught with probability π. This random monitoring technology, once we allow heterogeneity in π, captures the idea is that some agents are more visible than others, and hence less likely to get away with reneging on obligations. Thus, agent 1 delivers the goods in the second subperiod only if

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

where the RHS involves a deviation by liquidating the investment, which is detected with probability $1 - \pi$. Inserting $V^1(x, y)$ and letting $\delta \equiv \lambda (1 - \beta) / \pi \gamma \beta$, this simplifies to what we call the repayment constraint

$$U^1(x, y) \geq \delta \rho y. \tag{4}$$

Notice that ρy is the obligation – the promised payment – of type 1 when subperiod 2 rolls around, and δ is an effective discount rate he uses in contemplating whether to make good. Intuitively, a low monitoring probability π, a low rate of time preference β, a low stake in economic activity γ, or a high liquidation value λ all make δ big, and hence increase the temptation to renege. We think it is fair to call an agent more trustworthy when he has smaller δ, since this makes him more inclined to honor a given promise. More trustworthy agents can get bigger loans, precisely because they can credibly promise offer bigger repayments.

Let $\mathcal{F}_0$ denote the set of IF allocations with no commitment. Since (**??**) makes (**??**) redundant, $(x, y) \in \mathcal{F}_0$ satisfies the participation constraint (**??**) for type 2 and the repayment constraint (**??**) for type 1. For comparison, the IF set with partial commitment $\mathcal{F}_P$ satisfies (**??**)-(**??**), while the IF set with full commitment $\mathcal{F}_F$ only
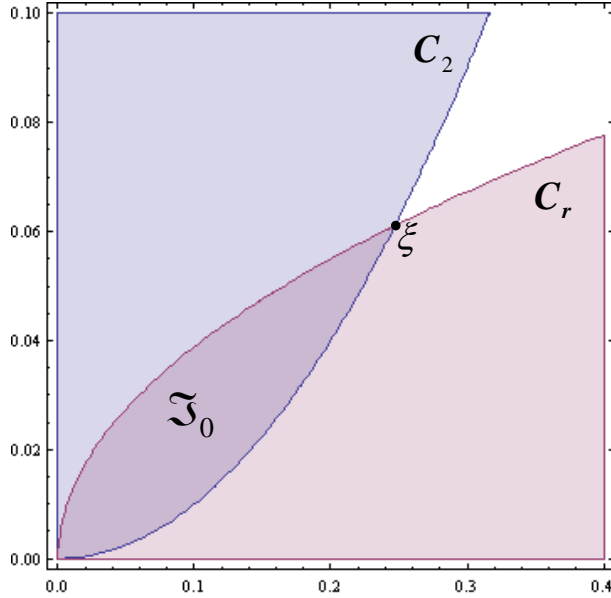
11

Figure 1: Incentive constraints.

requires $S(x, y) \geq 0$. Figure 1 shows $\mathcal{F}_0$ delimited by two curves defined by the relevant incentive conditions at equality,

$$C_2 \equiv \left\{ (x, y) : U^2 \left(\rho y, x \right) = 0 \right\} \tag{5}$$

$$C_r \equiv \left\{ (x, y) : U^1 \left(x, y \right) = \delta \rho y \right\}. \tag{6}$$

Clearly, $\mathcal{F}_0$ is convex and compact. It is nonempty, since $(0, 0) \in \mathcal{F}_0$, and it contains other points as long as there are gains from trade, which there are under the usual Inada conditions.

Let ξ be the unique point other than $(0, 0)$ where C_2 and C_r intersect, as shown in Figure 1. The following obvious result will be useful:

Lemma 1 *If* $\delta^b < \delta^a$, *then* ξ^b *lies northeast of* ξ^a *in* (x, y) *space.*

Proof: An increase in δ rotates C_r down, without affecting C_2. ∎

We now define some notions of Pareto optimal, or PO, allocations. The ex ante PO allocation, which would seem most relevant under full commitment, is the (x^o, y^o) that maximizes $S(x, y)$. With partial commitment, a welfare natural criterion is defined

by maximizing ex post welfare

$$\max_{(x,y)} \mathcal{W}(x, y) = \omega_1 U^1(x, y) + \omega_2 U^2(\rho y, x) \tag{7}$$

for some weights ω_1 and ω_2. As we vary these weights, we get the contract curve

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

The core, which is a natural set to study under partial commitment, unconstrained by repayment, is $\mathcal{K}_P = \mathcal{P} \cap \mathcal{F}_P$. One notion of the constrained core is $\mathcal{K}_0 = \mathcal{P} \cap \mathcal{F}_0$. While $\mathcal{K}_P$ is nonempty as long as there are gains from trade, $\mathcal{K}_0$ might not be. Of course, we can define an alternative notion, say $\mathcal{K}$, as the solution to (??) s.t. $(x, y) \in \mathcal{F}_0$ as we vary the weights. Since $\mathcal{F}_0$ is compact and $\mathcal{W}$ is continuous, $\mathcal{K} \neq \emptyset$.

The following rudimentary results are also useful:

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

Proof: (??) defines y as a function of x with

$$\frac{dy}{dx} = -\frac{\rho U_1^1 (U_{22}^2 U_1^2 / U_2^2 - U_{12}^2) - U_2^2 (U_{11}^1 U_2^1 / U_1^1 - U_{21}^1)}{\rho [U_1^2 (U_{22}^1 U_1^1 / U_2^1 - U_{12}^1) - U_2^1 (U_{11}^2 U_2^2 / U_1^2 - U_{12}^2)]}.$$

By the assumption that all goods are normal – i.e. in a standard utility maximization problem, for both types, consumption is increasing and production decreasing in wealth – all four terms in parentheses in dy/dx positive. ∎

Lemma 3 Let $(\hat{x}, \hat{y})$ maximize $\mathcal{W}(x, y)$ s.t. $(x, y) \in \mathcal{F}_0$. If the repayment constraint (??) is not binding then $(\hat{x}, \hat{y}) \in \mathcal{P}$.

Proof: Define the Lagrangian

$$\mathcal{L} = \omega_1 U^1(x, y) + \omega_2 U^2(\rho y, x) + \eta U^2(\rho y, x) + \varphi \left[U^1(x, y) - \delta \rho y \right] \tag{9}$$

where η and φ are multipliers. The FOCs are

$$\omega_1 U_1^1 + \omega_2 U_1^2 + \eta U_1^2 + \varphi U_1^1 = 0$$

$$\omega_1 U_2^1 + \omega_2 \rho U_2^2 + \eta \rho U_2^2 + \varphi \left(U_2^1 - \delta \rho \right) = 0$$
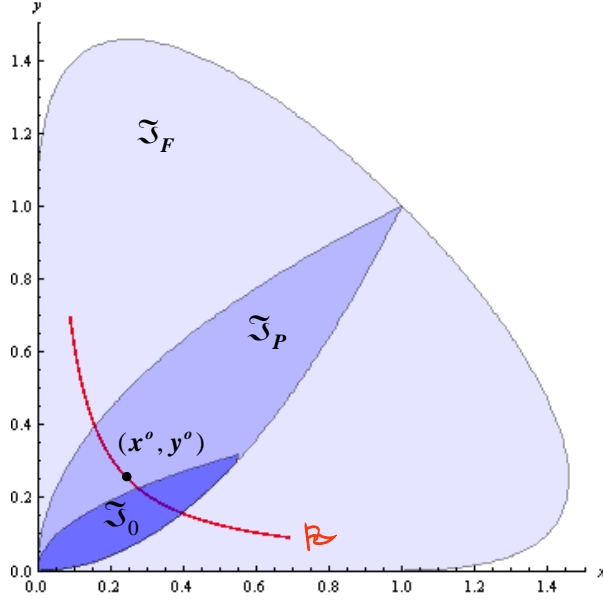
13

Figure 2: Pareto and incentive feasible sets.

