A Model of Interbank Settlement

Benjamin Lester *
Department of Economics
University of Pennsylvania
lester@econ.upenn.edu

July 30th, 2005

Abstract

A settlement system is a set of rules and procedures that govern when and how funds are transferred between banks in order to finalize their customers' transactions. Perhaps the most crucial feature of such a system is the frequency with which settlement occurs. A lower frequency allows banks to retain fewer idle reserves, and therefore economizes on financial opportunity costs. However, the accumulation of outstanding interbank liabilities under infrequent settlement introduces a large risk should a bank default on its payments. Moreover, when a bank retains fewer reserves, it is more vulnerable to external shocks. Conversely, a higher frequency of settlement decreases both the size and probability of defaults, but is more costly. This paper introduces a partial equilibrium model of a banking sector in which this tradeoff between cost and risk arises endogenously. We then complete the economy with a trading sector that has a micro-founded role for credit as a media of exchange. The result is a dynamic general equilibrium model that incorporates a sophisticated banking environment. Within this general equilibrium framework, we are able to conduct welfare and policy analysis. In particular, we study the optimal intra-day borrowing policy that the operator of a settlement system should impose on member banks.

---

[*] I am particularly indebted to Randy Wright for his guidance throughout the writing of this paper, and to Ed Green and Ludo Visschers for their suggestions. I thank Aleksander Berentsen, Ping He, Charles Kahn, David Mills, Deniz Selman, Skander Van den Heuvel, Neil Wallace, and members of the Penn Macro Lunch for helpful commentary. Finally, I'd also like to thank Stephen Millard and Matthew Willison for introducing me to this topic. All errors are my own.

1 Introduction

Many purchases, particularly of large value, are paid for with credit; that is, a buyer transfers to the seller a claim on funds held at his financial institution, which the seller can subsequently redeem at her financial institution. In such a transaction, the exchange of goods and the finalization of payment are separate events, and therefore require a specification for when and how the actual funds are delivered from the buyer's financial institution to the seller's. Such a specification, called a *settlement system*, is a "contractual and operational arrangement that banks and other financial institutions use to transfer funds to each other." (Zhou, 2000)

As a result of recent advances in technology and international financial integration, both the regularity and the value of these transactions have increased dramatically. For example, in the United States, the average *daily* value of transfers processed by the two primary settlement systems exceeds two trillion dollars (approximately 20% of annual GDP). Therefore, an efficient, stable settlement system is an increasingly essential component of a well-functioning modern economy. While there are many features to a settlement system, perhaps the most crucial is the frequency with which settlement occurs. This paper develops a framework to analyze the costs and risks associated with settlement frequency.

As the frequency of settlement increases, the settlement system becomes more costly. While there are several sources of this costliness, this paper focuses on *financial costs.* If banks do not have on hand sufficient reserves to finalize payments at the time of settlement, they must borrow the necessary funds. This borrowing is costly, as nearly all systems impose either interest charges or collateral requirements for borrowing, or penalties on banks that that are unable to meet their liabilities. Therefore, when a bank is forced to make settlements more often, it chooses to hold a larger amount of idle, liquid reserves. This sub-optimal portfolio allocation bears the opportunity cost of holding assets that could be generating revenue in the loan market.[1]

In a world with perfect commitment between banks, the optimal settlement system is trivially the least costly, and therefore frequent settlement would be undesirable. However, in reality we observe occasional events in which a financial institution is unable to deliver payments to finalize its customers' transactions, and therefore must default on its outstanding liabilities. The settlement system in place will be a crucial determinant of how the economy as a whole is affected by such events. Increased settlement frequency mitigates this risk in two important ways. First, as the frequency of settlement increases, the number of parties awaiting payment decreases. Therefore, a settlement system that reduces the time between the initial transaction and the finalization of payment decreases the potential size of a default. Secondly, since banks retain a higher level of reserves under systems with more frequent settlements, they are less vulnerable to exogenous shocks to net worth.

---

[1] A second type of costs that arise as settlement frequency increases are *resource costs.* These would include expenditures for the transfer, collection, accounting, and monitoring of payment finalization, among many others. We ignore these here. See Berger, Hancock, and Marquardt (1996) for a more complete exposition of costs inherent in payment systems.

Therefore, there exists tension between the risk associated with large defaults and the costs of controlling that risk. In Section 3, we begin by constructing a partial equilibrium model (in the sense that the price of loans is exogenous) of the banking sector in which a bank's intra-day net liabilities are stochastic, borrowing is costly, and banks are subject to random, exogenous shocks to net worth. Within this framework, the risk-cost tradeoff discussed above arises naturally. In Section 4, we introduce a simple example to illustrate the key results from the partial equilibrium model. We proceed with sensitivity analysis, which allows us to better understand how the magnitude of the risk-cost tradeoff is affected by changes to the parameters of the model economy.

The partial equilibrium example is a useful tool in that it clearly illustrates the tradeoff within the banking sector. However, it takes as given the demand for loans, as well as the willingness of buyers and sellers to trade, and therefore does not consider how their behavior might respond to changes in the costs or risks of banking. As a result, any attempt at welfare analysis within the partial equilibrium setting would require arbitrary weights on the costs and risks of settlement frequency. In Section 5, we complete the economy by introducing a trading sector in which agents conduct transactions using bank credit as the means of payment. This construction may be regarded as a significant contribution, as there are few examples in the literature of dynamic general equilibrium models that incorporate a rich, sophisticated framework for the analysis of banking behavior. The complete economy is a general equilibrium model in the following sense: there are two sectors, and variables that are endogenously determined in one sector are taken as given in the other. In equilibrium, of course, there is aggregate consistency between the behavior in the two sectors. The complete model allows us to analyze behavioral and welfare implications of the risk-cost tradeoff within the context of a monetary economy.

In Section 6, we extend the example introduced earlier to the general equilibrium framework. With a clear definition of welfare, we are now able to address certain policy issues that are often debated in the literature. Specifically, we parameterize the economy and consider the optimal borrowing policy that the central bank should impose on participants of the settlement system. One of the key insights, largely ignored in the literature, is that discussion of an "optimal" intra-day borrowing policy should be conducted within the context of a particular settlement frequency; that is, the optimal policy will generally *not* be the same under a real-time and deferred settlement system. With this in mind, we conclude by considering characteristics of an economy that might make one settlement frequency more appropriate than another. Section 7 concludes, and offers potential extensions of the model.

2 Literature Review

(To be written.)

3 A Partial Equilibrium Model of Interbank Settlement

Suppose there are a large number of ex-ante identical banks, each with an initial mass of depositors d. Upon receiving a deposit, a bank issues the depositor some form of receipt that can then be used to purchase goods in the marketplace. We loosely refer to this receipt as a "check," though it should be understood that the usual settlement system associated with checks does not necessarily apply. Banks face a fixed cost a to manage each deposit, and charge each depositor a fee ϕ. It is possible, however, that $\phi < 0$, in which case the bank must pay the depositor. This case has the natural interpretation of an interest-bearing bank account.

Before any trading begins, banks choose a fraction α of these initial deposits to retain as reserves, and loan out the remaining $1-\alpha$ in exchange for an up-front fee ρ. We assume that the probability of an agent redepositing a loan at the lending bank is zero, so that each bank has a mass d of active *buyers* in the market. We also assume that *sellers* are pre-assigned a bank to deposit any payments received, and that sellers are uniformly distributed across banks. Sellers deposit checks immediately after the transaction takes place. Therefore, throughout the day, a bank accrues *due-to*'s and *due-from*'s as its buyers and sellers, respectively, trade with other banks' customers in the marketplace. A bank's net liabilities at any time are equal to the difference between due-to's and due-from's.

Suppose that transactions occur at T discrete intervals throughout each day, and that a bank has cumulative intra-day net liabilities $l_t \in [\underline{L}_t, \bar{L}_t]$ at each $t \in \mathcal{T} \equiv \{1, 2, ..., T\}$, where $\bar{L}_t$ ($\underline{L}_t$) is the maximum (minimum) net position that occurs at time t with strictly positive probability, and $-d \leq \underline{L}_t < \bar{L}_t \leq d$ for all $t \in \mathcal{T}$. We denote the maximum possible net liability position $\bar{L} = \max_{t \in \mathcal{T}}\{\bar{L}_1, ..., \bar{L}_T\}$. Therefore, a bank's intra-day net liabilities is a T x 1 random variable $l \equiv (l_1, ..., l_T)$ with density $\Pi(\cdot)$, which we assume is continuous. Any outstanding liabilities are settled at the end of each day, so that each period begins with $l_0 = 0$. Moreover, each night deposits are reallocated across banks so that the following morning all banks begin with d deposits.

3.1 Notation, Definitions, and Key Concepts

A settlement system specifies the frequency with which banks must settle net liabilities incurred since the previous settlement. For a given frequency $F \leq T$, banks will settle at intervals of length $S = \frac{T}{F}$.[2] Let $\mathcal{S} \subset \mathcal{T}$ denote the subset of points in time in which banks settle, so that $\mathcal{S} \equiv \{S, 2S, ..., FS\}$, where $FS = T$. Also, for any given point in time t, let s_t denote the time of the previous settlement. That is,

$$s_t = \begin{cases} \max\{t' \in \mathcal{S} : t' < t\}, & \text{for } t > S \\ 0 & \text{, for } 0 \leq t \leq S \end{cases}$$

The settlement system is operated by a third party which we will refer to as *the central bank*, as this is both convenient and consistent with many settlement systems

---

[2] For simplicity, we consider frequencies for which S is an integer.

world-wide. The settlement system operates in the following manner: each bank has an account at the central bank, and we assume that banks begin each day with a zero balance.[3] At the time of the first settlement, a bank with positive net liabilities must transfer that amount to its account at the central bank, which then immediately distributes the appropriate sums to banks awaiting payment.[4] Conversely, a bank with negative net liabilities at the time of settlement will have its account credited with the appropriate amount. A bank cannot withdraw its balance at the central bank - for example, to make loans - until the following morning.[5] Therefore, at subsequent settlement periods throughout the day, a bank with a positive account balance at the central bank will first use these funds to pay off any net liabilities, and only use its reserves when it has a zero account balance at the central bank.

