The Dissertation Committee for Justen Rene Kok Lye Ghwee certifies that this is the approved version of the following dissertation:

ESSAYS ON INTERMEDIATION, THE PAYMENTS SYSTEM AND MONETARY POLICY IMPLEMENTATION

Committee:

____

David A. Kendrick, Supervisor

____

Beatrix Paal, Co-Supervisor

____

Russell W. Cooper

____

Vincent J. Geraci

____

Paula Hernandez-Verme

ESSAYS ON INTERMEDIATION, THE PAYMENTS SYSTEM AND MONETARY POLICY IMPLEMENTATION

by

Justen Rene Kok Lye Ghwee, B. A.; M.A.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

the University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2005

To Low Besar

ESSAYS ON INTERMEDIATION, THE PAYMENTS SYSTEM AND MONETARY POLICY IMPLEMENTATION

Publication No. _________

Justen Rene Kok Lye Ghwee, Ph.D.
The University of Texas at Austin, 2005

Supervisors: David A. Kendrick and Beatrix Paal

This first essay reconsiders how a central bank might tailor its monetary policy in response to a liquidity shortage problem that arises from payments system design. Short run monetary intervention that completely mitigates liquidity shortage achieves Pareto optimality. However, it is not Pareto improving: by inducing shifts in agents' portfolio choice, short run monetary policy alters the long term real interest rate, and consequently, the distribution of consumption goods among heterogeneous agents. A regime that pays interest on reserves could attain Pareto improving allocation, but is never Pareto optimal. Under the interest on reserves scheme, the central bank can pursue policy targeting the quantity of reserves balances for liquidity provision purpose independently of policy targeting the interest rate for other broad monetary policy objectives.

The second essay evaluates the performance of the quadratic linear programming (QLP) method in accounting for a bank's liquidity management over the ten-day reserves maintenance period (RMP). The QLP method reasonably captures the qualitative features of the bank's demand for excess reserves. The simulated demand schedule is weakly J-shaped, implying greater demand for reserves as the reserve settlement day approaches. While institutional features account for the cyclical patterns in the earlier days of the RMP, bank's reserves "locked-in" cost avoidance activity and uncertainty about the size of central bank refinancing rationalize the large surge in the demand for reserves towards the settlement day. However, the QLP method is less successful in emulating the magnitude of the reserves demand dynamics comparable to that observed in the data.

The third essay examines the nature of equilibrium credit rationing under different assumptions with regard to investment technologies available to entrepreneurs applying for loans. Lenders ration credit to borrowers with low-risk investment technology in the form of (i) the constrained size of loan allotment, or (ii) the uncertainty in loan granting, but not both. The realized type of rationing depends on how much the borrower perceives the value of not being the recipient of one type of rationing over the other. Different loan market structures also imply different equilibrium loan contracts.

Table of Contents

Chapter 2. Bank's Liquidity Management and Monetary Policy Tools 49

Chapter 3. Credit Rationing and Intermediation 99

Appendices

List of Tables

List of Figures

Chapter 1

Payments System and Monetary Policy Tools

1.1 Introduction

This paper is an attempt to investigate some of the salient issues that arise from the interdependence between payments system and monetary policy implementation. Specifically, it tries to expand hitherto research on payments system to welfare analysis that takes into consideration the effect of short term liquidity provision policy on long term asset prices. Equivalently, it addresses the question of whether short term monetary policies have any real effect on resource allocation *via* changes in the long term real interest rate. The analysis is then extended to investigate the welfare effect of the recent proposal to pay interest on reserves.[1] Relatedly, it also examines the potential gain of an extra monetary policy tool *via* the interest rate payment on reserves.

Recent years have witnessed two opposing trends in financial institutions' activities. The growth in the volume of large value interbank payment flows handled by financial intermediaries since the early 1990s has been significant and widespread. For example,

[1] As of November 25, 2003, the Interest on Business Checking Act of 2003 (Bill # S.1967) has been read twice and referred to the Committee on Banking, Housing, and Urban Affairs of the US Senate.

in the United States, the average daily payment values has almost doubled since the past fourteen years to US$3 trillion (see *Table 1-1*), while the ratio of annual payment values to GDP is invariably high for most developed countries (see *Table 1-2*). However, concurrently there is a contrasting decline in financial institutions' demand for central bank reserves (see *Figure 1-1* in the case of the US). Indeed, today's amount of average daily payment value covered by reserves has halved since 1990 (see *Table 1-1*).[2]

These developments matter because the world's major central banks implement monetary policy by targeting and manipulating short term interest rates. This focus on interest rate policy is possible due to the stability, and hence the predictability, of the demand for central bank reserves by financial institutions in order to meet their reserve requirements. The rapid growth in payments flow and its unpredictable nature, which results in greater volatility in the demand schedule for reserves, could threaten this central bank's leverage over short term interest rate *via* traditional tools such as reserve requirements.[3] Simultaneously, however, there has been a strengthening of link between the structure of payments system and the monetary policy operating procedure.[4]

A more significant consequence of the diminishing reserves holding in the face of increasing payment flows is the risk of liquidity shortages. Recent liquidity crises, epitomized by the Fall 1998 Russian debt default, have led to the pursuing of persistent low interest rate policy by the Fed even during expansionary conditions (see *Figure 1-2a* and *Figure 1-2b*).[5] This is in line with hitherto theoretical models which argue that

[2]Part of this decease in reserves demand is due the ongoing progress in banking technology, such as the advent of sweep accounts. Financial institutions use sweep accounts as a mean to shift depositors' funds between non-interest bearing reservable account and interest bearing non-reservable account in order to reduce the amount of reserves during the day while meeting the reserve requirements which is in effect fortnightly. This practice has significantly decreased individual banks' reserves holding since the inception of sweep accounts *circa* 1995.

[3]Empirical evidences of increases in day-to-day volatility of reserves demand are documented by Bartolini, Bertola, and Prati (2002) and Furfine (2000) for the US, and Prati, Bartolini and Bertola (2002) for the G-7 economies.

[4]This is the case when reserve requirements are no longer binding, i.e. when the amount of reserves needed for payment needs is greater than the amount required by reserve requirements.

[5]The discount rate remained below its October 1998 level for five quarters thereafter. This is despite the expansionary nature of the economy then, as indicated by the growing nominal growth rate, the increasing inflation rate and positive real GDP gap. See Goodfriend (2002) and Lacker (2004).

driving the short term interest rate to zero in order to overcome liquidity shortages is optimal. However, this induces a basic question: does this short term policy intervention affect long run real asset prices, and in consequence, does it have any real wealth effect? Furthermore, such a policy could further fuel inflationary pressure. This leads to the practical query: how could monetary policy be implemented in such a way as to fulfill the dual objectives of liquidity provision and broad-economy well being? In other words, is there any policy that provides the latitude to pursue liquidity objective independent of other broad macroeconomic objectives? If such a policy exists, what are its welfare implications?

In order to address these questions, I develop and solve a payments system model in which there are three types of agents, namely, banks (creditors), non-banks (debtors), and a central bank in an exchange economy. The environment incorporates two different assets which are consumption debt and payment debt. The former is generated by the underlying real resources transaction between creditors and debtors, while the later is created solely for clearing unredeemed consumption debt among creditors. An innovation is the inclusion of an explicit demand for central bank reserves by creditors for precautionary motives in anticipation of future liquidity shortages in the payment debt market. This gives rise to endogenous short and long term interest rates which are linked *via* creditor's portfolio choice.

The model shows that under certain ranges of parameter values, the competitive equilibrium could be liquidity constrained. The allocation under this constrained competitive equilibrium is not optimal *vis-à-vis* the allocation under full liquidity equilibrium. The Monetary Authority, by temporarily injecting additional reserves into the economy *via* open market operations or discount window policy such that the short run interest rate is zero, could achieve Pareto optimality. These are standard results in the payments system literature.

However, this transient liquidity provision is not Pareto improving: *short* term policy intervention has both *short* and *long* terms real welfare effects. Infusing additional reserves in the economy causes the short term interest rate to fall, which in turn engenders decline in the long term asset prices. The decrease in the long run asset prices effects cheaper consumption for all debtors, at the expense of some creditors whose profit-taking in the payment debt market is eliminated by the fall in the short term interest rate.

A Pareto improving allocation is achievable by combining full liquidity provision with a policy that pays interest on reserves held by all creditors only and is financed by growth in money supply. The creditors are now compensated by interest income due to their reserves holding, but whose cost is imposed on all agents. Hence, this wealth transfer could mitigate the benefit debtors partake at creditors' loss when liquidity shortage is overcome *via* short term monetary policy. For some positive interest rate ranges, creditors could be sufficiently compensated without making debtors worse off, and vice versa.

Paying proportional interest on reserves increases reserve demand more than is proportional. However, paying positive interest is never Pareto optimal since risk averse creditors will never substitute reserves for other assets to sufficiently overcome a liquidity constraint. This is despite the fact that for creditors, the interest income always dominates the interest cost (in the form of inflation). The model also formalizes an idea proposed by Goodfriend (2002) that by paying interest on reserves, the central bank gains an extra policy tool to implement monetary policy. Specifically, targeting the quantity of reserve balances for liquidity provision purpose could be pursued independently from policy targeting interest rate for other broad monetary policy objectives.

The result that it is efficient for monetary authorities, for example, a central bank, to provide liquidity such that the short run interest rate is zero is robust. Freeman (1996, 1999), Green (1997), Khan and Roberds (2001), Martin (2004), Mills (2004) and Zhou (2000) all arrives at the same conclusion despite differing in each of their environments. However, due to the absence of endogenous mechanism linking short and long term prices, a basic issue of how short run policies induced by liquidity shortages affect long run welfare has not been adequately addressed. There are long term prices in Khan and Roberds (2001), Lacker (1997), Martin (2004) and Zhou (2000). Lacker (1997) shows that when there is an explicit demand for reserves by agents, efficiency is achieved when the short run interest rate equals the long run interest rate which is not necessarily zero. Khan and Roberds (2001) and Lacker (1997) also consider a policy of paying interest on reserves held by agents. They find that such policy is inefficient since it could not undo the inherent liquidity constraint. Nevertheless, the long term interest rate in these models is derived from exogenous money demand, and hence they do not offer satisfactory explanation on the link between short and long term prices.[6]

The plan of the paper is as follows. Section 2 illustrates the environment and describes how liquidity frictions could arise in a payments system. Section 3 then asks what are efficient allocations subject to both resource and liquidity constraints. Section 4 defines a competitive equilibrium and characterizes the equilibrium allocations when the liquidity constraint is non-binding and when it is binding. Next, in the presence of liquidity shortage problem, I look at how the equilibrium allocations under different institutional arrangements for making payments compare in welfare terms. Firstly, Section 5 shows that while a zero short term interest rate is a necessary condition for an equilibrium to be efficient, it might have long run real distributional effect on agents. Then, Section 6 considers a regime in which interest is paid on reserves held by creditors. It shows that while a policy of full liquidity provision and interest payment on reserves might not be Pareto optimal, it is Pareto improving for some positive values of interest rate on reserves. Section 7 concludes.

1.2 The Environment

The 2-period overlapping generation model essentially follows that of Freeman's (1996). There are two periods and each period is further divided into two subperiods, namely *subperiod-1* and *subperiod-2*.[7] For notational purpose, time period is denoted as period-

[6] An exception is Khan and Roberds (2001). However, they do not address welfare issue when agents are heterogeneous.

[7] Accordingly, these can be considered as the morning and afternoon subperiods.

(s, t), with $s = \{1, 2\}$ and $t = \{1, 2, 3...\}$ referring to subperiods and time periods respectively. There are $2n + 1$(where n is large) islands in the economy consisting of $2n$ outlying islands and one central island. The outlying islands consist of pairs of islands, in which each pair comprises two types of island: a "debtor" island and a "creditor" island, which are solely inhabited by debtors and creditors correspondingly. On each outlying island, a continuum of agents with mass 1 are born in each period-t. The size of the total population is constant for all $t > 0$. For simplicity, n is normalized to 1.

Creditors and debtors differ in three aspects: (i) the time pattern of their endowments, (ii) the time pattern of their desired consumption, and (iii) their original locations and movement pattern across islands. The assumptions made about these three facets are designed to create an environment in which a payments system arises such that questions pertaining to liquidity frictions could be addressed. The specific assumptions are as follow.

A young creditor born in time-$(1, t)$ is endowed with e units of the consumption good in period-$(1, t)$ and wants to consume at time period-$(2, t + 1)$. In contrast, a young debtor born in time-$(1, t)$ is endowed with e units of the consumption good at period-$(2, t)$, but wants to consume in both subperiods $(1, t)$ and $(2, t)$. These endowments are non-storable across periods, but are costlessly storable from one subperiod to the next within the agent's young period.

There is a problem of providing consumption for a creditor in his old age, while for a debtor the problem is providing consumption in the first subperiod when she is young. This is because trade between locations is not feasible: agents could not move freely between islands. In fact, each agent has a specific "travel itinerary" as follows. In order to overcome the provision of consumption problem, a debtor would visit her creditor in the same pair of islands in the first subperiod when they are young. Since she does not have any tradeable good in this subperiod-$(1, t)$, she would issue IOU in exchange for parts of the creditor's endowment.[8] The debtor promises to repay her creditor upon his

[8] It is assumed that each creditor holds a diversified portfolio of IOUs issued by different debtors.

presentation of the IOU to her at the central island where all debtors and creditors would go to in the first subperiod when old. A loan issued at time period-(1, t) carries the gross nominal interest rate of R_t. R_t could be interpreted as the *overnight* or the *long-term* nominal interest rate.

How does a young debtor repay her debt? In order to provide for his old age consumption, a creditor, who has visited the central island and whose loan is repaid by the debtors within his own cohort, would continue his journey to either a creditor or a debtor's island to purchase $y_{2,t+1}$ unit of consumption good from the next cohort. Old creditors arrive at their respective destinations after all travel by young debtors has been completed. An old creditor is free to choose whether he travels to a young creditor or a young debtor island.[9] He purchases the consumption good with the fiat money accrued from his debt holding that has been redeemed by a debtor. If the old creditor purchases $y_{2,t+1}$ unit of consumption good from a young debtor, then that young debtor could repay her loan (which she issued in the previous subperiod) in the next time period when she meets with her creditors at the central island. If the old creditor purchases $y_{2,t+1}$ from a young creditor, then that young creditor would hold the fiat money as reserves in order to finance his old-period consumption.[10] Let ζ be the fraction of old creditors that ultimately purchase consumption good from a young debtor.

There is also a continuum of initial old creditors with mass 1 scattered in the outlying islands. These old creditors aggregately own M_0 unit of fiat money. Let the nominal price of the consumption good be denoted by $P^i_{s,t}$, where $i = \{C, D\}$ signifies the location where the transaction takes place.

Each agent has an additively separable logarithmic utility function. Debtors and creditors' endowment and consumption patterns are given in the following table:

However, each IOU for consumption transaction is bilateral, i.e. each creditor holds on debtor's debt after the trade. There is no indivisibility of IOU problem. This is to make the environment resembles more like a large value payment system problem.

[9] Old creditors view young agents consumption goods as perfect substitute of each other.

[10] Hence, a young creditor has the choice of (i) lending his endowment to young debtors from the same cohort, or (ii) selling his endowment to old creditors, or (iii) doing both.

Agents	Endowment (1,t ; 2,t ; 1,t+1 ; 2,t+1)	Consumption (1,t ; 2,t ; 1,t+1 ; 2,t+1)	Utility
Debtors	(- ; e ; - ; -)	($y^D_{1,t}$; $y^D_{2,t}$; - ; -)	$\ln(y^D_{1,t}) + \ln(y^D_{2,t})$
Creditors	(e ; - ; - ; -)	(- ; - ; $y^C_{2,t+1}$; -)	$\ln(y^C_{2,t+1})$
Initial Creditors	M_0	(- ; - ; $y^C_{2,t+1}$; -)	$\ln(y^C_{2,t+1})$

In the central island, there is an infinitely-lived Monetary Authority that issues non-counterfeitable, costlessly exchangeable and intrinsically useless fiat money.[11] The growth rate of fiat money is σ_t. For the base model, σ_t is set to one.

The provision of old-age consumption is further complicated by the fact that the debtors' arrival to the central island and the creditors' departure from it might differ. These arrival and departure rates are stochastic, exogenous and common knowledge.[12] However, agents learn about their own arrival and leaving probabilities only after they have left the islands where they were born at. The table below summarizes the probability of a creditor's departure and a debtor's arrival at the central island:

Fraction of Agents	Early-leaving Creditors $(1-\alpha)$	Late-leaving Creditors (α)
Early-arriving Debtors (λ)	$(1-\alpha)\lambda$	$\alpha\lambda$
Late-arriving Debtors $(1-\lambda)$	$(1-\alpha)(1-\lambda)$	$\alpha(1-\lambda)$

The table shows that there is a possibility of asynchronized timing of meeting on the central island between a creditor and a debtor. Specifically, an early-leaving creditor whose debtor is late-arriving would face the problem of not being able to receive repayment on the loan he extended. On the other hand, a late-leaving creditor might receive his loan repayment from his early-arriving debtor before his departure to the outlying island. In such a case, there could be beneficial trade of loans. That is, an early-leaving creditor whose debtor is late-arriving could resell his debt to a late-leaving creditor. Let Q_{t+1} be the par-value of nominal debt purchased by late-leaving creditors from early-leaving creditors at the discounted nominal value of ρ_{t+1} for each \$1 of debt at

[11]Fiat money is introduced into the model so that agents may trade goods for cash.
[12]Consequently, the utility for creditors (who consume only when old) is in expected utility term.

time-($1, t+1$). $\frac{1}{\rho_{t+1}}$ could be interpreted as the gross *daylight* or the *short-term* nominal interest rate.

Finally, there is neither informational nor enforcement problem at the central island.[13] Also, there is no falsification of *IOU*s by agents. The trading pattern is summarized in the table below and *Figure 1-3*.

	1,t	*1,t+1*
1	young debtors visit young creditors' islands	all old creditors arrive at the central island
2		a fraction λ of old debtors arrive at the central island
3		a fraction (1-α) of old creditors leave the central island
4		a fraction (1-λ) of old debtors arrive at the central island
5	young debtors return home	a fraction α of old creditors leave the central island
	2,t	*2,t+1*
6	young debtors arrive at their respective original islands	all debtors arrive at debtor's island, and all creditors arrive at either a creditor's or a debtor's island

1.3 Optimality

Let $y_{1,t}^D$ and $y_{2,t}^D$ be the consumptions in subperiods *1* and *2* respectively of a young debtor born in time-t. Correspondingly, let $y_{2,t}^C$ and $y_{2,t}^{C*}$ be the old age consumptions of an early and a late departing creditor born at time-t-*1*. Consider the stationary allocation in which $y_{1,t}^D = y_1^D$, $y_{2,t}^D = y_2^D$, $y_{2,t}^C = y_2^C$, and $y_{2,t}^{C*} = y_2^{C*}$ for all t. The Social Planner's

[13]In other words, Monetary Authority could perfectly and costlessly enforce all consumption debt contracts. Mills (2004) argues that such a strong assumption implies that fiat money is not essential as a mean to clear *IOU*s. By clearly defining such a strong enforcement technology in his counter example, he shows that debts are settled by circulating inside money in the form of claims on young debtors' endowments. Fiat money is used by agents only as a medium of exchange. Hence, the practical justification for the focus on Monetary Authority funds in this model is to appeal to its real world institutional ubiquity as a mean of settlement for payment debts in all the major payments system worldwide (see *Table 1-2*).

problem is given by

$$\max_{y_1^D\ ,\ y_2^D\ ,\ y_2^C\ ,\ y_2^{C*}\ ,\ \zeta} \quad (1-\varpi) \quad \left[\ln(y_1^D) + \ln(y_2^D)\right] \quad + \varpi \left[\ (1-\alpha)\ln(y_2^C) + \alpha \ln(y_2^{C*})\right]$$

$s.t$

$$y_1^D + (1-\zeta)[(1-\alpha)y_2^C + \alpha y_2^{C*}] \leq e\ , \tag{1.1}$$

$$y_2^D + \zeta[(1-\alpha)y_2^C + \alpha y_2^{C*}] \leq e\ , \tag{1.2}$$

and

$$0 \leq \zeta \leq 1\ ,$$

where ϖ is the Pareto weight associated with creditors, while (1.1) and (1.2) are the feasibility constraint on the creditor island and the debtor island respectively.[14] (1.1) states that the amount of goods consumed by all visiting young debtors at subperiod-*1* and a fraction $(1-\zeta)$ of visiting old creditors at subperiod-*2* is constrained by young creditors' consumption good endowment. Similarly, (1.2) asserts that the amount of good consumed by all young debtors and a fraction ζ of visiting old creditors at subperiod-2 is constrained by young debtors' endowment of consumption good.

From the first order conditions, the optimality conditions of the economy are given by

$$\widehat{\widehat{y}}_1^D = 2(\frac{1-\varpi}{2-\varpi})e = \widehat{\widehat{y}}_2^D\ , \tag{1.3}$$

$$\widehat{\widehat{y}}_2^C = (\frac{2\varpi}{2-\varpi})e = \widehat{\widehat{y}}_2^{C*}\ , \tag{1.4}$$

and $\widehat{\widehat{\zeta}} = \frac{1}{2}$. In other words, the marginal utility of the young debtor in the first subperiod is equal to that in the second subperiod. Similarly, the marginal utility in the old-period for both the early and late leaving creditors are also equal.

[14]Recall that ζ is the fraction of old creditors assigned to young debtor islands.

1.4 Competitive Equilibrium.

Due to the trading pattern, the price of consumption good in time-t might possibly be different, depending on the subperiod and location of trade: $P^C_{1,t}$, $P^C_{2,t}$ and $P^D_{2,t}$. In order to simplify the construction of a competitive equilibrium, the following result would be useful. Since the consumption good is costlessly stored within a period, the price in creditor's islands must be constant across subperiods: $P^C_{1,t} = P^C_{2,t}$. On account of old creditors have the choice of visiting either the creditor or debtor islands in the second subperiod, the arbitrage condition $P^C_{2,t} = P^D_{2,t}$ holds. Hence, $P^C_{1,t} = P^C_{2,t} = P^D_{2,t} = P_t$.

1.4.1 Debtor's Optimization Problem

Let D_t be the nominal value of debt issued by a debtor at time-$(1, t)$, and let Z^D_t be her nominal demand for fiat money at time-$(2, t)$. A debtor's optimization problem is given by

$$\max_{y^D_{1,t}\ ,\ y^D_{2,t}\ ,\ D_t\ ,\ Z^D_t} \quad \ln(y^D_{1,t}) + \ln(y^D_{2,t})$$

$$s.t$$

$$P_t\ y^D_{1,t} \leq D_t\ , \tag{1.5}$$

$$P_t\ y^D_{2,t} + Z^D_t \leq P_t e\ , \tag{1.6}$$

$$R_t\ D_t \leq Z^D_t\ , \tag{1.7}$$

and non-negative constraints

$$0 \leq y^D_{1,t} \quad , \quad 0 \leq y^D_{2,t} \quad , \quad 0 \leq \ D_t \quad , \quad 0 \leq Z^D_t .$$

(1.5) and (1.6) are the debtor's budget constraints in the two subperiods when she is young. In the first subperiod, her consumption is limited by the amount of IOUs accepted by young creditors. In the second subperiod, she consumes the remainder of her endowment after sales to visiting old creditors. (1.7) is the repayment constraint which states that only fiat money is accepted for loan repayment.[15]

The solution to this problem is to set:[16]

$$y_{1,t}^{D}=\frac{e}{2R_t} \ \text{ and } \ y_{2,t}^{D}=\frac{e}{2}\ , \tag{1.8}$$

$$D_t=\frac{P_t e}{2R_t}\ , \tag{1.9}$$

and

$$Z_t^{D}=\frac{P_t\ e}{2}. \tag{1.10}$$

1.4.2 Creditor's Optimization Problem

Let $y_{2,t+1}^{C}$ and $y_{2,t+1}^{C*}$ be the time period-(2, $t+1$) consumption of an early-leaving creditor and a late-leaving creditor accordingly. In addition, let L_t be the nominal value of the loan extended by the creditor to debtors, and Z_t^C be his demand for fiat money (reserves) in time-t. Q_{t+1} is the par-value amount of nominal debt a late-departing creditor purchase from an early-departing creditors at time period-(1, $t+1$). A creditor's optimization problem is given by

$$\max_{y_{2,t+1}^{C}\ ,\ y_{2,t+1}^{C*}\ ,\ L_t\ ,\ Z_t^{C}\ ,\ Q_{t+1}\ ,\ \phi} \quad (1-\alpha)\ln(y_{2,t+1}^{C})+\alpha\ln(y_{2,t+1}^{C*}) \tag{1.11}$$

$s.t$

$$L_t \le \phi P_t e\ , \tag{1.12}$$

$$Z_t^{C} \le (1-\phi)P_t e\ , \tag{1.13}$$

[15]That is, the settlement of debt is subject to the repayment constraint which is imposed by spatial separation of agents. Note that (1.7) binds the repayment of consumer loans, rather than directly binds the purchase of $y_{1,t}^{D}$.

[16]Note that $y_{2,t}^{D}$ and Z_t^{D} are not functions of R_t.

$$P_{t+1}\ y^{C}_{2,t+1} \leq \lambda R_t L_t + \rho_{t+1}(1-\lambda)R_t L_t + Z^C_t\ , \tag{1.14}$$

$$P_{t+1}\ y^{C*}_{2,t+1} \leq R_t L_t + (1-\rho_{t+1})Q_{t+1} + Z^C_t\ , \tag{1.15}$$

$$\rho_{t+1}\ Q_{t+1} \leq \lambda R_t L_t + Z^C_t\ , \tag{1.16}$$

and

$$0 \leq y^{C}_{2,t+1} \quad , \quad 0 \leq y^{C*}_{2,t+1} \quad , \quad 0 \leq L_t \quad , \quad 0 \leq Z^C_t \quad , \quad 0 \leq Q_{t+1} \quad , \quad 0 \leq\ \phi \leq 1\ ,$$

where ϕ signifies the fraction of his endowments loaned to young debtors in time-t. Here (1.14) and (1.15) are correspondingly the budget constraints of an early-departing creditor and a late-departing creditor. (1.14) states that an early-departing creditor's old-period consumption expenditure is bounded by the total amount of his IOUs made good by early arriving debtors($\lambda R_t L_t$), the amount of his late arriving debt resold in the secondary loan market ($(1-\lambda)R_t L_t$) at a discount price of $\rho_{t+1} \leq 1$, and his fiat money holding (Z^C_t). On the other hand, (1.15) says that a late-leaving creditor's old-period consumption outlay is constrained by the amount of debt repaid by all his debtors ($\lambda R_t L_t + (1-\lambda)R_t L_t = R_t L_t$), his potential profit from purchasing early-leaving creditors' unredeemed debt at a discount ($Q_{t+1}(1-\rho_{t+1})$), and his fiat money holding (Z^C_t). Finally, the liquidity constraint (1.16) asserts that the nominal value of secondary debt purchased by late-leaving creditors is limited by their cash holding before the arrival of late debtors.

The solutions to the creditor's problem depends on whether the liquidity constraint (1.16) binds or not. If $\rho_{t+1} = 1$ so that no profits could be made in the secondary debt market, then (1.16) would not bind. Creditors would choose a portfolio of fiat money and debt by simply comparing the two assets' rates of return.[17] An interior solution would

[17]The composition of their portfolios will depend on the value of R_t.

be optimal only when $R_t = 1$. In this case, the solution is to allocate

$$\begin{cases} y^C_{2,t+1} = \frac{P_t e}{P_{t+1}} = y^{C*}_{2,t+1} \\ \quad L_t = \phi P_t e \ , \qquad Z^C_t = (1- \ \phi) P_t e \ , \qquad \phi \in [0,1] \\ \quad Q_{t+1} \in [0, P_t e] \end{cases} .$$

Note that when $R_t = 1$, both early and late departing creditor's old-period consumptions are equal. In addition, loans and fiat money are perfect substitutes, and a creditor has an indeterminate portfolio. Finally, the amount of secondary loan traded is also undetermined. On the other hand, if $R_t > 1$, a creditor chooses $(L_t, Z^C_t) = (P_t e, 0)$, and if $R_t < 1$, he decides upon $(L_t, Z^C_t) = (0, P_t e)$. It could be shown that neither of these corner solutions could be an equilibrium outcome.

If $\rho_{t+1} < 1$, then there are profits to be made from procuring debt in the secondary market and the liquidity constraint (1.16) would bind. The solution in this case also depends on the value of R_t. The condition such that the creditor chooses a portfolio with positive amount of both loans and reserves is given by

$$\left[\lambda R_t + \rho_{t+1}(1-\lambda)R_t - 1\right] \left[\frac{(1-\alpha)}{P_{t+1} y^C_{2,t+1}} + \frac{\alpha}{\rho_{t+1} P_{t+1} y^{C*}_{2,t+1}}\right] = 0 \ , \tag{1.17}$$

which follows from the first order conditions of the creditor's optimization problem. Specifically, (1.17) holds *iff*

$$R_t = \frac{1}{\lambda + \rho_{t+1}(1-\lambda)}. \tag{1.18}$$

The arbitrage pricing condition (1.18) states that the rate of return on loans is equal to the expected rate of return on fiat money, which includes the potential profits of using cash to make purchases in the secondary debt market. When $R_t = \frac{1}{\lambda+\rho_{t+1}(1-\lambda)}$, the solution to the creditor's problem is to set

$$\begin{cases} y^C_{2,t+1} = \frac{P_t e}{P_{t+1}} & < & \frac{P_t e}{\rho_{t+1} P_{t+1}} = y^{C*}_{2,t+1} \\ L_t = \phi P_t e \ , & Z^C_t = (1 - \ \phi) P_t e \ , & \phi \in [0, 1] \\ Q_{t+1} = \left[\frac{\lambda R_t P_t e}{\rho_{t+1}}, \frac{P_t e}{\rho_{t+1}} \right] & & \end{cases} . \tag{1.19}$$

A creditor is still indifferent between holding *IOU*s and holding fiat money. However, if he emerges to be a late-leaver, his entire cash holding would be used to buy unredeemed debts from the early departing creditors. Note that the consumption of the late-leaving creditor is now greater than that of the early-leaving creditor. Finally, it is again true that if (1.18) does not hold, then either $Z^C_t = 0$ or $L_t = 0$. Nevertheless, neither of these scenarios could be an equilibrium.

The relationship in (1.18) also gives the general expression for the overnight interest rate as a function of early arrival rate of the debtors (λ) and the next period's nominal discount rate for \$1 of resale loans (ρ_{t+1}).

Remark 1 *Some properties of R_t are as of the following:*

1. *If $\rho_{t+1} = 1$, then $R_t = 1$. Hence, $R_t = 1 = \frac{1}{\rho_{t+1}}$.*

2. *$\frac{\partial R_t}{\partial \rho_{t+1}} < 0$. That is, an increase in tomorrow's discount rate causes a decrease in today's overnight interest rate. Alternatively, it means that today's overnight interest rate moves in the same direction as tomorrow's daylight interest rate which is given by $\frac{1}{\rho_{t+1}}$.*

3. *Properties 1 and 2 above imply that if $\rho_{t+1} < 1$, then $R_t > 1$.*

4. *For $\forall \lambda \in (0, 1)$, $R_t \neq \frac{1}{\rho_{t+1}}$. That is, unless all debtors are late arrivees or early arrivees, there is a wedge between the overnight and the daylight interest rates. This wedge is due to the uncertainty with respect to the arrival of debtors to the central island.*

5. *If* $\rho_{t+1} < 1$, *then* $R_t < \frac{1}{\rho_{t+1}}$.

[*Proof* : $R_t - \frac{1}{\rho_{t+1}} = \frac{1}{\lambda+\rho_{t+1}(1-\lambda)} - \frac{1}{\rho_{t+1}} = \frac{\lambda(\rho_{t+1}-1)}{\rho_{t+1}[\lambda+\rho_{t+1}(1-\lambda)]} < 0$ for $\rho_{t+1} < 1$.]

The relationship between the overnight and daylight interest rates is illustrated in *Figure 1-4*.

1.4.3 Market Clearing Conditions

The market clearing condition for the consumer good on creditor islands at subperiods $(1, t)$ and $(2, t)$ are

$$y^D_{1,t} = \phi e$$

and

$$(1-\zeta)[(1-\alpha)y^C_{2,t+1} + \alpha y^{C*}_{2,t+1}] = (1-\phi)e$$

respectively, which could be combined as

$$y^D_{1,t} + (1-\zeta)[(1-\alpha)y^C_{2,t+1} + \alpha y^{C*}_{2,t+1}] = e. \tag{1.20}$$

On the other hand, the market clearing condition for the consumer good on debtor islands at subperiod-$(2, t)$ is

$$y^D_{2,t} + \zeta[(1-\alpha)y^C_{2,t+1} + \alpha y^{C*}_{2,t+1}] = e. \tag{1.21}$$

The market clearing condition for loans at time-t is

$$D_t = L_t\ , \tag{1.22}$$

while that for resale loans at the central island is

$$\alpha Q_{t+1} = (1-\alpha)(1-\lambda) R_t L_t. \tag{1.23}$$

(1.23) states that the demand for unredeemed debt by late-departing creditors equals the supply for unredeemed debt held by early-departing creditors whose debtors are late-arriving. The market clearing condition for fiat money in creditor and debtor islands are given correspondingly by

$$Z_t^C = (1-\zeta)M_t \ , \tag{1.24}$$

and

$$Z_t^D = \zeta M_t \ , \tag{1.25}$$

which could be combined as

$$Z_t^D + Z_t^C = M_t. \tag{1.26}$$

1.4.4 Equilibrium

There are two types of potential equilibria in the economy, depending on whether the liquidity constraint is binding or not. The analysis shows that any given economy has a unique equilibrium. Which type that equilibrium is, depends on the parameter values of α and λ.

Case 1: Daylight Liquidity is Unconstrained in Equilibrium.

If there is an equilibrium in which $\rho_{t+1} = 1$ (for all t), the values of the other endogenous variables in this equilibrium must be as follows:

$$R_t = 1 \ ,$$

$$P_t = \frac{M_0}{e} \ ,$$

$$y_{1,t}^D = \frac{e}{2} = y_{2,t}^D \ ,$$

$$y_{1,t}^C = e = y_{2,t}^{C*} \ ,$$

$$\frac{D_t}{P_t} = \frac{Z_t^D}{P_t} = \frac{e}{2} = \frac{Z_t^C}{P_t} = \frac{L_t}{P_t} \ ,$$

$$0 \le \frac{Q_{t+1}}{P_t} \le \frac{(1-\alpha)(1-\lambda)}{\alpha} \frac{e}{2} \ ,$$

and

$$\phi = \zeta = \frac{1}{2}.$$

Evidently, such an equilibrium exists *iff* in this equilibrium the liquidity constraint (1.16) is met. This is the case *iff*

$$1 - \alpha \ \le \alpha + \lambda. \tag{1.27}$$

(1.27) simply states that the fraction of agents at the central island who demand immediate liquidity (i.e. the early-departing creditors) is less than the fraction that supply it (i.e. late-leaving creditors and early-arriving debtors).

Note that in the *case*-1 equilibrium, the rates of return on both loans and fiat money are 1.[18] These two assets are perfect substitute in spite of the fact that their maturities are different. This equivalence is possible because there is no liquidity risk. Each young creditor holds equal amount of loans and real balances respectively in their asset portfolio.[19] The amount of secondary debt traded at the central island is indeterminate. A young debtor's consumption in her subperiod is equal to that in her second subperiod. Early and late departing creditors have equal consumption. These consumption good allocations of debtors and creditors satisfy the optimality conditions given by (1.3) and (1.4) for some ϖ.

[18] That is,

$$R_t = 1 = \frac{P_t}{P_{t+1}} \ .$$

[19] Relatedly, half of the old creditors visit young debtors' islands and the remaining half visit young creditors' islands to procure consumption good.

Case 2: Daylight Liquidity is Constrained in Equilibrium.

On the other hand, if (1.16) binds, the equilibrium must have

$$\rho_{t+1} = \frac{2\alpha}{1-\lambda} < 1 \ , \tag{1.28}$$

$$R_t = \frac{1}{\lambda + 2\alpha} \ > 1 \ , \tag{1.29}$$

$$P_t = \left(\frac{2R_t}{3R_t - 1}\right)\frac{M_0}{e} = \left(\frac{2}{3-(\lambda+2\alpha)}\right)\frac{M_0}{e} \ ,$$

$$y^D_{1,t} = \left(\frac{1}{R_t}\right)\frac{e}{2} = (\lambda+2\alpha)\frac{e}{2} \ < \ \frac{e}{2} = y^D_{2,t} \ ,$$

$$y^C_{2,t+1} = e \ < \ \left(\frac{1-\lambda}{\alpha}\right)\frac{e}{2} = \left(\frac{1}{\rho_{t+1}}\right)\frac{e}{2} = y^{C*}_{2,t+1} \ ,$$

$$\frac{D_t}{P_t} = (\lambda+2\alpha)\frac{e}{2} \ < \ \frac{e}{2} = \frac{Z^D_t}{P_t} \ ,$$

$$\frac{L_t}{P_t} = (\lambda+2\alpha)\frac{e}{2} \ < \ [2(1-\alpha)-\lambda]\frac{e}{2} \ = \frac{Z^C_t}{P_t} \ ,$$

$$\frac{Q_{t+1}}{P_t} = \frac{(1-\alpha)(1-\lambda)}{\alpha}\frac{e}{2} \ ,$$

and

$$0 < \phi = \frac{\lambda+2\alpha}{2} < \frac{1}{2} \quad , \quad 0 < \zeta = \frac{1}{3-(\lambda+2\alpha)} < \frac{1}{2}.$$

Obviously, the implied value of ρ_{t+1} is less than 1 *iff* [20]

$$1-\alpha \ > \alpha+\lambda. \tag{1.30}$$

When the liquidity problem is severe, the gross intraday nominal interest rate $\frac{1}{\rho_{t+1}} > 1$. This is the case when the fraction of agents who have meeting problem (i.e. the early-departing creditors) is greater than the fraction that do not (i.e. late-leaving creditors and early-arriving debtors), resulting in an excess demand for fiat money at the central island.[21] The lower α or (and) λ is (are), the more likely it is for daylight liquidity shortage to occur. Moreover, the rate of return on loan is higher than that on fiat money: these two assets are no longer perfect substitutes as they were when there is no liquidity risk. The shortage of liquidity causes the daylight nominal interest rate to be higher than the overnight nominal interest rate (i.e. $\frac{1}{\rho_{t+1}} > R_t$).[22]

This equilibrium, evidently, results in consumption allocations that are inefficient. The lack of liquidity, *via* the change in R_t, affects agents' design in allocating real resources. Each young creditors now holds proportionately more cash than debts: anticipating a liquidity shortage in the next period, a young creditor prefers to sell more of his endowment to visiting old creditors for fiat money than to visiting young debtors for *IOU*s. This results in unequal intra-subperiod consumption for debtors and unequal old-age consumption for creditors. Specifically, the debtor's first subperiod consumption is discounted by the positive long term interest rate, while the late-departing creditor's

[20]Specifically, for ρ_{t+1}, from the first order conditions of the creditor's problem, it can be shown that

$$0 < \rho_{t+1} = \frac{\frac{\alpha}{P_{t+1} y^{C*}_{2,t+1}}}{\frac{\alpha}{P_{t+1} y^{C*}_{2,t+1}} + \delta_{t+1}} < 1 \ ,$$

where δ_{t+1} the Langrange multiplier for the liquidity constraint (1.16) in time-*t+1*.