plus the constraints. Rearranging implies

$$\frac{(\omega_1 + \varphi) U_1^1}{(\omega_1 + \varphi) U_2^1 - \varphi \delta \rho} = \frac{(\omega_2 + \eta) U_1^2}{\rho (\omega_2 + \eta) U_2^2}.$$

If (??) is not binding then $\varphi = 0$, and $\rho U_1^1/U_2^1 = U_1^2/U_2^2$, which means $(\hat{x}, \hat{y}) \in \mathcal{P}$. ∎

Figure 2 shows the IF set when we have no commitment $\mathcal{F}_0$, partial commitment $\mathcal{F}_P$, and full commitment $\mathcal{F}_F$, for an example with $U^1(x, y) = \sqrt{x} - y$, $U^2(\rho y, x) = \sqrt{\rho y} - x$, $\beta = 3/4$, $\gamma = \pi = 1/2$ and $\rho = \lambda = 1$. Clearly, $\mathcal{F}_0 \subset \mathcal{F}_P \subset \mathcal{F}_F$, showing how commitment matters. In this example, $\mathcal{K}_0 = \mathcal{P} \cap \mathcal{F}_0 \neq \emptyset$ (again, this is not always the case). Focusing on no commitment, since as we said an increase in δ rotates C_r down but does not affect C_2, $\mathcal{F}_0$ and $\mathcal{K}$ shrink when δ increases. Also, $\mathcal{F}_0$ and $\mathcal{K}$ shrink when ρ decreases.[9] We think this is an interesting, albeit stylized, model of credit with imperfect commitment, monitoring and collateral. So far, however, it has nothing to say about banks; that comes in the next Section, along with heterogeneity across groups. Before proceeding, we mention some ways in which we can change the basic environment without affecting the main results.

---

[9]This may be less obvious, since a change in ρ shifts both curves C_r and C_2 in (x, y) space; to easily verify the claim, redraw the picture in $(x, \rho y)$ space.

14

3.2 Alternative Specifications

Rather than having permanently different types, suppose each agent is randomly selected to be type 1 or type 2 at every date. Then the continuation payoff is the same for all agents, $V(x,y) = \gamma S(x,y)/(1-\beta)$, where S is the surplus (**??**). Each period, after types are realized, the participation constraints are

$$U^1(x,y) + \frac{\pi\gamma\beta}{1-\beta}S(x,y) \geq 0$$

$$U^2(\rho y, x) + \frac{\pi\gamma\beta}{1-\beta}S(x,y) \geq 0,$$

while the repayment constraint is

$$\lambda\rho y \leq \frac{\pi\gamma\beta}{1-\beta}S(x,y).$$

Now we can have $U^j < 0$ for one type, and still satisfy these conditions, if S is big. The participation constraint for type 1 is no longer implied by the repayment constraint, so we must check all three conditions in defining $\mathcal{F}_0$. Although the results go through, this complicates the analysis, and so we use permanent types in the base-line model.[10]

Next, for those who do not like production, we can reinterpret the setup as a pure exchange economy. Again assume two permanent types, where now type j agents have preferences and endowments in the first and second subperiod $u^j(x_1^j, x_2^j)$ and (e_1^j, e_2^j). Without loss in generality, label types so that at their endowment points, type 1 has a greater MRS: $u_1^1/u_2^1 \geq u_1^2/u_2^2$. Now efficient allocations involve a loan $L \geq 0$ from type 2 to type 1 in the first subperiod, with repayment R in the second subperiod. Defining

$$U^1(R,L) = u^1(e_1^1 + L, e_2^1 - R)$$

$$U^2(L,R) = u^2(e_1^2 - L, e_2^2 + R),$$

this becomes a special case of our general model, with the opportunity cost of trading one's endowment replacing the production cost. The participation condition for type

---

[10]See the working paper Mattesini et. al. (2009) for all the details of the random-type model. In addition to being less tractable, an unnatural feature of that model is that, once we introduce banking, the bankers will be chosen randomly every period.

j is again $U^j \geq 0$, while the repayment constraint is

$$u^1(e_1^1 + L, e_2^1) - u^1(e_1^1 + L, e_2^1 - R) \leq \beta\pi V^1(R, L).$$

When $u^1 = v(x_1^1) + \lambda x_2^1$, this reduces to $R \leq \delta U^1(L, R)$, with δ exactly as in the basic model. This gives us a simple consumption-loan model with limited commitment, and endogenous credit limits determined by patience, monitoring, and connection to the market. However, it has no investment or collateral, which we want for our discussion of banking. Therefore, consider the following twist on the above specification. Suppose now that type 1 does not consume $e_1^1 + L$, but invests it in the first subperiod at return ρ. In the second period, he gives R to type 2, who derives payoff $U^2(L, R)$, as above, while the investor consumes the residual for a payoff $U^1(L, R) = \lambda [\rho (e_1^1 + L) + e_2^1 - R]$. The repayment constraint is again $R \leq \delta U^1(L, R)$. Type 1 can be said to be borrowing for investment purposes, with the loan collateralized by the investment. Loan come out of endowments, and there is no production, per se.

We can also reinterpret the model as having neoclassical investment. Let type 1 be a firm with payoff $U^1(K, rK) = f(K) - rK$, where f is a standard production function, and r is the price of capital put in place in the first subperiod, paid in the second, after production is complete. Using the endowment version in the previous paragraph, we have $U^2(rK, K) = u^2(e_1^2 - K, e_2^2 + rK)$. Everything goes through as stated above, where now the participation constraint for type 1 can be interpreted as a non-negative profit condition, while his repayment constraint can be written

$$rK \leq \frac{\pi\beta\gamma f(K)}{1 - \beta(1 - \pi\gamma)}.$$

This says the repayment rK cannot exceed an appropriately discounted measure of the future profit flow. Also, γ can be interpreted as a productivity shock here. As these examples show, the framework is flexible.

Also, although we are generally interested in the entire IF set, one can impose various arbitrary mechanisms, and consider different notions of equilibrium. Gu and

Wright (2010) analyze equilibria with generalized Nash bargaining and Walrasian price taking. For the sake of illustration, suppose we partition active agents into pairs containing one agent of each type at the start of every period, and let type 2 make a take-it-or-leave-it offer. With partial commitment, equilibrium selects $(x, y) \in \mathcal{P}$ where the participation constraint for 1 holds at equality. With no commitment, this is not feasible: if $U^1(x, y) = 0$ then 1 reneges for sure. Instead, equilibrium maximizes $U^2(\rho y, x)$ s.t. $(x, y) \in \mathcal{F}_0$, which implies $(x, y) \in \mathcal{K}$ but $(x, y) \notin \mathcal{P}$.[11] For the rest of the paper we study IF allocations, more generally, but since equilibrium for an arbitrary mechanism must be IF, most conclusions apply to any equilibrium as well.

4 Multiple Groups

For the points we want to make, we can set $N = 2$. Labeling the groups a and b, for $i = a, b$ we have: group i has two types 1^i and 2^i, and two goods 1^i and 2^i, specialized as above; both types are active with probability γ^i; type 1^i has liquidation value λ^i; and we detect deviations with probability π^i. For now ρ, β and the cardinality of $\mathcal{A}$ are the same in each group. Let $\delta^i = \lambda^i (1 - \beta) / \pi^i \gamma^i \beta$ and assume $\delta^a > \delta^b$, so that type 1^a have more of a commitment problem than 1^b: if $\mathcal{F}_0^i$ is the IF set for group i, $\mathcal{F}_0^a \subset \mathcal{F}_0^b$. The IF set for the economy as a whole is given by (x^i, y^i) for each group, plus a description of liquidation, transfers and deposits, as discussed below, subject to the relevant incentive constraints. To be clear, recall type 1^i can invest the output of either group a or group b, and the return and liquidation value are the same.[12] Recall also that there are no gains from trade across groups for mercantile reasons, since goods 1^a and 2^a produced by group a do not enter the utility functions of agents in group b, and vice versa. Hence, any interaction across groups will be exclusively due to incentive considerations.