Our analysis thus requires us to keep track of the following variables.

Definition 1 *Let a bank's* ***central bank account surplus at time*** t *for any realization of net liabilities* l *and settlement frequency* F *be denoted* $A_t(l;F)$*, and be recursively defined:*

$$A_t(l;F) = \begin{cases} \max\{0, A_{s_t}(l;F) - (l_t - l_{s_t})\}, & \text{for } t \in \mathcal{S} \\ A_{s_t}(l;F) & , \text{for } t \notin \mathcal{S} \end{cases}$$

with $A_0(l;F) = 0$.

As the formula suggests, a bank's central bank account surplus is non-negative, increases at settlements in which the bank receives payment finalization (when $l_t - l_{s_t} < 0$ for $t \in \mathcal{S}$), and decreases when a bank must settle net liabilities (when $l_t - l_{s_t} > 0$ for $t \in \mathcal{S}$). Notice that for any t that does not correspond to a settlement period, the central bank account surplus does not change.

Next we define the amount a bank must transfer to the central bank at a given settlement period in order to finalize its customers' payments.

Definition 2 *Let a bank's* ***outstanding net liabilities at time*** t *for a given realization of net liabilities* l *and settlement frequency* F *be denoted* $\hat{l}_t(l;F)$ *and be recursively defined:*

$$\hat{l}_t(l;F) = \max\{0, l_t - l_{s_t} - A_{s_t}(l;F)\} \qquad \text{for all } t \in \mathcal{T},$$

with $\hat{l}_0(l;F) = 0$.

---

[3] We could easily assume that banks begin the day with a positive account balance at the central bank. However, this further complicates the algebra and has no substantive effect on the model's results.

[4] This coincides, effectively, with a bank exchanging so-called 'commercial bank money' for 'central bank money,' which is then used to finalize payment. See the August 2003 BIS report, 'The role of central bank money in payment systems,' for details.

[5] Since we later assume that depositors have priority over reserves (at the expense of other banks) in the incidence of default, this assumptions turns out to be crucial to our results. Essentially, it precludes the possibility of a bank with negative cumulative intra-day net liabilities siphoning money to depositors prior to defaulting on outstanding liabilities. However, any alternative assumption that assures that other members of the settlement system have priority over money received within the current period is sufficient for the results reported.

From the definition, we see that outstanding net liabilities are equal to any new liabilities incurred since the preceding settlement, net of any previous balances at the central bank. Finally, it is also necessary to keep track of a bank's reserves available for transfer to the central bank.

Definition 3 *Let a bank's* ***reserves at time*** t *for a given realization of net liabilities* l*, reserve ratio* α*, and settlement frequency* F *be denoted* $R_t(l, \alpha; F)$ *and be recursively defined:*

$$R_t(l, \alpha; F) = \begin{cases} \max\{0, R_{s_t}(l, \alpha; F) - \hat{l}_t(l; F)\}, & \text{for } t \in \mathcal{S} \\ R_{s_t}(l, \alpha; F) & , \text{for } t \notin \mathcal{S} \end{cases}$$

with $R_0(l, \alpha; F) = \alpha d$.

This definition reflects that a bank begins with reserves αd, and thereafter the level of reserves decreases as net outstanding liabilities are settled. Again, when no settlement occurs reserves remain unchanged.

Should a bank's outstanding net liabilities at $t \in \mathcal{S}$ exceed it's reserves, it must borrow the difference, $\hat{l}_t(l; F) - R_{s_t}(l, \alpha; F)$, from the central bank.

Definition 4 *Let a bank's* ***total intra-day borrowing requirement*** *for a given realization of net liabilities* l*, reserve ratio* α*, and settlement frequency* F *be denoted* $b(l, \alpha; F)$ *and be defined:*

$$b(l, \alpha; F) = \sum_{t \in \mathcal{S}} \max\{0, \hat{l}_t(l; F) - R_{s_t}(l, \alpha; F)\}.$$

At the end of the trading day, a bank is assessed a fee based on it's total intra-day borrowing requirement given by $\psi\big[b(l, \alpha; F)\big]$, net liabilities return to zero, and the process repeats the following day. In reality, these fees are imposed through a combination of interest charges, collateral requirements, and borrowing caps (in which a bank is penalized for exceeding a predetermined borrowing limit).[6] Moreover, many systems maintain deductibles for small intra-day overdrafts, while they levy far more severe penalties on large overdrafts. Therefore, the following assumption on the convexity of ψ is not only convenient in that it provides curvature to our problem, but also seems consistent with real world practice.

Assumption 1 *The cost function for borrowing fees,* $\psi(\cdot)$*, is* C^2*, strictly increasing, and strictly convex.*

Since banks choose their reserve levels before the start of the day, they must project borrowing fees for any given level of reserves.

Definition 5 *Let a bank's* ***expected borrowing fees****, given reserve ratio* α *and settlement frequency* F*, be denoted* $\Psi(\alpha; F)$ *and be defined:*

$$\Psi(\alpha; F) = \mathbb{E}_l\Big[\psi\big[b(l, \alpha; F)\big]\Big] = \int \psi\big[b(l, \alpha; F)\big]\Pi(l)dl.$$

---

[6] For a detailed comparison of these various instruments, see Mills(2005).

Now that we have defined the structure of the banking sector, we are ready to turn our analysis to the problem facing a profit-maximizing bank.

3.2 The Problem of the Bank

For each deposit, a bank receives a deposit fee (or pays interest) ϕ and earns additional revenue $(1-\alpha)r\rho$ in loans, where r is the rate of time preference of the bank's customers. However, a bank must spend a fixed amount a to manage each account, and an additional expected borrowing fee $\Psi(\alpha; F)$. Therefore, a bank's profits are given by

$$\pi(\phi, \alpha; F) = d[\phi + (1-\alpha)r\rho - a] - \Psi(\alpha; F). \tag{1}$$

Assuming perfect competition in the banking sector, we use the zero expected profit condition to get

$$\phi = a - (1-\alpha)r\rho + \frac{\Psi(\alpha; F)}{d}. \tag{2}$$

Now we are able to characterize $\alpha^*(F)$, the optimal level of reserves that a profit-maximizing bank will choose under a system with settlement system F, as the solution to the first order condition[7]

$$-\frac{\partial}{\partial \alpha}\left[\Psi(\alpha; F)\right] = dr\rho. \tag{3}$$

This first order condition has the usual interpretation: the marginal benefit from an increase in the reserve ratio, resulting from a reduction in expected borrowing fees, is equated to the marginal cost of foregone loan revenue.

3.3 Borrowing Costs, Reserves, and the Frequency of Settlement

We now turn to the relationship between the expected cost function Ψ, the reserve ratio α, and the frequency of settlement F for the special case of $T = 2$, as this captures most clearly the key results and intuition.[8]

For $T = 2$, there are only two possible settlement systems. The first, with settlement frequency $F_1 = 1$, is a *deferred* settlement system, as transactions occurring at $t = 1$ are not finalized until the end of the period (i.e at $t = 2$). The second settlement system, with frequency $F_2 = 2$, is a *real time* settlement system, as transactions are finalized as they take place.

---

[7] Here we restrict our attention to interior solutions, though we consider corner solutions in the example presented in Section 4. Also, notice that this is a strictly concave function over a convex set, so that we are assured of a unique solution.

[8] It should be understood, however, that all results are *essentially* without loss of generality in the following sense: for any arbitrary finite T with settlement periods specified by $\mathcal{S}' \subset \mathcal{T}$, all results presented below remain true if 'increased settlement frequency' is defined by an $\mathcal{S}'' \subset T$ with the property that $\mathcal{S}'$ is a *strict* subset of $\mathcal{S}''$.

In what follows, we restrict α to the relevant open set $(0, \frac{\bar{L}}{d})$, so that banks hold strictly positive reserves, but not amounts greater than or equal to the maximum net liability position.[9] All proofs are presented in the appendix.

Claim 1 *Increasing the reserve ratio decreases the expected borrowing fees at any settlement frequency. That is, for $F \in \{F_1, F_2\}$,*

$$\frac{\partial}{\partial \alpha}\left[\Psi(\alpha; F)\right] < 0.$$

Claim 2 *For any level of reserves, expected borrowing fees are larger under a real-time settlement system than they are under a deferred settlement system. That is, $\forall\ \alpha \in (0, \frac{\bar{L}}{d})$,*

$$\Psi(\alpha; F_1) \leq \Psi(\alpha; F_2).$$

The claims above allow us to formalize the relationship between settlement frequency and the endogenously determined reserve ratio.

Result 1 *Profit-maximizing banks choose to retain a larger fraction of deposits as reserves under a real-time settlement system than they do under a deferred settlement system. That is,*

$$\alpha^*(F_1) \leq \alpha^*(F_2).$$

Since banking fees are a function of the reserve ratio, and we have seen that the reserve ratio is endogenously determined as the optimal solution to the banking problem under settlement system F, we are able to express ϕ as a function of F:

$$\phi(F) = a - [1 - \alpha^*(F)]r\rho + \frac{\Psi(\alpha^*(F); F)}{d}. \tag{4}$$

Therefore, we are now ready to formalize the effect of a settlement system on the banking fees faced by agents in the partial equilibrium model.

Result 2 *For any loan price ρ, banking fees are higher under a real-time settlement system than they are under a deferred settlement system. That is,*

$$\phi(F_1) \leq \phi(F_2).$$

3.4 Bank default

In this paper, we take the view that defaults tend to be the result of sudden, unforeseen events that lead to a permanent decrease in a bank's net worth. Such events may be caused by poor investments (the Long Term Capital Management crisis), fraud (the collapse of Barings Bank), or other calamities (such as the San Francisco earthquake or September 11th). After any of these shocks, the financial

---

[9] The assumption that $\alpha^* > 0$ can be motivated by the assumption that ψ is sufficiently steep, while $\alpha^* < \frac{\bar{L}}{d}$ can be justified by assuming that ψ is not too steep. Therefore, bounding $D\psi$ is a sufficient condition for interiority.

institution must often come up with a certain amount of capital to remain solvent. In this section, we introduce a very simple notion of default; one that abstracts from many details of the bankruptcy process, but captures the key concept that a bank's solvency during a crisis depends primarily on the magnitude of the shock and the strength of its financial position.