[21]Note that $(1-\alpha)Z_t^C$, i.e. fiat money held by early-leaving creditors, is not part of the total fiat money supplied at time period-(*1*,*t+1*).

[22]Martin (2004) also shows that $R_t > 1$. However, in his model, $R_t = \frac{1}{\text{discount factor}}$, which is a constant. In contrast, here $R_t = \frac{1}{\lambda+2\alpha}$, which depends on the arrival and departure parameters. As will be shown below, R_t varies depending on the severity of the liquidity shortage and the resulting adjustment in agents' portfolio choice.

consumption is marked-up by the positive short term interest rate. Therefore, in a liquidity constrained economy, the debtor's second subperiod marginal utility is now lower than that of her first subperiod. Ex-ante homogenous creditors now have ex-post heterogeneous consumption in which an early-leaving creditor's consumption is lower than that of a late-leaving creditor. That is, the late-leaving creditors have a higher gross real rates of return than the early-departing creditors.

The non-optimality of this equilibrium is also demonstrated by the reselling of debt at a discount in the secondary debt market. A late-departing creditor now spends all his currency holdings to purchase these discounted debts to maximize his profit. Liquidity shortage occurs even when it is fully anticipated by agents, and might be observed even in the long-run. The temporary shortage of fiat money renders unfeasible any insurance arrangement that would have late-leaving creditors compensate early-leaving creditors ex-post. It also rules out debt contract with state-contingent R_t.[23] Finally, note that ρ_{t+1}, R_t,and P_t are all functions of exogenous variables.[24] The results thus far are summarized in the following proposition:

Proposition 2 *The competitive equilibrium is uniquely determined by the parameter values α and λ.*

(i) If $\alpha + \lambda \geq 1 - \alpha$, then there exists a stationary equilibrium in which the liquidity constraint is non-binding. $\rho_{t+1} = 1 = R_t$, and the resulting consumption allocation is Pareto optimal.

(ii) If $\alpha + \lambda < 1 - \alpha$, then there exists a stationary equilibrium in which the liquidity constraint binds. $\rho_{t+1} < 1 < R_t$, and the resulting consumption allocation is not Pareto optimal.

[23] For example, a debtor who turns out to be early arriving would end up with extra fiat money in her old age. Since she acquires fiat money at the cost of less consumption when young but does not consume when old, this is inefficient.

[24] $M_t = M_0$ is due to the assumption that there is no monetary growth. Observe that when liquidity is scarce, an increase in M_t causes a proportional but less than one-to-one increase in P_t. The decrease in P_t is a characteristic of most financial crisis that is accompanied by liquidity shortages.

Henceforth, I focus on the case when $\alpha + \lambda < 1 - \alpha$, so that the liquidity constraint is binding.

1.5 Payments System and Monetary Policy

The previous section demonstrates that the competitive equilibrium of the economy might not potentially be Pareto optimal (i.e. when $\alpha + \lambda < 1 - \alpha$). It is therefore natural to ask whether there are particularly beneficial government policies. This section considers two policies to overcome the liquidity shortage problem, and consequently, to achieve optimality. Specifically, the Monetary Authority on the central island could temporarily inject additional liquidity by increasing the supply of reserves in time period-$(1, t+1)$ *via* (1) open market purchases or/and (2) discount window policies.

Under both policies, the Monetary Authority is authorized to issue and lend reserves equal to the nominal amount of debt presented by any of the late-leaving creditors. Next, late-departing creditors use the reserves to purchase IOUs held by early-departing creditors whose debtors are late-arriving. After these late-arriving debtors turned up at the central island and redeemed their debts with fiat money, late-leaving creditors repay those Monetary Authority loans.[25,26]

[25] In this model both open market operations and discount window facilities' objective is to provide short run liquidity to support any shortfall, but they are operationally different. Under open market operation, given the existence of a perfectly functioning interbank market (i.e. the secondary loan market), the Monetary Authority could affect short term (i.e. the daylight) interest rate by adjusting the amount of its reserves supply via the buying or selling of late creditors' unredeemed loans in the secondary loan market. Since the creditors will eventually repurchase his loans from the Monetary Authority once the late debtors redeemed their debts, this is akin to the Monetary Authority and the late creditors entering into a repo contract. On the other hand, under discount window lending, the Monetary Authority extends (unlimited or rationed) collateralized loans to late creditors at a pre-determined discount window interest rate. Illustratively, on a (Z, r)-*plane*, the reserves supply schedule is perfectly inelastic under open market operations. In contrast, under an unlimited provision of collateralized loans from a discount window facility, the reserve supply schedule is perfectly elastic on the (Z, r)-plane.

[26] In practice, the objective of open market operation is supplying the whole banking system's forecasted daily reserves needs, while discount window functions as an emergency provider of liquidity to specifically affected banks. The major part of open market operation occurs before 9 a.m., while that of discount window at the end of the transaction day, typically after 6 p.m.

The effect of these policies on P_{t+1} (and hence, on real interest rates and consumptions) depends on whether the Monetary Authority retires the injected fiat money at the end of period-(*1,t+1*) or not, and how the creditors readjust their asset portfolios. In the case the injected outside money is withdrawn from circulation, the price level remains across time periods need not remains constant. The temporary fluctuation in the reserves supply could, in fact, achieve a Pareto optimal allocation. However, this allocation is not Pareto superior to the original competitive equilibrium. Specifically, these policies, while overcoming the liquidity shortage problem, have unintended real effects.

1.5.1 Open Market Operation

Let X_{t+1} be the discounted nominal value of unredeemed second-hand debt purchased by the Monetary Authority from late-leaving creditors on the central island at time-(*1,t+1)*. The market clearing condition for second-hand debt now becomes

$$\frac{X_{t+1}}{\rho_{t+1}} + \alpha Q_{t+1} = (1-\alpha)(1-\lambda) R_t L_t. \tag{1.31}$$

From (1.31), solve for Q_{t+1}, and substituting it into (1.16) yields

$$X_{t+1} \geq \left[\rho_{t+1}(1-\alpha)(1-\lambda) - \alpha\lambda\right] R_t L_t - \alpha Z_t^C \ , \tag{1.32}$$

where the first term in the bracket on the *RHS* represent the discounted value of unredeemed debts (i.e. the demand for liquidity). It consists of the discounted nominal value of loans held by early-leaving creditors whose debtors are late-arriving minus the nominal value of loans held by the late-leaving creditors whose debtors are early-arriving. The second term on the *RHS* is the supply of liquidity which is solely due to the reserves holding of late-leaving creditors. The intra-period and the inter-period interest rates are now

$$\left(\frac{1}{\rho_{t+1}}\right)' = \frac{(1-\lambda)(1+\frac{X_{t+1}}{M_t})}{2\alpha + (3-\lambda)\frac{X_{t+1}}{M_t}}$$

and

$$R_t^{'} = \frac{(1 + \frac{X_{t+1}}{M_t})}{\lambda + 2\alpha + \frac{3X_{t+1}}{M_t}}$$

respectively.[27]The amount of X_{t+1} such that the liquidity constraint (1.16) is non-binding is

$$X_{t+1} \geqslant (1 - \alpha - \lambda)L_t - \alpha Z_t^C = [1 - \lambda - 2\alpha]\frac{P_t e}{2}. \tag{1.33}$$

Suppose that at the end of period-(*1,t+1*), the Monetary Authority retires $\frac{X_{t+1}}{\rho_{t+1}}$ worth of fiat money from the economy. Therefore, the amount of Monetary Authority's profit/loss from the open market purchase is

$$\Pi_{t+1}^{OMO} = \frac{X_{t+1}}{\rho_{t+1}} - X_{t+1} \geq 0 \ , \tag{1.34}$$

since $\rho_{t+1} \leq 1$. The effect of open market operation on money supply and price level depends on the disposition of any positive profit. The total stock of fiat money in the economy at the end of time-$t + 1$ is given by

$$M_{t+1} = M_t + X_{t+1} - \frac{X_{t+1}}{\rho_{t+1}} = M_t - \Pi_{t+1}^{OMO}.$$

Unequivocally, $M_{t+1} = M_t \Leftrightarrow \frac{1}{\rho_{t+1}} = 1 \Leftrightarrow \rho_{t+1} = 1$. That is, the Monetary Authority, by temporarily increasing money supply such that the gross nominal daylight interest rate decreases to 1, could overcome liquidity shortages.[28] Therefore, $\frac{1}{\widehat{\rho}_{t+1}} = 1 = \widehat{R}_t$. Both early and late leaving creditors would consume the same amount of consumption good, i.e. $\widehat{y}_{2,t+1}^C = e = \widehat{y}_{2,t+1}^{C*}$, and debtors smooth their inter-subperiod consumption, i.e. $\widehat{y}_{1,t}^D = \frac{e}{2} = \widehat{y}_{2,t}^D$. This allocation is Pareto optimal. This is due to the fact that the private marginal cost, the social marginal cost and the social marginal benefit are all equal to

[27]Note that the wedge between daylight and overnight interest rates, even though liquidity is still constrained, has now become smaller.

[28]If the Monetary Authority withdraws the extra fiat money issued in time period-($1, t + 1$) upon repayment by old-leaving creditors, then the stock of money supply and price of consumption goods are constant. However, if $\rho_{t+1}^{'} < 1$, then $M_{t+1} < M_t$. Hence, $P_{t+1} < P_t$, resulting in all old creditors being better off.

zero. Private marginal cost is zero because there is no individual credit risk and collateral posting is costless to late-departing creditors.[29] Since the marginal cost of increasing the real quantity of money is virtually zero, welfare is maximized when real money balance is provided up to the point of satiety, where marginal benefit is also zero.[30] Note that while holding fiat money as an asset is not costless (the opportunity cost is $\widehat{R}_t - \frac{\widehat{P}_t}{\widehat{P}_{t+1}}$) for agents, what is costless is the extra liquidity supplied by the Monetary Authority.[31, 32]

However, starting from a liquidity constrained equilibrium and moving to an unconstrained equilibrium is not Pareto improving: debtors gain at late-departing creditors' expenses. A temporary increase in the fiat money causes the short run asset price (i.e. the intraday nominal interest rate) to decrease, thus making late-departing creditors worse off since there is less profit to be made in the secondary loan market. However, early-departing creditors are not better off. This is because the actual benefit of a higher discount rate is transmitted to debtors *via* lower long run nominal and real interest rates.[33, 34]

Long run asset prices change due to creditors portfolio adjustments, which in turn is induced by the movement in short run asset price. To see this, note that the agents'

---

[29]Collateral is costless even though $R_t > 1$, since the unredeemed debt cannot be liquidated for other purposes, and hence the opportunity cost is zero. This is in contrast to Khan and Roberds (2001) in which collateral posting is costly. In their model, agent has an inherent preference for early consumption, which is financed solely by fiat money, relative to later consumption. Thus, the collateral requirement in order to acquire daylight credit in the succeeding period forces the agent to hold more illiquid assets relative to fiat money in the preceeding early consumption period. As such, free intraday credit with full collateralization is inefficient.

[30]This result is similiar to Friedman's optimum quantity of money.

[31]Instead, if it is assumed that the Monetary Authority buys all late-arriving debts owned by the early-leaving creditors, $Q_{t+1} = 0$. Substituting this Q_{t+1} in (1.16) shows that the liquidity constraint never binds : $0 < \lambda R_t L_t + Z_t^C$, where the inequality folllows from the logarithmic utility functions. Hence, $R_t = \rho_{t+1} = 1$,and $X_{t+1} = (1-\alpha)(1-\lambda)L_t$.

[32]In practice, the Fed charges financial depositories a minute-based interest rate of 36 basis points (at annual rate) on daylight overdrafts. However, in a sense, it is still true that $\frac{1}{\rho_{t+1}} = 1$. That is, the private and social costs of liquidity could still be equated at zero through a free intraday credit policy. The positive daylight credit interest rate charge could then be considered as the risk premium to compensate the Fed for the risk incurred in its provision of intraday liquidity.

[33]See equation (1.18).

[34]This distinction between real and nominal interest rates becomes clearer in the next section where the money stock of fiat money grows at a positive rate.

marginal utility is a convex function of consumption good. By Jensen's inequality, the relaxation of the liquidity constraint due to short term reserves injection causes the creditors' old-period expected marginal utility of consumption to decrease. In order for the price arbitrage condition (1.18) to hold, the debtors' young-period marginal utility must also decline.[35] Consumption smoothing then requires a young debtor to increase her first-subperiod consumption relative to the second-subperiod consumption. Creditors now substitute *IOU*s for real reserve balances relative to the case when the liquidity constraint binds. Hence, long run asset prices and creditor's portfolio readjust to new equilibrium values. Finally, note that while the end-of period stock of fiat money is constant, the price level has increased, due to the decrease in the demand for real reserves balances by all creditors. The following proposition holds:

Proposition 3 *Let $\alpha+\lambda < 1-\alpha$. Suppose the Monetary Authority conducts open market purchase and selling such that $\frac{1}{\widehat{\rho}_{t+1}} = 1$. Then the resulting consumption allocation is*

(i) Pareto optimal, but

(ii) not Pareto improving relative to the competitive equilibrium allocation.

Proof. *(ii) Since $\alpha + \lambda < 1 - \alpha$, $\ln\left(\frac{e}{2}\right)^2 > \ln((\lambda + 2\alpha)(\frac{e}{2})^2)$ and $\ln e < \ln\left((\frac{1-\lambda}{2\alpha})^{\alpha} e\right)$.*

■

As an example, for the parameter values $(\alpha,\ \lambda,\ e) = (0.3, 0.2, 10)$, the allocation when $\frac{1}{\widehat{\rho}_{t+1}} = 1$ is denoted by the point *OMO* in *Figure 1-5.* In contrast, the competitive equilibrium allocation when the liquidity constraint is still binding is given by the point *CE* which is below the Pareto frontier. Clearly, by moving from point *CE* to point *OMO* there is redistribution of real resources among agents which makes debtors better off but creditors worse off.

[35] Since the creditors and debtors' consumption is period specific and does not overlapped, this consumption pattern could be viewed as a consumption pattern for a single representative agent that consumes in both time periods.

1.5.2 Discount Window

Now assume that the Monetary Authority stands ready to lend at some pre-announced discount rate only to late-leaving creditors that post collateral to get the loan in order to purchase unredeemed debts in the resale loan market.[36] Let X_{t+1} be reinterpreted as the discounted nominal amount of loan lent by the Monetary Authority to late-leaving creditors at time period-(*1,t+1).* The market clearing condition for second-hand debt remains as in (1.31). Let r^{DW} be the gross nominal interest rate charged on discount window loans within the time period-($1,t+1$). Late-leaving creditors would borrow from the discount window as long as r^{DW} is less or equal to the intraday nominal interest rate,

$$r^{DW} \leq \frac{1}{\rho_{t+1}}. \tag{1.35}$$

Using (1.16), (1.31), and rearranging, the nominal amount of discount window loan extended, as in open market operation, is given by (1.32) in general. The extent of how much the liquidity constraint binds depends on r^{DW}, which is the choice variable for the Monetary Authority. Assume the special case of $r^{DW} = 1$. Hence, $\frac{1}{\rho_{t+1}} = 1$. That is, (1.16) is no longer binding and the optimal equilibrium is achieved. Similar to the condition under open market purchase, the amount of loan extended is given by (1.33). In this specific case, it could be shown that

$$r^{DW} X_{t+1} < Q_{t+1}.$$

That is, late-leaving creditors are still able to consume good after repaying their discount window loans.[37]

[36]In practice, the end-of-the-day discount window policy could be seen as encompassing intraday credit policy liquidity in the sense that it is pre-ordained by the later needs.

[37]Intuitively, this is due to the fact that there is no other risk apart from liquidity risk which goes away when $r^{DW} = 1$.

The Monetary Authority's profit/loss from discount window activity is

$$\Pi_{t+1}^{DW} = r^{DW} X_{t+1} - X_{t+1} \ ,$$

which is zero if $r^{DW} = 1$.Hence, although the stock of money supply fluctuates within the period-(*1,t+1*), they remain constant over time. Finally, as in the open market purchase case, using discount policy tool to overcome the liquidity problem has real redistribution effect that causes the debtors to be better off at the expense of the creditors.

In either policy, it is optimal for the Monetary Authority to inject liquidity such that $\frac{1}{\rho_{t+1}} = 1$. Since both the open market operation and the discount window policies caused an equal temporary increase in the amount of reserves and there is no private cost, these two policies are equivalent. In the Goodfriend and King (1988) dichotomy, this temporary increase in reserves amount is a banking policy, as opposed to a monetary policy in which the change in the amount of reserves is unsterilized. Finally, creditors might anticipate that the Monetary Authority would pursue a liberal liquidity provision intervention and hence by-passing the secondary loan market. However, this adverse incentive is avoided here. Early-departing creditors could not enter into any credible contract with the Monetary Authority either directly or indirectly *via* late-departing creditors. This is due to the fact that these creditors would need to leave the central island earlier and never come back. Hence, they would definitely not meet the Monetary Authority again, and would not necessarily meet up with late-departing creditors either.[38]

1.6 Interest Payment on Reserves

This section describes the consequences of paying interest, θ, proportional on reserves held by creditors. This interest payment is financed by increasing the fiat money stock. Why

[38]Indeed, the Monetary Authority does not guarantee free intraday credit explicitly ex-ante. A possible practical reason is the open access nature of the payment system design: the Monetary Authority may be overwhelmed by demand for liquidity by infinitely large creditors ex-post.

might this policy be of interest? Section *4* and *5* demonstrated that when the liquidity constraint binds, an open market operation or a discount window policy could achieve a Pareto optimal allocation. However, this occurs at the expense of redistributing resources from creditors to debtors relative to the laissez-faire competitive equilibrium. On the other hand, paying interest on reserves would benefit creditors without adversely affecting debtors or vice-versa. Indeed, depending on the chosen interest payment value, both creditors and debtors could be made better off than they were in a liquidity constrained competitive equilibrium. Here I interpret the fiat money holdings of debtors as currency (to be used to pay their consumption good debt) and the fiat money holdings of the creditors as reserves (to be used to pay their payment debt, and is part of their optimal portfolio of IOUs and fiat money). Hence, while the interest cost is accrued to all agents, creditors earn interest income on their reserves holding but debtors do not. I now turn to the following question: Is it possible for this interest on reserves policy to compensate for the distributional effects of short run liquidity provision policies?

In order to answer it, the competitive equilibrium under the interest on reserves regime is derived. Then liquidity provision *via* open market operation (or discount window) is conducted. The interest on reserves scheme applies to all creditors only. Any reserves deposit by creditors with the Monetary Authority upon arrival at the central island, earns interest payment of θ per-unit reserves when withdrawn. Each creditor, irrespective of being early or late departing, is allowed to make only a single deposit and withdrawal. Given that the growth rate of fiat money is $\sigma > 1$, the Monetary Authority's budget constraint is

$$(\sigma - 1)M_t = (\theta - 1)Z_t^C. \tag{1.36}$$

By (1.24), fiat money growth rate is equivalent to

$$\sigma = \theta - \zeta(\theta - 1). \tag{1.37}$$

1.6.1 Liquidity Constrained Competitive Equilibrium

The debtor's optimization problem remains unchanged. The creditor's optimization problem is basically similar with the following changes. He chooses $\{y^C_{2,t+1},\ y^{C*\prime\prime}_{2,t+1},\ L_t,\ Z^C_t,$ $Q_{t+1}\}$ to maximizes (1.11) subject to given (1.12), (1.13), the new budget constraints

$$P_{t+1}\ y^C_{2,t+1} = \lambda R_t L_t + \rho_{t+1}(1-\lambda) R_t L_t + \theta Z^C_t\ , \tag{1.38}$$

$$P_{t+1}\ y^{C*}_{2,t+1} = R_t L_t + (1-\rho_{t+1}) Q_{t+1} + \theta Z^C_t\ , \tag{1.39}$$

and the augmented liquidity constraint

$$\rho_{t+1}\ Q_{t+1} = \lambda R_t L_t + \theta Z^C_t. \tag{1.40}$$

Solving the above problem using the same analysis as in Section 4, the gross nominal overnight interest rate is now given by

$$\widetilde{R}_t = \frac{\theta}{\lambda + \rho_{t+1}(1-\lambda)}. \tag{1.41}$$

Utilizing the various market clearing conditions in (1.20) - (1.26), it is straightforward to derive the gross nominal intraday and overnight interest rates as

$$\widetilde{\rho}_{t+1} = \frac{2\alpha\theta}{1-\lambda} \tag{1.42}$$

and

$$\widetilde{R}_t = \frac{\theta}{\lambda + 2\alpha\theta} \tag{1.43}$$

respectively. Evidently, $\widetilde{\rho}_{t+1} < 1$ *iff* $1 - \alpha\theta > \alpha\theta + \lambda$. That is, the intraday gross nominal interest rate $\frac{1}{\widetilde{\rho}_{t+1}} > 1$ *iff* at the central island the fraction of agents demanding prompt liquidity (i.e. early-departing creditors) is larger than fraction of potential suppliers (i.e. late-departing creditors and early-arriving debtors) whose reserves holding is now augmented by the interest payment θ. From (1.26) and (1.41), the price level is

$$\widetilde{P}_t = \left(\frac{2\widetilde{R}_t}{3\widetilde{R}_t - 1}\right)\frac{M_t}{e} = \left(\frac{2\theta}{3\theta - \lambda - 2\alpha\theta}\right)\frac{M_t}{e} \ ,$$

which implies that inflation rate between periods t and $t{+}1$ is now

$$\frac{\widetilde{P}_{t+1}}{\widetilde{P}_t} = \sigma = \theta - \widetilde{\zeta}(\theta - 1). \tag{1.44}$$

Therefore, the equilibrium consumption allocation is given by

$$\widetilde{y}^D_{1,t} = \left(\frac{1}{\widetilde{R}_t}\right)\frac{e}{2} = \frac{\lambda + 2\alpha\theta}{\theta}\frac{e}{2} \quad < \quad \frac{e}{2} = \widetilde{y}^D_{2,t} \ ,$$

$$\widetilde{y}^C_{2,t+1} = \left(\frac{\theta}{\sigma}\right)e = \frac{\theta}{\theta - \widetilde{\zeta}(\theta - 1)}e < \left(\frac{1-\lambda}{2\alpha}\right)\left(\frac{1}{\theta - \widetilde{\zeta}(\theta - 1)}\right) = \left(\frac{1}{\widetilde{\rho}_{t+1}}\right)\left(\frac{\theta}{\sigma}\right)e = \widetilde{y}^{C*}_{2,t+1},$$

with

$$\widetilde{\zeta} = \frac{\theta}{3\theta - 2\alpha\theta - \lambda}.$$

Assume that $1 < \theta < \frac{1-\lambda}{2\alpha}$. That is the liquidity constraint is still binding under a regime that pays positive interest on reserves. Paying positive interest on reserves held by creditors causes them to substitute real reserve balances for loan to debtors when young. This asset substitution continues until the expected rate of return on reserves is equal to the rate of return on IOUs, which is the long run interest rate, $\widetilde{R}_t$. Moving from a non-interest paying reserves regime, the net increase in real reserve balances under interest payment is

$$\left(\frac{\widetilde{Z^C}}{\widetilde{P}_t}\right) = (\theta - 1)(1 - \alpha)e = (\theta - 1)(2 - \lambda - 2\alpha)\frac{e}{2} + \frac{(\theta - 1)\lambda}{\theta}\frac{e}{2} > 0.$$

Here $(\theta - 1)(2 - \lambda - 2\alpha)\frac{e}{2}$ is the increment due to interest payment on original reserves holding, while $\frac{(\theta-1)\lambda}{\theta}\frac{e}{2}$ is the augmentation due to increase in the amount of reserves holding *per se*. This later term is equivalent to the decrease in the amount of real loan extended to young debtors, $\frac{\widetilde{L}_t}{\widetilde{P}_t}$. Although the increase in real reserve balances is still insufficient to overcome any liquidity shortage, the range in which liquidity constraint binds is reduced. Indeed, the intraday interest rate does not necessarily dominates the overnight interest rate: although $\frac{1}{\widetilde{\rho}_{t+1}} > 1$ due to shortages of liquidity, since $\widetilde{R}_t \geq \theta$, it is now possible for $\frac{1}{\widetilde{\rho}_{t+1}} < \widetilde{R}_t$.

Debtors' first subperiod level of consumption is now lower than that when there is no interest payment on reserves. On the other hand, both early and late departing creditors' consumption levels increase. This is because while the cost of interest payment is borne by all agents *via* inflation, only creditors benefit from the scheme (i.e. $1 < \sigma < \theta$). However, the wedge between a late-leaving creditor's consumption level and that of an early-leaving creditor is not as large as it is when no interest is being paid on reserves.[39] The equilibrium allocation is inefficient.[40,41]

1.6.2 Policy Intervention

As is suggested by Goodfriend(2002), in addition to paying interest on reserves held by creditors, let's assume that the Monetary Authority also pursues policy to overcome the

[39] Note that under interest on reserves regime, $\widetilde{y}^{C*}_{2,t+1} = \left(\frac{1}{\widetilde{\rho}_{t+1}}\right)\left(\frac{\theta}{\sigma}\right)e = \left(\frac{1}{\rho_{t+1}}\right)\left(\frac{1}{\sigma}\right)e$. Hence, $\widetilde{y}^{C*}_{2,t+1} - \widetilde{y}^{C}_{2,t+1} = \left(\frac{1}{\rho_{t+1}} - \theta\right)\left(\frac{1}{\sigma}\right)e$. In constrast, when reserves do not earn any interest, $y^{C*}_{2,t+1} - y^{C}_{2,t+1} = \left(\frac{1}{\rho_{t+1}} - 1\right)\left(\frac{1}{\sigma}\right)e$. Since $\frac{\theta}{\sigma} > 1$, $\left(\widetilde{y}^{C*}_{2,t+1} - \widetilde{y}^{C}_{2,t+1}\right) < \left(y^{C*}_{2,t+1} - y^{C}_{2,t+1}\right)$.

[40] Khan and Roberds (2001) arrive at the same non-optimality result. However, in their model, $\sigma > \theta$. That is although an agent desire to hold more reserves when interest is being paid to finance his preferred early consumption, he could not satiate himself with cash holding. This is due to the fact that he fully bears the cost of the interest payment. In contrast, here $\sigma < \theta$. There is no inhibition for creditors not to satiate himself with reserves and thus render the liquidity constraint non-binding.

[41] If $\theta > \frac{1-\lambda}{2\alpha}$, the liquidity constraint no longer binds. Hence, $\frac{1}{\widetilde{\rho}_{t+1}} = 1$ and $\widetilde{R}_t = \theta > \frac{1-\lambda}{2\alpha}$. Early and late leaving creditors now have the same marginal utility of consumption. However, $\widetilde{y}^{D}_{1,t} < \widetilde{y}^{D}_{2,t}$. Hence, it is not Pareto optimal. This is because for sufficiently high θ, the interest cost shouldered by debtors, reflected in long run asset prices increment, makes them worse off.

liquidity constraint by temporarily satiating the market with reserves (*via* open market operation or discount window, see *Figure 1-6*) such that $\frac{1}{\widehat{\widetilde{\rho}}_{t+1}} = 1$. Then it follows that $\widehat{\widetilde{R}}_t = \theta$ and $\widehat{\widetilde{P}}_t = \left(\frac{2\theta_t}{3\theta_t - 1}\right)\frac{M_t}{e}$. From the debtor's equilibrium consumption allocation (1.8), the market clearing condition for consumption good at debtor islands in period-($\mathit{2}, t$) (1.21), and the creditor's new budget constraints (1.38) and (1.39), the consumption allocations for debtors and creditors are respectively given by

$$\widehat{\widetilde{y}}^{D}_{1,t} = \frac{e}{2\theta} \quad < \quad \frac{e}{2} = \widehat{\widetilde{y}}^{D}_{2,t}$$

and

$$\widehat{\widetilde{y}}^{C}_{2,t+1} = \widehat{\widetilde{y}}^{C*}_{2,t+1} = \left(\frac{\theta}{\sigma}\right) e = \left(\frac{3\theta - 1}{2\theta}\right) e \ ,$$

since $\widehat{\widetilde{\zeta}} = \frac{\theta}{3\theta - 1}$ and $\sigma = \frac{2\theta^2}{3\theta - 1}$. Let $r_t = R_t \frac{P_t}{P_{t+1}}$ be the long run real interest rate. Clearly, $\widehat{\widetilde{r}}_t = \frac{3\theta - 1}{2\theta}$. Hence, $\widehat{\widetilde{y}}^{C}_{2,t+1} = \widehat{\widetilde{r}}_t e = \widehat{\widetilde{y}}^{C*}_{2,t+1}$.

If $\theta = 1$, then the allocation is the same as the one derived in the regime without interest payment on reserves, and is Pareto optimal. If instead $\theta > 1$, then the resulting allocation is never Pareto optimal. Debtors' first subperiod consumption is still being discounted by the positive long run *nominal* interest rate. However, the creditors' consumption is now augmented by the long run *real* interest rate. Again here, short run policy to overcome liquidity shortages affects short run asset price, which in turn causes shifts in long run nominal and real assets prices *via* creditors's readjustment of portfolio choice (and is reflected by the arbitrage pricing condition (1.18).

Since the interest paid on reserves is financed by growth in the stock of fiat money, its cost, σ, is shouldered by both creditors and debtors. Nevertheless, the interest income, θ, solely benefits creditors. This wealth transfer could potentially mitigate the debtors' gain at late-departing creditors' loss of profit in the secondary debt market when sufficient additional reserves are injected into the economy such that the liquidity constraint no longer binds. Indeed, for some ranges of θ, both creditors and debtors could be made better off *vis-à-vis* the competitive equilibrium in which reserves do not earn interest.

Specifically, if θ is in the range of

$$1 < \frac{1}{3 - 2(\frac{1-\lambda}{2\alpha})^{\alpha}} \leq \theta \leq \frac{1}{\lambda + 2\alpha} \ ,$$

then there could be Pareto improvement under the interest payment on reserves regime. This is because when $\theta \leq \frac{1}{\lambda+2\alpha}$, a creditor is definitely better off while a debtor could do no worse than her utility level attainable under the competitive equilibrium without interest payment on reserves. In contrast, when $\frac{1}{3-2(\frac{1-\lambda}{2\alpha})^{\alpha}} \leq \theta$, a debtor is made better off while a creditor could do no worse than his utility level achieved under the non-interest payment on reserves competitive equilibrium. Indeed, the greater the liquidity/coordination problem is, the greater is the range for Pareto improving intervention *via* interest payment on reserves. The following proposition follows:

Proposition 4 *Consider an interest paid on reserves regime. Assume that* $\alpha\theta + \lambda < 1 - \alpha\theta$. *Then, the competitive equilibrium allocation is not Pareto optimal.*

Next, suppose also that the Monetary Authority intervenes by injecting sufficient liquidity into the market such that $\frac{1}{\tilde{\tilde{\rho}}_{t+1}} = 1$ *and set the value of* θ.

(i) If $\theta = 1$, *the allocation is Pareto optimal.*

(ii) If $\theta > 1$, *then the allocation is never Pareto optimal.*

(iii) If $1 < \frac{1}{3-2(\frac{1-\lambda}{2\alpha})^{\alpha}} \leq \theta \leq \frac{1}{\lambda+2\alpha}$, *then the allocation is Pareto improving, but not Pareto optimal.*

Figure 1-7 illustrates the existence of a Pareto improving θ. Using the same parameter values as the example in Section 5, $(\alpha, \lambda, e) = (0.3, 0.2, 10)$, *Figure 1-8* shows that the new Pareto frontier under the interest on reserves regime is below the original Pareto frontier when there is no interest payment on reserves. Hence, even if short run reserves injection renders the liquidity constraint to be non-binding, the resulting equilibrium allocation is only optimal in the sense of second best. The golden rule allocation in Section *2* is only attainable when $\theta = 1$, i.e. by not paying any interest on reserves. However, by setting θ such that $\theta \in \left[\frac{1}{3-2(\frac{1-\lambda}{2\alpha})^{\alpha}} \; , \; \frac{1}{\lambda+2\alpha}\right]$, the allocation is Pareto improving *vis-à-vis* the competitive equilibrium under $\theta = 1$.

Note that when additional reserves is injected into the market such that the liquidity constraint does not bind, the gross long run nominal interest rate is equal to the gross interest rate paid on reserves, i.e. $\widehat{\widehat{R}}_t = \theta$. However, the value of θ is under the Monetary Authority's discretion. The Monetary Authority is free to change the value of θ in order to target the long run nominal interest rate. On the other hand, above the minimum amount of reserves required to keep $\widehat{\widehat{R}}_t = \theta$, the Monetary Authority is free to target any amount of reserves to overcome liquidity shortages. That is, the Monetary Authority, under the paying interest on reserves regime, could target the quantity of reserve balances for liquidity purposes independently from the targeting of interest rate for other broad monetary policy objectives.[42]

1.7 Conclusion

This paper is an attempt to study some of the pertinent issues that arise from the interrelation of the payments system and monetary policy implementation. It shows that short term policy in response to liquidity shortages might have unintended consequences beyond that of overcoming the liquidity problem.

By incorporating endogenous demand for reserves by agents into a payment model, it is shown that liquidity shortages, even if fully anticipated, could be an equilibrium phenomenon. Sufficient liquidity provision via open market operation or discount window policy eliminates this lack of liquidity and the resulting equilibrium is Pareto optimal. However, such short term monetary intervention has long term real welfare effects. Specifically, monetary injection into the economy lowers the long term interest rate, which in turn affects the distribution of consumption good such that debtors gain at creditors'

[42]For example, if liquidity shortages occur when expected inflation pressure is increasing, the Monetary Authority could simultaneously flood the market with reserves to provide liquidity and setting a higher θ to fend off the inflationary pressure.

expense.

In contrast, full liquidity provision under a monetary regime that pays positive interest on reserve balances held by all creditors and is financed by growth in money supply could achieve Pareto improving allocations *vis-à-vis* the liquidity constrained laissez-faire equilibrium. This is because the creditors are now remunerated by interest income on reserves whose cost is enforced on all agents. There exist some interest rate ranges for which creditors could be sufficiently compensated without making debtors worse off, and vice versa. However, a monetary regime that pays positive interest on reserves is never Pareto superior to a monetary regime that doesn't.

A testable prediction of this paper is whether a central bank's short term intervention in response to liquidity shortages in the financial market lowers long term interest rates. For example, does the movement in the Federal Funds rate during and immediately after a liquidity crisis affect the yield curve? What is the effect on banks' balance sheet? These empirical findings could potentially assist monetary authorities to make decision on whether to adopt redistribution policy such as subsidizing banks in time of liquidity crisis. To some extent, it could also shed light on market expectations in such circumstances on which policymakers could base their decisions upon.

The results here have immediate implication on the proposed Interest on Business Checking Act of 2003 that mandates the paying of interest on businesses' deposits at financial depositories and on financial depositories' reserves held at the Fed. A main advancement of the Act is to induce financial depositories to hold more Fed funds in an era of declining reserves holding and large payment flows and uncertainty. The analysis here predicts an increase in banks' reserves that is more than proportional than the per-unit interest rate paid on reserves. However, this result follows from the model specific assumptions. A more general utility function might potentially result in different conclusion. Hence, an empirical investigation along this line could provide a validation of the model, and consequently, the results therein.

More importantly, by paying interest on reserves, the Fed could potentially gain an extra policy tool in its monetary policy implementation. Specifically, under this regime, policy targeting the quantity of reserve balances for liquidity provision purpose could be pursued independently from policy targeting interest rate (which is the interest rate paid on reserves) for other broad monetary policy objectives. For example, a liquidity shortage in expansionary times, such as that during the aftermath of the Russian debt default, could be overcome by satiating the market with liquidity and simultaneously paying higher rates on reserves to mitigate inflationary pressure.

Finally, two qualifications. Firstly, there is no credit risk in the current environment. Nevertheless, it is anticipated that the optimal policy remains the same. However, since the Monetary Authority might now incur some losses in its loan provision, the open market operation and discount window lending equivalency need not hold. Secondly, there is only a single settlement asset in the model, namely funds issued by the Monetary Authority. How would the results hitherto change if there are multiple settlement assets? What are the consequences of foreign exchange risk to payments system design? Modeling multicurrency settlement asset would have profound policy implications in view of the fact that there is no prohibition against private money issuance (in the US) and the functionality of the Continuous Linked Settlement (CLS) bank in which international payment debts are cleared by various central bank monies. Fujiki (2003) and Hernandez-Verme (2004) are promising models in this direction.