---

[11] The relevant Lagrangian is given by (??) with $\omega_2 = 1$ and $\omega_1 = \eta = 0$. The first order conditions in this case rearrange to $\rho(U_1^1 + \rho y)/U_2^1 = U_1^2/U_2^2$, which means $(x, y) \notin \mathcal{P}$ (the indifference curve for type 1 is steeper than for type 2). The inability to commit – or, perhaps rather, the ability to not commit – means type 1 does better at bargaining.

[12] We could allow ρ or λ to depend on which group produced the good, but it adds little other than notation. It is more interesting to let ρ and λ differ according to who invests the good.

4.1 Transfers

In addition to producing to invest their own output, suppose we have all type 1^b agents produce an extra $t > 0$ units of good 2^b and transfer it to type 1^a, who invest it, and liquidate the returns for their own benefit. Since there are γ^b/γ^a active type 1^b agents for each active type 1^a agent, the payoffs are

$$\hat{U}^1\left(x^a, y^a, t\right) \equiv U^1\left(x^a, y^a\right) + \lambda^a \rho t \gamma^b/\gamma^a \tag{10}$$

$$\hat{U}^1\left(x^b, y^b, t\right) \equiv U^1\left(x^b, y^b + t\right). \tag{11}$$

One can think of t as a lump sum tax on type 1^b agents, the proceeds of which go to their counterparts in group a.[13] This scheme has incentive effects that we want to analyze 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, but instead of liquidating, they transfer the returns back to the first group for consumption. This delegated investment activity we claim is essential, in the sense that it can change the IF set. However, transfers also change the IF set. We claim that delegated investment can do more, and to make the case, we first analyze pure transfers.

The participation conditions for type 2^i agents are, as before,

$$U^2\left(\rho y^i, x^i\right) \geq 0, \, i = a, b, \tag{12}$$

but with transfers the repayment constraints for type 1^i change to[14]

$$\hat{U}^1\left(x^i, y^i, t\right) \geq \delta^i \rho y^i, \, i = a, b. \tag{13}$$

---

[13]Transfers in the other direction are given by $t < 0$, and it is never useful to have simultaneous transfers directions, given $U^1(x, y) + \lambda \rho y \leq U^1(x, 0)$. Also, note the tax is not compulsory – agents can always deviate to autarky.

[14]To be clear, (??) is the incentive condition for type 1^i to make a payment to type 2^i – i.e. to agents in their own group. For a type 2^a agent, who is meant to liquidate the return from investing the transfer from group b, this can be written

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

This condition, which simplifies to (??), says type 1^a do not want to renege on their obligation by liquidating the return from investing their own output, after liquidating the return from investing the transfer.

The IF set with transfer t satisfies (**??**) and (**??**). Notice t only enters these conditions only through $\hat{U}^1(x^i, y^i, t)$. Thus, when it comes time to settle their obligations, t affects the long-run (continuation) values for investors 1^a and 1^b, but not the short-run costs or benefits to reneging. Since type 1^b are better off and type 1^b worse off when $t > 0$, this relaxes the repayment constraints in group a and tightens them in group b. Whenever these constraints are binding in group a but not b, this expands the IF set.

To see just how much we can accomplish with this scheme, consider the biggest transfer from group b to a subject to (**??**) and (**??**). This is a standard maximization problem, with a unique solution $\tilde{t}$ and an implied allocation $(\tilde{x}^i, \tilde{y}^i)$ for each group i. Since the RHS of (**??**) is increasing in δ^b, $\tilde{t}$ rises as δ^b falls (when agents are more patient, more visible, or more connected to the system, we can extract more from them). By way of example, let $U^1(x, y) = x - y$, $U^2(\rho y, x) = u(\rho y) - x$, and, to make the case stark, $\lambda^b = 0$. Then IF allocations in group b solve

$$u\left(\rho y^b\right) - x^b \geq 0 \tag{14}$$

$$x^b - y^b - t \geq 0. \tag{15}$$

The maximum IF transfer 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 y^* solves $\rho u'(\rho y) = 1$. Notice (**??**)-(**??**) hold with equality.

In this example, with transfer $\tilde{t}$, production by type 1^b agents $\tilde{y}^b$ is efficient, type 2^b agents give all of their surplus to 1^b by producing $\tilde{x}^b$, and we tax away the entire surplus of group b, because with $\lambda^b = 0$ we do not have to worry about repayment.[15] Giving the proceeds of this tax to type 1^a agents allows us to relax the incentive constraint in group a, since 1^a now have more to lose if they are caught cheating. Giving them $\tilde{t}$ is the best we can do to relax their constraints, since if the tax were bigger there would be defection by group b. The main point is that with t we can

---

[15]We can easliy relax $\lambda^b = 0$. In general, in this example, the maximum transfer is $\tilde{t} = u(\rho\tilde{y}) - \left(1 + \delta^b\rho\right)\tilde{y}$, where $\rho u'(\rho\tilde{y}) = 1 + \delta^b\rho$. Notice $\partial\tilde{y}/\partial\delta^b < 0$ and $\partial\tilde{t}/\partial\delta^b < 0$.

change the IF set. It is no surprise that tax-transfer schemes have incentive effects; we present the results only so we can conclude below that deposits do even more.

4.2 Deposits

Let $d \geq 0$ denote deposits, defined as follows. Deposits are units of good 2^a produced by type 1^a and transferred to type 1^b, who invest it, but rather than liquidating the return as they did with pure transfers, now type 1^b agents pay it back to group a for consumption by type 2^a. As with pure transfers, we can consider deposits going the other way by setting $d < 0$. We call this delegated investment, since $d > 0$ entails 1^b investing the output of 1^a, Clearly, this has nothing to do with 1^b having better investment opportunities than 1^a, since for now ρ is the same for both groups. Instead, it is all about incentives.

We still face the participation conditions (??) for type 2^i in each group, but the repayment conditions change as follows. Since type 1^a is now only obliged in the second subperiod to pay $\rho (y^a - d)$, their constraint is

$$\hat{U}^1(x^a, y^a, t) \geq \delta^a \rho (y^a - d), \tag{16}$$

and since type 1^b is obliged to pay $\rho \left(y^b + d\gamma^a/\gamma^b\right)$, theirs is

$$\hat{U}^1 \left(x^b, y^b, t\right) \geq \delta^b \rho \left(y^b + \gamma^a d/\gamma^b\right), \tag{17}$$

where these conditions allow transfers, in addition to deposits, since they use the payoffs defined in (??) and (??). We also face a resource constraint

$$0 \leq d \leq y^a. \tag{18}$$

Putting these together, 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 (??) and (??)-(??).

Notice that we relax the repayment constraint in group a while tightening it in group b with $d > 0$, as we did before with $t > 0$. But it is critical to understand that deposits are different than transfers in how they impact incentives: t only affects

continuation payoffs, while d affects directly the within-period benefits to reneging by changing the obligations of types 1^a and 1^b. With this in hand, we can present the result that delegated investment is essential, in the sense that if we start with $d = 0$, and then introduce deposits, the IF set expands.[16]

Proposition 1 $\mathcal{F}_0 \subset \mathcal{F}_d$ *and for some parameters* $\mathcal{F}_d \backslash \mathcal{F}_0 \neq \emptyset$.