Suppose that at $T = 2$, prior to settlement, a bank-specific catastrophe occurs with probability σ. When such an event takes place, the bank's solvency requires a payment of size z to an institution outside the model, where z is a stochastic variable distributed on $\mathbb{R}_{++}$ according to the distribution $G(\cdot)$. We assume that if the solvency shock is small relative to a bank's financial position, that the central bank will "bail out" the bank by paying z from its own reserves. However, if the shock is sufficiently large, the central bank will not bail out the bank, z will go unpaid, and a default will ensue.[10] Following Kahn and Roberds (1998), we assume that when a bank is liquidated, priority is given to the depositors at the expense of the banks awaiting payment. Finally, we assume that the central bank insures those depositors not repaid through the liquidation of the bank's reserves, much like the FDIC in the U.S. system.

Formally, a bank's financial position at time T is equal to the funds it has available less outstanding liabilities, or

$$R_T(l,\alpha;F) + A_T(l;F) - b(l,\alpha;F) \equiv \alpha d - l_T,$$

and there exists a threshold $\bar{D} > \bar{L}$ such that a bank will default if $z-(\alpha d - l_T) > \bar{D}$.[11]

The frequency of settlement affects the size of cumulative defaults in two important ways. On the one hand, under a higher frequency of settlement, banks retain a larger level of reserves and therefore decrease the probability of default. On the other hand, a higher frequency of settlement also reduces the size of outstanding net liabilities at any time.[12] We formalize these two ideas below.

Definition 6 *Let a bank's* ***probability of default at time*** T *for a given realization of net liabilities* l *and reserve ratio* α *be denoted* $\Delta(l,\alpha)$ *and be defined*

$$\Delta(l,\alpha) = \sigma \int_z \mathbf{1}\Big\{ z - [(\alpha d - l_T) > \bar{D} \Big\} dG(z).$$

Integrating, we see that

$$\Delta(l,\alpha) = \sigma\Big[1 - G(\bar{D} + \alpha d - l_T)\Big].$$

---

[10] The question of whether or not the central bank should bail out a bank is an interesting question in and of itself. The current model could be extended to explore this issue by treating default as an endogenous decision (as in Kahn and Roberds (1998), Mills (2005), and others), and determining the optimal bail out policy. However, for the purpose of this paper, we take this policy as given.

[11] The assumption that $\bar{D} > \bar{L}$ is consistent with the notion that large banks don't default as a result of the usual intra-day liquidity needs. Instead, banks default when unforeseen circumstances cause a surprise shock to net worth.

[12] Kahn, McAndrews, and Roberds (2003) point out a third effect of settlement frequency on default that we do not consider. They point out that deferred settlement may decrease the risk of "gridlock", and thus defaults arising from liquidity shortages.

It is now easy to conclude that the probability of default is weakly decreasing in α for all l. Therefore, from Result 1, we can derive the following result characterizing the relationship between the probability of default and the frequency of settlement.

Result 3 *The probability of default is greater under a deferred settlement system than it is under a real-time settlement system. That is,* $\Delta\big(l, \alpha^*(F_1)\big) \geq \Delta\big(l, \alpha^*(F_2)\big)$ *for all* l*.*

We now turn to the relationship between the frequency of settlement and the size of net outstanding liabilities.

Result 4 *At any time* $t \in \mathcal{T}$ *and for any realization of net liabilities* l*, the size of net outstanding liabilities is smaller under a real-time settlement system than under a deferred settlement system. That is, for any* l*,*

$$\hat{l}_t(l; F_1) \geq \hat{l}_t(l; F_2).$$

Assuming that shocks are iid across banks, by the law of large numbers we are able to define the size of total defaults in the banking sector, making explicit the two key components: the probability of default and the size of outstanding net liabilities.

Definition 7 *Let the* ***total expected default under settlement system*** F *be denoted* $\delta(F)$ *and be defined*

$$\delta(F) = \mathbb{E}_l\Big[\Delta(l, \alpha^*(F)) \cdot \hat{l}_2(l; F)\Big]. \tag{5}$$

From Results 3 and 4, we arrive at the second major result.

Result 5 *The total expected default is larger under a deferred settlement system than it is under a real-time settlement system. That is,*

$$\delta(F_1) \geq \delta(F_2).$$

3.5 Key Results

From Results 1 - 5, we can capture the cost-risk trade-off of settlement frequency with the following two observations:

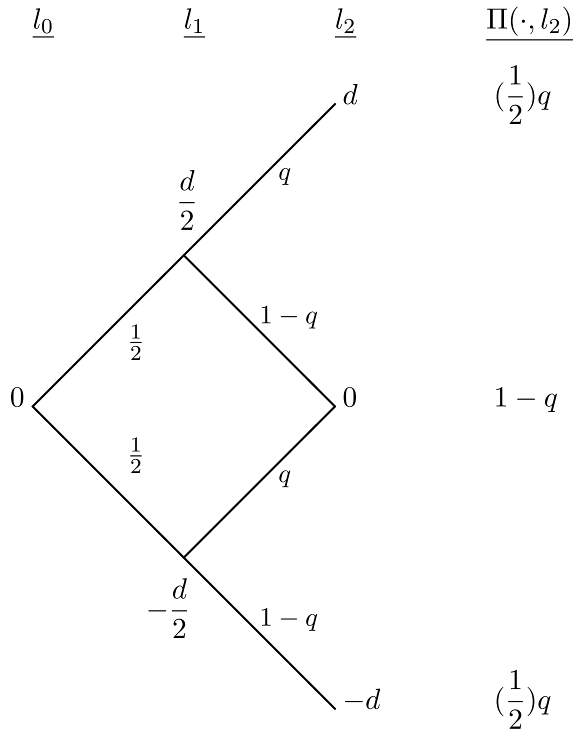
1. Under a real-time settlement system, banks choose to retain a larger fraction of initial deposits as reserves. On the one hand, this results in larger banking fees (or smaller interest payments) for depositors. However, on the other hand, a higher reserve ratio and a shorter interval between settlements imply that both the probability of default and the size of any particular default are decreased. As a result, a real-time system weakly decreases the size of any default, so that the risk of a seller's bank not receiving payment finalization decreases.

2. A deferred net settlement system is less costly for banks to operate, they choose a lower reserve ratio, and depositors therefore face smaller banking fees. However, both lower reserves and the accumulation of net liabilities throughout the trading period allow for larger defaults, and increase the risk that a seller's bank will not receive payment.

4 A Simple Example

Consider the following simple example to illustrate the key characteristics of our model setup. Suppose that at $t = 1$, net liabilities either increase or decrease by $\frac{d}{2}$ with equal probability $\frac{1}{2}$. If $l_1 = \frac{d}{2}$, then at $t = 2$ net liabilities return to zero with probability $1 - q$ and increase to d with probability q, where $q \in (0, \frac{1}{2})$. Similarly, if $l_1 = -\frac{d}{2}$, then at $t = 2$ net liabilities return to zero with probability $1 - q$ and decrease to $-d$ with probability q. Figure 1 below illustrates the distribution of net liabilities.

FIGURE 1: DISTRIBUTION OF l



Suppose further that $\psi : [0, d] \rightarrow \mathcal{R}_+$ has the form $\psi(b) = \eta b^2$. This function is clearly C^2, strictly increasing, and strictly convex. We also assume that $\eta < \frac{2r\rho}{dq}$, which guarantees an interior solution on the domain $(0, \frac{1}{2})$.[13]

---

[13] If $\alpha^*(F_1) \in [\frac{1}{2}, 1]$, then $\alpha^*(F_1) = \alpha^*(F_2)$, and we lose two margins of interest: both the banking fees and the probability of default are identical under F_1 and F_2.

4.1 Deferred Settlement

Consider first a deferred settlement system (F_1) in which settlement occurs only once at $t = 2$. Banks face ex-ante expected costs of borrowing

$$\Psi(\alpha, F_1) = (\frac{1}{2})q(d - \alpha d)^2\eta. \tag{6}$$

We can use this to solve the first order condition for the optimal reserve ratio under settlement system F_1,

$$\alpha^*(F_1) = 1 - \frac{r\rho}{\eta dq}. \tag{7}$$

We conclude that, under a deferred settlement system, depositors face a banking fee

$$\phi(F_1) = a - \frac{1}{2}r\rho\Big[\frac{r\rho}{\eta dq}\Big]. \tag{8}$$

Now consider the probability and size of default under a deferred settlement system. By construction, there is only one realization, $l = (\frac{d}{2}, d)$, for which a bank has positive net outstanding liabilities at $t = 2$, so we can restrict our attention to this realization. The probability of default under F_1 is given by

$$\Delta\Big((\frac{d}{2}, d), \alpha^*(F_1)\Big) = \sigma\Big[1 - G\Big(\bar{D} - \frac{r\rho}{\eta q}\Big)\Big]. \tag{9}$$

Moreover, net outstanding liabilities at $t = 2$ for $l = (\frac{d}{2}, d)$ are given by

$$\hat{l}_2\Big((\frac{d}{2}, d); F_1\Big) = d. \tag{10}$$

Therefore, the total expected default under F_1 is

$$\delta(F_1) = (\frac{1}{2})q\sigma\Big[1 - G\Big(\bar{D} - \frac{r\rho}{\eta q}\Big)\Big]d. \tag{11}$$

4.2 Real-Time Settlement

Now consider the settlement system F_2, in which settlement occurs at both $t = 1$ and $t = 2$. Under the parametric assumptions established above, specifically that $\eta < \frac{2r\rho}{qd}$, we now illustrate the results from the previous section.