Table 1-1

Total and Average Daily Payment Values

	Total				**FedWire**					
	Total Value (US$ tr)	Ratio of Total Transactions Value to GDP (at annual rate)	Ave. Daily Value (US$ tr)	Ave. Daily Value (% of GDP)	Total Value (US$ tr)	Ratio of Total Transactions Value to GDP (at annual rate)	Ave. Daily Value (US$ tr)	Ave. Daily Value (% of GDP)	Total Payment Value covered by Reserves (%)	Ave. Daily Payment Value covered by Reserves (%)
1990	421	73	1.7	0.29	199	34	0.8	0.14	0.031	7.7
1991	410	68	1.6	0.27	192	32	0.8	0.13	0.028	7.0
1992	437	69	1.7	0.27	199	31	0.8	0.12	0.029	7.3
1993	473	71	1.9	0.28	208	31	0.8	0.12	0.031	7.7
1994	507	72	2.0	0.29	211	30	0.8	0.12	0.031	7.8
1995	533	72	2.1	0.29	225	30	0.9	0.12	0.028	7.0
1996	581	74	2.3	0.29	249	32	1.0	0.13	0.024	5.9
1997	651	78	2.6	0.31	288	35	1.1	0.14	0.019	4.7
1998	679	78	2.7	0.31	329	38	1.3	0.15	0.016	4.0
1999	641	69	2.5	0.27	343	37	1.4	0.15	0.014	3.6
2000	672	68	2.7	0.27	380	39	1.5	0.15	0.012	3.1
2001	735	73	2.9	0.29	424	42	1.7	0.17	0.011	2.9
2002	721	69	2.9	0.27	406	39	1.6	0.15	0.012	3.1
2003	763	69	3.0	0.28	437	40	1.7	0.16	0.012	3.1
Ave.		72		28.5		35		13.9		

Table 1-2

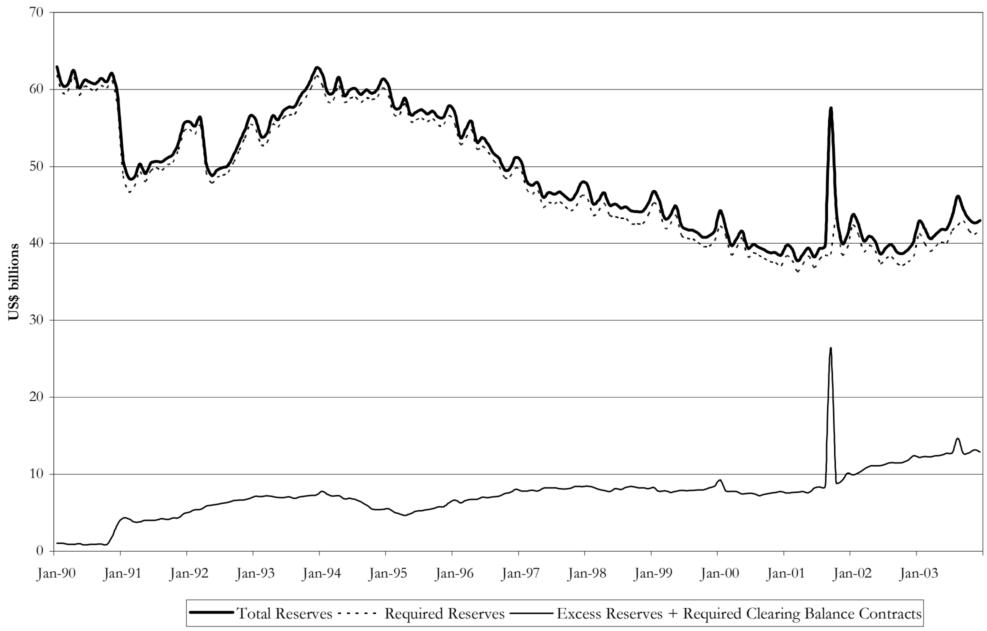
Features of Selected Interbank Funds Transfer Systems (2002)

Country	Name	Type	Setttlement Method is Central Bank Money	Value of Transactions (US$ tr)	Ratio of Total Wholesale Transactions Value to Total Payment	Ratio of Total Transactions Value to GDP (at annual rate)	Ave. Daily Payment Value covered by Reserves (%)
Belgium	ELLIPS	RTGS	yes	15.3	0.96	50.7	15.7
France	TBF	RTGS	yes	108.7	0.81	61.6	7.5
Germany	RTGS	RTGS	yes	145.1	0.98	60.4	8.3
Hong Kong	CHATS	RTGS	yes	12.6	n.a.	80.1	7.2
Italy	BI-REL	RTGS	yes	28.0	0.90	19.0	16.5
Japan	BOJ-NET	RTGS	yes	161.9	0.73	37.7	n.a.
Netherlands	TOP	RTGS	yes	24.1	0.92	46.4	16.5
Singapore	MEPS	RTGS	yes	5.7	0.95	61.9	17.2
Sweden	K&E-RIX	RTGS	yes	16.0	0.95	53.2	n.a.
Switzerland	SIC	RTGS	yes	33.2	0.99	103.1	4.5
UK	CHAPS	RTGS	yes	119.5	0.95	66.5	0.7
US	Fedwire	RTGS	yes	436.7	0.57	39.7	3.1
US	CHIPS	Net	yes	326.6	0.42	29.7	n.a.
EU	TARGET	RTGS	yes	475.0	0.90	n.a.	n.a.
EU	EURO1	Net	yes	50.5	0.10	n.a.	n.a.
Canada	LVTS	Net	yes	22.5	n.a.	25.1	0.1
Japan	FXYCS	Net	yes	41.6	0.19	9.7	n.a.

Note: In a Real Time Gross Settlement (RTGS) payments system, payments coincide with settlement. All settlement is paid in full and at real time, involving transfer of funds, normally central bank money. In a Net payment system, payment is based n net position of a payer to a payee and vice versa. For Hong Kong and UK, the values of transaction include both domestic currency and foreign currency payments. The percentages for domestic payment transactions are 89.2% and 73.4% forHong Kong and UK respectively. In addition, CHATS (Hong Kong) and SICS (Switzerland) deal with both large value and retail payments. The average business day in a given year is assumed to be 250 days.

Source: CPSS (Nov 2003), Red Book Statistical Update.

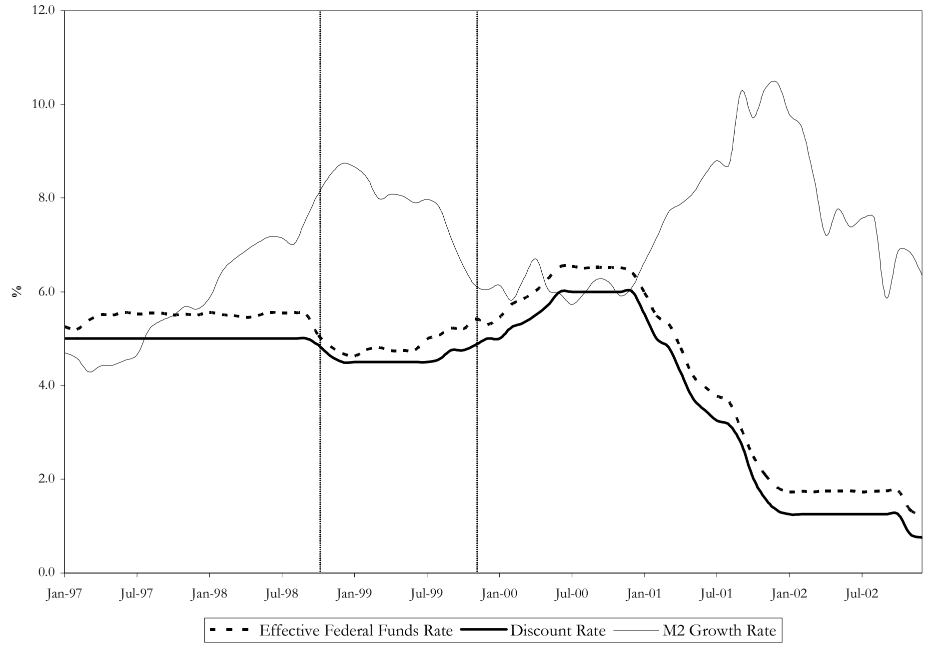
Figure 1-1
Reserves held by Depository Institutions (US)



Note : Required reserves consist of applied vault cash and required reserves balance. Clearing balance are balances held to meet clearing balance requirements. Clearing balance earns credit that could be used to pay for intraday liquidity service.

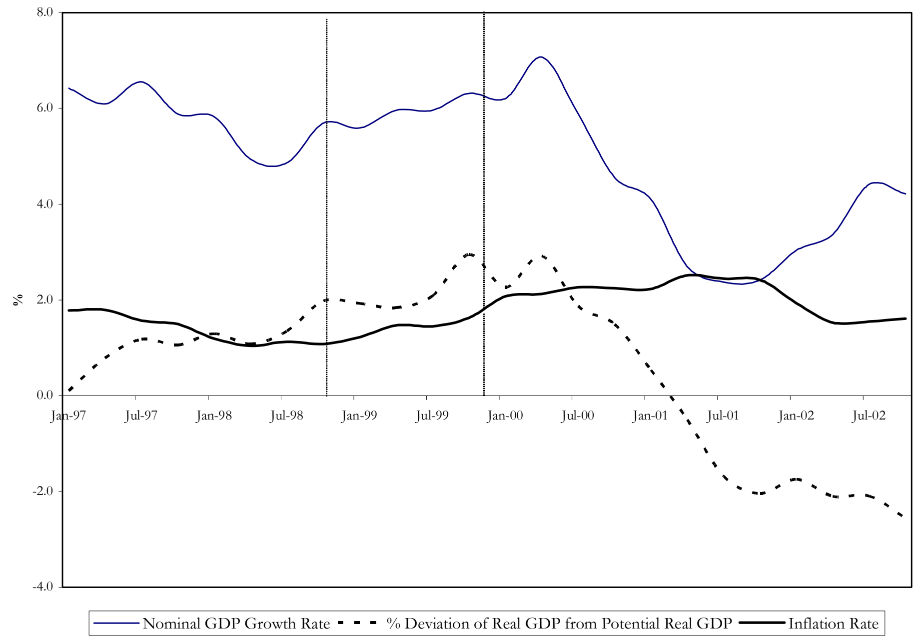
Source: Board of Governors of the Federal Reserve System

Figure 1-2a
Short Term Interest Rates and M2 Growth Rate



Source: Board of Governors of the Federal Reserve System.

Figure 1-2b
Growth and Inflation



Note: Nominal and real GDPs are in US$ bn of chained 2000 dollars. CPI index 2000=100.
Source: Bureau of Economic Analysis and Congressional Budget Office

Figure 1-3
Agents' Travel Sequence

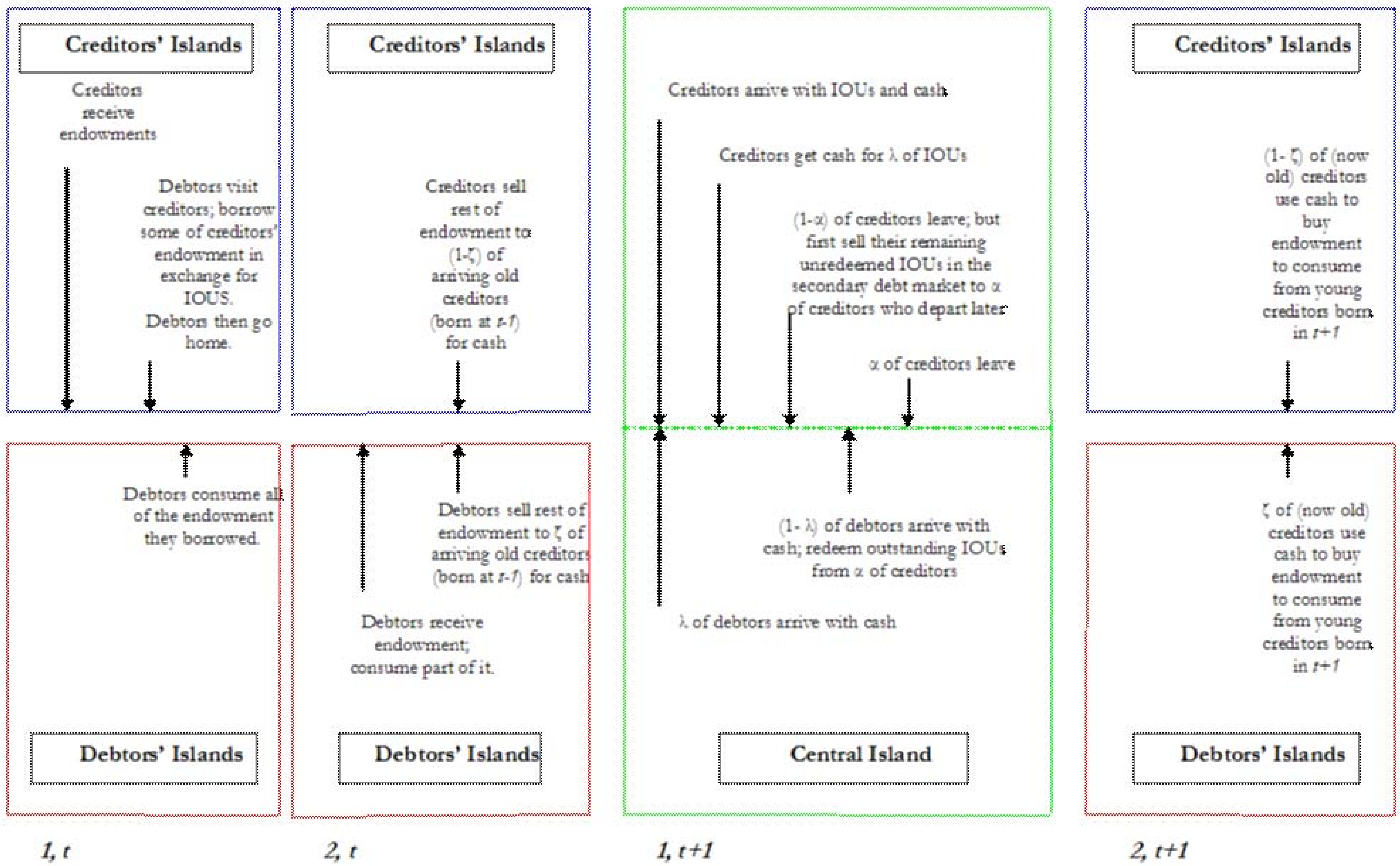


Figure 1-4
Overnight (R_t) and Daylight $(\frac{1}{\rho_{t+1}})$ Interest Rates

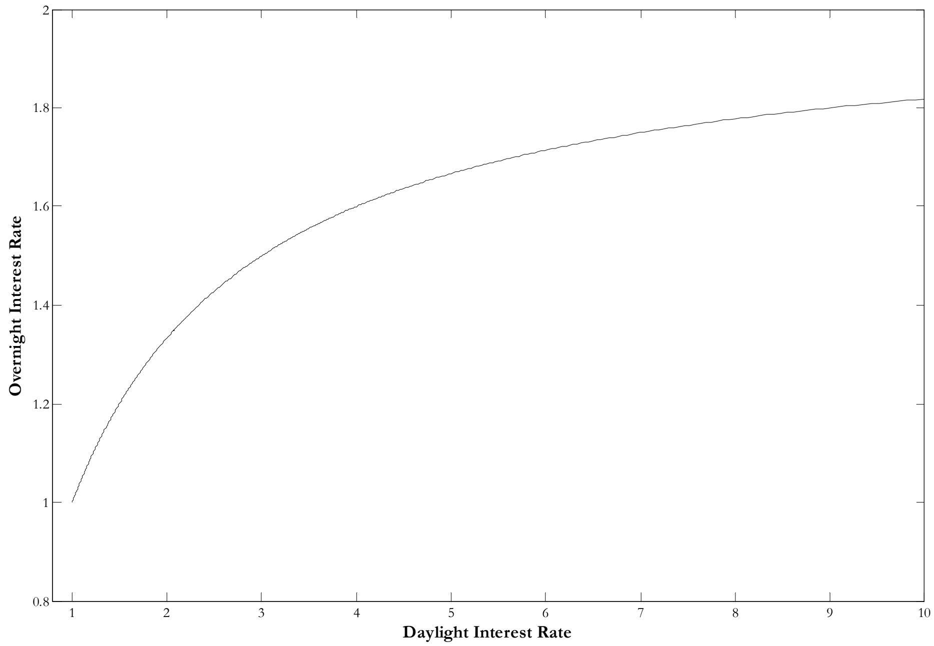
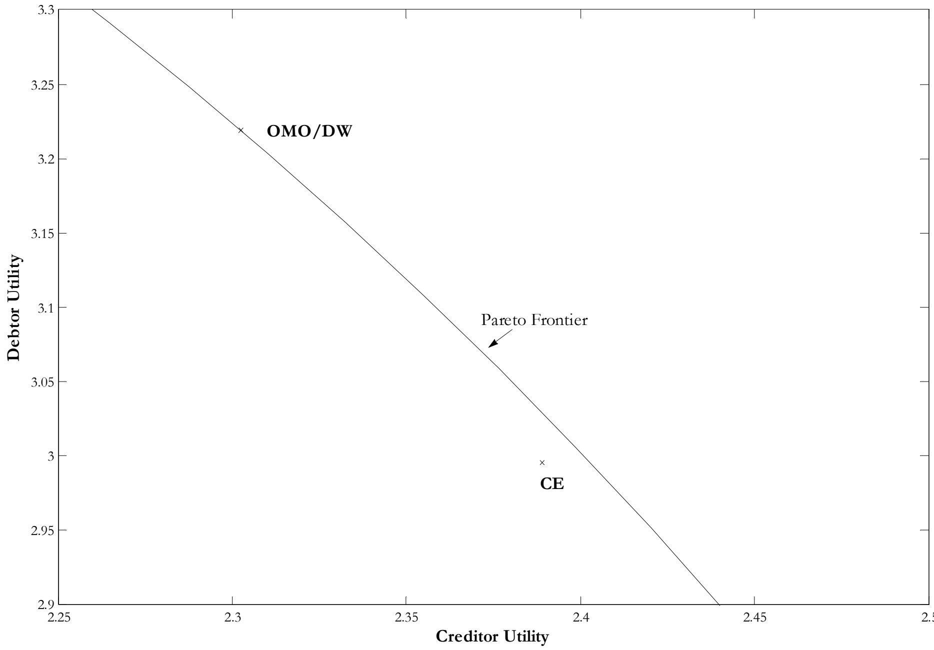


Figure 1-5
Full Liquidity Provision is Pareto Optimal, but not Pareto Improving *vis-à-vis* the Liquidity Constrained Competitive Equilibrium



Note: (α, λ, e) = (0.3, 0.2, 10). At $CE: 1 < R < \frac{1}{\rho}$, while at $OMO/DW: R = 1 = \frac{1}{\rho}$

Figure 1-6
Implementing an Interest Payment on Reserves Regime

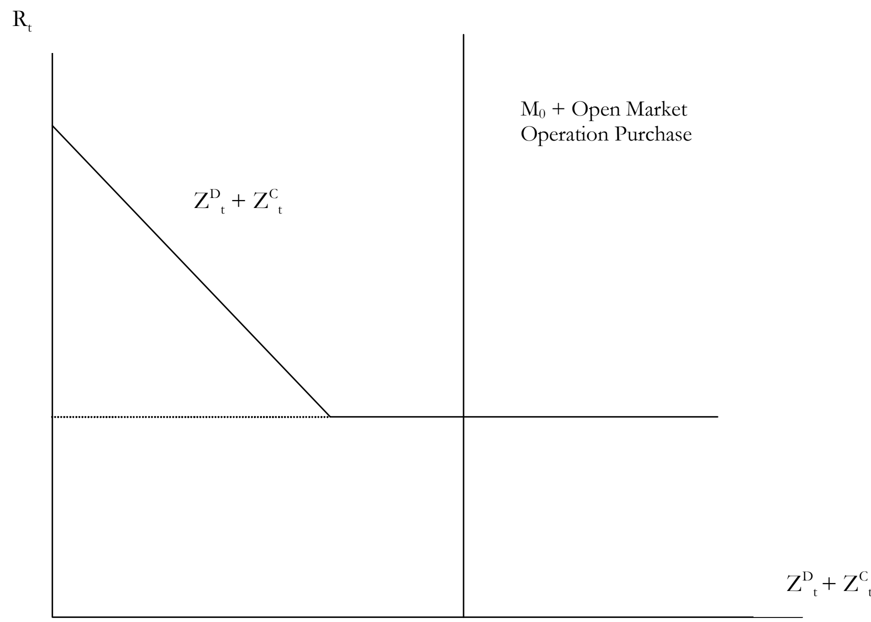
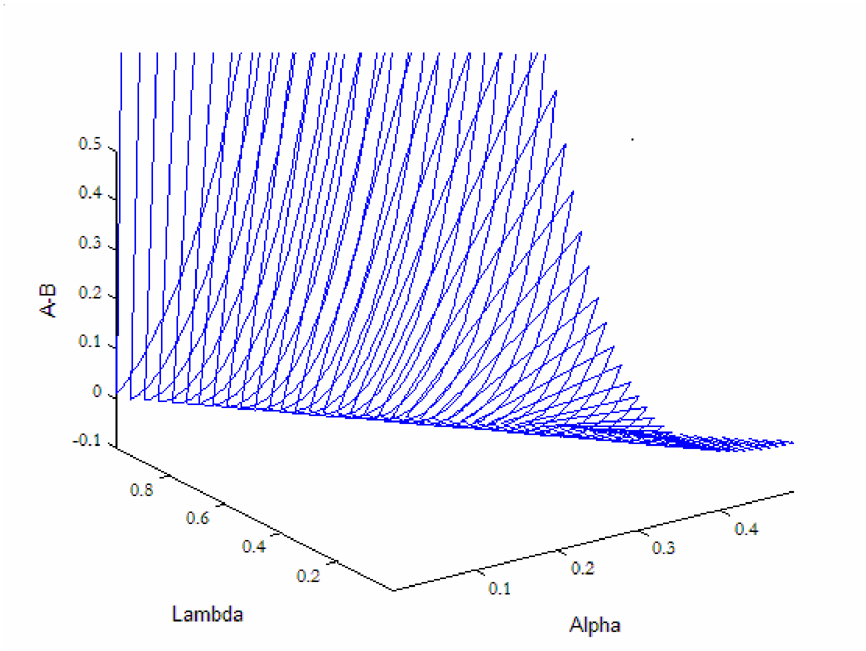


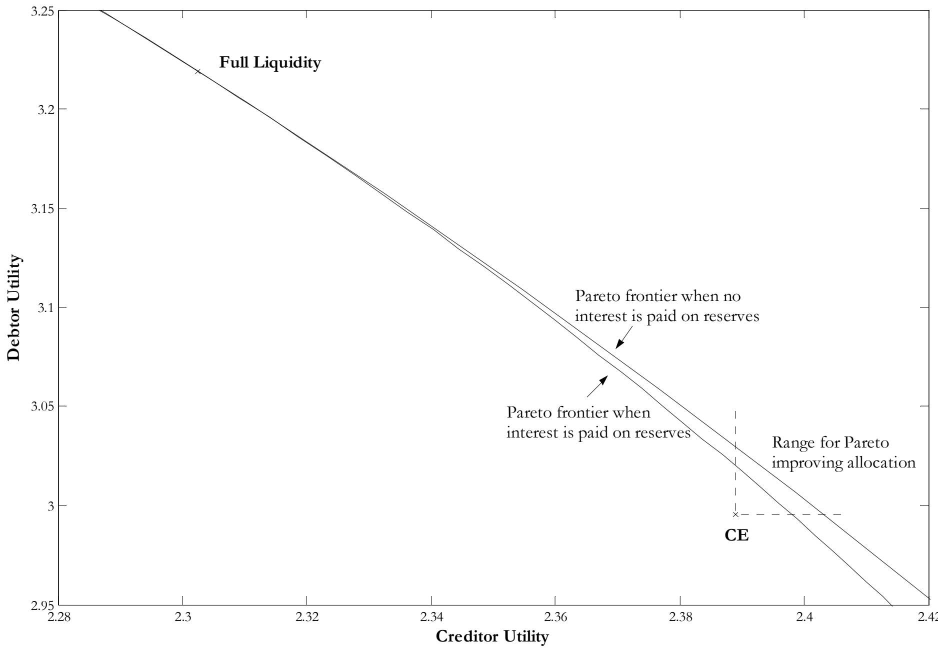
Figure 1-7
Existence of a Pareto Improving Equilibrium under the Paying Interest on Reserves Regime



Note: A-B = $\dfrac{1}{\lambda + 2\alpha} - \dfrac{1}{3 - 2(\frac{1-\lambda}{2\alpha})}$.

Figure 1-8

Full Liquidity Provision is not Pareto Optimal, but is Pareto Improving under the Paying Interest on Reserves Regime



Note: Pareto Improving θ is such that $1 < \dfrac{1}{3-2(\frac{1-\lambda}{2\alpha})} \leq \theta \leq \dfrac{1}{\lambda + 2\alpha}$.

Chapter 2

Bank's Liquidity Management and Monetary Policy Tools

2.1 Introduction

Reserves, in the form of central bank money, is important and ubiquitous in the settlement of payment debt in the payments system.[1,2] Reserves managers at depository institutions need to determine the amount of reserve balances to hold in order to meet their payment obligations. As much as possible, managers would like to minimize the amount of these non-interest bearing balances in their portfolios. However, financial institutions also demand reserves to avoid overnight overdraft on a daily basis and to fulfill their reserve requirements over the bi-weekly maintenance period. This "one instrument (i.e. reserve balances) and three targets" dilemma is further compounded by

[1]In general, payments system refers to arrangements which allow consumers, businesses and other organizations to transfer funds usually held in an account at a financial institution to one another. For example, *FedWire* is a real-time gross settlement payments system in which a paying bank fully settles its payment debt by directly transferring reserves from its account to the receiving bank's account which are both held at the central bank in real time throughout the day.

[2]The amount of total wholesale payment value in the United States was US$736 trillion for the year 2003. This is equivalent to almost seventy times the US GDP at the annual rate. 60% of this amount went through the *FedWire* transfer system. The corresponding average daily payment value was US$1.7 trillion (16% of the US GDP at the annual rate).

the uncertainties emanating from payment flows and the central bank's contemporaneous monetary policy stance that could potentially affect future reserve positions. In spite of this unpredictability, there is a discernible cyclical pattern in the aggregate demand for reserves within the reserve maintenance period. For example, in the case of the United States, banks have a proclivity to demand the greater part of their reserves (and the Fed to accommodate by increasing supply) on the last days of each maintenance period. A bank's reserve position also tends to dip on a Friday before surging on the following Monday.

In this essay I address the bank's liquidity management problem in the presence of payment flow uncertainty, reserves system design, and daily reserves market intervention by monetary authority. I investigate whether a simulated demand for reserves using the optimal control method could mimic, qualitatively and quantitatively, the empirically observed periodicity. Specifically, I examine the applicability of the quadratic linear programming with an additive noise to the bank's reserve management problem. I also analyze whether policies, conducted via monetary policy instruments, that affects the features of the operating environment induce changes to the hitherto anticipated pattern in the demand for reserves, and if so, how?

An understanding of the reserve management by financial depositories is imperative to monetary authority who needs to maneuver the various short term nominal interest rates at a high frequency. Since the early 1990s, central banks in the developed countries have adopted a relatively market-based monetary regime in which open market operation is conducted on a daily basis to supply reserves to the interbank market. Reserves are supplied at a level where the demand for it generates an interest rate, realized at or near the central bank's targeted interest rate. This policy of targeting interest rate on a daily basis requires high frequency estimates of the demand for reserves in order to direct the action of the monetary authority. In the United States, the Fed views the demand for and supply of reserves as primarily a function of the demand for required reserve balances.[3] However, these required balances are relatively demand inelastic. In practice, banks also demand reserves above the required amount to safeguard against unanticipated fluctuations that may engender insufficiency in their reserve positions, and hence, triggering substantial penalties. Modelling this precautionary demand for excess reserves is a conventional way to generate an elastic demand for balances to capture the short run dynamics of the reserves market and interest rate. Nevertheless, heretofore, due to its indirect role in the conduct of monetary policy, the analysis of demand for excess reserves has received little attention.

Here I model the reserve management of a representative bank as a stochastic control problem in discrete time. Given reserves and payments system structures, a bank demands reserves to fulfill its reserve requirements and is subject to payment obligation shocks. In order to capture the role of the monetary authority in the reserves market, the size of monetary intervention is also incorporated into the model, in two different cases, exogenously and endogenously. The inclusion of a tracking criterion function in the model is a logical representation of a bank targeting the level of reserves to hold on each day.

A closely related work is Clouse and Dow (2002) who model the demand for reserves as a dynamic programming problem. Their model captures many institutional features of the federal funds market and enables them to discuss the effects of various policy changes on the operating environment. However, there is no explicit role for monetary authority in their model. On the other hand, Bartolini, Bertola and Prati (2002) show how a bank's optimal response to the uncertainty with regard to central bank's market intervention manifest itself in the periodical volatility of the short term interest rate (and indirectly, its predictable variability in excess reserves holding behavior). In contrast, the

[3] In practice, the Fed supplies reserve balances such that the short term interest rate is most likely to return, and continue to, trade near the target as soon as feasible. That is, the Fed tries to ensure that an action on one day does not imperil the attainment of interest rate target on subsequent days. In addition, the Fed focuses on the short term interest rate's daily performance rather than on the average performance over a reserve maintenance period. Krieger (2002) provides further discussion on daily actions of the open market operation Desk.

model here is relatively flexible in the sense that it includes both features of the operating environment and monetary intervention that are exclusive to the Clouse and Dow (2002) and the Bartolini, Bertola and Prati (2002) models respectively. This is useful in light of the recent Prati, Bartolini and Bertola (2003) finding that the empirical evidences of the US federal funds market are not robust to changes in institutional environment and the approach of monetary authority's market intervention.[4]

The stochastic tracking model here exhibits qualitative dynamics of the bank's demand for excess reserves over the maintenance period similar to that observed in the empirical data. In particular, the demand schedule is weakly J-shaped, implying relatively greater bank's demand for reserves as the settlement day approaches *vis-à-vis* earlier days. Seemingly, there are two well-defined sub-periods within a given maintenance period: (i) the first seven days, and (ii) the last three days, each with its own cyclical patterns. The reserves operating environment features and payment flows could account for cyclical patterns in the first sub-period. For instance, the unequal pricing of overnight overdraft penalty on different days explains the observed lower demand for reserves on Fridays. On the other hand, the distinct increment in the demand for excess reserves in the last three days is attributable to their qualitative difference from the preceding days: accumulated reserves impose a "locked-in" cost on the holder, which is increasing in the length of days from the settlement day.[5] However, the built-in flexibility in meeting the reserve requirements enables the bank to defer the accumulation of reserves until the last few days. Consequently, the demand for excess reserves is higher in the second sub-period relative to the first sub-period.

However, the model, as it is, could not sufficiently rationalize a quantitative feature of the observed data: the model's last day increment rate (i.e. spike) in the demand for

[4]The same conclusion also holds true for the other interbank markets in the G7 economies and the Euro zone. The different monetary policy implementation regimes adopted in different countries partly accounts for this non-robustness result. For example, there is no reserve requirements in the Euro area (and most of the G7 countries). In addition, for the Euro area, as in New Zealand, reserves holding is remunerated.

[5]Section 2 provides an illustration of this so-called "locked-in" cost.

reserves is not as manifestly strong as that documented in the data. I ascribe this to the fact that the optimizing bank is more aggressive in smoothing its reserves position, and hence, does not fully embrace a "wait-and-see attitude until settlement day" to reposition its reserve holding. In order to remedy this deficiency, I augment the model with endogenous monetary intervention. That is, the bank now faces an additional uncertainty in the sense that there are deviations of monetary adds/drains from its historical daily values.[6] This additional uncertainty about the size of monetary refinancing, together with the erstwhile payment flows shock, generates a larger spike whose magnitude is closer to that observed in the data. The uncertainty about monetary refinancing condition, especially on the settlement day, inhibits the bank's previously optimal reserve holding patterns. That is, since the size of the monetary adds/drains is no longer certain, the bank could not ensure itself that sufficient funds will be available on demand. This effect is more prominent on the settlement day since there is practically no day left to unwind an up-to-then reserve position. Since, there is also a reserve requirement to meet, the bank tends to be more conservative, and thus, demand relatively more reserves on the last day.

I also analyze the effects of two other monetary policy tools on the demand for reserves. A lower reserve requirement ratio does not change the amount excess reserves as a percentage of reserves requirements, yet yields a less volatile demand for reserve balances. The consequence of paying interest on reserves depends crucially on the rate of the reserves interest. If the interest on reserve is lower than the overnight interest rate, then the effect is ambiguous. However, if interest on reserves is equal to the overnight interest rate, then there is an increase in the demand for reserves over the maintenance period and a lower volatility. This is due to the fact that both the opportunity cost and "locked-in" cost of holding accumulating reserves are now zero.

The seminal model on the demand for excess reserves is by Poole (1968). Poole's

[6] Of course, the optimizing bank could limit the magnitude of the uncertainty it is willing to tolerate at the expense of higher model misspecification.

representative bank targets a level of reserves such that the marginal benefit of avoiding the cost associated with not meeting the reserve requirements, is equal to, the marginal cost of holding non-interest bearing excess reserves. This approach leads to two underlying propositions: (i) if no interest is paid on excess reserves, then the demand for excess reserves is inversely related to the short term nominal interest rate, which is the opportunity cost of holding reserves; and (ii) since excess reserves are precautionary provision against uncertainty about reserve position, the demand for excess reserves is increasing with respect to increment in uncertainty.[7]

Since then, the related literature deals with the demand for excess reserves only indirectly. A large part of the related study on the federal funds market analyses why the federal funds rate does not exhibit the martingale property within the reserve maintenance period. That is, why banks do not treat reserves held on different days throughout a maintenance period as perfect substitutes. If the demand for reserves is essentially to meet the reserve requirements, then a bank can substitute reserves, via bidding up lower interest rate or pressuring down high interest rate such that the expected opportunity cost of holding reserves is equalized throughout a reserve maintenance period. However, for examples, Hamilton (1996) and Bartolini, Bertola and Prati (2001), find evidences that the level of the federal funds rate in the United States moves anticipatedly on different days of the reserve maintenance period.[8] This suggests that banks are reluctant to shift their demand for reserves across maintenance days, and hence, to take advantage of the predictable federal funds rate movements.

This failure to arbitrage is explored by Furfine (2000) who argues that banks hold

[7]The negative relationship between a measure of money, for example, reserves, and the short term interest rate is generally referred to as liquidity effect. It is a concept which is directly associated to the slope of the demand curve for reserves. Evidence for liquidity effect at a daily frequency is inconclusive. Hamilton (1997) finds liquidity effect in the federal funds market only on the last day of a maintenance period. Carpenter and Demiralp (2004) documented liquidity effect for other days that are prior to the last settlement day. They also show that liquidity effect is non-linear in the sense its existence depends on the size of the existing aggregate reserves in the banking system and the size of the supply of reserves. See Thornton (2001), for example, for evidence on the failure of daily liquidity effect.

[8]Prati, Bartolini and Bertola (2003) provide similar evidences for the G7 economies and the Euro zone.

reserves not only to meet reserve requirements but also for the liquidity utility that those reserves provide. His empirical evidence shows that daily patterns in payment flows, to an extent, could account for similar daily variations in the federal funds rates. Clouse and Dow (1999) and Bartolini, Bertola and Prati (2001) proffer fixed costs that occur when reserves are traded as another obstacle that prevents arbitrage. These indivisible costs include broker's fees and a bank's search cost for another bank with the same liquidity needs. Such transaction costs create a bank's predilection for late borrowing and lending of funds, which in turn, causes the last days upsurge in the demand for excess reserves. Banks are also concerned with the uncertainty about future monetary policy stance of the central bank. Bartolini, Bertola and Prati (2002) and Mitusch and Nautz (2001) propose that the volatility of the interbank rate within a reserve maintenance period is dependent on the banks' perception of the likelihood and magnitude of central bank's future monetary intervention in the reserves market. For example, if the future refinancing condition becomes more speculative or is expected to be more costly, then the amount and the volatility of demand for excess reserves by banks will increase (as is the case toward the end of a maintenance period).

The plan of this paper is as follows. The next section discusses background information related to the demand for reserves in the case of the United States. Section 3 then illustrates the basic model in which a representative bank demands for reserves in the presence of payment flows uncertainty. Section 4 presents the computational results. Section 5 examines the effects on the demand for excess reserves of policies involving monetary policy tools, namely, reserve requirements and the paying of interest on reserves. Section 6 considers a regime in which the bank endogenizes the uncertainty about the magnitude of monetary intervention in the reserves market. Section 7 concludes.

2.2 The Demand for Reserves

2.2.1 Motives and Patterns of the Demand for Reserves

A bank in the United States demands reserve balances to meet reserve requirement, to meet clearing balance requirements and to avoid overnight overdraft penalty. The first two motives are pre-committed demand for reserves that entail no uncertainty. While reserve requirements are mandated by legislation, the clearing balance requirements are made voluntarily by a bank to meet its payment needs from business transactions with other banks. Bank also hold excess reserves as a protection against daily reserves when facing unpredictable net inflow or outflow of funds. The monthly average value of excess reserves is US$1.3 billion (which represents 3.3% of reserve requirements).[9] However, as shown in *Figure 2-1*, the average value of excess balances varies from day-to-day within a reserve maintenance period. They amount from an average of US$0.84 billion for the first seven days, to US$2.7 billion for the two last non-settlement days, and to slightly above US$5 billion on the last day of the reserve maintenance period.[10]

2.2.2 Excess Reserves and Monetary Policy Implementation.

The role of excess reserves in the monetary policy operating procedure in the United States gained prominence in 1982. Owing to the high volatility of the federal funds rate over the preceding years, the Federal Reserve shifted from a growth of non-borrowed reserves target to a policy targeting the more stable borrowed reserves. In practice, the implementation of monetary policy effectively targeted the overnight interest rate, since the only way to attain the desired amount of borrowing was to maintain the appropriate spread between the discount rate and the federal funds rate.[11] Nevertheless, the demand for excess reserves was believed to be relatively inelastic then, and its response to the

[9] 1996-2005 figures (source: Board of Governors of the Federal Reserve, Washington, D.C.).