Proof: Since any allocation in $\mathcal{F}_0$ can be supported once deposits are allowed by setting with $d = 0$, it is trivial that $\mathcal{F}_0 \subset \mathcal{F}_d$. To show that more allocations may be feasible with deposits, it suffices to give an example. To make the example easy, set $\lambda^b = 0$, so that holding deposits does not affect the incentive constraints for group b. We claim that 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 group b to a, as discussed in the previous section. 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 (**??**) and (**??**) 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 and hence expands the IF set. ■

The example in the proof has $\lambda^b = 0$, which means type 1^b agents have no incentive to renege, and their investments are perfect collateral; but it is easy to construct examples with $\lambda^b > 0$ (see Mattesini et al. 2009). Also, as long as it does not violate the repayment constraint for type 1^b, which is certainly true if $\lambda^b = 0$, we could set $d = y^a$ and let 1^b invest *all* output of 1^a. In this case, the repayment constraint for 1^a reduces to $\hat{U}^1(x^a, y^a, \tilde{t}) \geq 0$, which is their participation condition. Thus, when λ^b is small, having agents in group a delegate all investment eliminates their commitment problem entirely. Similarly, suppose $\pi^a \approx 0$, which means 1^a would never repay a

---

[16]Notice that we are not claiming $\mathcal{F}_0 \subset \mathcal{F}_d$ for any fixed $d = \bar{d} > 0$, since then the repayment constraint in group b may be violated for some allocation in $\mathcal{F}_0$. The claim is rather that deposits are essential when we get to choose d.

loan. In this case, credit, investment and exchange cannot even get off the ground in group a unless type 1^a deposits some output with 1^b.

As an example, consider $U^1(x, y) = u(x) - y$ and $U^2(y, x) = y - x$, $\rho = 1$, $\delta^i = \delta$, $\lambda^i = \lambda < 1$ and $\omega_1^i = 1$. Since $\omega_1^i = 1$, type 2 get no surplus, and so we know $x^i = y^i$ in both groups. This allows us to represent the IF set for the economy as a whole in two dimensions, by $(x_a, x_b) \in \mathbb{R}_+^2$ satisfying

$$u(x^a) - x^a + \lambda t \geq \delta(x^a - d) \tag{19}$$

$$u(x^b) - x^b - t \geq \delta(x^b + d), \tag{20}$$

where now group b makes a transfer t and accepts deposits d. We claim that using d allows us to expand the IF set beyond what we can achieve with t. To see this, note that $t > 0$ relaxes the constant for group a (**??**) by λt while tightening (**??**) by t. Now consider deposits $d = \lambda t / \delta$. This relaxes (**??**) for group a by the same amount λt, but only tightens (**??**) by $\lambda t < t$.

Therefore, we can relax groups a's repayment constraint by a given amount with less tightening of the constraint for group b by using deposits. Figure 3 shows the IF sets for group in (x_a, x_b) space for several cases. First, $t = d = 0$ is shown by the red square. Then using transfers from group b to group a, but no deposits, the IF set expands by the dark blue area. Symmetrically, using transfers from group a to group b, but no deposits, the IF set expands by the dark red area. Using deposits from group a to b, the IF set expands more, now also including the light blue area. Symmetrically, and finally, the IF set expands by the light red area with deposits going from group b to a. This illustrates how deposits are essential.

In terms of economics, suppose you want consumption now, and pledge to deliver something in return out of your investments. When the time comes to make good, you are faced with a temptation to renege and liquidate the investments for your own benefit. This limits your credit. By depositing resources with a third party, who invests for you, the temptation is relaxed. Of course, we must consider the temptation of the third party, in general, although this is a non-issue if $\lambda^b \approx 0$. But as long as
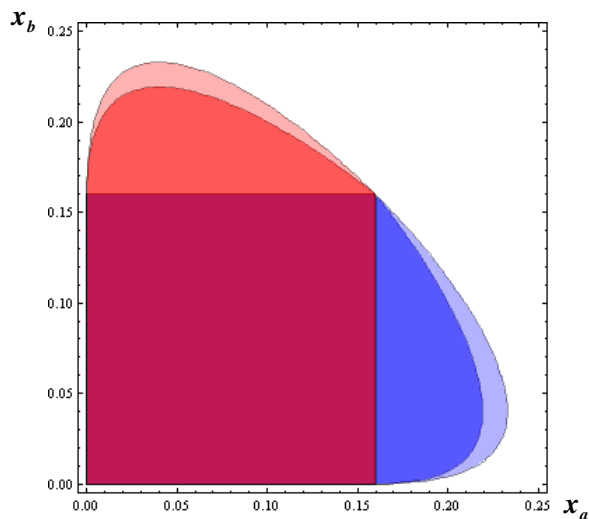
Figure 3: An example where deposits are essential.

the third party is more trustworthy, in the sense made precise earlier, $d > 0$ can be beneficial. We interpret the third party as a bank, since it accepts deposits and makes investments, and moreover its liabilities (claims on deposits) help facilitate transactions. They facilitate transactions in the sense that you get more today if your promise of repayment is backed by deposits – by your banker's good name, so to speak.

Although we are less concerned with details of implementation than with describing feasible allocations, the following discussion ought to make it even clearer that it makes sense to call type 1^b bankers. Suppose we ask, how should we keep track of the exchanges between agents? One way, which is especially nice when record keeping is costly, is this: When type 1^a wants to consume, in the first subperiod, he produces and deposits the output with type 1^b in exchange for a receipt. He gives type 2^a the receipt in exchange for good 1^a. Type 2^a takes this receipt, which is backed by a promise of 1^b, while he would less readily accept the personal promise of 1^a. Type 2^a carry the receipt until they want to consume, when they redeem it for good 2^a. Banker 1^b pays 2^a out of deposits – principle plus returns on investment – which "clears" the receipt. The same banker 1^b generally makes also payments to group

b depositors from the returns on investing his own output. This works because the bank is trustworthy.

To us, this evidently resembles banking, with receipts acting as inside money, as have various bank-issued instruments over time, from notes to checks to debit cards. When we say it resembles banking, we mean that this activity is consistent with what the general public and standard references regard as banking. As Selgin (2006) puts it, "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." The receipts described above play precisely this role.[17]

5 Extensions and Applications

Here we explore several issues. We first study who should hold deposits. We then analyze how to monitor when it is costly. Finally, we discuss rate-of-return dominance.

5.1 Who Should Hold Deposits?

Given two groups with $\delta^a > \delta^b$ and $\rho^a = \rho^b = \rho$, we claim that it may be desirable in a Pareto sense to have group a agents deposit resources with group b, but it is never desirable to have group b deposit with a. Let $(\hat{x}^i, \hat{y}^i)$ be the best IF allocation for group i with no transfers or deposits, solving

$$\max_{x^i, y^i} \mathcal{W}^i \left(x^i, y^i \right) \ \text{ s.t. } \ \left(x^i, y^i \right) \in \mathcal{F}_0^i \tag{21}$$

---

[17]We like the story about circulating receipts for several reasons, including their advantage in record keeping, but there could in principle be other ways to proceed. Commenting on an earlier version of our model, Chris Phelan suggested that 1^a could give output to 2^a, instead of a receipt, then 2^a could give it to 1^b to invest. This also looks something like banking, or at least delegated investment, although it does not have inside money, and it is really beneficial only because 1^b has better investment opportunities. In any case, following the tradition of determining which objects serve as means of exchange based on properties like storability and portability, going back at least to Menger (1892), we can rule out the above arrangement with the assumption that 2^a agents cannot transport first-subperiod goods. This means that receipts have to do the circulating.

with $\mathcal{W}$ defined in (**??**) for some welfare weights. At $(\hat{x}^i, \hat{y}^i)$, obviously, no IF allocation for group i makes 1^i better off without making 2^i worse off, and vice-versa. Then we ask, given $(\hat{x}^a, \hat{y}^a)$ and $(\hat{x}^b, \hat{y}^b)$, can we make one group better off without hurting the other? Clearly transfers cannot help in this regard, since the group making the transfer is worse off. So we set $t = 0$. Then, if $d \neq 0$ helps, we say deposits are Pareto essential, or PE.[18]

Consider the allocation $(\tilde{x}^i, \tilde{y}^i)$ that for some d solves (**??**) with $\mathcal{F}_d^i$ replacing $\mathcal{F}_0^i$. Then deposits are PE if there is d such that $\mathcal{W}^i(\tilde{x}^i, \tilde{y}^i) \geq \mathcal{W}^i(\hat{x}^i, \hat{y}^i)$ for both i with one strict inequality. A necessary condition for deposits from group i to group j to be PE is that the repayment constraint does not bind in group j, since otherwise deposits by tightening this constraint shrink the IF set and lower $\mathcal{W}^j$. Given these observations, we have:

Proposition 2 *Deposits are PE iff the repayment constraint binds for one group and not the other. Given $\delta^a > \delta^b$, other things equal, deposits from a to b can be PE, but not vice-versa.*

Proof: As above, $(\hat{x}^i, \hat{y}^i)$ is the best IF allocation for group i with $t = d = 0$. There are two possible cases for each group: either the repayment constraint binds or it does not. If it binds for both groups, then $d \neq 0$ tightens the repayment constraint in one, making it worse off, so deposits are not PE. If the repayment constraint does not bind in either group, then $d \neq 0$ can only make one group better off by making the other worse off, so again deposits are not PE. Finally, if the repayment constraint binds for one group but not the other, there is some $d \neq 0$ that relaxes the constraint in the former group without reducing welfare in the latter. In this case deposits are PE. Given $\delta^a > \delta^b$, it is possible for the repayment constraint to bind in group a only, but it is not possible for it to bind in group b only. ∎

This result says that when $\delta^a > \delta^b$ it can be useful to set up a bank in group b but not in group a. In Figure 4, we show how this works for a case where the welfare

---

[18]To explain this usage, recall that essential means the IF set becomes bigger or better. Here we mean better, according to the Pareto criterion, and so we use PE.
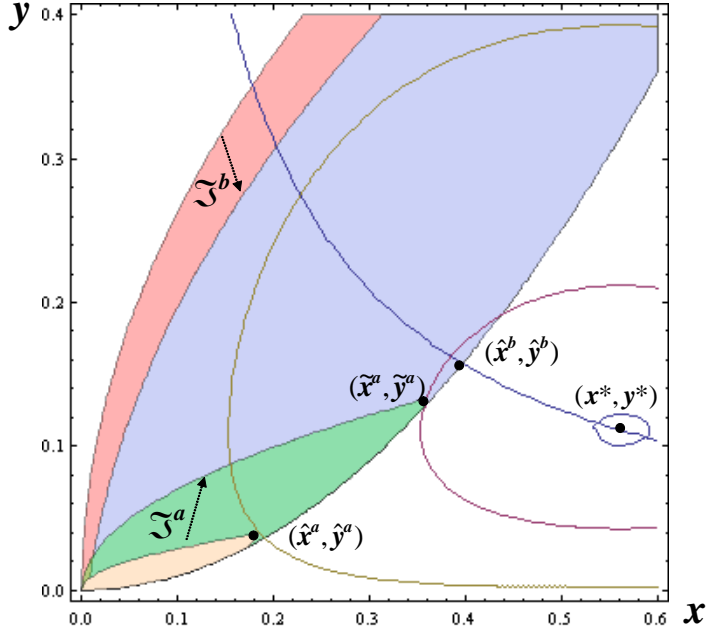
Figure 4: Illustration of Proposition **??**.

weights are the same in the two groups, $\omega_j^a = \omega_j^b$, so that the planner's indifference curves centered around (x^*, y^*) apply to either. When $d = 0$, $\left(\hat{x}^b, \hat{y}^b\right) \in \mathcal{P}$ solves (**??**) for group b, but a severe commitment problem in group a means we cannot do better than $(\hat{x}^a, \hat{y}^a)$. Introducing $d > 0$ shifts the repayment constraint in for group b in and out for a. As long as d is not too big, this has no effect on group b and allows a to do better. While many other results and applications along these lines could be presented, we think this makes the key point that, other things equal, good bankers must be more trustworthy.[19]

5.2 How Should We Monitor?

Suppose monitoring is costly: detecting opportunistic deviations with probability π^i implies a utility cost $\pi^i k^i$ in group i. Define a new benchmark with $d = 0$ as the

---

[19] One can also show that $\omega_j^a = \omega_j^b$, $\delta^a = \delta^b$ and $\rho^b > \rho^a$ implies it may be PE to set up a bank in group b but not group a, or that other things eqaul good bankers must have acess to better investments. Similarly, one can show it may be PE to set up a bank in b but not a if $\delta^a = \delta^b$, $\rho^b = \rho^a$ and $\omega_1^a < \omega_1^b$.

solution (x^i, y^i, π^i) to

$$\max_{(x,y,\pi)} \mathcal{W}^i(x, y) - \pi k^i \text{ s.t. } x \in \mathcal{F}_0^i \text{ and } 0 \leq \pi \leq 1. \tag{22}$$

It is immediate that the repayment constraint must be binding, since otherwise we could reduce monitoring costs. Also notice the following: if (x^*, y^*) maximizes $\mathcal{W}$ with full commitment, typically it does not maximize $\mathcal{W}$ when monitoring is endogenous, since reducing π implies a first order gain while moving away from (x^*, y^*) entails only a second order loss.

We are now interested in minimizing total monitoring costs rather than asking if deposits are PE, although of course with transferable utility we could always compensate agents, so any decrease in total cost can yield a Pareto improvement. Also, at every date let there now be exactly one active agent of each type in each group, so that there is a single candidate for banker in each group (later we discuss what happens more generally). If one group deposits with the other, we can reduce the cost of monitoring in the former only at the expense of increasing it in the latter, since as we said the repayment constraint must bind. As we now show, whether this is desirable depends on the differences in costs k^i and in the incentives to deviate as captured by γ^i or λ^i.[20] We also show that it may be efficient to have one agent handle all of the investments, so that we can concentrate exclusively on him and give up monitoring others altogether, as long as he has a high enough γ/λ.

Proposition 3 *Suppose $\gamma^b \geq \gamma^a$, $\lambda^b \leq \lambda^a$, and $k^b \leq k^a$. Then $d > 0$ (deposits from a to b) may be desirable, while $d < 0$ cannot be. Also, as long as γ^b/λ^b is above a threshold defined in the proof, it is desirable to set $d = y^b$ and $\pi^a = 0$.*

Proof: Given the assumptions on parameters, at the benchmark $d = 0$, we have $U^1(x^b, y^b) \geq U^1(x^a, y^a)$. Also, the repayment constraint at equality in group b implies

---

[20]Note that with one active agent of each type we can still have $\gamma^a \neq \gamma^b$, as long as we relax the assumption that $\mathcal{A}^a$ and $\mathcal{A}^b$ have the same cardinality. Also note that we do not use δ^i here, since it is a mongrel parameter that embodies λ^i, γ^i and π^i, and π^i is now endogenous.

$\lambda^b \rho \left(y^b + d \right) = \pi^b \gamma^b \frac{\beta}{1-\beta} U^1 \left(x^b, y^b \right)$. Therefore

$$\frac{\partial \pi^b}{\partial d} = \frac{(1-\beta)\rho}{\beta} \frac{\lambda^b}{\gamma^b U^1 \left(x^b, y^b \right)},$$

and similarly for group a. Therefore

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

which says $d > 0$ decreases and $d < 0$ increases total cost.