Under a real-time settlement system, a bank faces expected borrowing fees that clearly exceed those under deferred settlement:

$$\Psi(\alpha, F_2) = (\frac{1}{2})(1 - q)(\frac{d}{2} - \alpha d)^2\eta + (\frac{1}{2})q(d - \alpha d)^2\eta. \tag{12}$$

A profit-maximizing bank will find the optimal reserve ratio under F_2 to be

$$\alpha^*(F_2) = \frac{1 + q}{2} - \frac{r\rho}{\eta d}. \tag{13}$$

Simple algebra allows us to conclude that $\alpha^*(F_2) > \alpha^*(F_1)$. The resulting banking fees are given by

$$\phi(F_2) = a - \frac{1}{2} r\rho\Big[\frac{r\rho}{\eta d}\Big] + (1-q)\Big[\frac{\eta d q}{8} - \frac{r\rho}{2}\Big], \tag{14}$$

which can also be easily proven to be greater than the fees under deferred settlement described in (8).[14] Again, restricting our attention to the realization $l = (\frac{d}{2}, d)$, the probability of default under F_2 is

$$\Delta\Big((\frac{d}{2}, d), \alpha^*(F_2)\Big) = \sigma\Big[1 - G\Big(\bar{D} - \frac{d}{2}(1-q) - \frac{r\rho}{\eta}\Big)\Big]. \tag{15}$$

Since $G(\cdot)$ is a CDF and therefore has the property that $G'(\cdot) \geq 0$, and moreover

$$\bar{D} - \frac{d}{2}(1-q) - \frac{r\rho}{\eta} > \bar{D} - \frac{r\rho}{\eta q},$$

we conclude that

$$\Delta(l, \alpha^*(F_1) \geq \Delta(l, \alpha^*(F_2)).$$

Since settlement occurred at $t = 1$, net outstanding liabilities are given by only those liabilities incurred between $t = 2$ and $t = 1$. Therefore, as established in Result 4, the outstanding net liabilities are reduced under a real-time system:

$$\hat{l}_2\Big((\frac{d}{2}, d); F_2\Big) = \frac{d}{2}. \tag{16}$$

Therefore, the total expected default under F_2 is

$$\delta(F_2) = (\frac{1}{2})q\sigma\Big[1 - G\Big(\bar{D} - \frac{d}{2}(1-q) - \frac{r\rho}{\eta}\Big)\Big]\frac{d}{2}. \tag{17}$$

As we've proven that both components of $\delta(F_2)$ are smaller than the corresponding components under F_1, trivially $\delta(F_1) > \delta(F_2)$.

4.3 The Magnitude of the Risk-Cost Tradeoff

We've established above the existence of a tradeoff between risk and cost that arises within the analysis of settlement frequency. However, the magnitude of this tradeoff depends heavily on the parameters of the economy. In an effort to better understand the key forces at play, we now parameterize the economy, and consider the sensitivity of $\alpha(F_i)$, $\phi(F_i)$, and $\delta(F_i)$ to two key parameters: η and q.[15]

Let $a = .05$, $d = .15$, $r = .05$, $\rho = .1$, $\bar{D} = .5$, $\sigma = .02$, and $G(\cdot)$ be given by a uniform distribution over the interval $[0, \bar{D} + d]$. First, we will fix $q = .25$ and focus on how the policy parameter η, representing the cost of borrowing from the central bank, affects a bank's choice of reserves, along with the resulting fees and defaults. Figure 2A illustrates that, for small η, banks choose to retain the minimum level of

---

[14] See appendix

[15] Note that this is not a calibration, but rather a purely illustrative exercise.

reserves (zero) under either settlement system. However, notice in Figure 2C that even when banks hold zero reserves, the banking fees are larger under F_2 than they are under F_1, as the increased intra-day borrowing costs are passed along to the depositors. Moreover, due to the increased liquidity requirements under a real-time system, banks begin to hold strictly positive reserves for smaller values of η under F_2 than they do under F_1; at these values of η, the difference between banking fees under the two systems is most pronounced. As η increases towards its upper bound, the difference in banking fees diminishes. Figure 2B illustrates the two key effects of settlement frequency on total defaults: the *size effect* introduced in Result 4, and the *probability effect* introduced in Result 3. First, since banks settle at $t = 1$ under a real-time system, the maximum size of net outstanding liabilities at $t = 2$ is $\frac{d}{2}$, as opposed to a maximum size of d under a deferred system. This explains the fact that, even when banks hold zero reserves under both systems, default is smaller under F_2. The second effect - a decreased probability of default as reserves increase - accounts for the downward slope of the curves corresponding to regions where they retain strictly positive reserves.

[INSERT FIGURE 2A - 2C HERE]

We now fix $\eta = .2$, and concentrate on the magnitude of the risk-cost tradeoff for different values of q. Recall that q represents the conditional probability that a bank's intraday liabilities will increase to d , leaving a bank with positive outstanding net liabilities at $t = 2$. For small values of q, banks in a deferred settlement system will not retain any reserves, since the probability of needing them for settlement at $t = 2$ is small. However, as is evident in Figure 3A, banks in a real-time settlement system will always retain positive reserves to meet liabilities accrued at $t = 1$. Furthermore, notice that for small q, $\phi(F_2)$ is larger than $\phi(F_1)$, while $\delta(F_2)$ and $\delta(F_1)$ appear similar. One might conclude that for small q, a deferred settlement system would be preferred. As q increases, the tradeoff seems to skew toward a real-time settlement system, as the fees become similar while the difference in default sizes becomes more pronounced.

[INSERT FIGURE 3A - 3C HERE]

Overall, the partial equilibrium example above is a useful tool for illustrating the primary forces at work within the banking sector. However, since it takes the underlying transactions as exogenous, it ignores the behavior of buyers and sellers; and therefore does not consider how their behavior might respond to changes in the banking fees or default risks. Moreover, we have thus far ignored any profits or losses assumed by the central bank. As such, any definition of welfare (and thus optimality) would require arbitrary weights on the risks and costs of settlement. We now introduce a trading sector to complete the economy. In this general equilibrium model, we have a more clear definition of welfare and a framework to discuss optimal settlement policies.

5 The Trading Sector

5.1 The Basic Model

The trading sector resembles the framework established in He, Huang, and Wright (2005).[16] There is a $[0,1]$ continuum of agents, and a fraction M are initially endowed with one unit of fiat money. Money is indivisible and agents hold either zero or one unit. Before entering the market, an agent with money has the option to deposit his money at the bank in exchange for a "check," which is an equally acceptable form form of payment as cash. Again, this service has a cost ϕ, where $\phi > 0$ can be interpreted as a banking fee and $\phi < 0$ as interest earned on deposits. An agent with money will trivially choose to use banking if $\phi \leq 0$ and not if $\phi > 0$. Since this paper is concerned with payment systems, we will restrict our attention to the case of $\phi \leq 0$, when agents use checks as payment. An agent without money can take out a loan for an up-front fee ρ.[17] Therefore, when all agents choose to deposit their money at the bank, a fraction $M_1 = \frac{M}{\alpha}$ will enter the market with checks (these are *buyers*), while the remaining $1 - M_1$ enter the market without money (these are *sellers*). Naturally, M_1 is the total money supply.

Agents trade in a random, bilateral matching (or *decentralized*) market in which meetings occur at T discrete points in time throughout the day. When a consumer meets a producer, there is probability x that he likes the producer's specialized good. In this case, the seller produces the good at cost c, and the buyer immediately consumes the good and receives utility u. The probability of a double-coincidence of wants is zero. Goods are indivisible, non-storable, and they are produced, traded, and consumed immediately. When an agent sells a good in the decentralized market in exchange for a check, he immediately leaves the market and deposits the check at his bank. When his bank receives payment finalization, his account is credited one unit and he becomes a buyer the following day.

5.2 The Costs of Default

As we have seen in Section 3, each day there are defaults of size $\delta(F)$. Suppose that there is a cost of recovery specified by $\Omega\big(\delta(F)\big)$, where $\Omega(\cdot)$ is increasing. Depending on the specifications of the settlement system, these costs may be born (ultimately) by either the buyers or the sellers. One common practice, known as *survivors pay*, requires that the remaining solvent banks pay the recovery costs. Under the assumption of perfect competition in the banking sector, this implies that the buyers will absorb the burden of default through increased banking fees. On the other hand, if the system does not require the participants of the settlement system to pay for

---

[16] The key distinction here is that we do not introduce theft to motivate banking, as in He, Huang, and Wright (2005). Instead, we will concentrate on the case of interest-bearing bank accounts. Moreover, we do not require a Lagos-Wright (2005) framework, in which agents alternate between a centralized and decentralized market, because we limit our analysis to the case of $\{0,1\}$ money holdings.

[17] As He, Huang, and Wright (2005) point out, this fee can be interpreted as a coupon payment of $r\rho$ in each period.

recovery, then the costs may be born by the sellers. In such a case, we assume the sellers will pay an (actuarially fair) insurance premium to protect themselves from payment default. This practice, the so-called *third party pays*, is also fairly common.

Let us assume that a fraction of the recovery costs, $\xi \in [0,1]$, are paid by the buyers and the remaining $1-\xi$ by the sellers. Such generality nests the two most common practices as special cases ($\xi = 1$ and $\xi = 0$, respectively). Moreover, in addition to settlement frequency F and intra-day borrowing fees ψ, ξ provides a 3rd avenue for policy analysis, as it allows us to examine the effects of shifting risk from one side of the market to the other.

5.3 The Central Bank and Taxes

A general equilibrium model also requires that we account for the revenue and operating expenses of the central bank. In particular, any profits or losses should be incorporated into a welfare comparison of settlement frequency. To this end, suppose we require that the central bank break even.[18] Therefore, if the cost of bailing out banks and and reimbursing depositors exceeds the revenue from intra-day lending, a lump-sum tax is levied on all agents. Conversely, if revenue exceeds costs, agents receive a lump-sum subsidy.

We know from Section 3 that the central bank's revenue is

$$\Psi(\alpha;F) = \mathbb{E}_l\left[\psi\big[b(l,\alpha;F)\big]\right]. \tag{18}$$

Let $\bar{z}(l;F,\psi)$ be the shock z' such that $z' - (\alpha^*(F;\psi)d - l_T) = \bar{D}$. We have assumed that for all $z \leq \bar{z}(l;F,\psi)$, the central bank bails out the bank and pays z. Otherwise, for all $z > \bar{z}(l;F,\psi)$, the central bank must reimburse those depositors not covered by the bank's remaining reserves. At time T, the bank has $d - l_T$ depositors, and R_{T-1} reserves; so if $d - l_T > R_{T-1}$ the central bank reimburses depositors, whereas if $R_{T-1} > d - l_T$ the central bank collects the remaining reserves of the defaulted bank. Therefore, for a given frequency F and policy ψ, the central bank faces costs of

$$\mathbb{E}_l\left[\int_0^{\bar{z}(l;F,\psi)} zG(z)dz + \int_{\bar{z}(l;F,\psi)}^{\infty}(d - l_T - R_{T-1}(l,\alpha^*(F;\psi);F))G(z)dz\right]. \tag{19}$$

Therefore, the tax associated with frequency F and policy ψ, which we denote $\tau(F,\psi)$, is simply the difference between equations (19) and (18).