[10] The respective excess reserves as a percentage of reserve requirements are an average of 1.8% on non-settlement day and 9.7% on settlement day.

[11] See Walsh (1999).

changes in the targeted overnight interest rate was not considered seriously.

By the 1990s, however, the stable relationship involving borrowed reserves has disintegrated. Therefore, overnight interest rate targeting would have to rely on something else for interest elasticity in the federal funds market. Since excess reserves holding incurs opportunity cost, banks are more inclined to cut back on them whenever market interest rates rise. This provides the crucial elasticity needed. Thus, given a downward sloping demand schedule for reserves, the supply of reserves could be managed at a daily frequency to achieve a targeted overnight interest rate in the federal funds market.[12]

2.2.3 Reserve Requirements Accounting

Since July 1988, banks in the Federal Reserve system have operated under the lagged reserve requirement structure. A reserve maintenance period begins on a Thursday and ends on a Wednesday two week later. Since a bank is not expected to operate on a Saturday and a Sunday, the effective number of days where reserves position is calculated is ten days, with Fridays counted as three days (to account for Saturdays and Sundays). On the other hand, the reserves requirement ratio for the current reserve maintenance period is calculated from the bank's reserves position for a two-week period beginning from a Tuesday sixteen days prior to the beginning of the current reserve maintenance period. Hence, there is no uncertainty on the required reserves ratio during a maintenance period.

2.2.4 Substitutability

During the maintenance period, a bank hold reserve balances, less vault cash held during the computation period (i.e. the bank's applied vault cash) to meet its reserve requirements which are specified in terms of an average level of maintained balances over the ten

[12]Required reserves could potentially be interest elastic, too: an increase in the opportunity cost may bring about a substitution from non-interest bearing deposits, which in turn, cause a corresponding decline in the total amount of required reserves. However, this is likely to be time consuming, and hence, not feasible as the basis for daily short term interest rate targeting.

days. This confers on the bank considerable flexibility in managing its daily reserve position. For instance, given that the opportunity cost of holding idle reserves is positive, if a bank reserve balances climb higher unexpectedly on a particular day, it could hold lower amounts of reserves on the following days. Alternatively, if the bank's balances drops more than anticipated, it could simply demand higher balances over the remaining days. This substitutability of balances from one day to the next within a maintenance period helps to stabilize the overnight interest rate. For instance, if the overnight interest rate rise above the level expected to prevail over the remainder of the maintenance period, a bank could sell excess reserves in the federal funds market, and therefore, mitigating some of the upward pressure on the overnight rate.

However, it is worth noting that such flexibility is less pronounced toward the last days of the maintenance period, and specifically, on the last settlement day. Furthermore, banks are also obligated to have non-negative reserve position daily in lieu of an overnight overdraft penalty. Therefore, excess balances held on one day do not ensure protection against overdraft risk on future days, rendering reserves imperfect substitute across days in any given maintenance period.

2.2.5 Penalty Structure

The penalty rate for incurring an overnight overdraft is 400 basis point above the realized federal funds rate. At present, reserve requirements preclude clearing requirements and are both subject to pecuniary penalties for any shortfall from the required levels. The penalty rate for reserve requirement deficiency is 200 basis point above the discount rate. In addition, there is also a "locked-in" cost in holding excess reserves. To illustrate this "locked-in" cost, consider a bank with a period-average excess reserves of US$1 million. On settlement Wednesday, it must lend US$14 million worth of reserves overnight to avoid ending the maintenance period with forgone interest earning. However, the bank may not be able to do so without drawing its reserve position below zero. Hence, the excess balances may not be lent at any price. In contrast, a bank with a period-average deficiency of US$1 million in meeting its reserve requirements on settlement Wednesday must borrow US$14 million overnight. However, the bank may not have be able to find a lending bank with matching needs on such short notice, or it may not have sufficient unutilized lines of credit. Thus, the overnight interest rate could rise significantly until the bank go to seek funds at the discount window. In short, the reserves deficiency cost is not necessarily lower than that of the overnight overdraft.

2.3 The Environment

A representative bank undertakes daily transaction involving reserve balances across a ten-day reserve requirement maintenance period that ends on settlement day-10. Time is discrete so that all reserves operation that the bank carries out throughout a given day could be summarized into a single daily transaction. The bank starts day-t with the previous day cumulated level of reserve balances, Y_t. Bank demands reserves to meet daily payment obligations and a bi-weekly reserve requirements. Assume that the amount of required reserves, $\overline{rr}$, is constant throughout a given maintenance period. Hence, the bank faces a reserve requirements constraint given by

$$\sum_{t=1}^{10} Y_t \geq \overline{rr}. \tag{2.1}$$

Banks goes to the reserves (i.e. interbank) market to borrow or lend reserve funds to each other. The amount of reserves borrowed by a bank on day-t, R_t (> 0), will be returned to the lending bank on day-($t+1$) at the interest rate i_t.[13] There is also payment uncertainty, ε_t, such that the bank could not perfectly control its end-of-day-t reserve position. ε_t is Gaussian with *cdf* $\Phi(\cdot)$ and *pdf* $\phi(\cdot)$, is i.i.d. over time, and is uncorrelated across banks. That is,

$$\varepsilon_t \sim N(0, \sigma^2) \quad ; \quad E(\varepsilon_t, \varepsilon_{t+1}) = 0. \tag{2.2}$$

[13] That is, the opportunity cost per unit of reserves at day-t is i_t, which is the interest charged on interbank loans extended from day-t to day-$(t+1)$

and

$$E(\varepsilon_{t,i}, \varepsilon_{t,j}) = 0,$$

for the *i-th* bank. Since reserve requirements is constant, ε_t construes an innovation to both realized and excess reserves.[14]

In order to alleviate problem of market thinness in the interbank market that could circumscribe its ability to supply the desire amount of reserves, the monetary authority intervenes by injecting or draining funds to and from the reserves market.[15] Let $\overline{m}_t > 0$ (< 0) be the amount of adds (drains) by the monetary authority, which the bank takes as given. Therefore, the equation of motion for the end-of-day cumulative level of reserves on day-t over all the preceding days of the current maintenance period, Y_{t+1}, is

$$Y_{t+1} = Y_t + R_t - R_{t-1} + \overline{m}_t + \varepsilon_t, \tag{2.3}$$

for $t = 0, 2, 3, ...9$.[16]

Define

$$\widehat{Y}_t = Y_t + R_t - R_{t-1} + \overline{m}_t,$$

as the end-of-day-t cumulative reserve balances prior to a reserve position shock. Since reserve position cannot be negative by the end of the day's trading session, the overnight overdraft cost is in effect *daily* and is given by

$$o(Y_t) = \begin{cases} -(i_t + \gamma_o)Y_t & if \quad Y_t < 0 \\ 0 & otherwise \end{cases}.$$

[14] The infinite support assumption for ε_t implies that there exist a likelihood, however remote, of a reserve position shortfall. The representativeness of the bank and $E(\varepsilon_{t,i}, \varepsilon_{t,j}) = 0$ indicate that $E(\varepsilon_t)$ is non-random. Together, these implications suggest that i_t clears the interbank market for reserve balances.

[15] In practice, the Fed's intervention in the federal funds market occurs between 9:30-11 a.m.

[16] In summary, R_t captures the bank's interaction with other banks, through the payments system mechanism and the daily interbank market. Alternatively, m_t represents the bank's interdependence with the monetary authority, via the payments system structure and the operational framework of monetary policy.

Conversely, the reserve requirements deficiency cost applies only on *settlement day*-10 and is given by

$$d(Y_{1,}Y_{1,\ldots}Y_{10,}\overline{rr}) = \begin{cases} -(i_t+\gamma_d)Y_t & if \quad \sum_{t=1}^{10} Y_t - \overline{rr} < 0 \\ 0 & otherwise \end{cases}.$$

It follows that the expected overdraft and deficiency costs are respectively

$$E_t\left[o(Y_t)\right] = -(i_t+\gamma_o)\cdot Prob(\widehat{Y}_t+\varepsilon_t < 0), \tag{2.4}$$

and

$$E_{10}\left[d(Y_{1,}Y_{1,\ldots}Y_{10,}\overline{rr})\right] = -(i_9+\gamma_d)\cdot Prob\left(\sum_{t=1}^{10}\left(\widehat{Y}_t+\varepsilon_t\right) - \overline{rr} < 0\right). \tag{2.5}$$

At each day-t, the bank is assumed to choose a plan for the current and future amount of reserves to purchase and sell in the interbank market, $\{R\}_{t=0}^{9}$, in order to minimize the daily opportunity cost of holding non-interest bearing reserves (i_tR_t), the expected daily overnight overdraft cost (2.4), and the expected reserve requirements deficiency cost on the settlement day (2.5). In deciding the amount of funds to trade, the bank must takes into consideration its thus far accumulated reserves position, Y_t. That is, the bank's optimization program is

$$\underset{\{R\}_{t=0}^{9}}{min}\ E_t\left\{\sum_{t=0}^{9}\left[\beta_t R_t + \alpha_{t+1}o(Y_{t+1})\right] + d(Y_{1,}Y_{1,\ldots}Y_{10,}\overline{rr})\right\} \tag{2.6}$$

where

$$\beta_t = \begin{cases} 3 & for \quad t=1,6 \\ 1 & otherwise \end{cases}$$

and

$$\alpha_t = \begin{cases} 3 & for \quad t=2,7 \\ 1 & otherwise \end{cases}$$

account for the fact that the cost of holding reserves on Fridays is one-third of other days.[17,18]

Following Clouse and Dow (2002) and Furfine (2000), a key optimality condition that the solutions must satisfy is given by

$$(i_t + \gamma_o) \cdot Prob(\widehat{Y}_t + \varepsilon_t < 0) = \alpha_t(i_{t+1} + \gamma_o) \cdot Prob(\widehat{Y}_{t+1} + \varepsilon_{t+1} < 0). \qquad (2.7)$$

Namely, marginal costs due to overnight overdraft must be equalized across days. Since Friday's position (i.e. $t = 2, 7$) is carried over the weekend, a Friday overdraft is only penalized once but accounts for three days. If (2.7) is true, then the bank could not hold fewer reserves on one day and more reserves on another day such that the average amount of balances held remains constant, without increasing the cost.[19]

2.3.1 The Stochastic Tracking Model

This section reinterprets the analytical model above into a tracking model in a stochastic optimal control framework. Let $\mathbf{x}_t$ and $\mathbf{u}_t$ be the vectors for state and control variables respectively. Assume a quadratic cost function when the state vector deviates from a target level $\mathbf{x}^*$. Then, the criterion function is given by

$$\min_{\{\mathbf{u}\}_{t=0}^{T-1}} J = \left\{ L_T(\mathbf{x}_T) + \frac{1}{2} \sum_{t=0}^{T-1} L_t(\mathbf{x}_t, \mathbf{u}_t) \right\},$$

[17]Recall that on a day-t, the amount reserves bought and sold is R_t and the end-of-day cumulative balances is Y_{t+1}.

[18]The corresponding recursive formulation is

$$V_9(Y_{10}) = \min_{R_9} E_9 \left[i_9 R_9 + o(Y_{10}) + d(Y_1, Y_1, ... Y_{10}, \overline{rr}) \right]$$

subject to

$$Y_{10} = Y_9 + R_9 - R_8 + \overline{m}_9 + \varepsilon_9$$

on day-10, and

$$V_t(Y_{t+1}) = \min_{R_t} E_t \left[i_t R_t + o(Y_{t+1}) + V_{t+1}(Y_{t+2}) \right]$$

subject to (2.3) on all other days ($t = 0, 1, ...8$).

[19]As suggested by Clouse and Dow (2002), (2.7) could be considered as to characterize the general pattern of daily reserve holdings across the maintenance period. Correspondingly, the optimization problem's FOCs define the bank's optimal level of reserves across the maintenance period.

with

$$L_T(\mathbf{x}_T) = \frac{1}{2}(\mathbf{x}_T - \mathbf{x}_T^*)'\mathbf{W}_T(\mathbf{x}_T - \mathbf{x}_T^*),$$

and

$$L_t(\mathbf{x}_t, \mathbf{u}_t) = (\mathbf{x}_t - \mathbf{x}_t^*)'\mathbf{W}_t(\mathbf{x}_t - \mathbf{x}_t^*) + (\mathbf{u}_t - \mathbf{u}_t^*)'\mathbf{\Lambda}_t(\mathbf{u}_t - \mathbf{u}_t^*),$$

for $t = 0, 1, 2, ...T-1, T$, where $\mathbf{W}_t$ is a positive semidefinite penalty matrix for the state variable and $\mathbf{\Lambda}_t$ and is a positive definite penalty matrix for the control variables.

Next, define $S_t = \sum_{\tau=1}^{t} Y_\tau$ and $U_t = R_{t-1}$. Thus, the law of motion (2.3) and the reserve requirements constraint (2.1) could be jointly rewritten as

$$Y_{t+1} = Y_t + R_t - U_t + \overline{m}_t + \varepsilon_t, \tag{2.8}$$

$$S_{t+1} = S_t + Y_t,$$

$$U_{t+1} = R_t.$$

Then, the original optimization problems (2.6) subject to (2.3) could be rewritten as the following quadratic linear programming program of choosing a sequence of the control variable $\mathbf{u}_t = R_t$ to minimize the criterion function

$$\underset{\{\mathbf{u}\}_{t=0}^{9}}{min} \; J = \left\{ \begin{array}{c} \frac{1}{2}(\mathbf{x}_{10} - \mathbf{x}_{10}^*)'\mathbf{W}_{10}(\mathbf{x}_{10} - \mathbf{x}_{10}^*) \\ +\frac{1}{2}\sum_{t=0}^{9}[(\mathbf{x}_t - \mathbf{x}_t^*)'\mathbf{W}_t(\mathbf{x}_t - \mathbf{x}_t^*) + (\mathbf{u}_t - \mathbf{u}_t^*)'\lambda_t(\mathbf{u}_t - \mathbf{u}_t^*)] \end{array} \right\} \tag{2.9}$$

subject to

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{C}\overline{m}_t + \boldsymbol{\varepsilon}_t, \tag{2.10}$$

$$\mathbf{x}_0 \; : \; given,$$

and (2.2). The vectors for state variables and the additive uncertainty are $\mathbf{x} = (Y \quad S \quad U)'$

and $\boldsymbol{\varepsilon} = (\varepsilon \;\; 0 \;\; 0)'$ respectively. The coefficient matrices are given by

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & -1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}', \quad \mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}',$$

The (3×3) matrix $\mathbf{W}$ is the penalty matrix for the state variables, $\mathbf{x}$, and the scalar, λ, is the corresponding penalty for the control variable R. The time-varying penalty matrix $\mathbf{W}$ and penalty scalar λ capture the opportunity cost associated with reserve holdings, and the overdraft and deficiency costs pertaining to shortfall in reserve position. Together with the daily target variables $\mathbf{x}^*$ and R^*, they capture the effects of the operating environment on the bank's desire to hold reserve balances. Finally, for simplicity, I assume that the overnight interest rate is constant, i.e. $i_t = i^o$ for all t.[20]

Since time is discrete, the additive uncertainty (2.2) could be treated via the certainty equivalence principle. The certainty equivalence solution to this optimization problem is the linear feedback rule (see Kendrick (1981, 2002)),

$$\mathbf{u}_t = \mathbf{G}_t \mathbf{x}_t + \mathbf{g}_t,$$

where the feedback gain matrix $\mathbf{G}_t$ is

$$\mathbf{G}_t = -\left[\mathbf{B}'\mathbf{K}_{t+1}\mathbf{B} + \lambda_t\right]^{-1}\left[\mathbf{B}'\mathbf{K}_{t+1}\mathbf{A}\right],$$

and the feedback gain vector $\mathbf{g}_t$ is

$$\mathbf{g}_t = -\left[\mathbf{B}'\mathbf{K}_{t+1}\mathbf{B} + \lambda_t\right]^{-1}\left[\mathbf{B}'\left(\mathbf{K}_{t+1}\mathbf{C} + \mathbf{p}_{t+1}\right) - \lambda_t \mathbf{u}_t^*\right].$$

[20] I also assume that there is no discounting, no discount window borrowing, and no carry-over provision. Carry-over provision allows bank to carry over a small portion of surplus or deficiency in meeting requirements from one maintenance period to the next. However, a bank may not carry over two deficits consecutively, is limited to 4% or reserve requirements for carry in, and any positive carryover not used in the following period is lost. That is, a carry-over function is non-linear around the required reserves.

Note that the Ricatti matrix $\mathbf{K}_t$ is

$$\mathbf{K}_t = \mathbf{A}'\mathbf{K}_{t+1}\mathbf{A} + \mathbf{W}_t - [\mathbf{A}'\mathbf{K}_{t+1}\mathbf{B}] [\mathbf{B}'\mathbf{K}_{t+1}\mathbf{B} + \lambda_t]^{-1} [\mathbf{B}'\mathbf{K}_{t+1}\mathbf{A}],$$

and the Ricatti vector $\mathbf{p}_t$ is

$$\begin{aligned}\mathbf{p}_t &= -[\mathbf{A}'\mathbf{K}_{t+1}\mathbf{B}] [\mathbf{B}'\mathbf{K}_{t+1}\mathbf{B} + \lambda_t]^{-1} \left[\mathbf{B}'\left(\mathbf{K}_{t+1}\mathbf{C} + \mathbf{p}_{t+1}\right) - \lambda_t \mathbf{u}_t^*\right] \\ &\quad + \mathbf{A}'\left[\mathbf{K}_{t+1}\mathbf{C} + \mathbf{p}_{t+1}\right] - \mathbf{W}_t\mathbf{x}_t^*,\end{aligned}$$

with

$$\mathbf{K}_T = \mathbf{W}_T,$$

and

$$\mathbf{p}_T = -\mathbf{W}_T\mathbf{x}_T^*.$$

2.3.2 Parameterization

Foremostly, it must be noted that twelve time periods are needed in order to fit the preceding QLP formulation. However, not all time periods are relevant to a particular state or control variable. The following table provides the time relevancy of each variables.

Variable	Relevant t	Fridays
R	0~9	1,6
m	0~9	1,6
Y	1~10	2,7
S	2~11	n.a.
U	0~9	1,6

The initial values for state variables are $\mathbf{x}_0 = (Y_0, S_0, U_0)' = (100, 0, 0)'$. The desired levels for end-of-day cumulative reserve balances, Y_t^*, is set such that the excess reserves as a percentage of reserve requirements is 3%, for $t > 0$. In order to achieve this, the desired daily reserves borrowing and lending, R_t^* (and hence, U_t^*), is set to increase as a percentage of reserve requirements 3.7% daily from $t = 1$ onward, with $R_0^* = 8\%$ (see below). Hence, $U_t^* = 8\% + (t-1)(3.7\%)$. For simplicity, the average reserve requirement is normalized to 100. I assume that the bank desires to meet its reserve requirements daily, i.e. $S_t^* = 100t$. Thus, the desired state and control variables can be summarized as the following:

$$\mathbf{x}_t^* = (Y_t^* \;\; S_t^* \;\; U_t^*)' = \Big(\; 103, \;\; 100t, \;\; 8 + (t-1)(3.7) \; \Big)',$$

for $t = 1, 2, ...10$, and

$$R_0^* = 8 \quad ; \qquad R_{1-9}^* = 8 + (3.7)t.$$

I have also set the values of the variables in their each irrelevant time period such that they neither have much qualitative nor quantitative effect on the values during the relevant time periods.

The values for $\overline{m}_t$ is taken from Bartolini, Bertola and Prati (2001). The Fed typically withdraws liquidity, worth on average $5 \sim 5.5\%$ of reserves requirements, from the system on the first two days of the maintenance period. Then, the Fed progressively adds liquidity into the reserves system until the settlement day. Aggregately, however, the Fed is a net seller of securities on non-settlement days, and is a net buyer on settlement Wednesday. The following table give the values for $\overline{m}$.[21]

t	0	1	3	4	5	6	7	8	9	10
m	-5	-5.8	-3.3	-3.8	-4	-4.2	-2.7	-1.8	-1.4	-5

Note also that for the cumulated level of reserve balances, Y, $t = \{2, 7\}$ represents a Friday in which the penalty for an overdraft is a third of that on other days, $t \neq \{2, 7\}$. On the other hand, a bank incurs a reserve deficiency penalty only if it fails to meet its reserve requirements, S, on the settlement day, but not on any given non-settlement

[21] Hence, $R_0^* = Y_1 - (Y_0 + U_0 + m_0) = 103 - (100 + 0 - 5) = 8$.

day. Hence, the opportunity cost for reserve holding to meet reserve requirements is the overnight interest rate for $t = 11$ ($t = 10$ returns the same values, too). Finally, since $U_t = R_{t-1,}$ a Friday for the amount of reserves bought and sold by the bank is when $t = \{1, 6\}$. The opportunity cost of holding a positive U when $t \neq \{1, 6\}$ is three times that on Fridays. Likewise for the amount of reserves bought and sold on day-t, R_t. For simplicity, the overnight interest rate $i_t = i_{t+1} = i^o = 6\%$ is assumed. Hence, $(i^o + \gamma_o) = 10\%$ and $(i^o + \gamma_d) = 8\%$. Together with (2.7), these values are incorporated in the time-varying penalty matrix, $\mathbf{W}_t$, and the penalty scalar λ_t. Specifically,

$$\mathbf{W}_0 = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 30 \end{pmatrix} \quad \mathbf{W}_1 = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 10 \end{pmatrix} \quad \mathbf{W}_2 = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 30 \end{pmatrix}$$

$$\mathbf{W}_3 = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 30 \end{pmatrix} \quad \mathbf{W}_4 = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 30 \end{pmatrix} \quad \mathbf{W}_5 = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 30 \end{pmatrix}$$

$$\mathbf{W}_6 = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 10 \end{pmatrix} \quad \mathbf{W}_7 = \begin{pmatrix} 10 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 30 \end{pmatrix} \quad \mathbf{W}_8 = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 30 \end{pmatrix}$$

$$\mathbf{W}_9 = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 30 \end{pmatrix} \quad \mathbf{W}_{10} = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{W}_{11} = \begin{pmatrix} 0.001 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and

$$\lambda_t = \begin{cases} 10 & for \;\; t = 1, 6 \\ 0.001 & for \;\; t = 10 \\ 30 & otherwise \end{cases} .$$

A random number generator is used to create ten ε's from a normal distribution with variance, $\sigma^2 = 6$. Substituting these ε's into (2.10) and solving the parameterized cost minimization program generates a sequence of state variables $\{\mathbf{x}\}_{t=1}^{10}$, and control variables $\{\mathbf{u}\}_{t=0}^{9}$. This process is repeated for ten thousand monte carlo runs. The average value for $\mathbf{x}$ from these replications is reported. Finally, despite the existence of the twelve time periods, for ease of discussion, the relevant time periods always refer to $t \in [0, 9]$ for the control variable (R_t) and $t \in [1, 10]$ for the state variable (Y_t) and reserve maintenance day.

2.4 The Base Case

Table 2-1 summarizes the statistics for the base case and the following various experiments. *Figure 2-2* illustrates the bank's demand for excess reserves as a percentage of the reserve requirements in the base case. The average excess reserves is only 0.78% of the reserve requirements, which is below both the targeted level of 3% and the empirical level of 3.3%. Two evident patterns emerge. Firstly, the demand for excess balances is lower on both Fridays (i.e. $t = 2, 7$) than on other maintenance days, before surging on the following Mondays. Secondly, there is a pronounced general upward trend in excess reserve holding after the second Friday (i.e. $t = 7$), producing a weak J-shaped schedule. Indeed, after the reserves increase on the second Monday (i.e. $t = 8$), reserve demand may even slow down on the next Tuesday (i.e. $t = 9$). However, without fail, the demand for excess reserves spikes on settlement Wednesday (i.e. $t = 10$).

The lower demand for excess reserves on Fridays could be attributed to the fact that the cost of incurring an overdraft on a Friday is only a third of those drawn on other days, since a Friday also accounts for Saturday and Sunday. Specifically, a negative reserve position on a Friday is penalized for only one day. However, if the bank borrows reserves to avoid the overdraft, it must pay three days' worth of interest payment. Thus, cutting down a Friday's overdraft costs three times as much as reducing a non-Friday's overdraft. Since Friday's overdraft is relatively cheaper, it follows that the optimal solution is for the bank to demand relatively lower balances on a Friday than on other days. Furthermore, the bank is also disinclined in committing to holding extra reserves on Friday since that day's reserves provide liquidity needs only for a day but will be "locked-in" for three days.[22]

The rise in the demand for excess balances on Monday is partially rationalized by the effort of bank to avoid required reserves deficiency. The simultaneous need to meet the requirement ratio and the desire to avoid "locked-in" cost means that the bank will, on average, demand reserves slightly above the required level. Hence, the increase in reserves balances on a Monday, partly to make up the reserves insufficiency on the previous Friday.

The marked rise in the demand for excess reserves in the last three days, especially on the settlement day, could be accounted for by their qualitative difference from the preceding days due to the bank deferment in the accumulation of reserves. A bank's response to unexpected changes in its reserve position due to liquidity shocks on one day depends on its perception of the availability of reserves on ensuing days. As the end of the maintenance period draws nearer, the ability of the bank to counteract past shocks decreases. This causes the bank to be more sensitive to shocks as time passes such that the elasticity of the *supply* for reserves decreases over time. Hence, in the early days of the maintenance period, the bank is relatively indifferent among various reserve positions and not overly concern about the need to compensate any payment shocks. Monetary intervention by the monetary authority also helps to mitigate the bank's reaction to the shocks. In contrast, the supply of reserves is relatively inelastic in the last few maintenance days. On the settlement day, for example, there is no more monetary intervention after that day's liquidity shock is realized to alleviate the bank's response. There is less time to unwind a reserve position in response to payment shocks and to meet the reserve requirements in the later rather than the earlier days of the maintenance period. Thus, a larger response by the bank on the last days.

[22]Since the borrowing and lending for reserves at day-t is included in the same day cumulated reserve balances, the weak demand for reserves on Fridays is not due to the lack of financial activity over the weekend.

	Spike (%) t = 7~10	t = 8~10	t = 9~10	Std. Dev. First 7 days	Last 3 days
Observed	7.02	6.40	6.71	0.56	3.90
Base Case	5.77	3.77	2.38	0.84	1.92

This is further confirmed by the other evidences, as shown in the table above (which is taken from *Table 2-1*). The last day demand increment rate is approximately 2.4%, which is lower than the 6.7% that occurs in the empirical data. However, the demand for excess reserves increases to almost 5.8% from the second Friday to settlement Wednesday, closer to the corresponding 7% increase in the observed data. In addition, the volatility in the demand for excess balances over the last three maintenance days is twice that over the first seven days (i.e. standard deviation of 1.92 vs. 0.84).[23]

The qualitative, if sufficiently not quantitative, features of the demand for excess reserves emanated from the tracking model is similar to that observed in the empirical data. Specifically, there seems to be two distinct sub-periods within a given maintenance period: (1) from the first Tuesday to the second Friday, and (2) the last three days. In contrast, Bartolini, Bertola and Prati (2001) find that the last day is statistically different from the non-settlement days. This discrepancy may be due to the fact that the bank pursues a more aggressive demand smoothing in this optimal base case *vis-à-vis* the empirical model. For example, the bank meets its reserve requirements daily *and* minimizes its reserves "locked-in" cost by maintaining a lower daily average of balances. Indeed, the diminished dip on the second Friday observed in the data suggests that banks begin to turn away from the "wait-and-see" strategy and step up their demand for reserves from two or three days before settlement Wednesday. This supports the argument for two distinct sub-periods disjointed after day 7. Therefore, the control model here is a reasonable approximation to the observed data on excess reserves.

[23]Note also that although changes in σ^2 affect the time path of the cumulative reserve balances (Y_t) in each single monte carlo run, it does not affect the *average* value of Y_t over ten thousand monte carlo runs reported here. This is due to the fact that $E[\varepsilon_t] = 0$.

2.5 Operating Environment and Monetary Policy Instruments

2.5.1 Equal Overdraft Penalty

The first experiment is to analyze the extent of the fall in Friday's demand for excess reserves as the result of pricing decision in the operating environment. The overdraft penalty for Fridays is upped such that overdraft costs are equalized on all days (i.e. $\alpha = 1$ for $t = 1, 2, 3, \ldots, 10$).[24] *Figure 2-3* depicts the bank's new demand for excess reserves when the overnight overdraft penalty is constant throughout the maintenance period. Clearly, the decrease in the amount of excess reserves on Fridays has lessened. For example, demand for reserves no longer dip below the reserve requirements on the first Friday. The overall standard deviation for excess reserves has also moderated to 10.4% less than that under the base case, resulting in smoother daily average demand for reserve balances.[25] This observation acknowledges the importance of operating environment, such as the pricing structure of the payments systems and the interbank market, in influencing the bank's demand for reserves behavior.

2.5.2 Reserve Requirements

The documented increased volatility in the demand for excess reserves in recent years has, to an extent, impaired the implementation of the day-to-day monetary policy. A more

[24]The penalty matrix $\mathbf{W}_t$ and penalty scalar λ_t are now

$$\mathbf{W}_1 = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 30 \end{pmatrix} \quad \mathbf{W}_6 = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 30 \end{pmatrix} \quad \mathbf{W}_{2,7} = \begin{pmatrix} 30 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 30 \end{pmatrix}$$

and

$$\lambda_t = \begin{cases} 30 & for \quad t \neq 10 \\ 0.001 & for \quad t = 10 \end{cases} .$$

[25]This is mostly due to the lower volatility in the first seven days due to the diminishing Friday's dip.

volatile demand for excess reserves could make the estimation of the demand schedule for reserves more arduous and produce larger errors.[26] This increment in the volatility is due to the decline in the amount of reserves held, which in turn, is caused by reserves avoidance activity by banks. Since reserves earn no interest income, a bank is generally disinclined to demand more reserves than that is called for by the reserves requirements and seeks ways to avoid them. The advent of sweep accounts since the mid 1990s has further hastened this development. Sweep account allows a bank to move around its excess reserves from a non-interest bearing account that is subjected to reserve requirements (e.g. checking account) to an interest-bearing account that is not subjected to reserve requirement (e.g. savings account) throughout the business day. This enables banks to lower reservable deposits base for the calculation of reserves requirements.[27,28]

Here I evaluate the consequence of having a lower level of reserve requirements. That is, the required ratio is halved from that in the base case, while the desire level of excess reserves remains at 3% of reserve requirements. *Figure 2-4* shows the effect of the lower reserve requirements ratio on bank's demand for excess reserves on each day of the maintenance period. Evidently, the immediate cause is for the bank to decrease its reserves holding in accordance with the lower requirements. However, the bank still demands the same amount of excess balances as a percentage of reserve requirements

[26] See Sellon and Weiner (1996) for an illustration.

[27] Another significant consequence of the thinning of the reserves market is that it has made the payments system highly leveraged. For example, for the years 1994∼2003, the average daily value of wholesale payments transferred *via FedWire* has doubled to US$1.7 trillion. Within the same ten years period, the corresponding percentage of these daily payment debts backed by reserves held by the whole banking system has more than halved from 7.7% in 1990 to 3.1%.

[28] Yet, another effect of the proliferation in sweep accounts is that the demand for excess reserves is beginning to have less to deal with meeting reserve requirements and more to do with the desire to prevent overnight overdrafts. In light of this, in January 2003, the Fed took the innovative measures of (1) replacing the discount window with the so-called primary lending facility, (2) supplanting the discount rate with a modest borrowing rate (i.e. 100 basis point above the effective federal funds rate) and, (3) eliminating the non-pecuniary costs associated with borrowing from the discount window. The intended effect is to provide incentive for banks in danger of incurring overdrafts to borrow overnight from the primary lending facility without concern for additional regulatory oversight. The implicit objectives are to depress volatility and to smother any likelihood of interest rates surge in the federal funds rate. The model here does not incorporate this primary lending facility. See Madigan and Nelson (2002) for further discussion on the new discounting procedure.

(i.e. 0.78%).

An unanticipated result is that the overall volatility and the last day spikes in excess reserves holding have all decreased *vis-à-vis* the base case. This is despite the perceived increase in the risk of drawing overdraft penalty elicited by the lower total demand for balances. This decrease in volatility contradicts with the notion of lower reserve requirement causing more volatiltity. These outcomes could be put across by two facts. Firstly, since the required ratio is now lower, there is less need for the bank to adjust its reserve position on the settlement day. This results in smaller last day surge in the demand for excess reserves. Secondly, since the resulting total reserve balances is also pared down, the bank responds to the concomitant higher overdraft risk by demanding more precautionary balances on the days that it is most susceptible to reserves shortfall, i.e. Fridays, and maintaining the same level of excess reserves as the base case. Concurrently, the optimizing bank lowers reserves demand on non-Fridays to smooth its reserve position. This produces a flatter reserve demand schedule for the first sub-period *vis-à-vis* the base case.[29]

2.5.3 Paying Interest on Reserves

A proposal to counter reserves avoidance activity by banks (and hence, increasing the amount of reserves voluntarily held in the banking system) is to pay interest on reserves held.[30] I investigate two versions of the interest payment schemes, namely, (1) setting the interest on reserves less than the overnight interest rate and, (2) making them equal. Since $i^o = 6\%$, the interest on reserves, θ, is set to equal 3% in the former and 6% in the

[29] An experiment which eliminates reserve requirements results in bank demanding a positive level of balances, too. However, no explanation is offered here as the repeal of the requirements changes the operating framework which renders any comparison ambiguous.

[30] Some countries have enacted this policy. Reserve Bank of New Zealand pays explicit interest on reserves deposited by banks in its system. In the United States, implicit interest is paid on reserves held to meet clearing requirements in the form of daylight overdrafts credit from the Fed. In addition, the Interest on Business Checking Act of 2003, currently waiting for the Senate vote, mandates explicit interest payment on a bank's reserves held in its account at the Federal Reserves.

latter.[31,32]

Figure 2-5 exhibits the over time demand schedule for excess reserves under the paying interest on reserves regime for each interest rate. In the case $\theta < i^o$, the demand schedule under the new regime differs from the base case's demand schedule only slightly: the level and volatility of the demand for excess reserves and last days spikes are both qualitatively and (almost) quantitatively equal. In contrast, when $\theta = i^o$, the level of excess reserves demanded under the interest on reserves regime is 2.9% of reserve requirements, higher than that under the base case. Demand volatility and last days' surges also declined *vis-à-vis* the case when there is no interest payment on reserves.

These differences in level, volatility and spikes could be ascribed to the fact that unless $\theta = i^o$, the opportunity cost of holding idle reserves is positive. Hence, for $\theta < i^o$, holding reserves is still costly and the bank still faces the same trade-off as in the case

[31] I assume that interest is paid on all reserves.

[32] The penalty matrix $\mathbf{W}_t$ and penalty scalar λ_t are:
(1) for $\theta = 3\%$,

$$\mathbf{W}_0 = \begin{pmatrix} 15 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 15 \end{pmatrix} \quad \mathbf{W}_1 = \begin{pmatrix} 15 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{pmatrix} \quad \mathbf{W}_6 = \begin{pmatrix} 15 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$
$$\mathbf{W}_{2,7} = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 15 \end{pmatrix} \quad \mathbf{W}_{3,4,5,8,9} = \begin{pmatrix} 15 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 15 \end{pmatrix}$$
$$\mathbf{W}_{10} = \begin{pmatrix} 15 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{W}_{11} = \begin{pmatrix} 0.001 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \lambda_t = \begin{cases} 5 & for \;\; t = 1,6 \\ 0.001 & for \;\; t = 10 \\ 15 & otherwise \end{cases}.$$

(2) for $\theta = 6\%$,

$$\mathbf{W}_0 = \begin{pmatrix} 12 & 0 & 0 \\ 0 & 0.001 & 0 \\ 0 & 0 & 12 \end{pmatrix} \quad \mathbf{W}_1 = \begin{pmatrix} 12 & 0 & 0 \\ 0 & 0.001 & 0 \\ 0 & 0 & 4 \end{pmatrix} \quad \mathbf{W}_6 = \begin{pmatrix} 12 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$
$$\mathbf{W}_{2,7} = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 12 \end{pmatrix} \quad \mathbf{W}_{3,4,5,8,9} = \begin{pmatrix} 12 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 12 \end{pmatrix}$$
$$\mathbf{W}_{10} = \begin{pmatrix} 12 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{W}_{11} = \begin{pmatrix} 0.001 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \lambda_t = \begin{cases} 4 & for \;\; t = 1,6 \\ 0.001 & for \;\; t = 10 \\ 12 & otherwise \end{cases}.$$

with no interest on reserves. Whether the amount of reserves increases or not depends on the substitution and income effects brought about by the lower cost to hold reserves. However, when $\theta = i^o$, not only the opportunity cost is zero, but there is also no "locked-in" cost associated with accumulated reserves. That is, the bank need not adjourn its reserve holding toward the last few days to meet the end-of-period reserve requirements. Concurrently, this induces the bank to hold relatively more excess reserves daily, resulting in lesser demand volatility and last days' surges in reserves level, and lower probability of incurring overnight overdraft.[33] An upshot of this is the importance of setting θ prudently, if and when, the monetary authority decides to implement the interest on reserves regime.

2.6 Monetary Intervention

Thus far, the quadratic linear tracking model has reasonably captured the qualitatively aspects of the observed dynamic patterns in the demand for excess reserves. However, it is less successful in appropriating the magnitude of the dynamics. Specifically, the tracking model picks up only 35% of the last day spurt in the demand for excess reserves than what is suggested by documented data. A plausible reason is that, *vis-à-vis* a real world bank, the optimizing bank in the model is more aggressive in dealing with the "locked-in" cost. That is, in the optimal case, the bank does not fully embrace a "wait-and-see" attitude until settlement day to reposition its reserve holding. As is shown in *Table 2-2*, the optimizing bank's last four days (of the maintenance period) accumulation of reserves upward surge compares favorably to the observed last day spike in the data. The bank could pursue this reserves smoothing activity during these last days of the maintenance period since the only uncertainty it faces deals with the payment flows.