To prove the rest, let $(\breve{x}^a, \breve{y}^a)$ solve $\max \mathcal{W}^a (x, y)$ s.t. the participation constraint for 2^a only. Then it is easy to check that it is optimal to set $d = \breve{y}^a$, $\pi^a = 0$, and

$$\pi^b = \bar{\pi} \equiv \frac{1-\beta}{\beta} \frac{\lambda^b \left(y^b + \breve{y}^a \right)}{\gamma^b U^1 \left(x^b, y^b \right)} \tag{23}$$

as long as $\bar{\pi} \leq 1$. Clearly, $\bar{\pi} \leq 1$ iff γ^b / λ^b is above a threshold defined by (??). ∎

For the next part of the discussion, suppose that monitoring costs are borne by type 1^i in each group i. Consider what happens when 1^a deposits $d > 0$ with 1^b, and the former must compensate the latter for any increase in monitoring cost with a transfer t. For this demonstration, let $U^1 = x - y$ and $U^2 = u(y) - x$, so we have transferable utility. Also, set $\rho = 1$, and distinguish between the probability of monitoring participation, fixed at $\pi^i = \pi$ in both groups, and the probability of monitoring repayment, denoted p_d^i in group i when deposits are d. An allocation is feasible if there is a $p_d^i \leq 1$ such that all the constraints hold. Note that in this setup the participation constraints for agents 1^i are no longer redundant.[21]

When $d = t = 0$, it is easy in this example to characterize p_0^i and $c = p_0^a k^a + p_0^b k^b$. Consider $d > 0$ and $t > 0$. If d and t are not too big, p_d^i is still given by the repayment constraints at equality, and total cost is

$$c = \frac{y^a - d}{x^a - y^a} \delta^a p_d^a k^a + \frac{y^b + d}{x^b - y^b} \delta^b p_d^b k^b.$$

---

[21] The complete list of participation and repayment constraints in this example is

$$u(y^a) - x^a \geq 0 \qquad\qquad u(y^b) - x^b \geq 0$$
$$x^a - y^a - t - p_d^a k^a + \theta^a (x^a - y^a) \geq 0 \qquad x^b - y^b + \lambda^b t - p_d^b k^b + \theta^b (x^b - y^b) \geq 0$$
$$x^a - y^a - \delta^a (y^a - d) \geq 0 \qquad\qquad x^b - y^b - \delta^b (y^b + d) \geq 0$$

where $\delta^i = \lambda^i (1-\beta) / \beta p_d^i \gamma^i$ and $\theta^i = \pi \gamma^i \beta / (1-\beta)$.

One can check, as in the proof of Proposition 3, that $d > 0$ reduces c iff

$$\frac{\lambda^b k^b}{\gamma^b (x^b - y^b)} < \frac{\lambda^a k^a}{\gamma^a (x^a - y^a)}.$$

The reduction in cost for 1^a is $\bar{t} = \delta^a p_d^a k^a d / (x^a - y^a)$, which is enough to compensate 1^b iff

$$\lambda^b \bar{t} \geq \frac{\delta^b p_d^b k^b d}{x^b - y^b}. \tag{24}$$

One can now solve for the range of parameters where 1^a is willing to delegate $d > 0$ to 1^b and compensate him for any increase in his monitoring cost.

Although other applications along these lines can be considered, we close this section by briefly mentioning a trade off that arises when one considers the efficient number of bankers generally: fewer bankers reduces monitoring on the extensive margin, but since it increases deposits per banker it also means that we must increase monitoring on the intensive margin. In fact, even if there is only $N = 1$ group, if we consider asymmetric allocations then it can be desirable to designate a subset of the group as bankers and concentrate monitoring efforts on them. For illustration, consider U^i from previous example, but now assume the monitoring cost $k_0 + pk$ displays increasing returns.

Suppose there are n active type 1^b agents, and hence n potential bankers. Given an allocation and deposits d, the monitoring probability if there is a single banker is given by the binding repayment constraint

$$p_1 = \frac{\lambda^b k^b (ny^b + d)}{\gamma^b (x^b - y^b)},$$

and total monitoring costs are $k_0 + p_1 k$.[22] If we increase the number of banks to m, the monitoring probability becomes

$$p_m = \frac{\lambda^b k^b (ny^b + d)}{m\gamma^b (x^b - y^b)} = \frac{p_1}{m},$$

since investment is now divided across m separate bankers. But total monitoring costs are now $m(k_0 + p_1 k/m) = mk_0 + p_1 k$, higher than with a single banker.

---

[22] Notice the banker now has to invest his own production y^b, deposits from $n - 1$ other type 1^b agents, and deposits from type 1^a agents.

This illustrates the savings due to having fewer banks. However, it assumes that y^b is feasible with one banker, and with a single banker the temptation to behave strategically and liquidate d may be too strong. If this is the case, there is a minimum number of banks necessary to sustain y^b, say m^*. To guarantee y^b is feasible, we must satisfy $p_{m^*} = 1$, and the optimal number of bankers would be

$$m^* = \frac{\lambda^b k^b \left(ny^b + d\right)}{\gamma^b \left(x^b - y^b\right)}.$$

We can say e.g. that there should be fewer bankers when there are fewer resources to invest, when they have more of a stake in the economy, or when they have lower liquidation values. While much more could be done with this application, hopefully point is clear.

5.3   Rate Of Return Dominance?

In case it is not already obvious, we now show that good bankers need not have the best investment opportunities, if they are relatively good at commitment. Suppose ρ^i differs across groups. For some parameters, we claim that deposits in group b are PE, even if $\rho^a > \rho^b$.

Proposition 4 *For all $\delta^b < \delta^a$ there exists $\rho^b < \rho^a$ such that $d > 0$ is PE.*

Proof: Consider the case where $\rho^i = \rho$ for all i. By the discussion surrounding Theorem **??**, we know that $d > 0$ is PE when $\delta^b < \delta^a$. For small ε, by continuity, this is also true when $\rho^b = \rho - \varepsilon$. ∎

Notice there is a trade-off: when group a deposits with group b, they have to give up the return ρ^a, and when $\rho^a - \rho^b$ is large $d > 0$ may no longer be PE. In the working paper (Mattesini et al. 2009), we analyze some cases where we work out exactly how big the difference $\rho^a - \rho^b$ can be and still have $d > 0$ desirable (while in the above Proposition we only claim it is true for some ε difference). Rather than go into the algebra, here, we simply relay the key message. The best bankers are not necessarily agents with the highest investment opportunites. It is not necessarily because of high

returns that agents trade claims on certain investments; it may be due to commitment issues that plague alternative investments.

6   A Brief History of Banks

We have established that, because of incentive issues, it can be beneficial for an agent who wants resources from a second agent to make a deposit with a third party – an intermediary – who invests on his behalf until the returns are claimed by the second party. The reason is that the third party may be more credible, or more trustworthy, in terms of honoring obligations. He can be more trustworthy because he is more patient, because he is more visible, because has more at stake in the system, or because his gains from liquidating the investment and absconding with the returns are lower. This arrangement can be efficient, even if the third party does not have access to the highest-return investment opportunities, although other things equal a higher return is better. We think this resembles banking. In this section we go into a little more detail as to how the predictions of the model compare with the history of banking broadly.

We begin by mentioning that, although the deposit receipts discussed above constitute inside money, the theory here involves no outside money. Although it would be interesting to include outside money in future work, from the historical perspective, institutions that accepted deposits in goods came long before the invention of coinage in the 7th century. In ancient Mesopotamia and Egypt, e.g., mainly for security and to economize on transportation costs, goods were often deposited in palaces, temples, and, in later periods, private houses. As Davies (2002) describes the situation:

> 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.

We think that it is especially interesting that deposit receipts that long ago led to "transfers to the order of third parties" so they could facilitate transactions and payments, like our receipts. Relatedly, in his detailed description of the medieval Venetian bankers, Mueller (1997) describes the practice of accepting two types of deposits: regular deposits, which were specific goods that bankers had to deliver on demand; and irregular deposits, involving specie or coins that only had to be repaid with the same value, not the same specie or coins. The depositor making an irregular deposit tacitly agreed to the investment by the banker of the deposits. Like in modern times, when you put currency in the bank, you do not expect to withdraw the same currency later, only something of some specific value. This is true in the model too: the liability of the bank is not the deposit of goods, per se, but claims on the returns to its investments.[23]

Many regard the English goldsmiths as the first modern bankers. Originally, their depositors were again mainly interested in safe keeping, which is a simple type of investment. But early in the 17th century their deposit receipts began circulating in place of money for payment purposes – the first incarnation of British banknotes – and shortly thereafter they allowed deposits to be transferred by "drawn note" or cheque. Others call the Templars the first modern bankers. During the crusades, because of their skill as warriors, these knights became specialists in protecting and moving