5.4 Banking Equilibria

We denote the value function of an agent with zero money holdings (prior to paying the costs of default) V_0, and the value function of an agent with one unit of money

---

[18] What is important for welfare analysis is *not* the assumption that profits or losses within a specific settlement system equate to a given value, but rather that two systems being compared yield the *same* profits or losses.

(prior to paying banking fees or the costs of default) V_1. With this notation, we derive the flow Bellman equations:

$$\begin{aligned} rV_0 &= M_1x(V_1 - V_0 - c) - (1+r)[(1-\xi)\omega] - \tau && (20) \\ rV_1 &= (1-M_1)x(u + V_0 - V_1) - (1+r)[\phi + \xi\omega] - \tau && (21) \end{aligned}$$

where we've simplified notation by setting $\omega \equiv \Omega\big(\delta(F)\big)$.

An equilibrium in which agents use banks must satisfy the following conditions:

1. *Individual Rationality*: neither producers nor consumers would rather drop out of the economy and live in autarchy.

$$V_0 \geq 0 \tag{22}$$
$$V_1 \geq 0 \tag{23}$$

2. *Incentive Compatibility*: both producers and consumers will trade in a bilateral match.

$$V_1 - V_0 - c \geq 0 \tag{24}$$
$$u + V_0 - V_1 \geq 0 \tag{25}$$

3. *Banking*: agents with money choose to use banks.

$$\phi \leq 0 \tag{26}$$

These simplify[19] to the following conditions, expressed as constraints on the banking fee ϕ:

$$\begin{aligned} \phi &\leq 0 && (27) \\ \phi &\leq \frac{(1-M_1)x(u-c) - rc}{1+r} + \omega\Big[1 - 2\xi - \frac{(r+x)(1-\xi)}{M_1x}\Big] - \frac{(r+x)\tau}{(1+r)M_1x} && (28) \\ \phi &\geq \frac{M_1xc + (1+r)(1-2\xi)\omega - (r + M_1x)u}{1+r} && (29) \end{aligned}$$

The first and second conditions, respectively, ensure that banks pay interest on deposits, and that this interest is sufficiently large that a seller is willing to incur the costs of production in order to become a buyer. The third condition ensures that the interest earned on deposits is not so high that buyers are unwilling to withdraw their deposits from the bank.

In equilibrium, the loan price (or interest rate) that clears the market must satisfy $\rho = V_1 - V_0$. We solve and get that

$$\rho = \frac{(1-M_1)xu - M_1xc - (1+r)\phi + (1+r)(1-2\xi)\omega}{r+x}. \tag{30}$$

---

[19] It turns out that $V_0 \geq 0 \Rightarrow V_1 - V_0 - c \geq 0$, and that this further implies that $V_1 \geq 0$.

Definition 8 *Consider an economy* $\{a, d, r, c, x, T, M, \bar{D}, \tau, \Pi(\cdot), G(\cdot), \Omega(\cdot)\}$*. Given a settlement system* $\{F, \psi, \xi\}$*, a* ***banking equilibrium*** *is a fixed point* $\{\alpha^*, \rho^*\}$ *such that*

(I) *The Banking Sector: Given* ρ^**,* α^* *satisfies the bank's optimality condition in equation (3). Moreover,* α^* *uniquely determines banking fees* ϕ^* *according to (4), the size of default* δ^* *according to (5), and the total money supply* $M_1^* = \frac{M}{\alpha}$*.*

(II) *The Trading Sector: Given* $\{\phi^*, \delta^*, M_1^*\}$*, the constraints (27) - (29) are satisfied and the price of loans* ρ^* *is determined as in equation (30).*

(III) *Aggregate Consistency: Deposits equal the supply of fiat money, so that* $d = M$*.*

(IV) *Central Bank Balances Budget: Taxes* τ *are determined by the difference between (19) and (18).*

5.5 Welfare

Assuming equal weight on all agents, we can define equilibrium welfare

$$W^* = (1 - M_1^*)V_0^* + M_1^* V_1^*.$$

Solving equations (18) and (19), we see that welfare is inversely related to both banking fees and default costs:

$$W^* \propto M_1^* \Big[(1 - M_1^*)x(u - c) - (1 + r)\phi^*\Big] - (1 + r)\Big[M_1^*(2\xi - 1) + 1 - \xi\Big]\omega^* - \tau. \quad (31)$$

In the next section, we consider how settlement frequency affects equilibrium prices, behavior, and welfare within the context of the simple example introduced in section 4.

6 Extending the Example

We now specify the remaining parameter values and solve the example in the general equilibrium framework. Let us assume that $d = .15, \bar{D} = .3, r = .05$ and $G \sim \mathcal{U}[0, \bar{D}]$. Furthermore, we specify $a = 0, q = .05, u = .4, c = .05, x = .2$, $\sigma = .1$, and $\Omega(\delta) = \kappa\delta$. We begin with baseline values of $\kappa = 30$ and $\xi = .5$, though we will later analyze how the equilibrium changes as these values vary.

We derive a *demand curve* for each F_i by substituting the appropriate equations for $\phi(F_i)$ (equations (8) and (14), respectively) and $\delta(F_i)$ (equations (11) and (17), respectively) into equation (30). This demand curve relates the price of money, ρ, to the supply of money, $\frac{M}{\alpha}$, and is downward sloping. The *supply curve*, given by the bank's optimality condition (equations (7) and (13), respectively), illustrates an upward sloping relationship between the price of loans and a bank's willingness to supply them. We compute the intersection to find $\{\alpha^*(F_i), \rho^*(F_i)\}$, which we can use to pin down the banking fees $\phi^*(F_i)$, the costs of default $\omega^*(F_i)$, and the welfare

$W^*(F_i)$. Though closed form solutions are not practical, we continue our analysis of welfare and optimality graphically.[20]

6.1 The Optimal Borrowing Policy

A topic of great debate within the settlement system literature is the central bank's optimal policy for intra-day borrowing. The analysis below suggests something that has been mostly overlooked: the optimal policy varies across settlement frequency, and thus any discussion of borrowing policies must be made within the context of a specific settlement frequency.

We now consider the effect of the central bank's borrowing policy on equilibrium reserves, fees, loan prices, default size, and welfare within the context of the example introduced in Section 4. Figure 4A illustrates a bank's equilibrium reserve ratio, $\alpha^*(F_i)$, across different borrowing fees, represented here by η. As expected, we find that for any η a bank retains larger reserves under F_2 than under F_1. Recall that in the partial equilibrium model, we show that for any given ρ, $\phi(F_1) < \phi(F_2)$. However, in the general equilibrium framework, ρ is an endogenous variable and will be different under F_1 and F_2. As we see in Figure 4B, for any borrowing policy η, the price of loans is higher under real-time settlement than it is under deferred settlement. The effect of increased revenue from larger values of ρ is so strong that it dominates the effects of decreased revenue from larger intra-day liquidity requirements, and the result is in fact *decreased fees* (or increased interest payments) under F_2, as illustrated in Figure 4C. Therefore, in equilibrium, the costs of increased settlement frequency are passed to agents in the form of higher *loan* fees, and not banking fees. As we would expect, increased frequency decreases the costs of default. In Figure 4D, we see that for any η, the costs of default under F_1 are greater than those under F_2.

[INSERT FIGURE 4A - 4E HERE]

We now turn our attention to equilibrium welfare as a function of the policy parameter η. From Figure 4E, we see that the optimal policy under F_2 implies a smaller η than under F_1. Moreover, under the current (arbitrary) parameterization, welfare under a real-time settlement system with an optimal borrowing policy is greater than welfare under a deferred settlement system with an optimal borrowing scheme. The intuition here is simple: since a bank's liquidity needs are greater under a real-time system, reserve management is more sensitive to changes in the borrowing fee at low levels of η. In turn, the central bank is able to achieve the proper balance of risk and cost, via the effect of η on α, at a cheaper price (i.e for a smaller η).

As a result, the banking costs that are passed along to depositors are small relative to the benefits of decreased risk from default. Of course, under an alternative parameter specification, the reverse may be true: the costs of achieving the efficient

---

[20] All parameter values considered below support equilibria that satisfy the constraints (20) - (24), as well as the example-specific constraint that $\eta < \frac{2r\rho}{dq}$.

level of reserves under a real-time system may be large relative to the gains from decreased default risk, and thus a deferred settlement system would be preferred.

6.2 Welfare and the Costs of Default

The equilibria described above will no doubt be sensitive to the specific parameterization that has been imposed on the model. In light of our ad hoc approach to assigning parameter values, some sensitivity analysis seems appropriate in order to understand how equilibria change under alternative specifications. Therefore, we first set $\eta = 1.001$, which corresponds to the borrowing fee for which welfare is equivalent under F_1 and F_2, and proceed to graph the sensitivity of equilibrium welfare to three key parameters: κ, $G(\cdot)$, and ξ.

In Figure 5A, we see that for smaller levels of κ deferred settlement is the preferred system, while real-time settlement is more appropriate for large levels of κ. This result is natural: a settlement system that reduces defaults should be more efficient when defaults are more costly. For example, a system that is more vulnerable to contagion when a crisis occurs may correspond to a higher κ, and therefore prefer real-time settlement.