In a related strand of the literature on the dynamics of short term interest rates, Bartolini, Bertola and Prati (2001, 2002) contend that consideration for the size of the

[33] In the experiment where $\theta = i^o$, there is no required reserves deficiency on any day.

Fed's daily open market operation, and its associated uncertainty, could explain away the weak J-shaped demand schedule for excess reserves over the maintenance period documented in the case of the United States.[34] *Figure 2-6* represents the consequence of subtracting the amount of monetary adds and drains on the bank's demand for excess reserves over the maintenance period. In particular, *Figure 2-6a* depicts the observed case while *Figure 2-6b* shows the simulated base case model.[35] In the former, the observed periodical pattern for reserve holding on settlement day seemingly reflects the cyclical behavior of funds provided by the Fed to the federal funds market. The over time schedule for excess reserves net Fed's intervention is flatter and resembles a white noise process. This is even markedly apparent in the optimal base case.[36] This suggests that the failure to take into consideration the uncertainty about the size of the open market intervention by monetary authority could potentially account for the base case model's shortfall in the settlement day spike.

In order to capture the bank's uncertainty with respect to the size of the open market intervention, I endogenize the size of the monetary injection/withdrawal, m_t, in the stochastic tracking model. Specifically, the bank is concerned about the uncertainty deviation from the historical values of the magnitude of the open market operation intervention, $\overline{m_t}$. This uncertainty with regard to the refinancing stance of the monetary authority arises from the fact the monetary authority may be "unwilling" or "unable" to fully accommodate the needed liquidity to clear the reserves market.[37] Hence, the

[34]Mitusch and Nautz (2001) provide the corresponding theoretical arguments for the Euro case.

[35]The data points for the empirical case here, i.e. *Figure 2-6a*, are taken from Bartolini, Bertola and Prati (2001).

[36]However, the difference between the observed and optimal demand for excess reserves on settlement day remains. There are two possible alternative reasons. Firstly, as noted earlier, the optimizing bank is relatively more aggressive in minimizing reserves "lock-in" cost. Secondly, the last day payment shocks in the model is relatively smaller (i.e. enters positively) or is adverse to that day's reserve position (i.e. negatively). Hence, even though the optimizing bank is plausibly as aggressive as a real world bank suggested by the data, the last day spike is less pronounced.

[37]Monetary authority is "unwilling" due to some predisposed objectives, or is "unable" because of some inherent and market limitations. An example of the former is the monetary authority's pursuance of tight monetary policy stance even though market liquidity is constrained. An example for the later is the difficulty in implementing unusually large repurchase agreement operations due to the bank's insufficient collateralizable securities.

bank's challenge is to design an augmented tracking model whose policy rule minimizes deviations from the its original approximating tracking model.

2.6.1 The Augmented Model with Endogenous Monetary Intervention

Following Rustem (1988) and Hansen and Sargent (2001), the bank adopts a bounded worst-case strategy, optimizing under the additional constraint that the monetary authority contrives to produce the most disadvantageous parameterization of the augmented model.[38] The bank, though is unsure about the size of adds/drains, is confident that its hitherto model is a good approximation to the observed data in the sense that the approximation error is bounded, i.e.

$$\sum_{t=0}^{9}(m_t - m_t^*)^2 \mid \mathbf{x}_0 < \eta, \tag{2.11}$$

where m_t is now a control variable that is fed back on the history of $\mathbf{x}$. η measures the size of the maximal specification error tolerated by the bank in the face of uncertainty about the size of m_t. Hence, for all $\eta > 0$, the bank has preference for committing initially to a worst policy so as to minimize its total expected costs for the worst possible deviation. In order to assure stability, let $0 < \eta << \infty$.[39]

Assume a Markov perfect equilibrium of two-player game. That is, both bank and the monetary authority chooses sequentially and simultaneously in every period, taking the each other's decision rule as given. Then, the augmented criterion function is

$$\underset{\{\mathbf{u}\}_{t=0}^{9}}{min}\ \underset{\{m\}_{t=0}^{9}}{max}\ J = \left\{ L_{10}(\mathbf{x}_{10}) + \frac{1}{2}\sum_{t=0}^{9} L_t(\mathbf{x}_t, \mathbf{u}_t) - \frac{\mu}{2}\left(\sum_{t=0}^{9}(m_t - m_t^*)^2\right) \right\}, \tag{2.12}$$

[38] Hansen and Sargent (2001) term their approach as robust control. Kendrick and Tucci (2001) and Gonzales (2003) apply the idea to issues solvable via quadratic linear programming.

[39] If $\eta = 0$, then the bank is back with the original base case model.

subject to

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{C}m_t + \boldsymbol{\varepsilon}_t, \tag{2.13}$$

and (2.2), with $\mathbf{x}_0$ given, and where μ is the Langrange multiplier for (2.11).[40] The deviations in $(m_t - m_t^*)$ are now fed back on the endogenous state vector, $\mathbf{x}_t$. Allowing this feedback is part of the way that the cost minimizing bank designs a rule that acknowledges the uncertainty about m_t.[41] μ is a parameter that measures the bank's tolerance with respect to the uncertainty about m_t. For example, m_t is chosen such that the deviations in $(m_t - m_t^*)$ could precipitate the criterion value (3.27) to approach a very large positive number. In order to restraint this possibility, the minimizing bank sets μ at a relatively higher value. Thus, the lower the value of μ, the higher is the bank's concern for uncertainty about m_t, with increases in the deviation of $(m_t - m_t^*)$. In contrast, if $\mu \to \infty$, then the minimizing bank is not concerned with uncertainty with respect to m_t, and the base case model is realized.

2.6.2 The Augmented Stochastic Tracking Model

It is now straightforward to rewrite the program (3.27) subject to (2.13), (2.2) and $\mathbf{x}_0$ into a quadratic linear tracking program of the following form:

$$\underset{\{R\}_{t=0}^{9}}{min} \quad J = E_t \left\{ \begin{array}{c} \frac{1}{2}(\widetilde{\mathbf{x}}_{10} - \widetilde{\mathbf{x}}_{10}^*)'\widetilde{\mathbf{W}}_{10}(\widetilde{\mathbf{x}}_{10} - \widetilde{\mathbf{x}}_{10}^*) \\ +\frac{1}{2}\sum_{t=0}^{9}[(\widetilde{\mathbf{x}}_t - \widetilde{\mathbf{x}}_t^*)'\widetilde{\mathbf{W}}_t(\widetilde{\mathbf{x}}_t - \widetilde{\mathbf{x}}_t^*) + (\widetilde{\mathbf{u}}_t - \widetilde{\mathbf{u}}_t^*)'\boldsymbol{\Lambda}_t(\widetilde{\mathbf{u}}_t - \widetilde{\mathbf{u}}_t^*)] \end{array} \right\} \tag{2.14}$$

$$\widetilde{\mathbf{x}}_{t+1} = \widetilde{\mathbf{A}}\widetilde{\mathbf{x}}_t + \widetilde{\mathbf{B}}\widetilde{\mathbf{u}}_t + \boldsymbol{\varepsilon}_t, \tag{2.15}$$

$$\widetilde{\mathbf{x}}_0 \;:\; given,$$

and (2.2). The state and additive uncertainty vectors remain the same, i.e. $\widetilde{\mathbf{x}} = \mathbf{x}$ and $\boldsymbol{\varepsilon} = (\varepsilon \quad 0 \quad 0)'$. The new control vector is $\widetilde{\mathbf{u}} = (R \quad m)'$and the new desired level of

[40]The constant $\mu\eta$ can be dropped, without any loss of generality.

[41]In other words, the policy rule takes into consideration the possibility that the approximating base case model's dynamic is misspecified.

monetary adds/drains is $m_t^* = \overline{m}_t$. The parameter matrices are $\widetilde{\mathbf{A}} = \mathbf{A}$ and

$$\widetilde{\mathbf{B}} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

The time-varying penalty matrix for the state variables, $\mathbf{W}_t$, remains a (3×3) matrix, while the new time-varying penalty matrix for the control variables

$$\widetilde{\mathbf{\Lambda}}_t = \begin{pmatrix} \lambda_t & 0 \\ 0 & -\mu \end{pmatrix}$$

is a negative semidefinite penalty matrix for the control variables.

2.6.3 Results

Figure 2-7 describes the demand for excess reserves when the bank takes into consideration the uncertainty with regard to the size of monetary authority intervention, m_t, for different values of μ. For $\mu = 10000$, the augmented model's over time demand schedule parallels closely the base case model's schedule. However, lower values for μ, implying more uncertainty with regard to m_t, elicit a greater deviation in the respective demand schedules for excess balances from that under the base case. More importantly, for all values of μ considered here, there is a larger last day upsurge in reserve holding *vis-à-vis* the base case.

	Spike (%)			Std. Dev.	
	t = 7~10	t = 8~10	t = 9~10	Last 3 days	Last 4 days
Observed	7.02	6.40	6.71	3.90	3.46
Base Case	5.77	3.77	2.38	1.92	2.40
μ = 80	7.47	5.29	3.13	2.65	3.11

Table 2-2 and *Figure 2-8* compare bank's demand for excess reserves from the empirical data with those from the theoretical tracking model under exogenous and endogenous monetary authority intervention. For example, as shown in the table above for $\mu = 80$, the last day spike in excess reserve position resulted from the augmented model is larger than that under the base case and picks up 47% of the observed surge. Likewise, a greater pick-up rates for the three and four days' spikes at 83% and almost 100% respectively. The last-three day volatility under which m_t is incorporated endogenously are also more aligned to the empirical data than the corresponding base case computed volatility.[42] Similar conclusions could also be drawn for the differences from mean data. Therefore, the bank is attentive about the uncertainty in the size of the monetary intervention, more so as the settlement day approaches. The bank readjusts its optimizing behavior to minimize costs, including "locked-in" cost, with its concerns about uncertainty surrounding the size of monetary authority intervention. This results in less smoothing of reserve balances than that which is in accordance with the optimal base case behavior (where there is no uncertainty about m_t).

The resulting data from model simulations seems to suggest that there are two distinct dominating behaviors in the bank's demand for excess reserves. In particular, the effect of monetary intervention uncertainty is more pronounced toward the last three days of the maintenance period. This observation seems to convey the idea that μ may be time-varying. This is akin to setting different weights for different days. Specifically, the bank is more wary of uncertainty with respect to the size of monetary intervention towards the end of the maintenance period relative to preceding days. A higher penalty weight (i.e. lower value of μ_t) on these last days *vis-à-vis* earlier days (i.e. higher value of μ_t) captures this relative aversion to uncertainty.[43]

[42]The last three and four days' volatilities under the augmented model are 68% and almost 90% of the corresponding observed volatilities. The associated base case percentages are only 49% and 69% respectively.

[43]I consider five different combinations of μ's with regard to deviation in μ_t on the first seven days and the remaining three days, and on the first nine days and the settlement day. Namely, (i) $\mu_{0-6} = 80$ & $\mu_{7-9} = 50$, (ii) $\mu_{0-8} = 80$ & $\mu_9 = 50$, (iii) $\mu_{0-6} = 110$ & $\mu_{7-9} = 50$, (iv) $\mu_{0-8} = 10000$ & $\mu_9 = 40$,

	Spike (%) t = 8~10	t = 9~10	Std. Dev. Last 3 days	Last 4 days
Observed	6.40	6.71	3.90	3.46
Base Case	3.77	2.38	1.92	2.40
μ(t = 9) = 30	8.05	6.60	4.34	4.39

Table 2-3 and *Figure 2-9* contrast the bank's demand for excess reserves under different time-varying values for μ. As could be seen, the increment rates for the last three days and on the settlement days are all closer to the empirical case than those recorded under constant μ. From the above table for instance, in the case where the bank worries about the deviation in the size of monetary intervention only on the last day, m_{10}, the last day surge under $\mu_{0-8} = 10000$ & $\mu_9 = 30$ (i.e. 6.6%) is almost equivalent to the increment rate observed in the data (i.e. 6.7%). Furthermore, in contrast to the cases when μ is constant, the volatilities for all subsets of days throughout the maintenance period under time-varying μ are also closer to the volatilities observed in the data. These results reinforce the argument that considerations for (i) the uncertainty surrounding the size of monetary intervention, and (ii) the distinction of two sub-periods within a reserve maintenance period, rationalize the qualitative and quantitative aspects of the demand for excess reserves by an optimizing bank in the context of a linear quadratic control problem with a tracking function.[44]

2.7 Conclusion

This essay analyzes bank's liquidity management problem in the presence of payment flow uncertainty, reserves system pricing structure, and monetary authority's daily inter-

and (v) $\mu_{0-8} = 10000$ & $\mu_9 = 30$. The values for $\mathbf{W}_t$ and λ_t remain as in the original parameterization.

[44]However, more specification is needed to validate the augmented model as a better approximation to the observed data *vis-à-vis* the base model. Indeed, uncertainty about the size of monetary intervention per se may turn out to be neither sufficient nor necessary to explain the time path dynamics of the demand for excess reserves.

vention in the reserves market and utilization of other monetary policy tools. In spite of these uncertainties, there are observed periodicities in excess reserves demand over the reserves maintenance period. The quadratic linear tracking model has reasonably encapsulated both the qualitative and quantitative (albeit, less successfully) features of the bank's demand for excess reserves over the maintenance period similar to that observed in the empirical data.

Specifically, the demand schedule is weakly J-shaped, suggesting relatively greater bank's demand for reserves as the settlement day approaches *vis-à-vis* earlier days. In view of this, there are two distinctive sub-periods within a given maintenance period: (1) the first seven days, and (2) the last three days, each with its own predictable patterns. The bank's response to the exogenous flow of payments and the operating environment of the reserves market could explain the expected demand behavior in the first sub-period. On the contrary, the obvious escalation in the demand for excess reserves in the last three days is a result of the bank's desire to minimize "locked-in" cost in the preceding days. The flexibility inherent in meeting the reserve requirements enables the bank to put off the accumulation of reserves until the last few days. As a consequence, the demand for excess reserves is higher in the second sub-period relative to the first sub-period. Further consideration for uncertainty about the size of the monetary authority's intervention in the reserves market produces a settlement day surge in the demand for excess reserves whose magnitude is closer to that observed in the data, *vis-à-vis* the base case with $\overline{m}_t$.

I also examine the effects of two other monetary policy tools on the demand for reserves. While alterations in the reserve requirement ratio lead to only a quantitative effect, paying interest on reserves brings about both quantitative and qualitative implications on reserves demand.

The stochastic linear quadratic programming (QLP) with a tracking function utilized here is a suitable representation of a bank's inventory-like day-to-day reserve management problem. The QLP model is flexible enough to integrate the main features of the market for central bank reserves, and is numerically easy to solve that the curse of dimensionality problem loses most of its forces. Thus, the model could potentially be extended to include additional institutional aspects such as carry-over provision and the recently established primary lending facility discussed earlier.[45] Potential non-linearity in the operating environment such as the one that accompanies the carry-over provision could be approximated by a first-order expansion of the equations of motion and a second-order expansion of the criterion function. Then, the QLP method could solve the approximated problem. This operation is iterated such that the criterion and system equations are expanded each time around the solution attained on the preceding iteration. The iterations continue until tolerable convergence is achieved.

However, in embracing each facet of the liquidity management problem, most of the design specifications must be incorporated into a quadratic cost functional. That is, the QLP method is optimal in only a narrow sense, and actual reserve management specifications (e.g. the reserve requirements constraint and unequal daily overdraft penalty) must be translated or reinterpreted so as to fit into the framework of this method. Moreover, although the certainty equivalence principle allows convenient algorithm for solving the dynamic programming problem in the sense that it separates the parameter estimation problem from the control problem, it does not characterize optimal control problems in general.

A concern is the fact that the outcome for the QLP model with additional uncertainty about the size of monetary intervention is sensitive to the values assigned to μ's. That is, different ranges of μ produce different dynamics in the demand for reserves. This is an example of the "discontinuity in μ" problem examined by Gonzalez and Rodriguez (2004). They find a unique discontinuity point for the value of μ that causes different dynamics for a state variable across the two regions. In contrast, there are many discontinuity points in the present QLP model. It is possible that the reason is the non-monotonic nature of the bank's demand for reserves.

Another apparent weakness of the model is the ad-hoc manner in which the size of

[45] See footnote 28.

monetary intervention is incorporated into the model. In truth, the size of the monetary intervention affects the bank's reserve position indirectly via the overnight interest rate. Instead, a better interpretation for m_t is as a net payment inflow to the bank. Then, it is possible that the bank takes into consideration of the uncertainty pertaining to the size of the net inflow (which is dependent on the bank's opposite transacting party) which affects its day-to-day decision-making.

Next, two qualifications. Banks differ in, among other attributes, in size and in being a net seller or a net buyer in the reserves market. However, since the overnight interest rate is assumed to be constant, the model here is a partial equilibrium analysis. This abstracts the effect of assorted experiments on heterogeneous banks, which may react variedly to a common policy action. However, as is mentioned earlier, the monetary authority's (for concreteness, the Fed) day-to-day monetary policy with regard to the interbank market rate is to stabilize it as much as it is feasible. The monetary authority remedies any deviation from a targeted rate by conducting open market operation, a variable captured by the size of the monetary intervention and, to an extent, endogenized in the model. Certainly, the targeted interest rate does not often change during the short spanned reserves maintenance period.[46]

Finally, the tracking model here incorporates the uncertainties in a simplest way. Both innovation about the payment flows and uncertainty regarding the size of monetary intervention enter the model additively, which is then treated via the certainty equivalent principle. Furthermore, the latter does not involve time sensitive strategic roles for the min-max programming. In contrast, by appealing to a Stackelberg-type game, the program could be transformed into a more naturally interpretable interaction between a lead central monetary authority and a following bank. However, there is the possibility that the certainty equivalence property no longer holds. Moreover, there is also uncertainty about the value of parameters, which is absent in this model. This multiplicative-type

[46]At any rate, as mentioned in footnote 7, the evidence of a liquidity effect is still inconclusive in the literature.

of uncertainty opens up the possibility of using learning algorithm, both passive and active, which may lead to better understanding of the reserves market dynamics. A QLP tracking model that addresses all these shortcomings and extensions is left for future undertaking.

Table 2-1
Statistics for Experiments ($\overline{m}_t$)

	Observed	Base Case	Equal Overdraft Penalty	Lower Reserve Requirements	Interest on Reserves < *i*	Interest on Reserves = *i*
% Excess Reserves						
1	2.00	1.27	1.29	0.65	1.25	2.02
2	0.90	-0.96	-0.02	-0.47	-0.99	-0.04
3	1.50	0.08	-0.23	0.01	0.06	0.59
4	1.00	0.42	0.16	0.21	0.40	0.89
5	1.30	0.42	0.18	0.21	0.39	0.84
6	1.50	0.21	0.06	0.10	0.16	0.65
7	2.50	-1.13	-0.18	-0.57	-1.22	-0.66
8	3.10	0.77	0.61	0.39	0.70	1.01
9	2.80	2.14	2.04	1.09	2.05	2.34
10	9.70	4.57	4.48	2.28	4.46	4.67
Spike (%)						
t = 2~3	0.59	1.05	-0.21	0.97	1.06	0.63
t = 7~10	7.02	5.77	4.67	5.77	5.75	5.37
t = 8~10	6.40	3.77	3.85	3.75	3.73	3.62
t = 9~10	6.71	2.38	2.39	2.33	2.36	2.28
Average						
All	2.63	0.78	0.84	0.39	0.73	1.23
First 6 days	1.37	0.24	0.24	0.12	0.21	0.83
First 7 days	1.53	0.04	0.18	0.02	0.01	0.61
First 8 days	1.73	0.14	0.23	0.07	0.09	0.66
First 9 days	1.84	0.36	0.43	0.18	0.31	0.85
Last 4 days	4.53	1.59	1.74	0.80	1.50	1.84
Last 3 days	5.20	2.49	2.38	1.25	2.40	2.67
Std. Dev.						
All	2.60	1.64	1.47	0.82	1.62	1.49
First 6 days	0.40	0.72	0.54	0.36	0.72	0.67
First 7 days	0.56	0.84	0.51	0.42	0.85	0.83
First 8 days	0.76	0.82	0.50	0.41	0.83	0.78
First 9 days	0.80	1.01	0.76	0.51	1.01	0.92
Last 4 days	3.46	2.40	2.05	1.20	2.39	2.25
Last 3 days	3.90	1.92	1.96	0.96	1.90	1.85

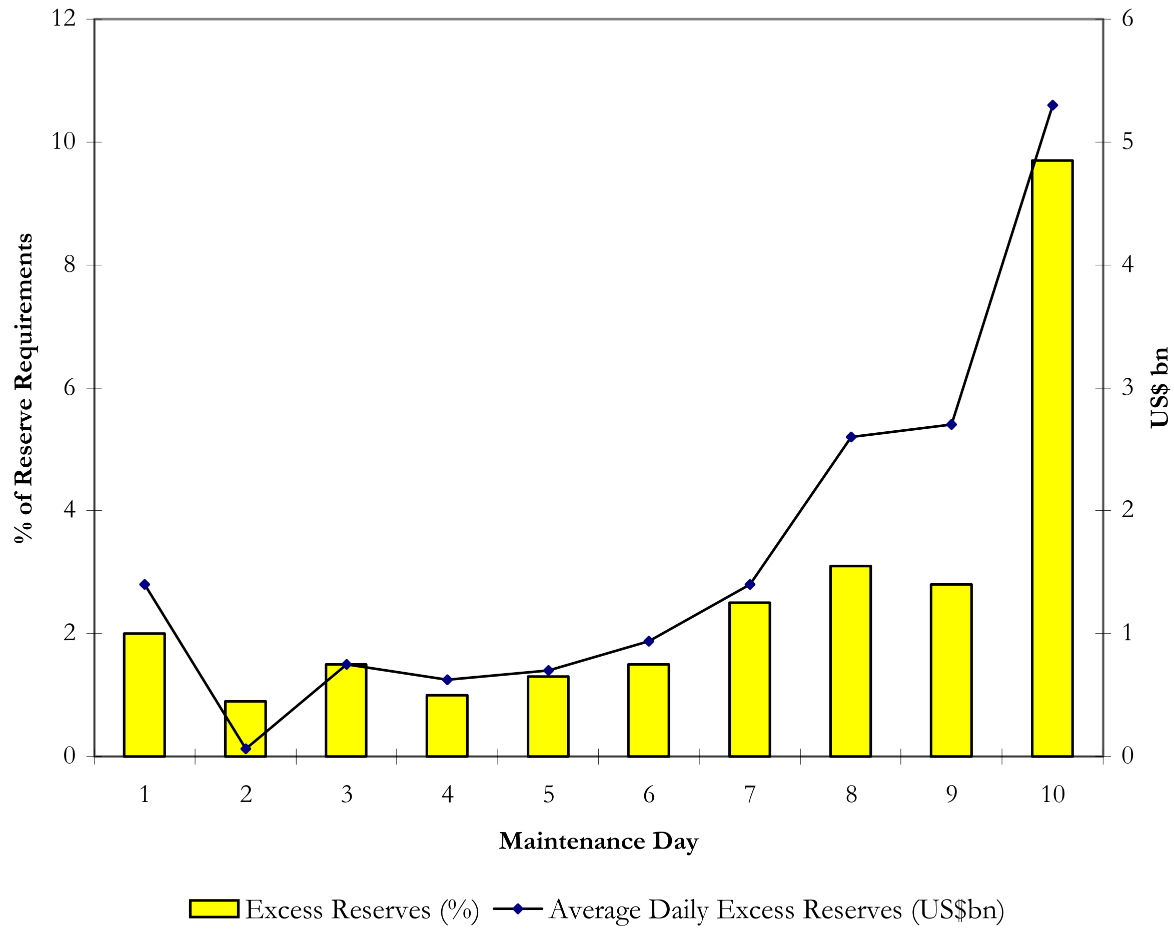
Table 2-2
Statistics for Uncertain Monetary Intervention with Constant μ $(m_t, \bar{\mu})$

	Observed	Base Case	$\mu = 70$	$\mu = 80$	$\mu = 100$	$\mu = 110$
Spike (%)						
t = 2~3	0.59	1.05	0.17	0.28	0.52	0.55
t = 7~10	7.02	5.77	8.84	7.47	6.54	6.42
t = 8~10	6.40	3.77	6.02	5.29	4.78	4.69
t = 9~10	6.71	2.38	3.17	3.13	3.10	3.06
Std. Dev.						
All	2.60	1.64	2.46	1.97	1.67	1.64
First 6 days	0.40	0.72	1.34	0.98	0.76	0.73
First 7 days	0.56	0.84	1.61	1.35	1.23	1.20
First 8 days	0.76	0.82	1.50	1.25	1.16	1.14
First 9 days	0.80	1.01	1.71	1.32	1.14	1.11
Last 4 days	3.46	2.40	3.69	3.11	2.73	2.68
Last 3 days	3.90	1.92	3.01	2.65	2.41	2.37
Differences from Mean						
1	-0.63	0.49	1.43	1.36	1.25	1.20
2	-1.73	-1.74	-0.02	-0.19	-0.33	-0.34
3	-1.13	-0.70	0.15	0.09	0.19	0.21
4	-1.63	-0.36	-0.82	-0.39	0.03	0.05
5	-1.33	-0.36	-2.05	-1.18	-0.58	-0.53
6	-1.13	-0.57	-1.97	-1.34	-0.93	-0.88
7	-0.13	-1.91	-3.32	-2.94	-2.74	-2.70
8	0.47	-0.01	-0.73	-0.92	-1.10	-1.08
9	0.17	1.36	2.03	1.17	0.52	0.49
10	7.07	3.79	5.28	4.35	3.65	3.58

Table 2-3
Statistics for Uncertain Monetary Intervention with Time-Varying μ (m_t, μ_t)

	Observed	Base	$\mu_{0\sim6}=80$ $\mu_{7\sim9}=50$	$\mu_{0\sim8}=80$ $\mu_9=50$	$\mu_{0\sim6}=110$ $\mu_{7\sim9}=50$	$\mu_{0\sim8}=10^4$ $\mu_9=40$	$\mu_{0\sim8}=10^4$ $\mu_9=30$
Spike (%)							
t = 2~3	0.59	1.05	0.37	0.31	0.63	1.10	1.12
t = 7~10	7.02	5.77	7.72	8.03	6.74	8.31	10.27
t = 8~10	6.40	3.77	5.87	5.85	5.12	6.17	8.05
t = 9~10	6.71	2.38	3.59	3.64	3.56	4.73	6.60
Std. Dev.							
All	2.60	1.64	2.01	2.10	1.72	2.33	2.89
First 6 days	0.40	0.72	0.98	1.01	0.74	0.65	0.66
First 7 days	0.56	0.84	1.51	1.42	1.29	0.78	0.80
First 8 days	0.76	0.82	1.45	1.32	1.24	0.78	0.80
First 9 days	0.80	1.01	1.42	1.37	1.18	1.02	1.02
Last 4 days	3.46	2.40	3.24	3.34	2.82	3.50	4.39
Last 3 days	3.90	1.92	2.94	2.95	2.61	3.27	4.34
Differences from Mean							
1	-0.63	0.49	1.46	1.36	1.22	-0.07	-0.26
2	-1.73	-1.74	0.10	-0.12	-0.23	-2.01	-2.20
3	-1.13	-0.70	0.47	0.19	0.40	-0.92	-1.09
4	-1.63	-0.36	0.02	-0.34	0.23	-0.56	-0.73
5	-1.33	-0.36	-0.85	-1.21	-0.43	-0.55	-0.72
6	-1.13	-0.57	-1.30	-1.42	-0.89	-0.83	-1.02
7	-0.13	-1.91	-3.25	-3.11	-2.85	-2.17	-2.40
8	0.47	-0.01	-1.55	-1.11	-1.34	-0.18	-0.37
9	0.17	1.36	0.63	1.01	0.15	1.20	1.00
10	7.07	3.79	4.26	4.71	3.73	6.04	7.74

Figure 2-1
Excess Reserves (1996~2003)



Sources: Bartolini, Bertola and Prati (2001), and Carpenter and Demiralp (2004).