---

[23]Because they are making investments, our bankers are more than pure storage facilities. We emphasize this because Chris Phelan in another comment on an earlier version of the paper said that, according to our theory, coat-check girls at restaurants are bankers. Not quite. When you give your hat to a coat-check girl you expect the same hat back after dinner; when you give your income to a banker you expect something different. Still, there is something to his comment, since most people agree that the origin of deposit banking did revolve around safe keeping and convenience, like coat-check facilities. It may be interesting to introduce sake-keeping considerations formally, as in He et al. (2005, 2008) or Sanches and Williamson (2008), or relatedly counterfeiting considerations, as discussed in Nosal and Wallace (2007) and references therein.

money and other valuables. At some point, rather than e.g. shipping gold from point A to point B for one party and shipping different gold from point B to point A for another, they saved on security and transportation costs simply by reassigning the parties' claims to gold in different locations; we are aware, however, of no evidence that their liabilities were used as media of exchange, the way goldsmiths' receipts were used.[24]

It is also interesting to note that other institutions that engaged in the type of banking we have in this paper – accepting deposits of goods that facilitate other transactions and credit arrangements – were still common after the emergence of modern banks. In colonial Virginia, e.g., tobacco was commonly used in transactions because of the scarcity of precious metals (Galbraith 1975). The practice of depositing tobacco in public warehouses and then exchanging authorized certificates, attesting to its quality and quantity, was extremely common and survived for over 200 years. Similarly, in the 19th century, to facilitate transactions and credit arrangements between cocoon producers and silk weavers, banks established warehouses that stored dried cocoons or silk and issued warrants that could be used to pledge for credit.[25] So while it may be interesting to discuss banking in modern monetary economies, we think it is also interesting and historically relevant to discuss banking without outside money, as we did here.

The main thing to take away from the above examples is that an early development in the evolution of banking was for deposits to be used to facilitate exchange. As in the model, throughout history, a second party is more likely to give you something if you can use in payment the liability of a credible third party, rather than your own promise. As we said, notes, cheques, debit cards, and related instruments issued by commercial banks have this feature. Returning to Venice, Mueller (1997) explains how

---

[24]Nice discussions of the goldsmith bankers can be found in *Encyclopedia Britannica* (we looked at the 1941 and 1954 editions). For specialized treatments, see Joslin (1954) and Quinn (1997). See Weatherford (1997) or Sanello (2003) for more on the Templars.

[25]As attested by Federico (1997), the first of these warehouses was funded by a group of entrepreneurs in Lyons in 1959. The Credit Lyonnais established its own warehouse in 1877 and was soon imitated by a series of Italian banks.

deposit banking came to serve "a function comparable to that of checking accounts today; that is, it was 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 of doing business. In short, the current account constituted 'bank money,' money based on the banker's promise to pay."[26]

Such a system can only work well if bankers are trustworthy. Our theory says that the more patient or visible an agent is, or the more he has at stake, the more credible he becomes. The Rialto banks in medieval Venice 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 pratimony 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 1997, p. 97). It is also interesting to point out that, although the direct evidence for this is scant, Venetian bankers seem to have been subject to occasional monitoring, as in the theory: "In order to maintain 'public faith,' the Senate in 1467 reminded bankers of their obligation to show their account books to depositors upon request, for the sake of comparing records." (Mueller 1997, p. 45). While it may have been prohibitively costly for depositors to continuously audit the books, one can imagine monitoring every so often. And if caught cheating, the punishment was indeed lifetime banishment from any banking activity in Venice, although apparently this happened rarely.[27]

---

[26]According to some, these early deposits did not actually circulate, in the sense that transferring funds from one account to another "generally required the presence at the bank of both payer and payee" (Kohn 1999). This is the argument for regarding the goldsmiths the first modern bankers. See also Quinn (2002). But even if they did not circulate, in this sense, the deposits of the earlier bankers cleary still facilitated payments. And Spufford (1988) documents that the Florentines were already using cheques in 1368.

[27]We think it is obvious that visibility and monitoring have always been crucial for good banking, but if one wants more evidence, 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

We also mention that many bankers historically started as merchants, who almost by definition have a greater connection to the market than a typical individual. As Kohn (1999) describes it, 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 wealthiest group; wealth was concentrated in the hands of landowners, who controlled agriculture, forests, and mineral rights. But the merchants probably had the most to lose from reneging on obligations. Thus, "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 first bankers being individual who had a great connection to the market is given by Pressnell (1956) in its study of country banking in England during the Industrial Revolution. Almost all of the early country banks grew up as a by-product of some other main activity, usually some kind of manufacturing.

Also, returning one last time to Venice, consider:

> 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. Their organizational form was generally either that of the fraterna or that of the partnership, the latter often concluded between a citizen and a noble. (Mueller 1997, p. 82)

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."

As in the model, there seem to have been interesting issues concerning the efficient number of bankers, revolving around trade offs related to credibility and commitment and to the amount of deposits.

Finally, does our theory have anything to say about banking panics, in general, and the recent financial crisis in particular? Gorton (2009) argues that the recent banking crisis is a wholesale panic, whereby some financial firms ran on others by not renewing sale and repurchase agreements. This resembles a retail panic in which the depositors withdraw rather than continuing their demand deposits. By analogy, depositors in the current crisis were firms that lent money in the repo market. The location of subprime risks among their counterparties was unknown. Depositors were confused about which counterparties were really at risk, and consequently ran all banks. While our framework is obviously too rudimentary to get at all of the intricacies related to the recent financial situation, a perturbation to the model can highlight one fundamental mechanism.

Suppose that the probability of being active each period γ is subject to shocks. One can then imagine that the uncertainty surrounding these shocks could induce agents to not renew their deposits (or deposit less) to reestablish the banker's incentives. What is more, our analysis implies that this is efficient. Such shocks may depend on the nature of the firm's business – say, γ could be affected by the housing market if that business is to originate mortgage loans, or it could be affected by political events. More generally, whenever visibility goes down, in the sense that monitoring becomes more difficult as captured by a reduction in the probability π, the model predicts that credit will be hampered, as will any exchanges that are facilitated by credit. But, again, this is efficient – mechanism design tells us that when π becomes much lower, not only can the credit market cease up, it *should* cease up. We are not advocating the position that there were no market failures associated with recent events. We are merely saying that it can be interesting to look at them through the lens of efficient mechanisms.

7 Conclusion

We studied banking using mechanism design. We began by describing an environment with preferences, technologies, and certain frictions – including temporal separation, commitment issues, imperfect monitoring. We then characterized incentive-feasible and optimal allocations. We did not start with prior assumptions about what banks are, who they are, or what they do. Rather, we looked at feasible or efficient outcomes and tried to interpret them in terms of arrangements that look like banks. It can be efficient for certain agents in the model, chosen endogenously, to accept deposits and make investments. This facilitates exchange, and can be efficient even if these agents do not have the best investment opportunities, as long as they are relatively trustworthy. This activity resembles banking, both in modern and historical contexts. And it is essential – if we were to rule it out, the set of feasible allocations would be inferior. We also discussed who would make a good banker, how many bankers we should have, and how to monitor them when it is costly.

We think the mechanism design approach is useful for thinking about these issues. We also think the underlying environment is interesting even if one does not take this approach. One can impose a particular mechanism – bargaining, competitive markets, or whatever – and look for equilibria. What is the nature of the equilibrium set? Does it involve deterministic or stochastic credit cycles? Are these equilibria efficient? The environment is a good one for this exercise because it captures important aspects of credit, such as limited commitment, monitoring, collateral, and so on, in a tractable way. It is tractable mainly because much of the economic activity takes place across subperiods within a period, making the analysis similar to that in two-period models, yet the model is genuinely dynamic, and one can make good use of the infinite horizon (without which credit, like money, is hard to get off the ground). The model also can be generalized in many directions, including uncertainty or private information. All of this is left to future research.

References

Andolfatto, D. (2008) "The Simple Analytics of Money and Credit in a Quasi-linear Environment," mimeo, Simon Fraser University.

Andolfatto, D. and E. Nosal (2008) "A Theory of Money and Banking," *Federal Reserve Bank of Cleveland Working Paper* 03-10.