[INSERT FIGURE 5A - 5C HERE]

A second crucial determinant of the optimal settlement system is $G(\cdot)$, the distribution of exogenous shocks that hit a bank. In Figure 5B, we consider equilibrium welfare as we increase the upper bound, say $zmax$, on shocks (maintaining the assumption that G is uniform). Again, the result is intuitive: as the upper bound on shocks becomes larger, the effect of responsible reserve management diminishes. Therefore, the probability effect discussed earlier vanishes, and the costs of real-time settlement are no longer balanced by the benefits of decreased default probability. The question arising here, which seems consistent with real-world concerns, is: what sort of shocks is the central bank worried about? If they are worried about small shocks that can be avoided with more responsible reserve management, then perhaps the liquidity requirements of a real-time system would be beneficial. However, if they are more concerned with large shocks that will overwhelm a bank regardless of its previous financial position, then some of the appeal of a real-time system is lost (though not all), and a deferred system may be preferred.

Finally, we consider the effects of shifting the burden of default, represented by ξ, from the sellers to the buyers. From Figure 5C, we find that, for the current set of parameters, $W^*(F_2)$ increases as we shift the burden of default to the buyers (i.e as $\xi \to 1$), while $W^*(F_1)$ remains relatively flat. The reason is quite simple: the reserve ratio under F_2 implies that a larger proportion of the population are sellers, while there is a more even split of buyers and sellers under F_1. However, this result seems to be more a result of modeling abstraction than real-world phenomena. To better understand the consequences of recovery practices would likely require modification of the model.

7 Conclusion

This paper develops a partial equilibrium model of reserve management in the presence of stochastic realizations of intra-day liabilities and costly borrowing from the central bank. Within this framework, increased settlement frequency implies that banks face greater financial costs. In a perfectly competitive banking sector, these costs are transferred to the customers via higher banking fees. However, in the presence of exogenous shocks to net worth, increased settlement frequency also mitigates the risk to the economy by decreasing both the probability and the size of potential defaults. In the context of a simple example, we illustrate this tradeoff between cost and risk, and examine its sensitivity to key model parameters. One finding is that banks may respond very differently to a given borrowing policy under alternative settlement frequencies. Specifically, banks are very responsive to small borrowing costs under a real-time system, while they are not under a deferred system. We also find that if the distribution of intra-day liabilities is mean-reverting (to zero), a deferred settlement system may be preferred. On the other hand, if there is a larger probability of ending the day with large outstanding liabilities, a real-time system may be more appropriate.

We then complete the model by introducing a trading sector in which agents endogenously choose to use credit to conduct bilateral transactions. In this general equilibrium framework, the effects of banking fees and default risk on the behavior of both buyers and sellers are taken into account. We observe the sensitivity of steady-state welfare to policy parameters (such as η and ξ), as well as parameters intrinsic to the economy (such as κ and G). We confirm that the optimal borrowing policy for a real-time settlement system imposes smaller borrowing fees on banks than the optimal borrowing policy for a deferred settlement system. Moreover, we find that economies which face larger costs of recovery from a default (for example, when the banking sector is vulnerable to contagion) and economies that are more prone to external shocks prefer a real-time settlement system. Conversely, in economies in which these risks are not as substantial, a less expensive deferred settlement system may be welfare-improving.

The primary contribution of this paper is the development and synthesis of a partial equilibrium model of a banking sector with an existing model of trade. The result is a dynamic general equilibrium model that captures key elements of *both* bank's and consumer's behavior. The policy experiments and sensitivity analysis we perform represent only a small fraction of the potential topics that this model could address. For example, heterogeneity with respect to the costs (a) that banks face may provide an interesting mechanism to explore *why* certain banks choose to join a settlement system and others do not. Moreover, a true inter-temporal banking framework and more careful consideration of the central bank's bail-out policies might provide new insights into the moral hazard that arises when banks don't internalize the risks of default. Finally, by considering the case of divisible money within the trading sector, we might find that equilibrium prices differ for agents that pay with credit as opposed to cash (when both are circulating). These are areas for future work.

References

Allen Berger, Diana Hancock, and Jeffrey Marquardt (1996), "A Framework for Analyzing Efficiency, Risks, Costs, and Innovations in the Payments Systems." *Journal of Money, Credit and Banking* 28, 696 - 732.

Committee on Payment and Settlement Systems (2003), "The Role of Central Bank Money in Payment Systems," Bank for International Settlements.

Committee on Payment and Settlement Systems (2005), "Statistics on Payment and Settlement Systems in Selected Countries - Figures for 2003," Bank for International Settlements.

Scott Freeman (1996), "The Payment System, Liquidity, and Rediscounting," *American Economic Review* 86, 1126 - 1138.

Edward Green (1999). "Money and Debt in the Structure of Payments," *Quarterly Review*, Federal Reserve Bank of Minneapolis.

Ping He, Lixin Huang, and Randall Wright (2003), "Money and Banking in Search Equilibrium." University of Pennsylvania, *mimeo.*

Charles Kahn and William Roberds (2001), "Real-time Gross Settlement and the Costs of Immediacy." *Journal of Monetary Economics* 47, 299 - 319.

Charles Kahn and William Roberds (1998), "Payments System Settlements and Bank Incentives." *Review of Financial Studies* 11, 845 - 870.

Thorsten Koeppl, Cyril Monnet, and Ted Temzelides (2005), "Mechanism Design and Payments. " *mimeo.*

Charles Kahn, James McAndrews, and William Roberds (1999), "Settlement Risk Under Gross and Net Settlement ." *Journal of Money, Credit, and Banking* 35 (4), 591 - 608.

Ricardo Lagos and Randall Wright (2005), "A Unified Framework for Monetary Theory and Policy Analysis." *Journal of Political Economy* 113 (3), 463 - 484.

Benjamin Lester, Stephen Millard, and Matthew Willison (2004), "Settlement Systems in Search Equilibria with Bank Default." *mimeo*, Bank of England.

Antoine Martin (2004), "Optimal Pricing of Intraday Liquidity." *Journal of Monetary Economics* 51 (2), 401 - 424.

Jean-Charles Rochet and Jean Tirole (1996), "Controlling Risk in Payment Systems." *Journal of Money, Credit and Banking* 28 (4), 832-862.

Ruilin Zhou (2000) "Understanding Intraday Credit in Large-Value Payment Systems." *Economic Perspectives*, Federal Reserve Bank of Chicago, third quarter, 29-44.

8 Appendix

8.0.1 Proof of Claim 1

Proof. First, note that

$$\begin{aligned} b(l;\alpha,F_1) &= \max\{0, l_2-\alpha d\} \\ b(l;\alpha,F_2) &= \max\{0, l_1-\alpha d, l_2-\alpha d\} \end{aligned}$$

This implies that, for $F \in \{F_1, F_2\}$, if $\alpha_2 > \alpha_1$, then

$$\begin{aligned} b(l;\alpha_1,F) &\geq b(l;\alpha_2,F) && \text{for all } l \\ b(l;\alpha_1,F) &> b(l;\alpha_2,F) && \text{for } l = (\bar{L}_t, t \in \mathcal{S}). \end{aligned}$$

Since ψ is strictly increasing, this implies that for $F \in \{F_1, F_2\}$, if $\alpha_2 > \alpha_1$, then

$$\begin{aligned} \psi[b(l;\alpha_1,F)] &\geq \psi[b(l;\alpha_2,F)] && \text{for all } l \\ \psi[b(l;\alpha_1,F)] &> \psi[b(l;\alpha_2,F)] && \text{for } l = (\bar{L}_t, t \in \mathcal{S}). \end{aligned}$$

We conclude that

$$\begin{aligned} \alpha_2 > \alpha_1 &\Rightarrow \mathbb{E}_l\Big[\psi[b(l;\alpha_1,F)]\Big] > \mathbb{E}_l\Big[\psi[b(l;\alpha_2,F)]\Big] \\ &\Rightarrow \Psi(\alpha_1;F) > \Psi(\alpha_2;F). \end{aligned}$$

■

8.0.2 Proof of Claim 2

Proof. Again, since

$$\begin{aligned} b(l;\alpha,F_1) &= \max\{0, l_2-\alpha d\} \\ b(l;\alpha,F_2) &= \max\{0, l_1-\alpha d, l_2-\alpha d\} \end{aligned}$$

we know that, for any $\alpha \in (0, \frac{\bar{L}}{d})$,

$$b(l;\alpha,F_2) \geq b(l;\alpha,F_1) \qquad \text{for all } l.$$

By the same logic as above, we conclude that

$$\Psi(\alpha;F_1) \leq \Psi(\alpha;F_2).$$

■

8.0.3 Toward a Proof of Result 1

We require two intermediate results to prove Result 1:

Result 6 *Expected borrowing fees exhibit convexity in α at any settlement frequency. That is, for $F \in \{F_1, F_2\}$,*

$$\alpha_1 > \alpha_2 \qquad \Rightarrow \qquad \frac{\partial}{\partial\alpha}\Big[\Psi(\alpha_1; F)\Big] > \frac{\partial}{\partial\alpha}\Big[\Psi(\alpha_2; F)\Big].$$

Proof. Suppose that $\alpha_1 > \alpha_2$. Since ψ is strictly increasing and strictly convex, we know that

$$b_2 > b_1 \qquad \Rightarrow \qquad \frac{\psi(b_2 + \Delta b) - \psi(b_2)}{\Delta b} > \frac{\psi(b_1 + \Delta b) - \psi(b_1)}{\Delta b}$$

for $\Delta b > 0$ sufficiently small. It follows that

$$b_2 > b_1 \qquad \Rightarrow \qquad \psi(b_2 + \Delta b_2) - \psi(b_2) > \psi(b_1 + \Delta b_1) - \psi(b_1) \tag{32}$$

if $\Delta b_2 \geq \Delta b_1 > 0$. Recall that a positive perturbation in α is a negative perturbation in b. Therefore, let

$$\begin{aligned} b_1 &= b(l; \alpha_1 + \Delta\alpha, F) = \max\{0, l_2 - (\alpha_1 + \Delta\alpha)d\} \\ b_1 + \Delta b_1 &= b(l; \alpha_1, F) = \max\{0, l_2 - \alpha_1 d\} \\ b_2 &= b(l; \alpha_2 + \Delta\alpha, F) = \max\{0, l_2 - (\alpha_2 + \Delta\alpha)d\} \\ b_2 + \Delta b_2 &= b(l; \alpha_2, F) = \max\{0, l_2 - \alpha_2 d\} \end{aligned}$$

for sufficiently small $\Delta\alpha > 0$. We know that $b_2 \geq b_1$ for all l, and strictly greater for some l that occurs with positive probability. We proceed with the proof for F_1, as the corresponding proof for F_2 is essentially identical.