Figure 2-2
The Demand for Excess Reserves

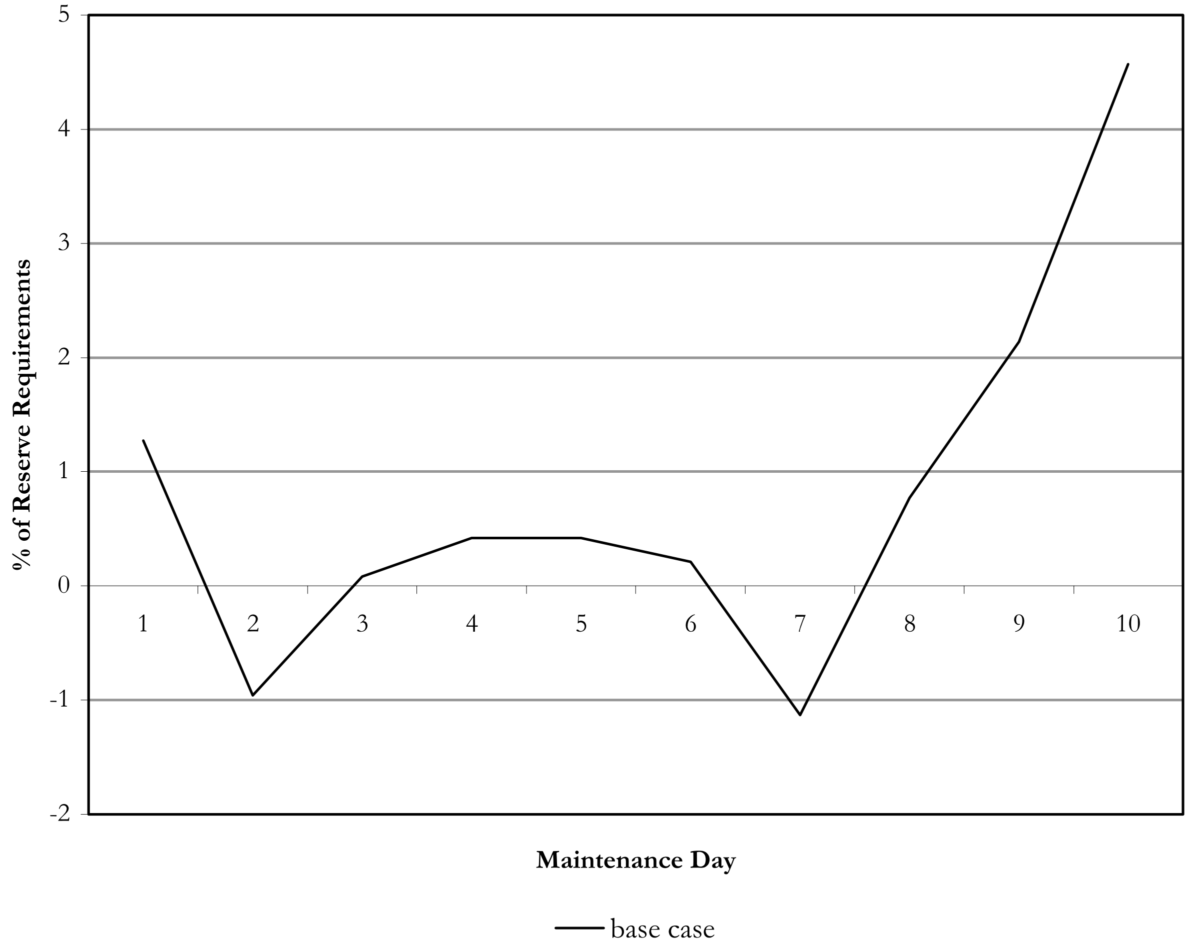


Figure 2-3
Equal Daily Overdraft Penalty Structure

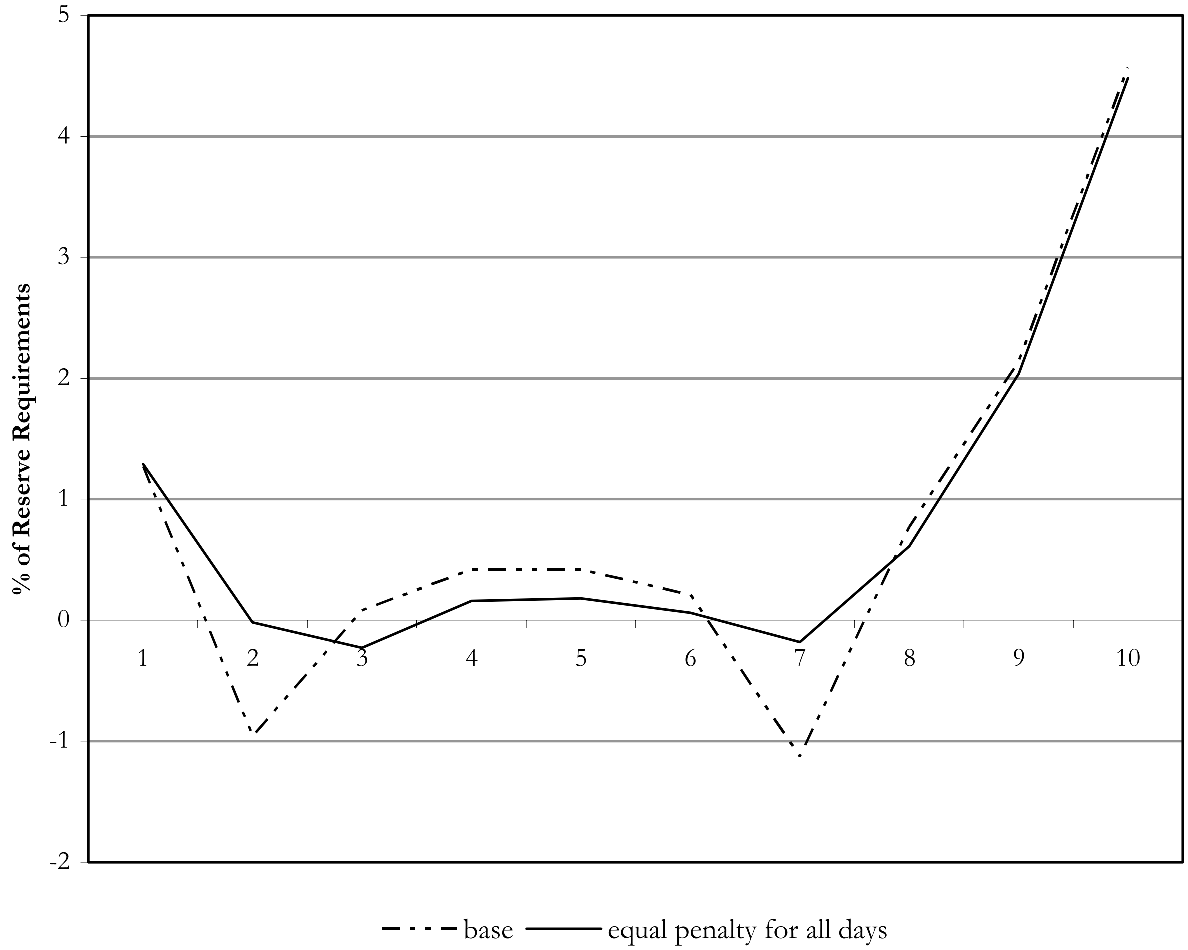


Figure 2-4
The Effect of Decreasing Reserve Requirements

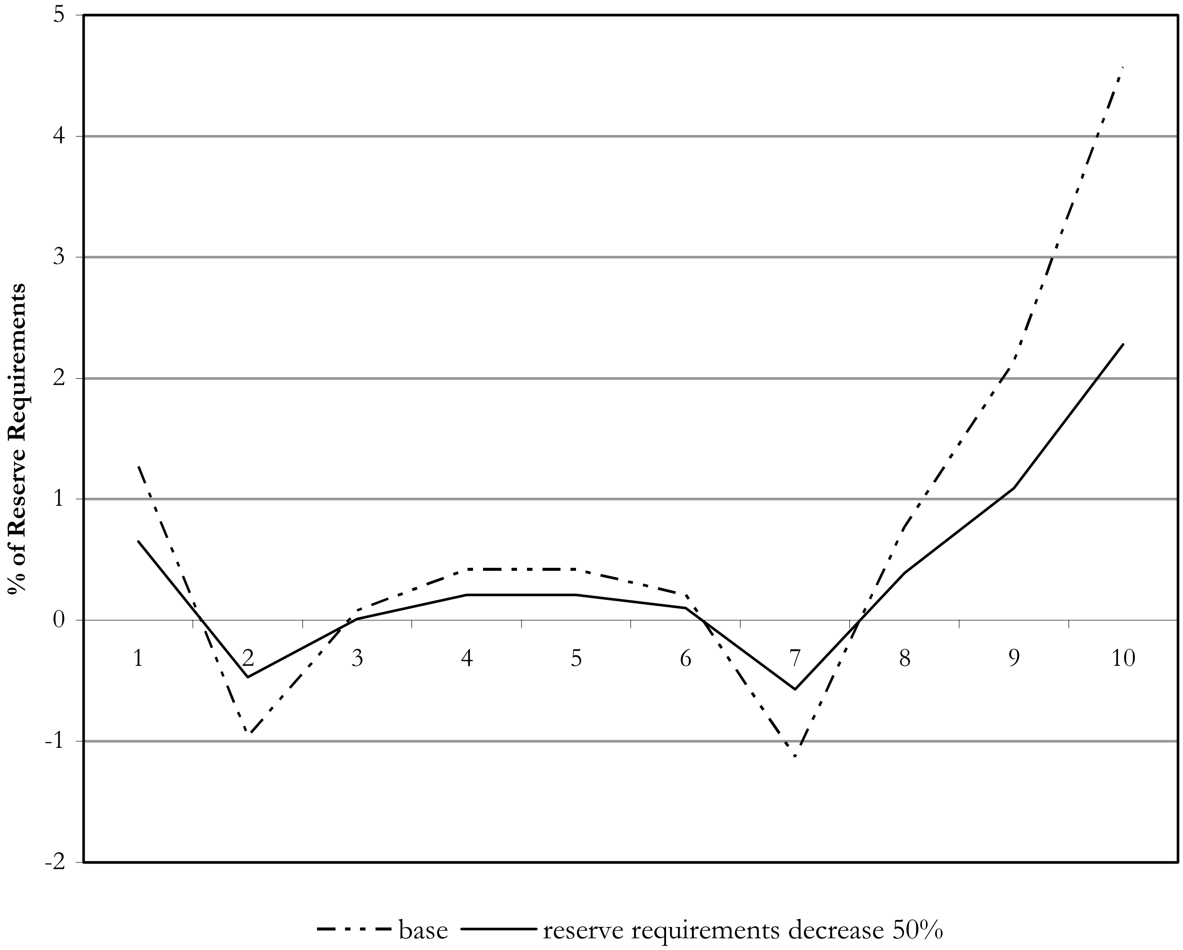


Figure 2-5
Paying Interest on Reserves

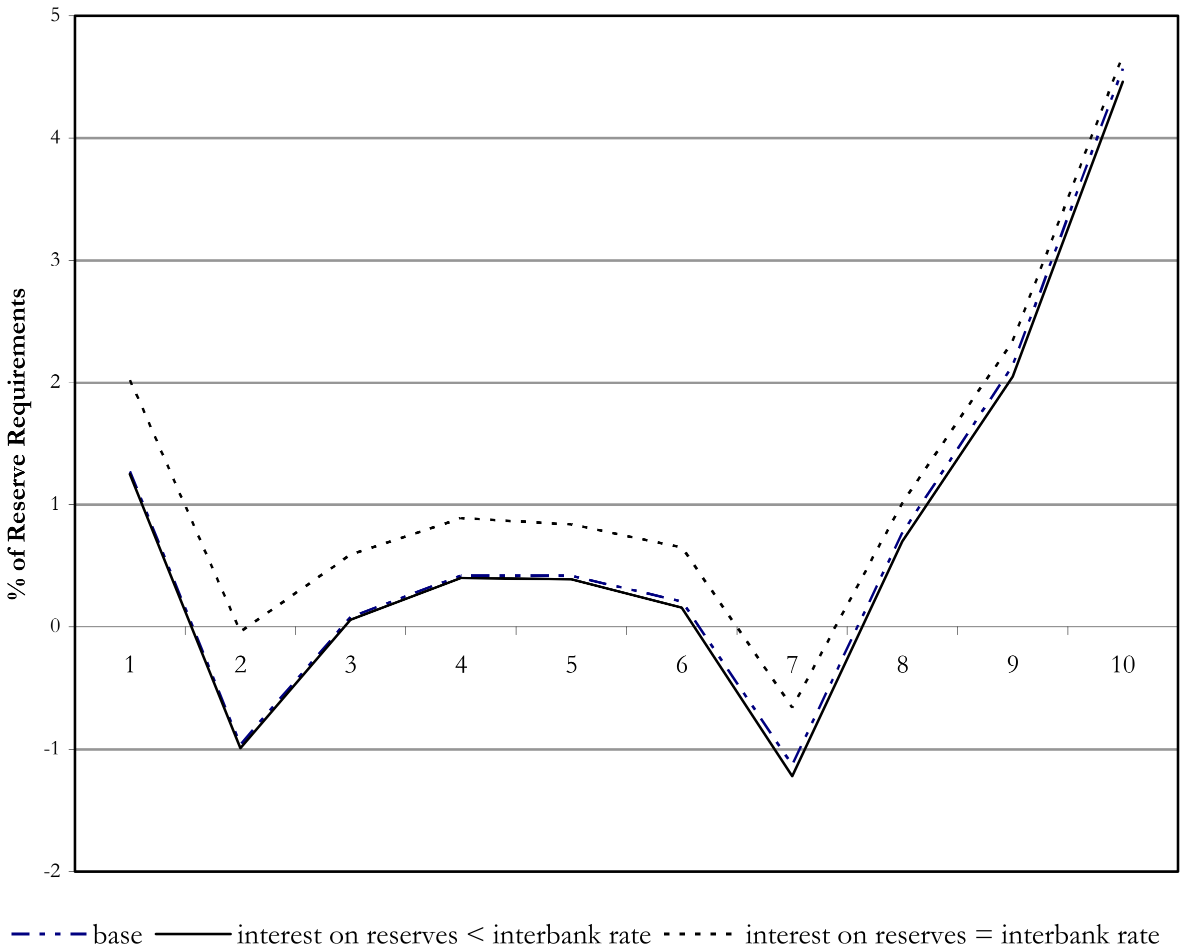
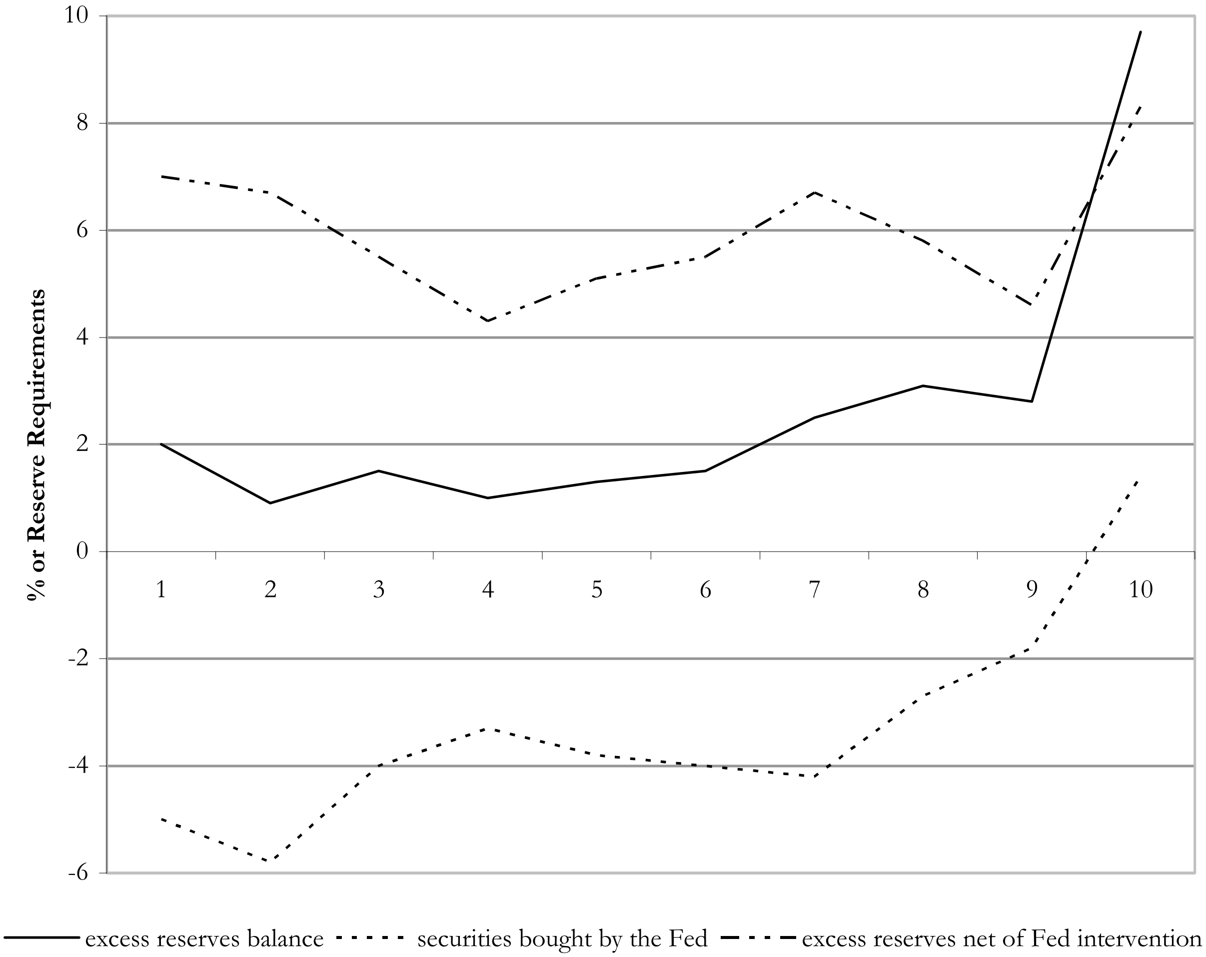


Figure 2-6a
Excess Reserves and Fed's Intervention



Source: Bartolini, Bertola and Prati (2001).

Figure 2-6b
Excess Reserves and the Monetary Authority's Intervention

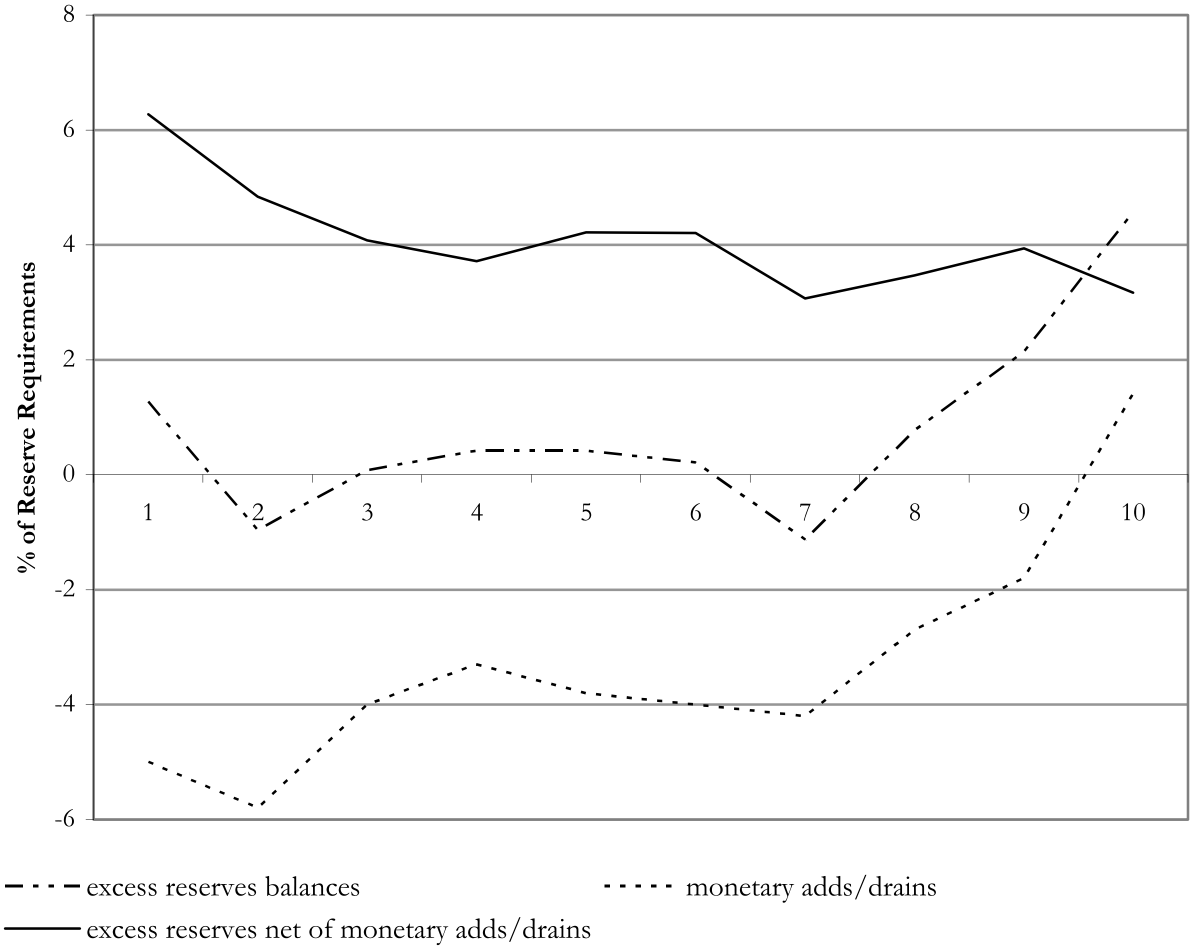


Figure 2-7
Augmented Model - Uncertain Monetary Intervention

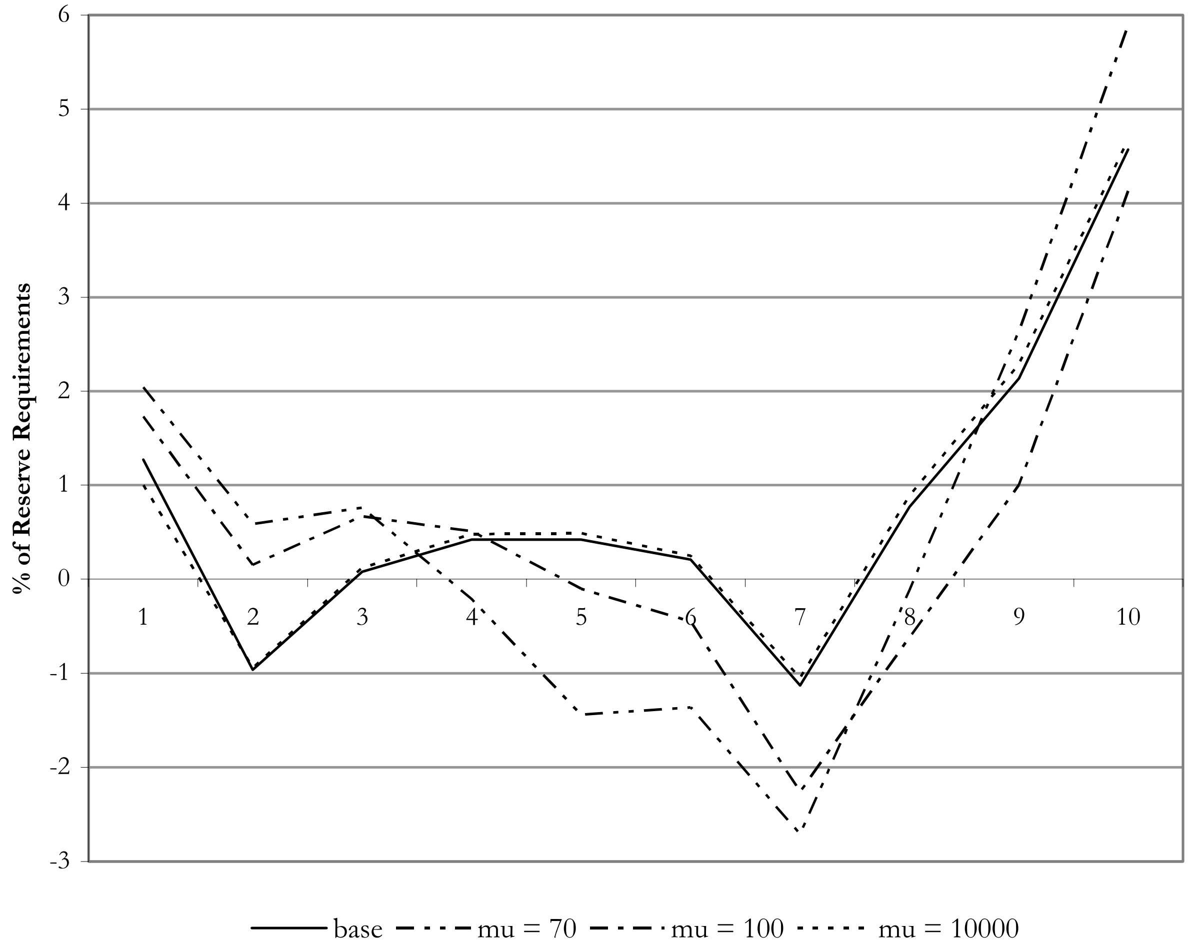


Figure 2-8
Augmented Model - Comparisons
Uncertain Monetary Intervention with Constant *mu*

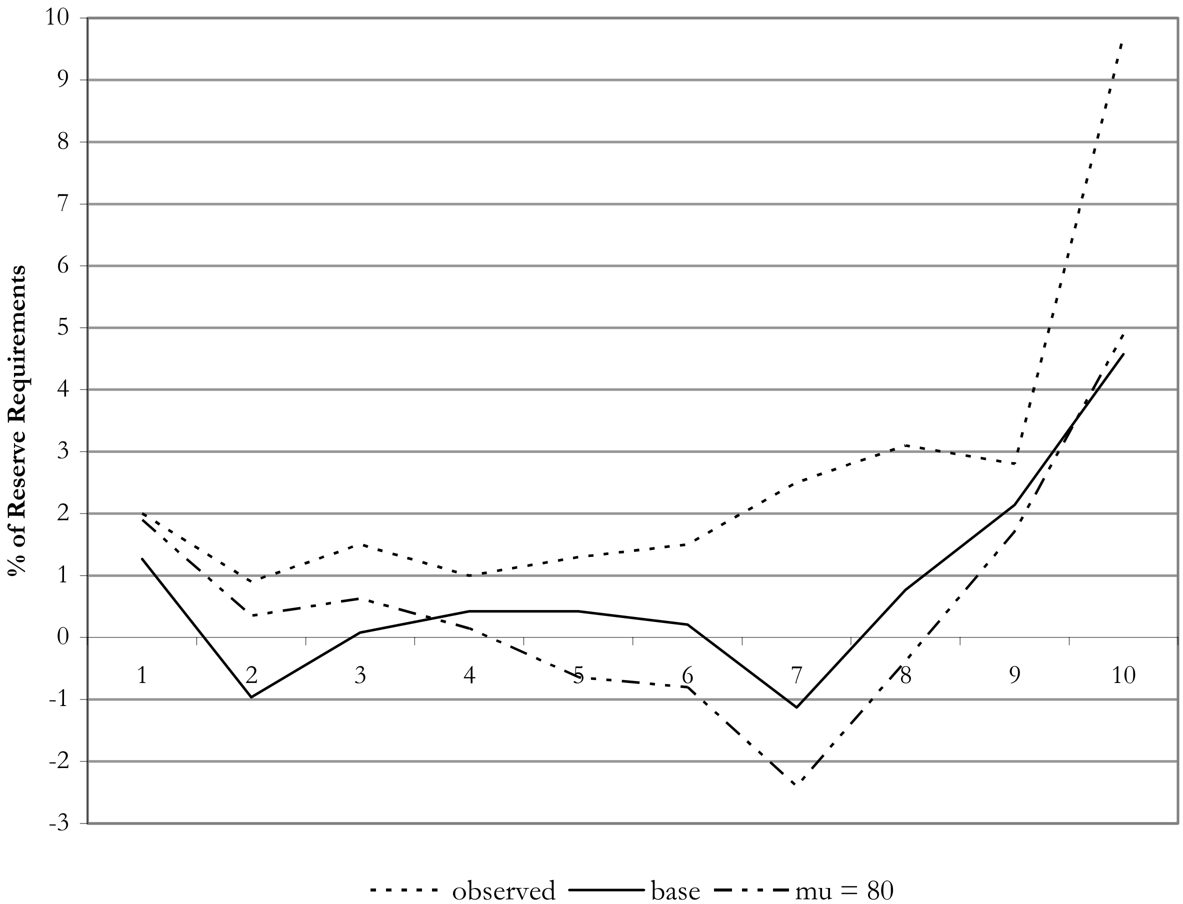
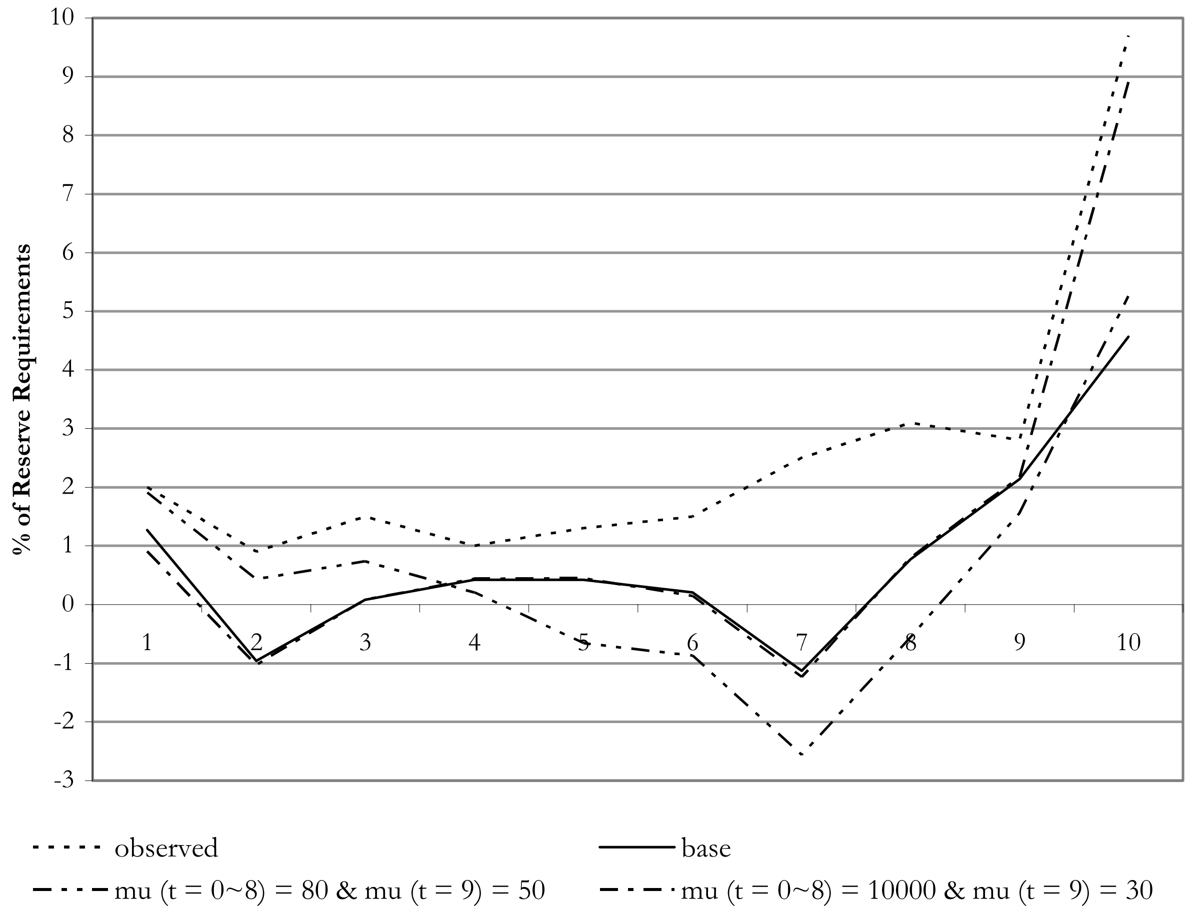


Figure 2-9
Augmented Model -
Uncertain Monetary Intervention with Time-Varying *mu*



Chapter 3

Credit Rationing and Intermediation

3.1 Introduction

It is not money that makes the world go round, but credit.

... Joseph Stiglitz [1]

Credit is central to any economy where investment performs a pivotal role, and in which entrepreneurs differ in their inherent ability to carry out the investment projects. Heterogeneities among individual aptitudes and technological aspects imply that concentration of the resources in the economy to a limited number of investment plans is potentially advantageous. That is, it might be beneficial for a provider of resources such as credit to ration some, but not other, agents seeking them. However, is this true if the economy is not wealth-constrained, all investment projects positive expected gross returns, and there is universal risk neutrality?

Given such an economy, in this essay I address some further consequences of rationing in the credit markets with imperfect information under different assumptions with regard to investment technologies available to entrepreneurs applying for a loan and the structure of the loan market. Firstly, I examine the characteristic of rationing if a separating

[1] The Economic Record, December 1988.

equilibrium in which rationing occurs exists. Specifically, what are the features of credit rationing when there are two self-selection instruments, namely, the probability of loan granting and the size of loan? Secondly, in view of the issue of financial liberalization in the developing countries, I investigate what is the relationship between credit market structure and equilibrium credit rationing. In particular, I compare the nature of credit rationing between a competitive and a monopolistic intermediated loan markets.

The model here considers both ex-ante and ex-post informational asymmetries. That is, there are both adverse selection and ex-post verification problems in the economy. Ex-post verification problem is employed in order to derive the optimal financial arrangement endogenously. Following seminal papers by Diamond (1984) and Gale and Hellwig (1985), the optimal contract is shown to be a standard debt contract. Then, I proceed to show that credit rationing arise endogenously in equilibrium. The environment is closest in spirit to Boyd and Smith (1992) who also consider both ex-ante and ex-post informational asymmetries, and credit rationing arise endogenously in their model. However, the intermediary in their model is an endogenous response to comparative locational dominance in monitoring, whilst here the advent of intermediaries is motivated by the advantage in monitoring cost saving.

Following Keeton (1979), there are two types of rationing, namely, (i) Type-1 rationing occurs when there is a partial or complete rationing of all the borrowers within a given group. It is also known as loan size rationing. (ii) Type-2 rationing occurs within a group of borrowers that is ex-ante indistinguishable from the lender's point of view, so that ex-post some borrowers of this group obtain the loan they demand fully while others are completely denied any loan. It is also known as loan granting or loan quantity rationing.

The analysis shows that if rationing occurs, the equilibrium is characterized by either type-1 or type-2 rationing (but not both). In the former, the probability of obtaining loan is one, but credit rationing occurs in the sense that the low-risk borrowers cannot get their unconstrained investment loan. In the latter, the low-risk borrowers can get their unconstrained quantity of loan, but the probability of acquiring loan becomes less than one. These differences are due to the heterogeneity in the borrowers' investment project technologies. For instance, a type-2 rationing occurs if the high-risk investment technology is sufficiently superior to the low-risk investment technology. This is the case when the low-risk borrower's production technology second-order stochastically dominates that of the high-risk borrower.[2] This in turn tends to aggravate the informational frictions further, which means that it may not be socially optimal to carry out all projects by employing type-1 rationing. In order to counter this effect, lenders choose type-2 rationing even though an investment project is divisible.[3]

Next, I demonstrate that a competitively intermediated loan market has only quantitative effect over direct lending. That is, while the type-2 rationing remains, the level of rationing under competitive intermediaries is less than that under direct lending. This is due to the saving in verification cost under intermediation: the portion of credit rationing attributed to ex-post informational problem tends to get smaller with the emergence of intermediation.[4] In contrast, the loan market populated by a monopolistic intermediary has both quantitative and qualitative effects. Although the high-risk borrowers with the superior production technology are still free from being rationed, the low-risk borrowers with the inferior investment project are also being financed by the lenders. This is due to the fact that the monopolistic intermediary absorbs all net social surplus accrued by the borrowers' investment activities, and hence it has the incentive to finance all projects as long as the expected net social surplus is sufficiently large.

The plan of the paper is as follows. Section 2 describes the environment. Section 3 considers direct lending relationship between lenders and creditors under public and private information. The case when there exist endogenous intermediaries in the loan

[2]That is, I assume that borrowers' production technology risk is characterized by the fact that the superior technological advantage of the high-risk borrowers necessarily implies that their probability of success is much lower *vis-à-vis* the technology of the low-risk borrowers, even though their respective expected returns are equal.

[3]Intuitively, since there is also a fixed verification cost, lenders or financial intermediaries prefer to employ type-2 rationing in order to minimize the expected verification cost.

[4]At the same time it also elucidates the fact that adverse selection is the fundamental element causing the imperfection in credit market, justifying the constant attention it receives in the literature.

market is analyzed in Section 4. In order to compare the market structure of the loan market, Section 5 considers a loan market that is intermediated by a monopolistic intermediary. Section 6 concludes. All proofs are in the Appendices.

3.2 The Environment

The size of the population is countably infinite, and every agent is indexed $s = 1, 2, 3, ..., \infty$. The agents in the economy comprises of risk-neutral lenders and borrowers. Each individual has an early-of-period endownment w. In addition, a borrower also owns an investment project that provides stochastic payoffs. A lender has the choice of either (i) saves w in a riskless storing technology with a real gross rate of return, $r > 1$, or (ii) loans w to borrowers at the loan interest rate R.

Borrowers consist of two types: those who are endowed with a relatively risky investment project and those who are endowed with a relatively safer investment plan. These debtors are identified by the projects that they are endowed with. The fraction of borrowers who own high-risk investment projects (type-H) is α, while the fraction of borrowers who own low-risk investment plans (type-L) is $(1-\alpha)$. Each borrower has the following options: (i) saves w in a riskless storing technology with a rate of return of 1, or (ii) invests w in her endowed investment plan. However, in case (ii), she has to seek an additional loan of the size q_i from lenders to carry out the project, where $i = \{H, L\}$. A loan of the size q_i produces a payoff of $k_i = \zeta_i f_i(q_i) q_i$ with probability P_i, and zero with probability $(1 - P_i)$, with $P_H < P_L$. Assume that for $\forall q_i$,

$$\zeta_i[f_i(q_i) + f_i'(q_i)q_i] > 0,$$

and

$$\zeta_i[2f_i'(q_i) + f_i''(q_i)q_i] < 0.$$

The project's realization, $\widetilde{k}_i$, is a random variable which is ex-post observable but costly to verify. The borrower announces $k_i^d = \{0, k_i\}$ to the lender at the end of period t. The lender then has to decide whether to verify k_i^d or not. Verification decision is ex-post non-stochastic and entails a cost of Cq_i which is paid by the lender.[5] In order to generate endogenous credit rationing in one of the cases below, let's assume that the safer investment project second order stochastically dominates the risky investment project. That is, for $\forall q_i$,

$$P_H \zeta_H f_H(q) = P_L \zeta_L f_L(q). \tag{3.1}$$

The values of α, P_H and P_L are assumed to be common knowledge. In contrast, the type of a borrower, $i = \{H, L\}$, and the realization of $\widetilde{k}_i$ are known only to the borrowers. Therefore, there are ex-ante informational asymmetry due to i and ex-post informational asymmetry due to $\widetilde{k}_i$, which give rise to an adverse selection and an ex-post verification problems respectively.[6] In the credit market, I assume that the demand for loan is not constrained by the amount of the economy's loanable wealth, i.e. $q < w$.[7]

With regard to the ex-post verification problem, a lender will have to decide at the end of the loan period whether to verify k_i^d or not. Since $\widetilde{k}_i$ can only take on two values, it is straightforward to see that no verification occurs when $k_i^d = k_i$. When $k_i^d = k_i$, the borrower pays a total of $T(k_i)$ to her lenders. Since no borrower will ever choose to pay the lender more than the minimum amount necessary to prevent verification, $T(k_i)$ must be the minimum of all possible repayment schemes when $k_i^d = k_i$. Thus, $T(k_i)$ is a non-contingent constant and the interest rate on loan can be defined as $R_i = \frac{T(k_i)}{q_i}$. Lenders monitor whenever a borrower claims she is unable to repay $R_i q_i$. Otherwise, borrowers will have an incentive to default and keep the proceed from her investment all to herself even when the project is successful. That is, when $k_i = 0$, verification will be carried out and a borrower repays her lenders $R_i(0)q_i$.

[5]The verification cost, Cq_i, is often considered in the literature as a monitoring or bankruptcy cost.

[6]For simplicity, I assume that P_i is not a choice variable so that incentive problem due to increases in loan interest rate (price effect) is abstracted away. That is, there is no moral hazard problem.

[7]Equivalently, this assumption states that the number of possible lenders far outnumbers the number of borrowers. Thus, in the absence of asymmetric information, credit will not be rationed.

I consider the Rothschild-Stiglitz's type-wise break-even separating contract and restrict attention to cases in which it exists.[8] The loan granting process is divided into two stages. In stage-1, given contract announcements by other lenders at period t, a lender announces loan contracts $\{q_i, R_i, R_i(0), \pi_i\}$, where q_i is the loan size ($q_i > 0$), R_i is the non-contingent loan interest rate, $R_i(0)q_i$ is the repayment when $k_i = 0$, and π_i is the probability of granting credit to borrowers. The feasibility constraints for the loan interest rate and the probability of granting a loan are

$$R_i(0) \leq R_i q_i \leq \zeta_i f_i(q_i) q_i,$$

and

$$0 \leq \pi_i \leq 1 \tag{3.2}$$

respectively.

In stage-2, the borrowers will select among the announced loan contract. Since the fixed cost nature of Cq_i makes the diversification of loans costly, lenders choose to contract with as few borrowers as possible, and they will choose these borrowers randomly. Due to the concavity of the investment project production function, a borrower will only enter into a contract with a finite number of lenders. All activities in the credit market cease after stage-2, preventing those borrowers who are being rationed from joining the credit market as lenders.[9] I assume that all contracts are enforceable.[10]

[8]Rothschild and Stiglitz (1976) argue that there may not exist any pooling equilibrium in which all borrowers receive the same contract since such a contract is always vulnerable to contracts that attract the low-risk borrowers away. They also show that there could only be one separating equilibrium which give complete insurance to the high risk borrowers. However, this separating equilibrium may not be stable if there are relatively few high-risk borrowers *vis-à-vis* low-risk borrowers. This is because it is possible that cross subsidization which makes both types of borrowers better off could be achieved through a pooling contract.

[9]Therefore, different types of borrowers will perceive different costs of being denied credit. This ensures that an equilibrium exists.

[10]Even if $\pi_{it} < 1$ is derived as the optimal contract, a lender will have the incentive to always offer a loan when a borrower submits his loan application. This is because the expected return from the investment project is above the market safe interest rate (see below). To avoid such potential inconsistencies, contracts are assumed to be enforceable.

As a result, a type-i borrower is denied credit with probability $(1-\pi_i)$ and is granted credit with probability π_i. Those who are denied credit store their endowment and consume w. Those who receives loan extension will invest in their investment plans, yielding the expected payoffs $P_i\left[\zeta_i f_i(q_i) - R_i\right] q_i - (1-P_i)R_i(0)q_i$. Her expected utility then, is

$$E\left[U_i^{DL}\right] = \pi_i\left\{P_i\left[\zeta_i f_i(q_i) - R_i\right] q_i - (1-P_i)R_i(0)q_i\right\} - (1-\pi_i)w.$$

In order to ensure self selection, the equilibrium must satisfy the following incentive compatibility(IC) constraints:

$$\begin{aligned}&\pi_i\{P_i[\zeta_i f_i(q_i) - R_i]q_i - (1-P_i)R_i(0)q_i - w\}\\ &\quad \geq \pi_j\{P_i[\zeta_i f_i(q_j) - R_j]q_j - (1-P_i)R_j(0)q_j - w\},\end{aligned}$$

for $i, j = \{H, L\}$, $i \neq j$. That is, when a type-i borrower mimics a type-j and gets a type-j contract, her amount of gain must not exceed her amount of loss from changing her contract.

If $k_i = 0$ and verification occurs, a type-i borrower pays $R_i(0)$. However, due to resource constraint, $R_i(0)$ cannot be positive. For feasibility, $R_i(0)$ must be less than or equal to w. If $R_i(0) < 0$, then it is possible to raise $R_i(0)$ whilst lowering R_i such that a borrower's expected repayment, $P_iR_i + (1-P_i)R_i(0)$, remains the same. This leaves the borrower no worse off, but yields a gain to the lenders by reducing his expected monitoring costs. Thus, $R_i(0) = 0$ holds. That is, a borrower surrenders whatever she produced (in this case, zero) to the lender when verification occurs. Together with the non-contingent payment R_i, this repayment scheme constitutes a standard debt contract.[11,12,13] Therefore, the loan contract can be simplified to $\{q_i, R_i, \pi_i\}$. The expected payoffs for a type-i borrower is

$$E[U_i^{DL}] = \pi_i\{P_i[\zeta_i f_i(q_i) - R_i]q_i\} + (1-\pi_i)w, \tag{3.3}$$

while that for the lender is

$$E[V_i^{DL}] = \pi_i q_i \left[P_i R_i - (1-P_i)C - r\right] \geq 0. \tag{3.4}$$

The feasibility constraint for the loan interest rate is now

$$0 \leq R_i \leq \zeta_i f_i(q_i). \tag{3.5}$$

Finally, let's assume that a borrower prefers to carry out the investment project than using the storage technology. That is,

$$E\left[U_i^{DL}\right] > w \tag{3.6}$$

[11] It is difficult to see the similarity when there are only two possible realizations and when the failure outcome is zero. However, the likeness is clearer when there are many possible returns.

[12] When there is only ex-post verification problem, the only consideration in selecting a contract is the minimization of the expected verification cost. Under this restriction, Diamond (1984), Gale and Hellwig (1985) and Williamson (1986) derive the optimality of the standard debt contract. However, when adverse selection problem exists alongside the ex-post verification problem, the ex-ante nature of adverse selection may provide incentives for low-risk borrowers to depart from pure debt contract as a mean to signal their type (which involves more complicated payment schedules than would be observed under the standard debt contract). Borrowers will then compare the cost incurred and the gain received from this departure from the standard debt contract. When these gains (e.g. a lower loan interest rate) outweigh the costs, the standard debt contract may not be an optimal arrangement.

However, Boyd and Smith (1993) provide some necessary conditions for the optimality of the standard debt contract under such an environment. Specifically, if the verification cost is sufficiently large, then the standard debt contract dominates any deviation from it. Furthermore, Wijkander (1992) makes the general claim that when the verification cost is sufficiently large, a standard debt contract is locally optimal.

[13] In the present model, the difference among borrowers' payoffs only occurs in the good state. However, in this good state there is no incentive problem. This suggests that there is no need for a low-risk borrower to engage in signaling. It also implies that it makes no difference whether it is the lender or the borrower that offers the loan contract.

holds.

3.3 Direct Lending

3.3.1 Full Information

As a benchmark, I first examine a competitive loan market equilibrium under full information about borrowers' types. Lenders optimize borrower's expected payoffs (3.3) subject to his participation constraint (3.4) and the feasibility constraints (3.2) and (3.5). *Contract 1* summarizes the equilibrium contract (refer to Appendix 1 for the proof).

Contract 1:

$$\begin{aligned}
&(i) \quad R_i^{\dagger} = \tfrac{r}{P_i} + \tfrac{(1-P_i)C}{P_i}, \\
&(ii) \quad q_i^{\dagger} = \underset{q_i}{arg\ max}\ \pi_i\{P_i[\zeta_i f_i(q_i) - R_i]q_i - w\}, \\
&(iii) \quad \pi_i^{\dagger} = 1,
\end{aligned}$$

Define $n = \frac{q_L^{\dagger}}{w} << \infty$ as the number of lenders a low-risk borrower contracts with in the full information economy.