Andolfatto, D., E. Nosal and N. Wallace (2007) "The Role of Independence in the Green-Lin Diamond-Dybvig Model," *Journal of Economic Theory* 137, 709-715.

Boyd, J.H. and E.C. Prescott (1986) "Financial Intermediary Coalitions," *Journal of Economic Theory* 38, 211-232.

Calomiris, C. and C. Kahn (1991) "The Role of Demandable Debt in Structuring Optimal Banking Arrangements," *American Economic Review* 81, 497-513.

Cavalcanti, R. and N. Wallace (1999a) "A Model of Private Bank Note Issue," *Review of Economic Dynamics* 2(1), 104-136.

Cavalcanti, R. and N. Wallace (1999b) "Inside and Outside Money as Alternative Media of Exchange," *Journal of Money Credit and Banking* 31(3), 443-457.

Davies, G. (2002) *A History of Money From Ancient Times to the Present Day*, 3rd. ed. Cardiff: University of Wales Press.

Diamond, D. and P. Dybvig (1983) "Bank Runs, Deposit Insurance, and Liquidity," *Journal of Political Economy* 91, 401-419.

Diamond, D. (1984) "Delegated Monitoring," *Review of Economics Studies* L1, 393-414.

Diamond, D. and R. Rajan (2001) "Liquidity Risk, Liquidity Creation and Financial Fragility: A Theory of Banking," *Journal of Political Economy* 109 (2), 287-327.

Ennis, H. and T. Keister (2008) "Commitment and Equilibrium Bank Runs," *Staff Report* 274, Federal Reserve Bank of New York, January.

Freixas, X. and Rochet, J.C. (2008) *Microeconomics of Banking.* Cambridge: MIT Press.

Gorton, G. and A. Winton (2002) "Financial Intermediation," in *Handbook of the*

Economics of Finance, Constantinides, Harris and Stulz, eds. Amsterdam: North Holland.

Gorton, G. (2009) "Slapped in the Face by the Invisible Hand: Banking and the Panic of 2007," mimeo.

Green, E.J. and P. Lin (2003) "Implementing Efficient Allocations in a Model of Financial Intermediation," *Journal of Economic Theory* 109, 1-23.

Gu and Wright (2010) "Endogenous Credit Cycles," mimeo.

He, P., L. Huang and R. Wright (2005) "Money and Banking in Search Equilibrium" *International Economic Review* 46, 637-670.

He, P., L. Huang and R. Wright (2008) "Money, Banking and Monetary Policy" *Journal of Monetary Economics* 55, 1013-24.

Hicks, John R. (1935) "A Suggestion for Simplifying the Theory of Money," *Economica* 2, 1-19.

Huangfu, X. and H. Sun (2008) "Private Money and Bank Runs," Mimeo, Queen's University.

Jacklin, C. (1987) "Demand Deposits, Trading Restrictions, and Risk Sharing," in E. Prescott and N. Wallace, eds., *Contractual Arrangements for Intertemporal Trade*. Minneapolis: University of Minnesota Press.

Joslin, D.M. (1954) "London Private Bankers, 1720-1785" *Economic History Review* 7, 167-86.

Kehoe, T.J., and D.K. Levine (1993) "Debt-Constrained Asset Markets,"*Review of Economic Studies* 60, 865-88

Kehoe, T.J., and D.K. Levine (2001) "Liquidity Constrained vs. Debt Constrained Markets,"*Econometrica* 69, 575-598.

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

Kocherlakota, N. and N. Wallace (1998) "Incomplete Record-Keeping and Optimal Payment Arrangements," *Journal of Economic Theory* 81, 272-289.

Koeppl, T., C. Monnet and T. Temzelides (2008) "A Dynamic Model of Settlement",

Journal of Economic Theory 142, 233-246.

Kohn, M. (1999) "Early Deposit Banking," Working Paper 99-03, Department of Economics, Dartmouth College.

Kiyotaki, N. and J. Moore (2005) "Financial Deepening," *Journal of the European Economic Association* 3, 701-712.

Kiyotaki, N. and J. Moore (2008) "Liquidity, Business Cycles, and Monetary Policy," Mimeo, Princeton University.

Lagos, R. and R. Wright (2005) "A Unified Framework for Monetary Theory and Policy Analysis," *Journal of Political Economy* 113, 463-484.

Leland, H.E and D.H. Pyle (1977) "Informational Asymmetries, Financial Structure and Financial Intermediation," *Journal of Finance* 32, 371-387.

Mattesini, F., C. Monnet and R. Wright (2009) "Banking: A Mechanism Design Approach" Working Paper, Federal Reserve Bank of Philadelphia.

Menger, K. (1892) "On the Origin of Money," *Economic Journal* 2, 239-255.

Mills, D.C. (2008) "Imperfect Monitoring and the Discounting of Inside Money," *International Economic Review* 49.

Mills, D.C. and R.R. Reed, (2008) "Default Risk and Collateral in the Absence of Commitment," Mimeo, Board of Governors.

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 R. Rajan (1998) "The Paradox of Liquidity," *Quarterly Journal of Economics*, 113 (3), 733-771.

Nosal, E.and G. Rocheteau (2009) "Money, Payments, and Liquidity," Mimeo, Federal Reserve Bank of Chicago and University of California - Irvine.

Nosal, E.and N. Wallace (2007) "A Model of (the Threat of) Counterfeiting," *Journal of Monetary Economics* 54, 994-1001.

Orsingher, R., (1967) *Banks of the World.* New York: Walker and Company.

Peck, J. and K. Shell (2003) "Equilibrium Bank Runs," *Journal of Political Economy* 111, 103-123.

Quinn, S. (1997) "Goldsmith-Banking: Mutual Acceptance and Interbanker Clearing in Restoration London," *Explorations in Economic History* 34, 411-42.

Quinn, S. (2002) "Money, Finance and Capital Markets," in *The Cambridge Economic History of Britain since 1700, vol. 1, 1700-1860*, 3rd ed., R. Floud and D. N. McCloskey, eds. Cambridge: 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 (3), 524-550.

Sanches D. and S. Williamson (2008) "Money and Credit with Limited Commitment and Theft," Mimeo, Washington University in St. Louis.

Sanello, F. (2003) *The Knights Templars: God's Warriors, the Devil's Bankers.* Lanham, MD: Taylor Trade Publishing.

Selgin, G. (2006) "Banking",forthcoming in *Encyclopedia Britannica.*

Townsend, R. (1979) "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory* 21, 265-93.

Townsend, R. (1988) "Models as Economies," *Economic Journal* 98, 1-24.

Townsend, R. (1989) "Currency and Credit in a Private Information Economy," *Journal of Political Economy* 97, 1323-1345.

Townsend, R. (1987) "Arrow-Debreu Programs as Microfoundations of Macroeconomics," in T.F. Bewley, ed., *Advances in Economic Theory, Fifth World Congress. Econometric Society Monographs Series* 12. Cambridge: Cambridge University Press.

Townsend, R. (1990) *Financial Structure and Economic Organization.* Basil Blackwell.

Wallace, N. (1988) "Another Attempt to Explain an Illiquid Banking System: The Diamond and Dybvig Model with Sequential Service Taken Seriously," *Federal Reserve Bank of Minneapolis Quarterly Review* 12, 3-16.

Wallace, N. (1990) "A Banking Model in which Partial Suspension is Best," *Federal Reserve Bank of Minneapolis Quarterly Review* 14 , 11-23.

Wallace, N. (2005) "From Private Banking to Central Banking: Ingredients of a Welfare Analysis," *International Economic Review* 46, 619-636.

Wallace, N. (2001) "Whither Monetary Economics?," *International Economic Review* 42, 847-869.

Wallace, N. (2008) "The Mechanism Design Approach to Monetary Economics," Mimeo.

Weatherford, J. (1997) *The History of Money.* New York: Crown Publishers.

Williamson, S. (1987) "Recent Developments in Modeling Financial Intermediation," *Federal Reserve Bank of Minneapolis Quarterly Review* 1132.