$$\begin{aligned} \Delta b_1 &= (b_1 + \Delta b_1) - b_1 = \max\{0, l_2 - \alpha_1 d\} - \max\{0, l_2 - (\alpha_1 + \Delta\alpha)d\} \\ \Delta b_2 &= (b_2 + \Delta b_2) - b_2 = \max\{0, l_2 - \alpha_2 d\} - \max\{0, l_2 - (\alpha_2 + \Delta\alpha)d\} \end{aligned}$$

Consider three cases:

$$\begin{aligned} l_2 - \alpha_2 d \; > l_2 - \alpha_1 d \; > 0 \quad &\Rightarrow \quad \Delta b_1 = \Delta b_2 = \Delta\alpha d. \\ l_2 - \alpha_2 d \; > 0 > l_2 - \alpha_1 d \quad &\Rightarrow \quad \Delta b_1 = 0 < \Delta\alpha d = \Delta b_2. \\ 0 > l_2 - \alpha_2 d \; > l_2 - \alpha_1 d \quad &\Rightarrow \quad \Delta b_1 = \Delta b_2 = 0. \end{aligned}$$

Clearly, $\Delta b_1 \leq \Delta b_2$ for all l, and moreover $\Delta b_1 > 0$ for some l with $\Pi(l) > 0$. So by (32) we have that $\alpha_1 > \alpha_2$ implies

$$\psi[b(l; \alpha_1 + \Delta\alpha, F)] - \psi[b(l; \alpha_1, F)] \geq \psi[b(l; \alpha_2 + \Delta\alpha, F)] - \psi[b(l; \alpha_2, F)]$$

for all l and strictly greater than for some l with $\Pi(l) > 0$. Therefore

$$\begin{aligned} \alpha_1 > \alpha_2 \qquad &\Rightarrow \qquad \frac{\Psi(\alpha_1 + \Delta\alpha; F) - \Psi(\alpha_1; F)}{\Delta\alpha} > \frac{\Psi(\alpha_2 + \Delta\alpha; F) - \Psi(\alpha_2; F)}{\Delta\alpha}, \\ &\Rightarrow \qquad \frac{\partial}{\partial\alpha}\Big[\Psi(\alpha_1; F)\Big] > \frac{\partial}{\partial\alpha}\Big[\Psi(\alpha_1; F)\Big] \end{aligned}$$

or, equivalently,

$$\frac{\partial^2}{\partial\alpha^2}\Big[\Psi(\alpha; F)\Big] > 0,$$

the desired result. ■

Result 7 *The impact of a marginal increase in the reserve ratio is larger (more negative) under a real-time settlement system than a deferred settlement system. That is,*

$$\frac{\partial}{\partial\alpha}\Big[\Psi(\alpha;F_1)\Big] > \frac{\partial}{\partial\alpha}\Big[\Psi(\alpha;F_2)\Big].$$

Proof. We follow the same logic employed in the proof of Result 1. Let

$$\begin{aligned}
b_1 &= b(l;\alpha+\Delta\alpha,F_1) = \max\{0, l_2-(\alpha+\Delta\alpha)d\} \\
b_1+\Delta b_1 &= b(l;\alpha,F_1) = \max\{0, l_2-\alpha d\} \\
b_2 &= b(l;\alpha+\Delta\alpha,F_2) = \max\{0, l_1-(\alpha+\Delta\alpha)d, l_2-(\alpha+\Delta\alpha)d\} \\
b_2+\Delta b_2 &= b(l;\alpha,F_2) = \max\{0, l_1-\alpha_2 d, l_2-\alpha_2 d\}
\end{aligned}$$

for a sufficiently small $\Delta\alpha > 0$. We have shown above that $b_2 \geq b_1$ for all l and strictly greater for some l with $\Pi(l) > 0$. We now consider, again, Δb_1 and Δb_2:

$$\begin{aligned}
\Delta b_1 &= \max\{0, l_2-\alpha d\} - \max\{0, l_2-(\alpha+\Delta\alpha)d\} \\
\Delta b_2 &= \max\{0, l_1-\alpha d, l_2-\alpha d\} - \max\{0, l_1-(\alpha+\Delta\alpha)d, l_2-(\alpha+\Delta\alpha)d\}
\end{aligned}$$

Clearly, for $l_2 > l_1$, we have that $b_1 = b_2$ and $\Delta b_1 = \Delta b_2$ so that $\psi(b_1+\Delta b_1)-\psi(b_1) = \psi(b_2+\Delta b_2)-\psi(b_2)$. For the case of $l_1 > l_2$, we have three subcases:

$$\begin{aligned}
l_1-\alpha d &> l_2-\alpha d > 0 & \Rightarrow \quad & \Delta b_1 = \Delta b_2 = \Delta\alpha d. \\
l_1-\alpha d &> 0 > l_2-\alpha d & \Rightarrow \quad & \Delta b_1 = 0 < \Delta\alpha d = \Delta b_2. \\
0 > l_1-\alpha d &> l_2-\alpha d & \Rightarrow \quad & \Delta b_1 = \Delta b_2 = 0.
\end{aligned}$$

Therefore, by (32), we conclude that

$$\psi[b(l;\alpha,F_2)] - \psi[b(l;\alpha+\Delta\alpha,F_2)] \geq \psi[b(l;\alpha,F_1)] - \psi[b(l;\alpha+\Delta\alpha,F_1)]$$

for all l and strictly greater than for some l with $\Pi(l) > 0$. Therefore, we can rearrange to conclude that

$$\frac{\Psi(\alpha+\Delta\alpha;F_1)-\Psi(\alpha;F_1)}{\Delta\alpha} > \frac{\Psi(\alpha+\Delta\alpha;F_2)-\Psi(\alpha;F_2)}{\Delta\alpha},$$

the desired result. ■

8.0.4 Proof of Result 1

Proof. By the bank's first order condition, $\alpha^*(F_1)$ and $\alpha^*(F_2)$ satisfy

$$\frac{\partial}{\partial\alpha}\Big[\Psi(\alpha^*(F_1);F_1)\Big] = \frac{\partial}{\partial\alpha}\Big[\Psi(\alpha^*(F_2);F_2)\Big] = dr\rho.$$

Suppose, towards a contradiction, that $\alpha^*(F_1) > \alpha^*(F_2)$. By Result 6,

$$\frac{\partial}{\partial\alpha}\Big[\Psi(\alpha^*(F_1);F_1)\Big] > \frac{\partial}{\partial\alpha}\Big[\Psi(\alpha^*(F_2);F_1)\Big].$$

But by Result 7, we have

$$\frac{\partial}{\partial\alpha}\left[\Psi(\alpha^*(F_2);F_1)\right] > \frac{\partial}{\partial\alpha}\left[\Psi(\alpha^*(F_2);F_2)\right],$$

implying that

$$\frac{\partial}{\partial\alpha}\left[\Psi(\alpha^*(F_1);F_1)\right] > \frac{\partial}{\partial\alpha}\left[\Psi(\alpha^*(F_2);F_2)\right],$$

a contradiction. ■

8.0.5 Proof of Result 2

Proof. Since $\alpha^*(F_1)$ is the optimal solution to the bank's problem,

$$[1-\alpha^*(F_1)]r\rho - a - \frac{\Psi(\alpha^*(F_1),F_1)}{d} \geq [1-\alpha^*(F_2)]r\rho - a - \frac{\Psi(\alpha^*(F_2),F_1)}{d}.$$

This simplifies to

$$[\alpha^*(F_2)-\alpha^*(F_1)]r\rho \geq \frac{\Psi(\alpha^*(F_1),F_1)-\Psi(\alpha^*(F_2),F_1)}{d}.$$

By Claim 2 $\Psi(\alpha;F_2) > \Psi(\alpha;F_1)$, so that

$$[\alpha^*(F_2)-\alpha^*(F_1)]r\rho > \frac{\Psi(\alpha^*(F_1),F_1)-\Psi(\alpha^*(F_2),F_2)}{d}.$$

This means that

$$a-[1-\alpha^*(F_2)]r\rho + \frac{\Psi(\alpha^*(F_2),F_2)}{d} \geq a-[1-\alpha^*(F_1)]r\rho + \frac{\Psi(\alpha^*(F_1),F_1)}{d},$$

or, equivalently,

$$\phi(F_2) \geq \phi(F_1).$$

■

8.0.6 Proof of Claim 3

Proof. The proof is trivial, once it's noted that

$$R_T(l,\alpha;F) + A_{s_T}(l;F) - (l_T - l_{s_T}) = \max\{l_1-l_2, \alpha d - l_2\}.$$

■

8.0.7 Proof of Result 4

Proof. For an arbitrary T, we can solve backwards to get that

$$\hat{l}_t(l;F) = \max\left\{0, \min\{l_t-l_{s_t}, l_t-l_{s_{t-1}}, ...l_t-l_{s_1}, l_t\}\right\}.$$

Therefore, for $T=2$,

$$\begin{aligned}\hat{l}_1(l;F_1) = max\{0,l_1\} &= \hat{l}_1(l;F_2)\\ \hat{l}_2(l;F_1) = max\{0,l_2\} &\geq \max\left\{0,\min\{l_2,l_2-l_1\}\right\} = \hat{l}_2(l;F_2)\end{aligned}$$

for all l. ■

8.0.8 Proof that $\phi(F_2) > \phi(F_1)$ in simple example.

Proof. The proof is completed in three steps. First, note that

$$\phi(F_2) > \phi(F_1) \Leftrightarrow \frac{1}{2}(r\rho)(\frac{r\rho}{\eta dq} - 1) + \frac{\eta dq}{8} > 0.$$

Secondly, it is simple to show that

$$\frac{\partial}{\partial \eta}\Big[\frac{1}{2}(r\rho)(\frac{r\rho}{\eta dq} - 1) + \frac{\eta dq}{8}\Big] = -\frac{1}{2}(\frac{r\rho}{dq})^2(\frac{1}{\eta^2} + \frac{qd}{8} < 0 \Leftrightarrow \eta < \frac{2r\rho}{dq}.$$