The quoted loan interest rate is made up of three items: (i) the market rate of return, r, (ii) a default risk premium, $\frac{1}{P_i}$, and (iii) the expected average verification cost, C. The risk premium reflects the fact that only non-defaulting contracts actually pay R_i to the lenders. $R_H > R_L$ holds, since $P_H < P_L$. This reflects the fact that under competitive loan market, borrowers receive the entire social surplus. No rationing occurs in equilibrium and borrowers receive the loan quantity that maximizes output from their investment projects. That is, the lenders prefer to finance all borrowers' investment projects because their expected returns, by (3.6), are non-negative. Since lenders know a borrower's type perfectly, they design contracts that are acceptable to both parties. Thus, when information is symmetrically shared, lenders are always better off by approving all loan applicants.

3.3.2 Private Information

Since $R_H > R_L$, a high-risk borrower has an incentive to misrepresent herself as a low-risk borrower. It is straightforward to show that *Contract 1* is not incentive compatible under private information. In order to avoid trivial cases, assume that

$$q_i \leq nw \tag{3.7}$$

where $n = \frac{q_L^\dagger}{w} << \infty$ is the number of lenders a low-risk borrower contracts with in the full information economy. The optimization problem for the lender is now:

Program 2

$$\begin{aligned}\max_{q_i, R_i, \pi_i} \quad & \alpha \left\{\pi_H \left(P_H[\zeta_H f_H(q_H) - R_H]q_H\right) + (1-\pi_H)\, w\right\} \\ & \quad + (1-\alpha)\left\{\pi_L \left(P_L[\zeta_L f_L(q_L) - R_L]q_L\right) + (1-\pi_L)\, w\right\},\end{aligned}$$

s.t

$$\pi_i q_i [P_i R_i - (1-P_i)C - r] \geq 0,$$

$$\begin{aligned}&\pi_H\{P_H[\zeta_H f_H(q_H) - R_H]q_H - w\} \\ &\quad \geq \pi_L\{P_H[\zeta_H f_H(q_L) - R_L]q_L - w\},\end{aligned} \tag{3.8}$$

$$\begin{aligned}&\pi_L\{P_L[\zeta_L f_L(q_L) - R_L]q_L - w\} \\ &\quad \geq \pi_H\{P_L[\zeta_L f_L(q_H) - R_H]q_H - w\},\end{aligned} \tag{3.9}$$

$$0 \leq R_i \leq \zeta_i f_i(q_i),$$

$$0 \leq \pi_i \leq 1,$$

$$q_i \leq nw,$$

where (3.8) and (3.9) are the incentive compatibility constraints for the high-risk and low-risk borrowers respectively and $i = \{H, L\}$.[14] Following Rothschild and Stiglitz (1976), for a Nash equilibrium, it must be the case that: (i) all equilibrium must display self selection, (ii) contracts earn lenders zero expected profits in equilibrium, and (iii) given the contracts received by type-j borrowers, the contracts received by type-i borrowers must be maximal for them among the set of all contracts that satisfy (ii) and self selection conditions (3.8) and (3.9). Condition (ii) implies

$$R_i^{\dagger\dagger} = \frac{r}{P_i} + \frac{(1 - Pi)C}{Pi},$$

for $i = \{H, L\}$.

The high-risk borrower's incentive compatibility constraint, (3.8), binds in the equilibrium. Since there are two unknown variables (i.e. π_L and q_L) in (3.8), the equilibrium values for π_L and q_L are not independent of each other. For any value of π_L within a certain range, there is a corresponding value of q_L which makes (3.8) binds, and vice-versa. This means q_L itself is a function of π_L. The following proposition summarizes the two extreme equilibria (see Appendix 2 for the proof).

Proposition 5 *(a)*

$$If \quad \left.\frac{\partial \pi_L}{\partial q_L}\right|_H < \left.\frac{\partial \pi_L}{\partial q_L}\right|_L, \quad then \quad \pi_L^{\dagger\dagger 1} = 1 \quad and \quad q_L^{\dagger\dagger 1} < nw.$$

That is, loan-size (type-1) rationing occurs.

(b)

$$If \quad \left.\frac{\partial \pi_L}{\partial q_L}\right|_H > \left.\frac{\partial \pi_L}{\partial q_L}\right|_L, \quad then \quad \pi_L^{\dagger\dagger 2} < 1 \quad and \quad q_L^{\dagger\dagger 2} = nw.$$

[14] For simplicity, the maximand in Program 2 can be written as

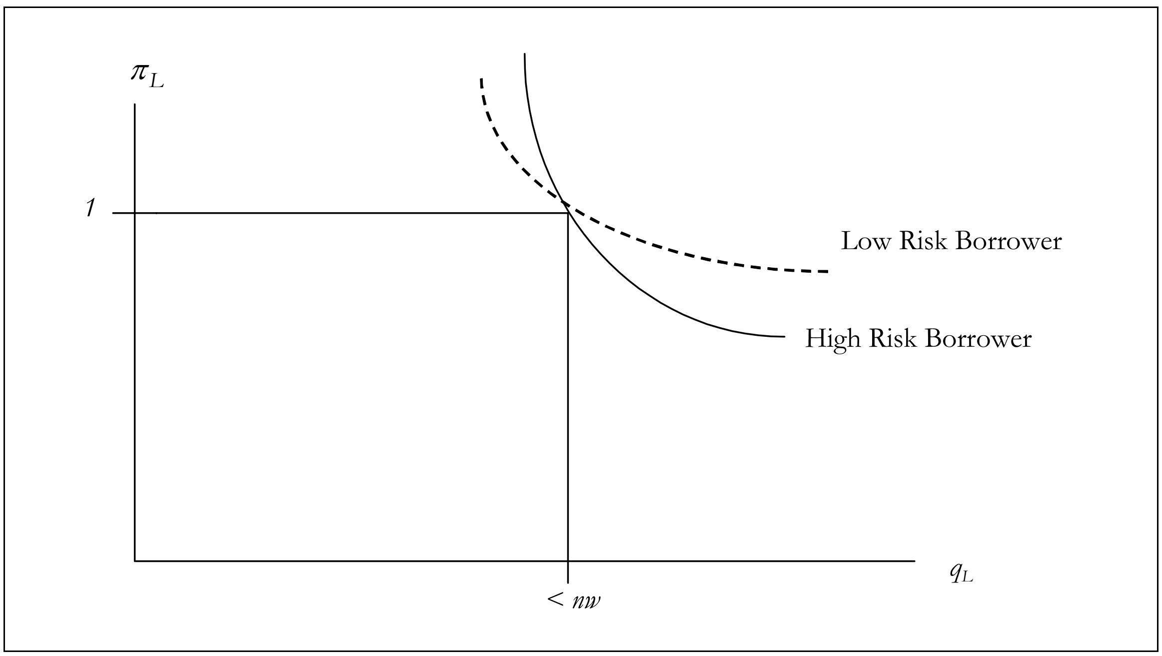
$$\begin{aligned} &\alpha\pi_H \{P_H[\zeta_H f_H(q_H) - R_H]q_H - w\} \\ &\quad + (1 - \alpha)\pi_L \{P_L[\zeta_L f_L(q_L) - R_L]q_L - w\}. \end{aligned}$$

This convention is followed thenceforth.

That is, loan-granting (type-2) rationing occurs.

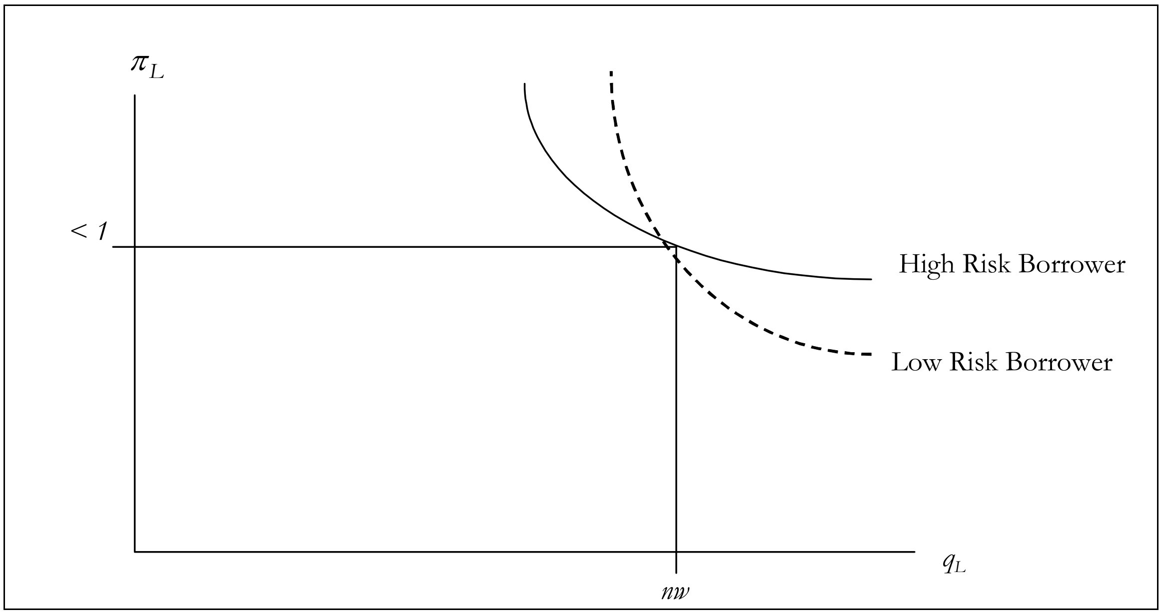
$\left.\frac{\partial \pi_L}{\partial q_L}\right|_H$ is the high-risk borrower's marginal rate of substitution from π_L to q_L, evaluated at $q_H^{\dagger\dagger} = q_L$, and defined it as (MRS_H). Similarly, $\left.\frac{\partial \pi_L}{\partial q_L}\right|_L$ is the low-risk borrower's marginal rate of substitution from π_L to q_L, evaluated at $q_L^{\dagger\dagger} = q_L$, and denote it as (MRS_L). When $MRS_H < MRS_L$, in order to get the same increase in π_L, a low-risk borrower is willing to pay more (i.e. willing to forgo some amount of loan) than a high-risk borrower would. That is, a low-risk borrower prefers a smaller amount of loan with a higher probability of obtaining loan to a larger amount of loan but with a lower probability of obtaining loan. Since (3.8) binds, if $\pi_L^{\dagger\dagger} = 1$, then $q_L^{\dagger\dagger}$ must be less than nw in order to recover the self selection condition. This case is given in *Figure 3-1.*

Figure 3-1
Type-1 Credit Rationing



On the other hand, when $MRS_{LB} < MRS_{HB}$, in order to get the same increase in π_L, a low-risk borrower is less willing to pay more than a high-risk borrower would. In other words, a low-risk borrower will prefer the larger amount of loan but with a lower probability of obtaining loan to the smaller amount of loan with a higher probability of obtaining loan. Again, since (3.8) holds with equality, if $q_L^{\dagger\dagger 2} = nw$, then $\pi_L^{\dagger\dagger 2}$ must be less than 1 in order to recover the self selection condition. This loan-granting (type-2) rationing is given in *Figure 3-2*.

Figure 3-2
Type-2 Credit Rationing



Next, holding R constant, differentiating borrowers' expected payoffs (3.3) w.r.t. P_i, and applying (3.1) gives

$$\left.\frac{\partial E[U_i^{DL}]}{\partial P_i}\right|_{\overline{R}} < 0. \tag{3.10}$$

This means that, as the loan interest rate increases, it is the low-risk borrowers who first exit from the loan market. Hence, the loan interest rate is not an efficient tool to sort borrowers.[15]

[15] Gale (1996) argues that in a loan market with adverse selection and ex-post verification problem, a commodity (loan) is not defined independently from its price (loan interest rate). An increase in R will cause low-risk borrowers to drop out first and attract high-risk borrowers into the loan market. Gale suggested that the loan size (q), the loan granting probability (π) and the loan interest rate (R) should not be defined as separate entities. Instead, the commodity *loan contract*, and not the loan itself should be considered as a commodity. Since R is embodied in the loan contracts which are traded, it must be ruled out as an equilibrium mechanism.

Furthermore, from (3.6), (3.10) and the investment project technology,

$$\frac{\partial\left(\partial U/\partial P\right)}{\partial\pi}<0$$

and

$$\frac{\partial\left(\partial U/\partial P\right)}{\partial q}<0$$

hold. This is the Spence-Mirless condition which enable sorting of agents types by means of type-1 or type-2 rationing.[16] Generally, in equilibrium, either type-1 or type-2 rationing, but not both is observed. However, under the second order stochastic dominance assumption (3.1), only type-2 rationing occurs.

Proposition 6 *If the high-risk borrower's expected payoff's elasticity w.r.t.* q_L *evaluated at* $q_H^{\dagger\dagger}=q_L$ *is less than the low-risk borrower's expected payoff's elasticity w.r.t.* q_L *evaluated at* $q_L^{\dagger\dagger}=q_L$*, then* $\pi_L^{\dagger\dagger}<1$ *and* $q_L^{\dagger\dagger}=nw$ *(see Appendix* 3 *for the proof).*

Intuitively, a low-risk borrower is subject to type-2 rationing if a high-risk borrower's investment project technology is sufficiently superior than that of the low-risk borrower's.

The equilibrium contract is given by

Contract 2:

$$\begin{aligned}&(i)\quad R_i^{\dagger\dagger}=\tfrac{r}{P_i}+\tfrac{(1-P_i)C}{P_i},\\&(ii)\;\; q_i^{\dagger\dagger}=nw,\\&(iii)\; \pi_H^{\dagger\dagger}=1;\quad \pi_L^{\dagger\dagger}=\frac{P_H\left[\zeta_H f_H(q_H)-\frac{r}{P_H}\right]-(1-P_H)C-\frac{1}{n}}{P_H\left[\zeta_H f_H(q_L)-\frac{r}{P_L}\right]-\frac{P_H(1-P_L)C}{P_L}-\frac{1}{n}}<1.\end{aligned}$$

As being mentioned above, in the equilibrium, (3.8) binds but (3.9) is slack.[17] That is, the contract $\{q_H, R_H, \pi_H\}$ for a type-H borrower is not affected by consideration of

[16]Intuitively, the Spence-Mirless condition is equivalent to the situation where the agents' indifference curves under the (π_L, q_L) plane crosses each other only once.

[17]Refer to Appendix 4 for the proof.

self selection.[18] Furthermore, a separating equilibrium exists only if

$$\alpha^{DL} > \frac{P_L}{P_L - P_H}\left\{1 - \frac{r + C}{(1-\pi_L)[P_L(\zeta_L f_L(q_L)) - \frac{1}{n}] + \pi_L r + P_L C}\right\},$$

where α^{DL} is the fraction of high-risk borrower in the economy under direct lending. This is equivalent to the situation in which there are relatively more high-risk borrowers than low-risk borrowers. This is because when the fraction of high-risk borrowers is sufficiently higher than that of the low-risk borrowers, pooling equilibrium is unattractive to low-risk borrowers.[19]

Clearly, the high-risk borrowers' expected payoffs under private information are the same as those under public information. However, the low-risk borrowers' expected payoffs under private information are less than the amount attained under public information. On the other hand, the low-risk borrowers' expected payoffs are higher than those of the high-risk borrowers' under both informational structures.

Note that credit rationing occurs among borrowers who are ex-ante indistinguishable. Since $\frac{\partial E[U_i^{DL}]}{\partial R_i} < 0$, there is a possibility that a high-risk borrower chooses the loan contract intended for a low-risk borrower's. Lenders respond to this incentive problem by reducing the probability of extending loan to the low-risk borrowers. Since $R_L < R_H$, the low-risk contract is still acceptable to the low-risk borrower. However, Besanko and Thakor (1987a) argue that the lower loan interest rate is of a lesser value to the high-risk borrower because the probability of her paying it is lower. Thus, the high-risk borrowers are coaxed away from the contract designed for the safer borrower. This fact is also implied by $\frac{\partial(\partial U/\partial P)}{\partial \pi} < 0$, which states that the low-risk borrower is less willing to pay relative to the high-risk borrower for a given increment in the probability of obtaining loan. Here, the probability of granting credit plays the role of an equilibrating

[18]This can also be seen from the fact that $\pi_{Ht}^{\dagger\dagger}$ and $q_{Ht}^{\dagger\dagger}$ do not depend on the low-risk borrower's parameters.

[19]This result, which is proved in Appendix 5, is similiar to that proved by Rothschild and Stiglitz (1976).

mechanism.[20]

The fact that there is no type-1 rationing seems to run counter to the established propositions on credit rationing in a competitive loan market. In Besanko and Thakor (1987b), a risk neutral lender will choose a type-1 rationing if a borrower's amount of collateral is insufficient. Furthermore, when lenders are risk averse, Schmidt (1997) argues that if (i) funding all projects is efficient, or (ii) funds are not scarce, or (iii) both, then there will be no type-2 rationing in equilibrium; only type-1 rationing will occur. Bencivenga and Smith (1993) avoided this particular problem by assuming linearity in the production function and project's indivisibility. However, neither play any role in the model here. Instead, since verification cost is assumed to be large and fixed, there is increasing return in financing an investment project fully. Therefore, it is not efficient to use type-1 rationing as a sorting tool. In the present model, the possibility of type-1 rationing exists from the outset. However, it does not occur in the equilibrium when investment project is characterized by (3.1).

3.4 Competitive Intermediated Loan Market

3.4.1 Existence of Intermediation

Thus far, the monitoring cost is assumed to be linear in the size of loan. Hence, intermediation does not matter. Now, instead, assume that the verification entails a fixed cost. Then, under direct lending, the equilibrium entails duplication of effort in that each debtor borrows from n lenders, and each lender verifies the debtor when she defaults. If a lender or a group of lenders act as an intermediary, however, then, this duplication of monitoring can be averted. When a single lender acts as an intermediary, he will contract with M independent borrowers and $M(n-1)$ lenders. The problem of duplication in monitoring will then shift from the lender-borrower relationship to the lender-intermediary

[20]See Gale (1996).

relationship, manifested in the form of delegation cost. However, this cost of delegation can be circumvented if the monitoring of the intermediary becomes unnecessary.

Following Diamond (1984) and Williamson (1986), the expected payoffs of the intermediary is

$$F = q_i \;\times\; \Sigma_{z=1,\ldots M} \;\; R_{i,z},$$

where $i = \{H, L\}$. Invoking the weak law of large numbers, it can be shown that

$$\underset{M\to\infty}{plim}\frac{1}{Mq}F = \alpha P_H R_H + (1-\alpha)P_L R_L.$$

If the number of defaulting borrowers is N, then the intermediary incurs a monitoring cost of the amount NC. Again from the law of large numbers

$$\underset{M\to\infty}{plim}\frac{1}{Mq}NC = \frac{C}{q}\left[\alpha(1-P_H) + (1-\alpha)(1-P_L)\right].$$

If the intermediary willingly contract with borrowers, then

$$\alpha P_H R_H + (1-\alpha)P_L R_L - \frac{C}{q}[\alpha(1-P_H) + (1-\alpha)(1-P_L)] - D \geq r, \tag{3.11}$$

must hold, and where D is the cost of delegated monitoring. As the intermediary finances a larger number of M (i.e. $M \to \infty$), it can guarantee at least the market rate of return, r, to its depositors who are the ultimate lenders that it contracts with. Given the contract, a finite-sized intermediary must write contracts with its depositors which involve monitoring. These depositors must be compensated for the monitoring costs by the intermediary. However as $M \to \infty$, the cost of delegated monitoring, D, goes to zero in the limit.[21]

In contrast, the expected payoffs under direct lending is given by

$$\alpha P_H R_H + (1-\alpha)P_L R_L - C[\alpha(1-P_H) + (1-\alpha)(1-P_L)] \geq r \tag{3.12}$$

[21]See Diamond(1984).

Since $q_i \leq nw$ and n is assumed to be sufficiently large, (3.11) < (3.12) holds. Thus, intermediation dominates direct lending as the optimal institutional arrangement in the credit market. Intuitively, under direct lending, equilibrium requires that each contract break-even and duplicated monitoring occurs.[22,23]

However, in equilibrium, the intermediary grows infinitely large, resembling a natural monopoly. In order to analyze intermediation in a competitive loan market, following Boyd and Prescott (1986), I assume that intermediaries are large in the sense that each intermediary contracts with countably infinite number of lenders; while intermediaries are small in the sense that the probability that all agents deal with only one intermediary is zero. I also drop the time subscript for the analysis from here onward.

3.4.2 Full Information

In an intermediated loan market, since the duplication of verification effort is circumvented, the intermediary expected profit is

$$E[F] = \pi_i \left[P_i R_i - \frac{(1-P_i)C}{q_i} - r \right]. \tag{3.13}$$

Contract 3 summarizes the equilibrium contract for intermediation under full information.

[22]Furthermore, since individual's resources is strictly much lesser than nw and monitoring incurs a sufficiently large fixed cost, it is not efficient for a lender to diversify his lending. In contrast, not only is duplicated monitoring avoided in intermediation, by diversifying, an intermediary does not need to break even on each individual contract since there is a possibility of internal cross subsidization among contracts.

[23]The intermediary shares many of the same characteristics as a real-world depositary financial intermediary. In particular, the financial intermediaries here:

(i) issue securities with return characteristics which are different from those they hold (asset transformer)

(ii) manage a diversified portfolio (risk diversifier, insurance provider)

(iii) gather and process information (information producer)

Moreover, as is shown below, the financial intermediaries also ration credit to borrowers.

Contract 3:

$$
\begin{aligned}
&(i) \quad R_i^{\ddagger} = \tfrac{r}{P_i} + \tfrac{(1-P_i)C}{P_i q_i},\\
&(ii) \quad q_i^{\ddagger} = \underset{q_i}{arg\ max}\ \pi_i\{P_i[\zeta_i f_i(q_i) - R_i]q_i - w\},\\
&(iii) \quad \pi_i^{\ddagger} = 1.
\end{aligned}
$$

Contracts under full information for both intermediated and direct lending cases are similar except that the loan interest rates are lower under intermediation due to the cost saving in verification cost duplication.

3.4.3 Private Information

The intermediary maximizes the borrowers' expected payoffs (3.3) subject to the incentive compatibility constraints (3.8) and (3.9), the feasibility constraints (3.5) and (3.2), and its participation constraint (3.13). *Contract 4* gives the equilibrium contract under competitively intermediated loan market when information is private (see Appendix 6 for the proof).

Contract 4:

$$
\begin{aligned}
&(i) \quad R_i^{FI} = \tfrac{r}{P_i} + \tfrac{(1-P_i)C}{P_i q_i},\\
&(ii) \quad q_i^{FI} = nw,\\
&(iii) \quad \pi_H^{FI} = 1, \quad \pi_L^{FI} = \frac{P_H(\zeta_H f_H(q_H) - \frac{r}{P_H}) - \frac{(1-P_H)C}{nw} - \frac{1}{n}}{P_H(\zeta_H f_H(q_L) - \frac{r}{P_L}) - \frac{P_H(1-P_L)C}{P_L nw} - \frac{1}{n}}.
\end{aligned}
$$

The realized loan interest rates under competitive intermediation is lower than those that prevailed under direct lending. This is due to the cost saving effect in verification cost, of which the benefits accrue to borrowers. In addition, the probability of a low-risk borrower being denied credit is lower under an intermediated competitive loan market than that under a direct lending regime (refer to Appendix 7 for the proof).

Proposition 7

$$If \quad \frac{\partial U_i^{DL}}{\partial \pi_i} \geq 0, \quad then \quad \pi_L^{FI} \geq \pi_L^{DL} = \pi_L^{\dagger\dagger}.$$

The level of type-2 rationing with respect to the low-risk borrower is lower than the realized level under direct lending. The larger is the fixed verification cost, the greater is the potential of cost saving that can be generated. Since the size and the structure of the population are constant, this implies a better probability of the low-type borrower in obtaining loan. However, the level of rationing due to adverse selection does not decrease under intermediation. This suggest that adverse selection is a fundamental problem that will not go away unless some ex-ante information production activities are carried out.[24]

3.5 Monopolistic Intermediated Loan Market

3.5.1 Full Information

Assume that the optimal contract is still a standard debt contract. In a full information monopolistic intermediated loan market, the intermediary maximizes its own expected payoffs (3.4), subject to the borrowers' participation constraints (3.3), and the feasibility constraints (3.5), (3.2) and (3.7). The equilibrium contract is given by

Contract 5:

$$\begin{array}{ll} (i) & R_i^{\ddagger\ddagger} = \zeta_i f_i(q_i) - \frac{1}{nP_i}, \\ (ii) & q_i^{\ddagger\ddagger} = nw, \\ (iii) & \pi_i^{\ddagger\ddagger} = 1. \end{array}$$

[24]Similiar to the case under direct lending, it can be shown that (i) the incentive compatibility constraint for the low-risk borrower is non-binding, and (ii) a separating equilibrium exist only if

$$\begin{aligned} \alpha^{FI} &> \frac{P_L}{P_L - P_H}\left\{1 - \frac{r + \frac{C}{nw}}{(1-\pi_L)\left(P_L\left[\zeta_H f_H(q_L)\right]\frac{\beta}{n}\right]\right) + \pi_L r + P_L \frac{P_L C}{nw}}\right\} \\ &> \alpha^{DL}. \end{aligned}$$

In particular, the region of α for the existence of a separating equilibrium becomes more severe under intermediation. Since the high-risk borrower's investment project is relatively technological superior than that of the low-risk borrower's, a higher value of α needed to guarantee the existence of the intermediated equilibrium means that the economy must be at a certain level of the development process for it to be feasible for intermediation to emerge endogenously. This is compatible with the general agreement that there is a causal relationship economic development and financial development.

The loan rate paid by the borrowers is equivalent to their respective reservation utility (expected payoffs) level. A monopolistic financial intermediary can efficiently extract borrowers' surplus.[25]

3.5.2 Private Information

The full information contract is not incentive compatible under an asymmetric information economy. In addition to the borrower's participation constraint and feasibility constraints, the intermediary maximizes its expected payoffs subject to the borrower's incentive compatibility constraint. The optimization program is given below.

Program 6:

$$\max_{q_i,\, \pi_i,\, R_i} \quad \alpha\pi_H q_H[P_H R_H - (1-P_H)C - r] + (1-\alpha)\pi_L q_L[P_L R_L - (1-P_L)C - r],$$

s.t

$$\pi_i\left(P_i\left[\zeta_i f_i(q_i) - R_i\right] q_i - w\right) \geq 0,$$

$$\pi_i\{P_i[\zeta_i f_i(q_i) - R_i]q_i - w\} \geq \pi_j\{P_i[\zeta_i f_i(q_j) - R_j]q_j - w\},$$

$$(3.5), \quad (3.2) \quad and \quad (3.7),$$

for $i = \{H, L\}$, $i \neq j$.

Contract 6 and *Proposition 5* summarize the equilibrium under monopolistic intermediated loan market.

[25] This is to be contrasted with the intermediary in the competitive market that takes the return to borrowers as a market determined variable, which may or may not equal to the reservation utility (expected payoffs) level.

Contract 6:

$$(i)\quad q_H^M = nw,$$

$$q_L^M \begin{cases} = 1 & if \quad \alpha P_H \left\{ \zeta_H f_H(q_L) - \zeta_L f_L(q_L) + \zeta_H f_H'(q_L)\rho q_L - \zeta_L f_L'(q_L)\rho q_L \right\} \\ & \qquad \leq (1-\alpha) \left\{ P_L \zeta_L f_L(q_L) + P_L \zeta_L f_L'(q_L)\rho q_L - (1-P_L)C - r \right\} \\ = 0 & if \quad otherwise \end{cases} .$$

$$(ii)\quad \pi_H^M = 1,$$

$$\pi_L^M \begin{cases} = 1 & if \quad \alpha \left\{ P_H[\zeta_H f_H(q_L) - \zeta_L f_L(q_L)]q_L - \left(\frac{P_L - P_H}{P_L} \right) w \right\} \\ & \qquad \leq (1-\alpha)\{ (P_L \zeta_L f_L(q_L) - (1-P_L)C - r)\, q_L - w \} \\ = 0 & if \quad otherwise \end{cases} .$$

$$(iii)\quad R_H^M \begin{cases} R_H^M = \zeta_L f_L(q_L) - \frac{w}{P_L q_L^M} & \text{if } \pi_L^M = 1 \\ R_H^M = \zeta_H f_H(q_H) - \frac{w}{P_H q_H^M} & \text{if } otherwise \end{cases} ,$$

$$R_L^M = \zeta_L f_L(q_L) - \frac{w}{P_L q_L^M}.$$

Proposition 8 *In a monopolistic loan market, if the social surplus from the low-risk investment project is sufficiently larger than the corrsponding social cost, then all low-risk borrowers will be granted loan at the full amount.*

Standard Principal-Agent model makes low-risk borrowers worse off, but leaves the expected payoffs of the high-risk borrowers unchanged. This difference in the expected payoffs is necessary to make self selection possible. However, in the present model, the high-risk borrower's participation constraint slacks if $\pi_L^M < 1$, while that of the low-risk borrower binds. This means that the expected payoffs of the low-risk borrowers are unchanged while those of the high-risk borrowers rise under private information. This is due to the fact that under full information, both the high-risk and low-risk borrowers have identical reservation expected payoffs, and the monopolistic intermediary forces the borrowers down to their reservation expected payoffs. Therefore, in order to induce self selection under private information, the contract must improve the high-risk borrowers' expected payoffs but leaves the expected payoffs of the low-risk borrowers unchanged. If the expected social surplus for a low-risk borrower's investment project is less than its social cost, then a monopolistic financial intermediary prices low-risk borrowers out of the loan market. This case corresponds to $\pi_L^M = 0$ or $q_L^M = 0$. Such rationing is inefficient because positive social surplus would be generated if loans were granted to the low-risk borrowers. Nevertheless, this surplus cannot be realized because the intermediary cannot distinguish between borrowers' types ex-ante. However, if the opposite is true, then all low-risk borrowers recievethe full amount of loan applied.[26]

3.6 Conclusion

This essay examines the functioning of credit markets with imperfect information, under different assumptions with respect to investment technologies available to entrepreneurs applying for a loan and the structure of the loan market. Specifically, I analyze the structure of loan contracts comprising interest rate, volume of investment and loan granting probability, in both competitive and monopolistic loan markets with adverse selection and ex-post verification problems under universal risk neutrality.

I show that even when the competitive economy is not wealth-constrained and all investment projects yield positive expected gross returns, lenders ration credit to a borrower with low-risk investment technology in the form of (i) the constrained size of loan allotment, or (ii) the uncertainty in loan granting, but not both. The realized type of rationing depends on how much the borrower perceives the value of not being the recipient of one type of rationing over the other. If she values the unconstrained loan size more relative to the certainty of obtaining credit, then the type-2 rationing (i.e. all or nothing rationing) characterizes the separating equilibrium in the credit market, and vice-versa. Loan granting rationing is a consequence of the low-risk borrower's investment technol-

[26]Note also that although the high-risk borrowers pay a lower loan interest rate under private information than that under full information, this entails no deadweight loss. This is because the reduction in R_H is a pure transfer between the high-risk borrowers and the monopolistic financial intermediary.

ogy second order stochastically dominates that of the high-risk borrower's. Loan size rationing is possible because an investment project is divisible.

Focusing on the type-2 rationing, I then address the implications of financial intermediation for an equilibrium rationing in a competitive loan market. Unsurprisingly, the level of rationing in a competitively intermediated loan market is lower than that under the direct lending regime. This is because financial intermediaries save on monitoring cost which is a response to the ex-post verification problem. On the other hand, in a monopolistic credit market, if the social surplus that emanates from the low-risk borrower's investment project is greater than its concomitant social cost, then she receives the unconstrained loan size with certainty. This is due to the fact that the monopolistic intermediary extracts all realized net social surpluses, and hence will not ration any low-risk borrower that has a positive net surplus from carrying out the investment project.

Thus, credit rationing is an outcome of an optimizing lender's behavior in a credit market with stochastic nature of investment technology and characterized by problems due to asymmetric information. However, as the comparison between competitive intermediation and direct lending demonstrates, not all forms of informational problem are necessary to generate credit rationing.[27] As Prescott and Townsend (1984) imply, what matters is the presence of borrowers whose commonly unknown heterogeneous attributes is ex-ante indistinguishable by lenders when they enter the loan market. Furthermore, the equilibrium with credit rationing is not robust to attributes of the investment technology and the structures of the loan market. As is shown in the analysis, different assumptions with regard to the payoff and risk of the investment project potentially yield different types of rationing, and different market structure leads to different characterizations of credit rationing.

The inclusion of collateral requirement which serves as an additional device to sort

[27]For another example, the Stiglitz and Weiss (1981) model of adverse selection becomes a model of favorable selection if the investment project payoffs project payoffs differ in means rather than in spreads.

borrowers' risk can mitigate rationing in the credit market (or even eliminate rationing if a borrower is not wealth-constrained with regard to the collateral posting).[28] However, in the present model, the inclusion of collateral requirement renders the informational problem trivial. That is, the sorting of risk by loan size rationing and loan granting rationing leads to self selection of loan applicants. The inclusion of collateral requirement means lenders will have three screening instruments at their disposal, which exceeds the dimensionality of the borrowers' privately observed, and hence, no rationing in equilibrium.[29]

Finally, credit rationing is widely perceived to have adverse effect on the economy. Financial intermediation alleviates the incentive frictions due to asymmetric information, and hence rationing, albeit imperfectly. Even if this is the case, there are a number of problems in using the results here and those in the literature to justify credit market intervention. Firstly, as mentioned above, the results are not robust to risk, market structure and other assumptions. Secondly, the appropriate benchmark for accessing the efficacy of intervention is the second-best efficiency since this is what a social planner who does not have access to private information can achieve. It is not enough simply to show that the market solution does not attain the full information efficiency. Thirdly, the effect of credit rationing is not necessarily monotonic.[30] The alleviation of credit rationing may not lead to the amelioration of the economic growth.

[28]See Besanko and Thakor (1987a, 1987b).

[29]See Jaffee and Stiglitz (1990).

[30]For instances, see Beck, Levine and Loayza (2000), Boyd, Levine and Smith (2001), Deidda and Fattouh (2002), Rousseau and Wachtel (2002), and Rioja and Valev (2004).

Appendices

Appendix 1: Direct Lending under Full Information

Program 1

$$\max_{q_i, R_i, \pi_i} \pi_i\{P_i[\zeta_i f_i(q_i) - R_i]q_i - w\},$$

s.t

$$\pi_i q_i \left[P_i R_i - (1 - P_i)C - r\right] \geq 0,$$

$$0 \leq \pi_i \leq 1.$$

Solution: Solve for type-L borrowers. The Langrangean is

$$\begin{aligned} L_L^{DL} &= \pi_L\{P_L[\zeta_L f_L(q_L) - R_L]q_L - w\} \\ &\quad + \lambda \pi_L[P_L R_L - (1 - P_L)C - r] \end{aligned}$$

where λ is the Langrange multiplier associated with (3.4). Differentiating the Langragean w.r.t. q_L,and π_L yields

$$\frac{\partial L_L^{DL}}{\partial q_L} = \pi_L \left\{P_L \left[\zeta_L f_L(q_L) - R_L\right] + P_L \zeta_L f_L(q_L) \rho q_L\right\},$$

and

$$\frac{\partial L_L^{DL}}{\partial \pi_L} = \{P_L[\zeta_L f_L(q_L) - R_L]q_L - w\} + \lambda\{P_L R_L - (1 - P_L)C - r\}.$$

From (3.6), it follows that $\frac{\partial L_L^{DL}}{\partial \pi_L} > 0$. This implies that $\pi_L^\dagger = 1$. Differentiating the Langragean w.r.t. R_L gives

$$\frac{\partial L_L^{DL}}{\partial R_L} = 0 \iff \lambda = q_L > 0.$$

That is, the lender's participation contraint binds. Thus

$$R_L^\dagger = \frac{r}{P_L} + \frac{(1-P_L)C}{P_L}.$$

Substituting equilibrium values into $\frac{\partial L_L^{DL}}{\partial q_L}$ and rearranging terms bring forth

$$\zeta_L f_L(q_L) + \zeta_L f_L(q_L)\rho q_L = R_L. \tag{3.14}$$

The left hand side represents the value of marginal production of investment with respect to loan size, while the right hand side is the corresponding marginal cost. Assume that when (3.14) holds with equality, the amount of loan granted is

$$q_L^\dagger = nw,$$

where $0 < n = \frac{q_L^\dagger}{w} << \infty$ is the number of lenders a low-risk borrower contracts with in the full information economy. The contract for a type-H borrower can be derived similarly.

Appendix 2: Direct Lending under Private Information

Program 2

$$\begin{aligned}\max_{q_i,R_i,\pi_i} \quad & \alpha\left\{\pi_H\left(P_H[\zeta_H f_H(q_H) - R_H]q_H\right) + (1-\pi_H)\,w\right\} \\ & + (1-\alpha)\left\{\pi_L\left(P_L[\zeta_L f_L(q_L) - R_L]q_L\right) + (1-\pi_L)\,w\right\},\end{aligned}$$

s.t

$$\begin{aligned} \pi_i q_i[P_i R_i - (1-P_i)C - r] \geq 0, \\ \pi_H\{P_H[\zeta_H f_H(q_H) - R_H]q_H - w\} \\ \geq \pi_L\{P_H[\zeta_H f_H(q_L) - R_L]q_L - w\}, \end{aligned} \tag{3.15}$$

$$\begin{aligned}&\pi_L\{P_L[\zeta_L f_L(q_L) - R_L]q_L - w\}\\&\quad\geq \pi_H\{P_L[\zeta_L f_L(q_H) - R_H]q_H - w\},\end{aligned} \tag{3.16}$$

$$0 \leq R_i \leq \zeta_i f_i(q_i),$$

$$0 \leq \pi_i \leq 1,$$

$$q_i \leq nw,$$

for $i = \{H, L\}$.