Finally, we substitute to find

$$\eta = \frac{2r\rho}{dq} \Rightarrow \frac{1}{2}(r\rho)(\frac{r\rho}{\eta dq} - 1) + \frac{\eta dq}{8} = 0.$$

Therefore, we conclude that, $\forall \eta < \frac{2r\rho}{dq}$, $\phi(F_2) > \phi(F_1)$. ■

Figure 2A - Reserves (eta)

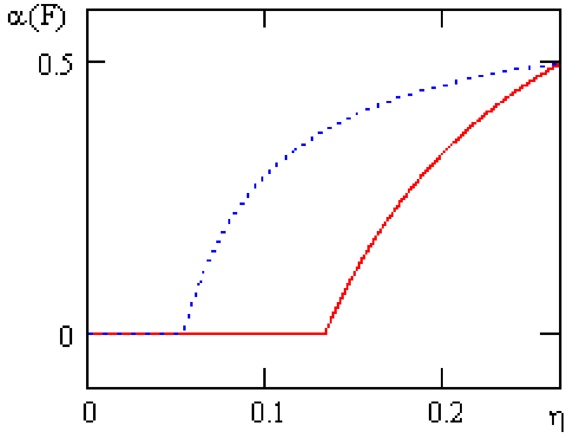


Figure 2B - Default (eta)

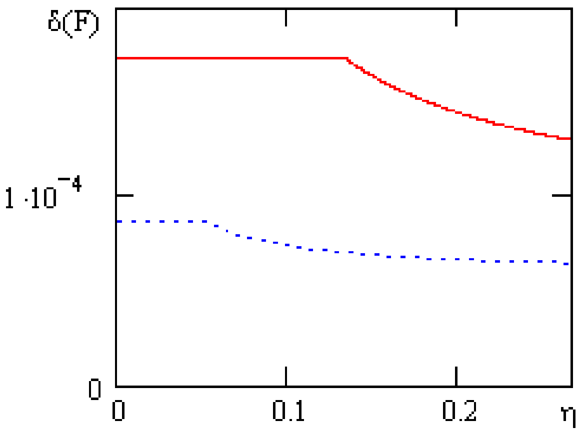


Figure 2C - Banking Fees (eta)

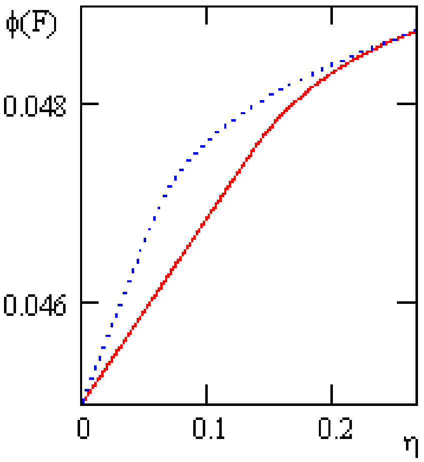


Figure 3A - Reserves (q)

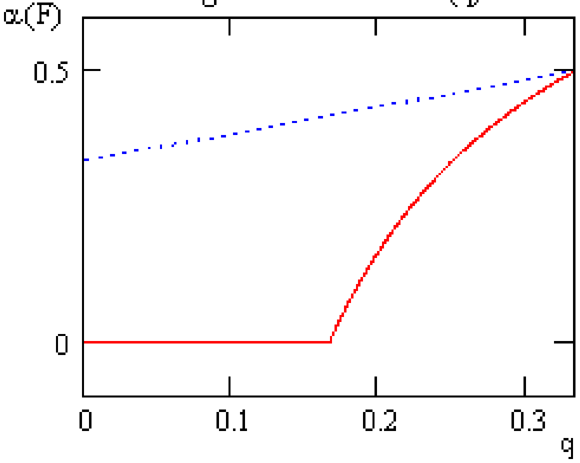


Figure 3B - Default (q)

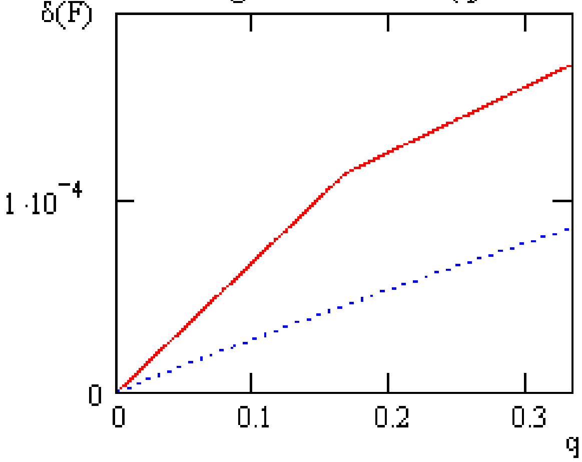
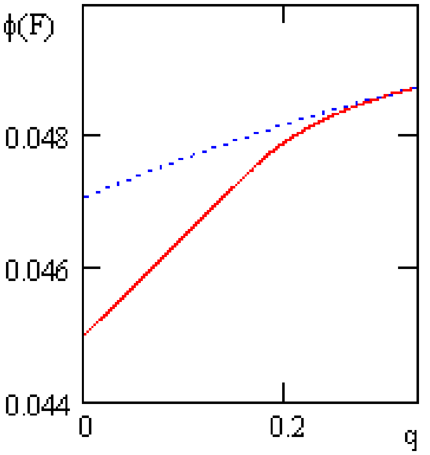


Figure 3C - Banking Fees (q)



F = 1

F = 2

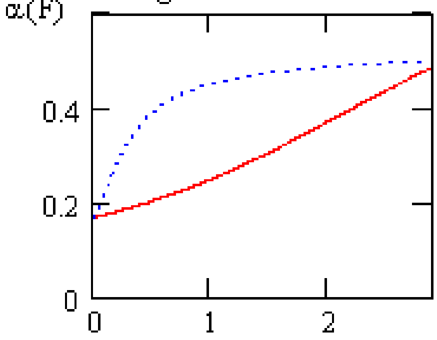


Figure 4A - Reserves

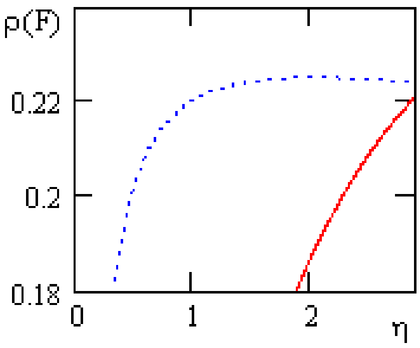


Figure 4B - Price of Loans

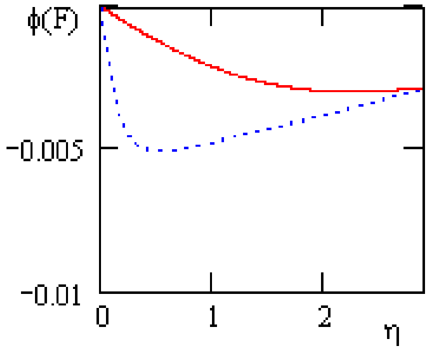


Figure 4C - Banking Fees

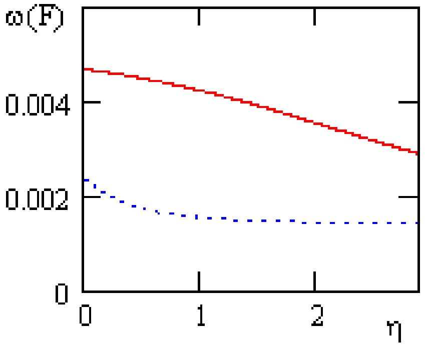


Figure 4D - Default Costs

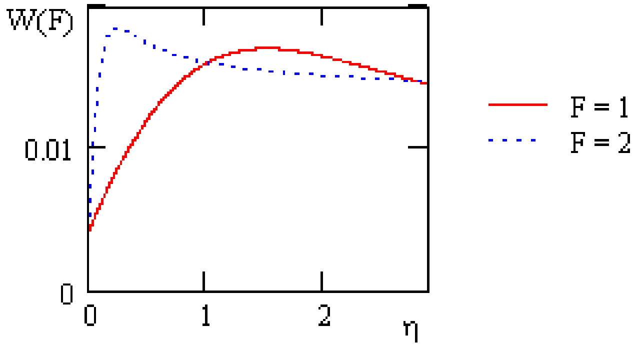


Figure 4E - Welfare

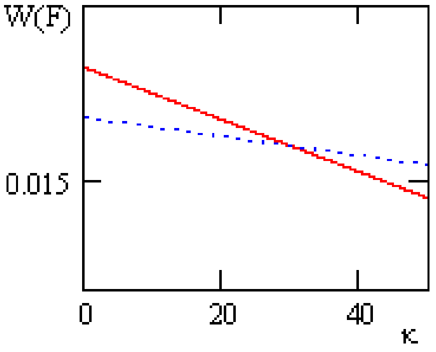


Figure 5A - Welfare (kappa)

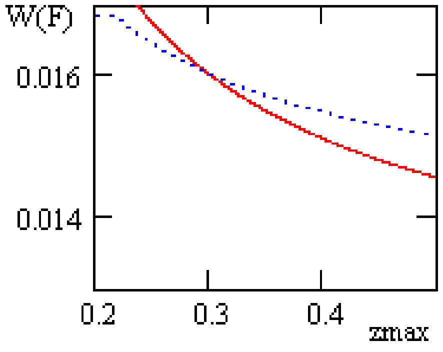


Figure 5B - Welfare (zmax)

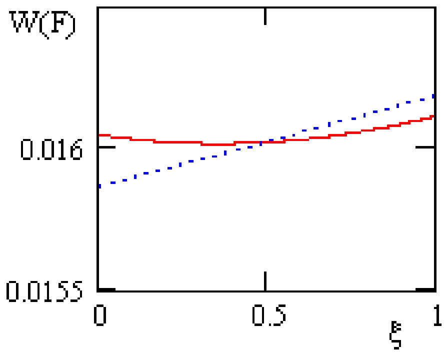


Figure 5C - Welfare (xi)