Solution: Since the participation constraints for both types of agents bind in the equilibrium,

$$R_i^{\dagger\dagger} = \frac{r}{P_i} + \frac{(1-Pi)C}{Pi},$$

for $i = \{H, L\}$. The contract $\{q_H, R_H, \pi_H\}$ for a type-H borrower is not affected by consideration of self selection. High-risk borrower's incentive compatibility constraint binds but that of the low-risk borrower's slacks. Ignoring the low-risk borrower's incentive compatibility constraint and substituting $R_H^{\dagger\dagger}$ and $R_L^{\dagger\dagger}$ into *Program 2*, the Langrangean is

$$\begin{aligned}L^{DL} \;=\;& \alpha\pi_{Ht}\left\{P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H}\right]q_H - (1-P_H)Cq_H - w\right\}\\&+(1-\alpha)\pi_L\left\{P_L\left[\zeta_L f_L(q_L) - \frac{r}{P_L}\right]q_L - (1-P_L)Cq_L - w\right\}\\&-\mu\left\{\pi_L\left[P_H\left(\zeta_H f_H(q_L) - \frac{r}{P_L}\right)q_L - \frac{P_H(1-P_L)Cq_L}{P_L} - w\right]\right.\\&\quad\left.-\pi_H\left[P_H\left(\zeta_H f_H(q_H) - \frac{r}{P_H}\right)q_H - (1-P_H)Cq_H - w\right]\right\}.\end{aligned}$$

Differentiating the Langrangean w.r.t q_H and π_H gives

$$\begin{aligned}\frac{\partial L^{DL}}{\partial q_H} \;=\;& \alpha\pi_H\left\{P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H}\right] + P_H\zeta_H f_H'(q_H)\rho q_H - (1-P_H)C\right\}\\&+\mu\pi_H\left\{P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H}\right] + P_H\zeta_H f_H'(q_H)\rho q_H - (1-P_H)C\right\}\end{aligned}$$

and

$$
\begin{aligned}
\frac{\partial L^{DL}}{\partial \pi_H} &= \alpha\left\{P_H\left[\zeta_H f_H(q_H)-\frac{r}{P_H}\right]q_H-(1-P_H)Cq_H-w\right\} \\
&\quad +\mu\left\{P_H\left[\zeta_H f_H(q_H)-\frac{r}{P_H}\right]q_H-(1-P_H)Cq_H-w\right\}.
\end{aligned}
$$

From (3.6), it follows that $\frac{\partial L^{DL}}{\partial q_H} > 0$. Hence, $q_H^{\dagger\dagger} = nw$. Since $\frac{\partial L^{DL}}{\partial \pi_H} > 0$, $\pi_H^{\dagger\dagger} = 1$. Therefore, $q_H^{\dagger\dagger}$ and $\pi_H^{\dagger\dagger}$ do not depend on any of the lower risk borrower's parameters.

Next, assume that $\frac{\partial L^{DL}}{\partial q_L} = 0$ and rearranging terms returns

$$
0 < \mu = \frac{(1-\alpha)\left\{P_L\left[\zeta_L f_L(q_L)-\frac{r}{P_L}\right]P_L\zeta_L f_L'(q_L)\rho q_L-(1-P_L)C\right\}}{P_H\left[\zeta_H f_H(q_L)-\frac{r}{P_L}\right]P_H\zeta_H f_H'(q_L)\rho q_L-\frac{P_H(1-P_L)C}{P_L}} < 1.
$$

Hence, the high-risk borrower's incentive compatibility constraint, (3.15), binds in the equilibrium. Since there are two unknown variables (i.e. π_L and q_L) in (3.15), the equilibrium values for π_L and q_L are not independent of each other. For any values of π_L within a certain range, there is a corresponding value of q_L which makes (3.15) binds, and vice-versa. This means q_L itself is a function of π_L.

Substituting the thus far derived equilibrium values, $R_H^{\dagger\dagger}$, $R_L^{\dagger\dagger}$, $\pi_H^{\dagger\dagger} = 1$ and $q_H^{\dagger\dagger} = nw$, into (3.15) and rearranging terms, bring forth

$$
\pi_L = \frac{P_H\left[\zeta_H f_H(q_H)-\frac{r}{P_H}-\frac{(1-P_H)C}{P_H}\right]nw-w}{P_H\left[\zeta_H f_H(q_L)-\frac{r}{P_L}-\frac{(1-P_L)C}{P_L}\right]q_L-w}. \tag{3.17}
$$

Differentiating (3.17) w.r.t. q_L gives

$$
\begin{aligned}
\left.\frac{\partial \pi_L}{\partial q_L}\right|_H &= -\frac{P_H\left[\zeta_H f_H(q_L)-\frac{r}{P_L}-\frac{(1-P_L)C}{P_L}+\zeta_H f_H'(q_L)\rho q_L\right]}{P_H\left[\zeta_H f_H(q_L)-\frac{r}{P_L}-\frac{(1-P_L)C}{P_L}]q_L -_H w\right]}\pi_L \\
&= (MRS_H)\cdot \pi_L,
\end{aligned}
$$

where MRS_H is the the high-risk borrower's marginal rate of substitution from π_L to q_L, evaluated at $q_H^{\dagger\dagger} = q_L$.

On the other hand, substituting $R_H^{\dagger\dagger}$, $R_L^{\dagger\dagger}$, $\pi_H^{\dagger\dagger}$ and $q_H^{\dagger\dagger}$ into the objective function of the *Program 2* results in

$$\begin{aligned} E[U_i^{DL}] &= \alpha\left\{P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H} - \frac{(1-P_H)C}{P_H}\right]nw - w\right\} \\ &\quad +(1-\alpha)\pi_L\left\{P_L\left[\zeta_L f_L(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right]q_L - w\right\}. \end{aligned}$$

Differentiating $E[U_i^{DL}]$ w.r.t. q_L and dividing the resulting derivative by $E[U_i^{DL}]$ derivation w.r.t. π_L yields

$$\begin{aligned} -\frac{\partial U_i^{DL}/\partial q_L}{\partial U_i^{DL}/\partial \pi_L} &= -\left.\frac{\partial \pi_L}{\partial q_L}\right|_L \\ &= -\frac{P_L\left[\zeta_L f_L(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L} + \zeta_L f_L'(q_L)\rho q_L\right]}{P_L\left[\zeta_L f_L(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right]q_L - w}\pi_L \\ &= (MRS_L)\cdot\pi_L, \end{aligned}$$

where MRS_L is the low-risk borrower's marginal rate of substitution from π_L to q_L, evaluated at $q_L^{\dagger\dagger} = q_L$.

By comparing equilibria, there are two extreme cases:

Case 1: Type-1 rationing occurs.

$$If \quad \left.\frac{\partial \pi_L}{\partial q_L}\right|_H < \left.\frac{\partial \pi_L}{\partial q_L}\right|_L, \; then \quad \pi_L^{\dagger\dagger 1} = 1 \quad and \quad q_L^{\dagger\dagger 1} < nw.$$

Proof: From (3.15) and $\pi_L \leq 1$, at the equilibrium

$$\frac{P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H} - \frac{(1-P_H)C}{P_H}\right]nw}{P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right]} \leq q_L \tag{3.18}$$

holds. The optimization problem now becomes

Program 2a:

$$\max_{q_L,\, \pi_L} \quad \pi_L \left\{ P_L \left[\zeta_L f_L(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L} \right] q_L - w \right\},$$

s.t

$$(3.17) \quad and \quad (3.18).$$

Substituting (3.17) into the objective function, the corresponding Langragean is defined by

$$\begin{aligned} L_L^{DL} \quad = \quad & \left\{ \frac{P_H \left[\zeta_H f_H(q_H) - \frac{r}{P_H} - \frac{(1-P_H)C}{P_H} \right] nw - w}{P_H \left[\zeta_H f_H(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L} \right] q_L - w} \right\} \\ & \times \left\{ P_L \left[\zeta_L f_L(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L} \right] q_L - w \right\} \\ & - \tau_{q_L} \left\{ \frac{P_H \left[\zeta_H f_H(q_H) - \frac{r}{P_H} - \frac{(1-P_H)C}{P_H} \right] nw}{P_H \left[\zeta_H f_H(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L} \right]} - q_L \right\}, \end{aligned}$$

where τ_{q_L} is the multiplier for (3.18). Differentiating the Langrangean w.r.t. q_L gives

$$\begin{aligned}
\frac{L_L^{DL}}{\partial q_L} &= \left\{ \frac{P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H} - \frac{(1-P_H)C}{P_H}\right] nw - w}{\left(P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right] q_L - w\right)^2} \right\} \\
&\quad \times \Bigg\{ \left(P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right] q_L - w\right) \\
&\qquad \times \left(P_L\left[\zeta_L f_L(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L} + \zeta_L f_L'(q_L)\rho q_{Lt}\right]\right) \\
&\quad - \left(P_L\left[\zeta_L f_L(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right] q_L - w\right) \\
&\qquad \times \left(P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L} + \zeta_H f_H'(q_L)\rho q_{Lt}\right]\right) \Bigg\} \\
&\quad + \tau_{q_L} \left\{ \frac{\left(P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H} - \frac{(1-P_H)C}{P_H}\right] nw\right)\left(P_H \zeta_H f_H'(q_L)\right)}{P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right]^2} + 1 \right\}
\end{aligned}$$

Since $\frac{L_L^{DL}}{\partial q_L} \geq 0$, if

$$\begin{aligned}
&\frac{P_L\left[\zeta_L f_L(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L} + \zeta_L f_L'(q_L)\rho q_{Lt}\right]}{P_L\left[\zeta_L f_L(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right] q_L - w} \\
< \; &\frac{P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L} + \zeta_H f_H'(q_L)\rho q_{Lt}\right]}{P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right] q_L - w}
\end{aligned} \tag{3.19}$$

holds, then $\tau_{q_L} > 0$. Therefore, (3.18) binds and the equilibrium value for q_L is

$$q_L^{\dagger\dagger 1} = \frac{P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H} - \frac{(1-P_H)C}{P_H}\right] nw}{P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right]} < nw = q_L^{\dagger},$$

since $\frac{(1-P_H)}{P_H} > \frac{(1-P_H)}{P_L}$. Substituting $q_L^{\dagger\dagger 1}$ into (3.17) yields

$$\pi_L^{\dagger\dagger 1} = 1 = \pi_L^{\dagger}.$$

However, when (3.19) holds, it is equivalent to

$$MRS_H < MRS_L,$$

or

$$\left.\frac{\partial \pi_L}{\partial q_L}\right|_H < \left.\frac{\partial \pi_L}{\partial q_L}\right|_L.$$

Case 2: Type-2 rationing occurs.

$$If \quad \left.\frac{\partial \pi_L}{\partial q_L}\right|_H > \left.\frac{\partial \pi_L}{\partial q_L}\right|_L, \; then \;\; \pi_L^{\dagger\dagger} < 1 \;\; and \;\; q_L^{\dagger\dagger} = nw.$$

Proof: Substituting π_H and q_H into (3.15) and rearranging terms yields

$$q_L = \frac{\left\{P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H}\right] - (1-P_H)C\right\} nw - (1-\pi_L)w}{\pi_L\left\{P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L}\right] - \frac{P_H(1-P_L)C}{P_L}\right\}}. \tag{3.20}$$

From (3.15) and $q_L \leq nw$,

$$\frac{P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H}\right] - (1-P_H)C - \frac{1}{n}}{P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L}\right] - \frac{P_H(1-P_L)C}{P_L} - \frac{1}{n}} \leq \pi_L \tag{3.21}$$

holds. Then, the *Program 2* can be rewritten as

Program 2b:

$$\max_{q_L,\, \pi_L} \quad \pi_L\left\{P_L\left[\zeta_L f_L(q_L) - \frac{r}{P_L} - \frac{(1-P_L)C}{P_L}\right] q_L - w\right\},$$

s.t

$$(3.20) \;\; and \;\; (3.21).$$

Substituting (3.20) into the objective function, the Langrangean is

$$\begin{aligned} L^{DL} &= \pi_L \Bigg\{ \left(P_L \left[\zeta_L f_L(q_L) - \frac{r}{P_L} \right] - (1-P_L)C \right) \\ &\quad \times \left(\frac{\left[P_H \left(\zeta_H f_H(q_H) - \frac{r}{P_H} \right) - (1-P_H)C \right] nw - (1-\pi_L)w}{\pi_L \left[P_H \left(\zeta_H f_H(q_L) - \frac{r}{P_L} \right) - \frac{P_H(1-P_L)C}{P_L} \right]} - w \right) \Bigg\} \\ &\quad - \tau_{\pi_L} \left\{ \frac{P_H \left[\zeta_H f_H(q_H) - \frac{r}{P_H} \right] - (1-P_H)C - \frac{1}{n}}{P_H \left[\zeta_H f_H(q_L) - \frac{r}{P_L} \right] - \frac{P_H(1-P_L)C}{P_L} - \frac{1}{n}} - \pi_L \right\}. \end{aligned}$$

where τ_{π_L} is the multiplier for (3.21). Differentiating the Langrangean w.r.t. π_L brings forth

$$\frac{\partial L^{DL}}{\partial \pi_L} = \left\{ \frac{\left(P_L \left[\zeta_L f_L(q_L) - \frac{r}{P_L} \right] - (1-P_L)C \right) w}{P_H \left[\zeta_H f_H(q_L) - \frac{r}{P_L} \right] - \frac{P_H(1-P_L)C}{P_L}} \right\} - w + \tau_{\pi_L}.$$

Assume that $\frac{\partial L^{DL}}{\partial \pi_L} \geq 0$. Then,

$$\tau_{\pi_L} \geq - \frac{\left(P_L \left[\zeta_L f_L(q_L) - \frac{r}{P_L} \right] - (1-P_L)C \right) w}{P_H \left[\zeta_H f_H(q_L) - \frac{r}{P_L} \right] - \frac{P_H(1-P_L)C}{P_L}} + w.$$

(3.1) guarantees that $\tau_{\pi_L} > 0$. Hence,

$$\pi_L^{\dagger\dagger 2} = \frac{P_H \left[\zeta_H f_H(q_H) - \frac{r}{P_H} \right] - (1-P_H)C - \frac{1}{n}}{P_H \left[\zeta_H f_H(q_L) - \frac{r}{P_L} \right] - \frac{P_H(1-P_L)C}{P_L} - \frac{1}{n}} < 1,$$

since $P_H < P_L$. Substituting $\pi_L^{\dagger\dagger 2}$ into (3.20), yields

$$q_L^{\dagger\dagger 2} = nw.$$

Since $\pi_L^{\dagger\dagger 2} < 1$ and $q_L^{\dagger\dagger 2} = nw$, from *case (1)*,

$$\left. \frac{\partial \pi_L}{\partial q_L} \right|_H > \left. \frac{\partial \pi_L}{\partial q_L} \right|_L \Longleftrightarrow MRS_H > MRS_L,$$

holds.

Appendix 3: Proof for Proposition 6

The statement: "the high-risk borrower's expected payoff's elasticity w.r.t. q_L evaluated at $q_H^{\dagger\dagger} = q_L$ is less than the low-risk borrower's expected payoff's elasticity w.r.t. q_L evaluated at $q_L^{\dagger\dagger} = q_L$," is equivalent to

$$\left(\frac{\partial U_{H,q_L}}{U_{H,q_L}}\right)\left(\frac{q_L}{\partial q_L}\right) < \left(\frac{\partial U_{L,q_L}}{U_{L,q_L}}\right)\left(\frac{q_L}{\partial q_L}\right),$$

where U_{H,q_L} is the high-risk borrower's payoff function evaluated at $q_H^{\dagger\dagger} = q_L$ (and correspondingly for U_{L,q_L}). Rearranging terms and dividing both sides by q_L results in

$$U_{L,q_L}\left(\frac{\partial q_L}{\partial U_{L,q_L}}\right) < U_{H,q_L}\left(\frac{\partial q_L}{\partial U_{H,q_L}}\right).$$

This inequality can be transformed into

$$\begin{aligned} & -\tfrac{1}{MRS_{L,q_L}} < -\tfrac{1}{MRS_{H,q_L}} \\ \Longleftrightarrow \quad & MRS_{L,q_L} < MRS_{H,q_L} \\ \Longleftrightarrow \quad & \left.\tfrac{\partial \pi_L}{\partial q_L}\right|_L < \left.\tfrac{\partial \pi_L}{\partial q_L}\right|_H \end{aligned}$$

From *case 2*, $\pi_L^{\dagger\dagger} < 1$ and $q_L^{\dagger\dagger} = nw$ holds.

Appendix 4: The Incentive Compatibility Constraint for the Low-Risk Borrower is Slack

Since the high-risk borrower's incentive compatibility constraint (3.15) binds,

$$\begin{aligned} &\pi_H\{P_H[\zeta_H f_H(q_H) - R_H]q_H - w\} \\ &\quad = \pi_L\{P_H[\zeta_H f_H(q_L) - R_L]q_L - w \\ &\qquad + P_L[\zeta_L f_L(q_L) - R_L]q_L - P_L[\zeta_L f_L(q_L) - R_L]q_L\}. \end{aligned}$$

Substituting $\pi_H = \pi_H^{\dagger\dagger} = 1$ into the above equation and rearranging terms bring forth

$$\begin{aligned} &\pi_L\{P_L[\zeta_L f_L(q_L) - R_L]q_L - w\} \\ &\quad = P_H[\zeta_H f_H(q_H) - R_H]q_H - w \\ &\qquad + \pi_L\{P_L[\zeta_L f_L(q_L) - R_L]q_L - P_H[\zeta_H f_H(q_L) - R_L]q_L\}. \end{aligned} \tag{3.22}$$

On the other hand, the following equation also holds:

$$\begin{aligned} &P_H[\zeta_H f_H(q_H) - R_H]q_H - w \\ &\quad = P_H[\zeta_H f_H(q_H) - R_H]q_H - w \\ &\qquad + P_L[\zeta_L f_L(q_H)]q_H - P_L R_H q_H - P_L[\zeta_L f_L(q_H)]q_H + P_L R_H q_H \\ &\quad = P_L[\zeta_L f_L(q_H) - R_H]q_H - w \\ &\qquad - P_L[\zeta_L f_L(q_H) - R_H]q_H + P_H[\zeta_H f_H(q_H) - R_H]q_H. \end{aligned} \tag{3.23}$$

Substituting (3.23) into (3.22) yields

$$\begin{aligned} &\pi_L\{P_L[\zeta_L f_L(q_L) - R_L]q_L - w\} - \{P_L[\zeta_L f_L(q_H) - R_H]q_H - w\} \\ &\quad = \pi_L\{P_L[\zeta_L f_L(q_L) - R_L]q_L - P_H[\zeta_H f_H(q_L) - R_L]q_L\} \\ &\qquad - P_L[\zeta_L f_L(q_H) - R_H]q_H + P_H[\zeta_H f_H(q_H) - R_H]q_H. \end{aligned}$$

Furthermore, since $q_H^{\dagger\dagger} = q_L^{\dagger\dagger} = nw$ and $P_H \zeta_H f_H(nw) = P_L \zeta_L f_L(nw)$,

$$\begin{aligned} &\pi_L\{P_L[\zeta_L f_L(q_L) - R_L]q_L - w\} - \pi_H\{P_L[\zeta_L f_L(q_H) - R_H]q_H - w\} \\ &\quad = (P_L - P_H)[R_H - \pi_L R_L] \\ &\quad > 0. \end{aligned}$$

Therefore, the incentive compatibility constraint for the low-risk borrower (3.16) does not bind.

Appendix 5: Existence of a Separating Equilibrium under Direct Lending

If a type-2 rationing equilibrium exists, it satisfies

$$\{R_H^{\dagger\dagger}, q_H^{\dagger\dagger}, \pi_H^{\dagger\dagger}\} \quad where \quad q_H^{\dagger\dagger} = nw, \quad \pi_H^{\dagger\dagger} = 1,$$
$$\{R_L^{\dagger\dagger}, q_L^{\dagger\dagger}, \pi_L^{\dagger\dagger}\} \quad where \quad q_L^{\dagger\dagger} = nw, \quad \pi_L^{\dagger\dagger} < 1.$$

Given the rental rate, , let's consider a pooling contract $\{\widetilde{R}, \widetilde{q}, \widetilde{\pi}\}$, where

$$\begin{aligned}\widetilde{R} &= \frac{r}{\widetilde{P}} + \frac{(1-\widetilde{P})C}{\widetilde{P}} \\ &= \frac{r}{\alpha(P_H - P_L) + P_L} + \frac{[1 - P_L - \alpha(P_H - P_L)]C}{\alpha(P_H - P_L) + P_L}.\end{aligned}$$

($\widetilde{P} = \alpha P_H + (1-\alpha)P_L$). I want to show that there is no contract $\{\widetilde{R}, \widetilde{q}, \widetilde{\pi}\}$ that: (i) rations no one, and (ii) is preferred to $\{R_L^{\dagger\dagger}, q_L^{\dagger\dagger}, \pi_L^{\dagger\dagger}\}$ by a low-risk borrower.

Proof: The optimization problem is now given by

$$\max_{\widetilde{R},\, \widetilde{q},\, \widetilde{\pi}} \quad \widetilde{\pi}\{P_L[\zeta_L f_L(\widetilde{q}) - \widetilde{R}]\widetilde{q} - w\}$$

s.t

$$\widetilde{q} \leq nw,$$

$$0 \leq \widetilde{\pi} \leq 1.$$

Since $P_L[\zeta_L f_L(\widetilde{q}) - \widetilde{R}]\widetilde{q} > 0$ and constant monitoring cost are assumed,

$$\widetilde{q} = q_L^{\dagger\dagger} = nw.$$

With this value, the Langrangean is given by

$$L = \widetilde{\pi}\{P_L[\zeta_L f_L(\widetilde{q}) - \widetilde{R}]nw - w\}.$$

Differentiating it w.r.t. $\widetilde{\pi}$ yields

$$\frac{\partial L}{\partial \widetilde{\pi}} = P_L[\zeta_L f_L(\widetilde{q}) - \widetilde{R}]nw - w.$$

There are two cases to be considered:

(i) $If \quad \frac{\partial L}{\partial \widetilde{\pi}} \leq 0, \quad then \quad \widetilde{\pi} = 0$. In this case there is no pooling contract that can attract the low-risk borrower and earn a non-negative expected profit. A low-risk borrower consumes w.

(ii) $If \quad \frac{\partial L}{\partial \widetilde{\pi}} > 0, \quad then \quad \widetilde{\pi} = 1$. In this case the low-risk borrower's optimization problem is

$$\max_{\widetilde{R},\, \widetilde{\pi}} \quad \widetilde{\pi}\{P_L[\zeta_L f_L(\widetilde{q}) - \widetilde{R}]\widetilde{q} - w\}$$

Then, there is no pooling contract that attracts all borrowers and earns non-negative profits if

$$\pi_L^{\dagger\dagger}\{P_L[\zeta_L f_L(q_L^{\dagger\dagger}) - R_L^{\dagger\dagger}]nw - w\} \geq \widetilde{\pi}\{P_L[\zeta_L f_L(\widetilde{q}) - \widetilde{R}]\widetilde{q} - w\}$$

Substituting $R_L^{\dagger\dagger} = \frac{r}{P_L} + \frac{(1-P_L)C}{P_L}$ and $\tilde{\pi} = 1$ into the above inequality yields

$$\pi_L^{\dagger\dagger}\{P_L[\zeta_L f_L(q_L^{\dagger\dagger}) - \frac{r}{P_L}]nw - (1 - P_L)Cnw - w\} \geq \{P_L[\zeta_L f_L(\widetilde{q}) - \widetilde{R}]\widetilde{q} - w\}.$$

This is equivalent to

$$\pi_L^{\dagger\dagger} \geq \frac{P_L[\zeta_L f_L(\widetilde{q}) - \widetilde{R}]\widetilde{q} - w}{P_L[\zeta_L f_L(q_L^{\dagger\dagger}) - \frac{r}{P_L}]nw - (1 - P_L)Cnw - w}. \tag{3.24}$$

which holds with a strict inequality if $P_L\widetilde{R}$ is sufficiently large. In turn, $\widetilde{R}$ is large if α is sufficiently large. Indeed, as $\alpha \to 1$, $\widetilde{R} \to \frac{r}{P_H} + \frac{1-P_H}{P_H}$. That is, if there are too many high-risk borrowers relative to low-risk borrowers, pooling contract becomes unattractive to low-risk borrowers. On the other hand, if there are too few high-risk borrowers relative to low-risk borrowers, then a low-risk borrower does well with a pooling contract. Using (3.24), it is straightforward to establish that the critical value of α in order to guarantee the existence of a separating equilibrium is given by

$$\alpha^{DL} > \frac{P_L}{P_L - P_H}\left\{1 - \frac{r + C}{(1-\pi_L)[P_L(\zeta_L f_L(q_L)) - \frac{1}{n}] + \pi_L r + P_L C}\right\}.$$

Appendix 6: Competitive Intermediation under Private Information

Program 4a:

$$\max_{q_i,\, \pi_i,\, R_i} \quad \alpha\{\pi_H\left(P_H\left[\zeta_H f_H(q_H) - R_H\right] q_H - w\right)\} \\ + (1-\alpha)\{\pi_L[P_L\left(\zeta_L f_L(q_L) - R_L\right) q_L - w]\}$$

$s.t$

$$E[F] = \pi_i\left[P_i R_i - \frac{(1-Pi)C}{q_i} - r\right] \geq 0,$$

$$\pi_i\{P_i[\zeta_i f_i(q_i) - R_i]q_i - w\} \geq \pi_j\{P_i[\zeta_i f_i(q_j) - R_j]q_j - w\},$$

$$0 \leq R_i \leq \zeta_i f_i(q_i),$$

$$0 \leq \pi_i \leq 1,$$

$$q_i \leq nw,$$

for $i = \{H, L\}$, $i \neq j$.

Solution: Since the loan market is competitive, the equilibrium loan interest rates are given by

$$R_i^{FI} = \frac{r}{P_i} + \frac{(1-P_i)C}{P_i q_i},$$

for $i = \{H, L\}$.

Ignoring the incentive compatibility constraint for the low-risk borrower and substituting R_H^{FI} and R_L^{FI} into the objective function, the Langrangean is

$$
\begin{aligned}
L^{FI} \;=\; & \alpha\left\{\pi_H\left(P_H\left[\zeta_H f_H(q_H)-\frac{r}{P_H}\right]\right)q_H-(1-P_H)C-w\right\} \\
& +(1-\alpha)\left\{\pi_L\left(P_L\left[\zeta_L f_L(q_L)-\frac{r}{P_L}\right]\right)q_L-(1-P_L)C-w\right\} \\
& -\mu\left\{\pi_L\left[P_H\left(\zeta_H f_H(q_L)-\frac{r}{P_L}\right)q_L-\frac{P_H(1-P_L)C}{P_L}-w\right]\right. \\
& \left.-\pi_H\left[P_H\left(\zeta_H f_H(q_H)-\frac{r}{P_H}\right)q_H-(1-P_H)C-w\right]\right\}.
\end{aligned}
$$

Using similar argument as in the direct lending case, the first order conditions $\frac{\partial L^{FI}}{\partial q_H}$ and $\frac{\partial L^{FI}}{\partial \pi_H}$ yield

$$q_H^{FI}=nw,$$

and

$$\pi_H^{FI}=1.$$

Furthermore, from $\frac{\partial L^{FI}}{\partial q_L}$, it is straightforward to verify that $\mu>0$. Hence, the incentive compatibility constraint for the high-risk borrower (IC_H) binds.

Since second order stochastic dominance (3.1) is assumed, I concentrate on the case when $MRS_H > MRS_L$. Substituting equilibrium values into IC_H brings forth

$$q_L=\frac{P_H\left[\zeta_H f_H(q_H)-\frac{r}{P_H}\right]nw-(1-P_H)C+\frac{\pi_L P_H(1-P_L)C}{P_L}-(1-\pi_L)w}{\pi_L P_H\left[\zeta_H f_H(q_L)-\frac{r}{P_L}\right]}. \tag{3.25}$$

Since $q_L \leq nw$,

$$\frac{P_H\left[\zeta_H f_H(q_H)-\frac{r}{P_H}\right]-\frac{(1-P_H)C}{nw}-\frac{1}{n}}{P_H\left[\zeta_H f_H(q_L)-\frac{r}{P_L}\right]-\frac{P_H(1-P_L)C}{P_L nw}-\frac{1}{n}} \leq \pi_L \tag{3.26}$$

holds. The optimization problem is then rewritten as

Program 4b:

$$\max_{q_L,\,\pi_L} \quad \pi_L\left(P_L\left[\zeta_L f_L(q_L)-\frac{r}{P_L}\right]q_L-(1-P_L)C-w\right)$$

s.t

$$(3.25) \quad \text{and} \quad (3.26).$$

Substituting (3.25) into the criterion function, the Langrangean is

$$\begin{aligned}
L^{FI} \;=\; & \pi_L\left\{\left(P_L\left[\zeta_L f_L(q_L)-\frac{r}{P_L}\right]\right)\right. \\
& \left.\left(\frac{P_H\left[\zeta_H f_H(q_H)-\frac{r}{P_H}\right]nw-(1-P_H)C+\frac{\pi_L P_H(1-P_L)C}{P_L}-(1-\pi_L)w}{\pi_L P_H\left[\zeta_H f_H(q_L)-\frac{r}{P_L}\right]}\right)\right\} \\
& -\pi_L[(1-P_L)C-w] \\
& -\tau_{\pi_L}\left(\frac{P_H\left[\zeta_H f_H(q_H)-\frac{r}{P_H}\right]-\frac{(1-P_H)C}{nw}-\frac{1}{n}}{P_H\left[\zeta_H f_H(q_L)-\frac{r}{P_L}\right]-\frac{P_H(1-P_L)C}{P_L nw}-\frac{1}{n}}-\pi_L\right).
\end{aligned}$$

Differentiating the Langrangean w.r.t. π_L produces

$$\begin{aligned}
\frac{\partial L^{FI}}{\partial \pi_L} \;=\; & \frac{P_L\left[\zeta_L f_L(q_L)-\frac{r}{P_L}\right]\left[\frac{P_H(1-P_L)C}{P_L}+w\right]}{P_H\left[\zeta_H f_H(q_L)-\frac{r}{P_L}\right]} \\
& -[(1-P_L)C+w]+\tau_{\pi_L}.
\end{aligned}$$

Assuming that $\frac{\partial L^{FI}}{\partial \pi_L}\geq 0$,

$$\begin{aligned}
\tau_{\pi_L} \;\geq\; & \left(\frac{P_L-P_H}{P_L}\right)rw \\
& +[\zeta_H f_H(q_L)-\zeta_L f_L(q_L)]\rho P_H(1-P_L)C \\
\;>\; & 0.
\end{aligned}$$

Hence, (3.26) binds and the equilibrium values of the probability of extending loan to a low-risk borrowers and her loan quantity are

$$\pi_L^{FI} = \frac{P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H}\right] - \frac{(1-P_H)C}{nw} - \frac{1}{n}}{P_H\left[\zeta_H f_H(q_L) - \frac{r}{P_L}\right] - \frac{P_H(1-P_L)C}{P_L nw} - \frac{1}{n}} < 1,$$

since $P_H < P_L$, and

$$q_L^{FI} = nw$$

respectively.

Appendix 7: Proof for Proposition 7

On the contrary assume that $\pi_L < \pi_L$. This is equivalent to

$$\frac{P_H(\zeta_H f_H(q_H) - \frac{r}{P_H}) - \frac{(1-P_H)C}{nw} - \frac{1}{n}}{P_H(\zeta_H f_H(q_L) - \frac{r}{P_L}) - \frac{P_H(1-P_L)C}{P_L nw} - \frac{1}{n}} < \frac{P_H[\zeta_H f_H(q_H) - \frac{r}{P_H}] - (1-P_H)C - \frac{1}{n}}{P_H(\zeta_H f_H(q_L) - \frac{r}{P_L}) - \frac{P_H(1-P_L)C}{P_L} - \frac{1}{n}}.$$

It is straightforward (but lengthy) to show that this inequality leads to

$$\zeta_H f_H(q_H) < r,$$

which still holds if

$$\zeta_H f_H(q_H) < \frac{r}{P_H} + \frac{(1-P_H)C}{P_H} + \frac{1}{n}.$$

Rearranging terms and multiplying both sides with q_H^{FI} and P_H yield

$$P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H} - \frac{(1-P_H)C}{P_H}\right] nw - P_H w < 0.$$

The above inequality also implies that

$$P_H\left[\zeta_H f_H(q_H) - \frac{r}{P_H} - \frac{(1-P_H)C}{P_H}\right] nw - w < 0,$$

which is a contradiction.

Appendix 8: Monopolistic Intermediation under Private Information

Program 6:

$$\max_{q_i,\pi_i,R_i} \quad \alpha\pi_H q_H[P_H R_H - (1-P_H)C - r] + (1-\alpha)\pi_L q_L[P_L R_L - (1-P_L)C - r],$$

s.t

$$\pi_i\left(P_i[\zeta_i f_i(q_i) - R_i]q_i - w\right) \geq 0,$$

$$\pi_i\{P_i[\zeta_i f_i(q_i) - R_i]q_i - w\} \geq \pi_j\{P_i[\zeta_i f_i(q_j) - R_j]q_j - w\},$$

$$0 \leq R_i \leq \zeta_i f_i(q_i),$$

$$0 \leq \pi_i \leq 1,$$

$$q_i \leq nw,$$

for $i = \{H, L\}$, $i \neq j$.

Solution: Ignoring the high-risk borrower's participation constraint and the low-risk borrower's incentive compatibility constraint, the Langrangean is

$$\begin{aligned} L^M &= \alpha\pi_H\left[P_H R_H q_H - (1-P_H)C - rq_H\right] \\ &\quad +(1-\alpha)\pi_L[P_L R_L q_L - (1-P_L)C - rq_L] \\ &\quad +\lambda\pi_L\left(P_L\left[\zeta_L f_L(q_L) - R_L\right]q_L - w\right) \\ &\quad -\mu\{\pi_L\left(P_H\left[\zeta_H f_H(q_L) - R_L\right]q_L - w\right) - \pi_H\left(P_H\left[\zeta_H f_H(q_H) - R_H\right]q_H - w\right)\}. \end{aligned} \tag{3.27}$$

Treating π_H and π_L as parameters, the first order conditions w.r.t. R_H and R_L are

$$\frac{\partial L^M}{\partial R_H} = 0 \iff \mu = \alpha > 0,$$

and

$$\frac{\partial L^M}{\partial R_L} = 0 \iff = \lambda = \frac{(1-\alpha)P_L + \alpha P_H}{P_L} > 0.$$

Therefore, the high-risk borrower's incentive compatibility constraint and the low-risk borrower's participation constraint bind. The loan interest rates are

$$R_L^M = \zeta_L f_L(q_L) - \frac{w}{P_L q_L}$$

and

$$\begin{aligned} R_H &= \zeta_H f_H(q_H) - \frac{w}{P_H q_H} \\ &\quad - \frac{\pi_L}{\pi_H P_H q_H}\left\{P_H[\zeta_H f_H(q_L) - \zeta_L f_L(q_L)]q_L - \left(\frac{P_L - P_H}{P_L}\right)w\right\} \end{aligned}$$

Since R_H is a function of π_H and q_L these variables cannot be determined independent of each other.

Substituting R_H into the Langrangean (3.27) and differentiate it w.r.t. q_H and π_H, yield

$$\frac{\partial L^M}{\partial q_H} = \alpha\pi_H[P_H\zeta_H f_H(q_H) + P_H\zeta_H f_H'(q_H)\rho q_H - (1-P_H)C - r]. \tag{3.28}$$

and

$$\frac{\partial L^M}{\partial \pi_H} = \alpha[P_H\zeta_H f_H(q_H)\rho q_H -_H w - (1-P_H)Cq_H - rq_H].$$

Since $\frac{\partial L^M}{\partial q_H} > 0$, $q_H^M = nw$, and since $\frac{\partial L^M}{\partial \pi_H} > 0$, $\pi_H^M = 1$.

Substituting the R_H and equilibrium values into the Langrangean (3.27) and differentiate it w.r.t. π_L bring forth

$$\begin{aligned} \frac{\partial L^M}{\partial \pi_L} &= -\alpha\left\{P_H[\zeta_H f_H(q_L) - \zeta_L f_L(q_L)]q_L - \left(\frac{P_L - P_H}{P_L}\right)w\right\} \\ &\quad +(1-\alpha)\{(P_L\zeta_L f_L(q_L) - (1-P_L)C - r)\, q_L - w\}. \end{aligned} \tag{3.29}$$

There are two cases:

$$\begin{array}{llll} (i) & \frac{\partial L^M}{\partial \pi_L} \geq 0 & \Rightarrow & \pi_L^M = 1, \\ (ii) & \frac{\partial L^M}{\partial \pi_L} < 0 & \Rightarrow & \pi_L^M = 0. \end{array}$$

If the loan granting's expected social surplus for a low-risk borrower (i.e. which is the second term on the right hand side of (3.29)), is less than the social cost (i.e. the first term on the right hand side of (3.29)), then a monopolistic financial intermediary prices low-risk borrowers out of the loan market. This case corresponds to $\pi_L = 0$. Such rationing is inefficient because positive social surplus would be generated if loans were granted to the low-risk borrowers. Nevertheless, this surplus cannot be realized because the intermediary cannot distinguish between borrowers' types ex-ante. However, if the opposite is true, then the low-risk borrowers may or may not be rationed in the loan market.

Next, substituting equilibrium values into the Langrangean (3.27) and differentiating it w.r.t. q_L yield

$$\begin{aligned} \frac{\partial L^M}{\partial q_L} &= -\alpha P_H \left\{ \zeta_H f_H(q_L) - \zeta_L f_L(q_L) + \zeta_H f_H'(q_L) \rho q_L - \zeta_L f_L'(q_L) \rho q_L \right\} \\ &\quad + (1-\alpha) \left\{ P_L \zeta_L f_L(q_L) + P_L \zeta_L f_L'(q_L) \rho q_L - (1-P_L)C - r \right\}. \end{aligned}$$

There are also two cases, with the same interpretation as the above, i.e.

$$\begin{array}{llll} (i) & \frac{\partial L^M}{\partial q_L} \geq 0 & \Rightarrow & q_L^M = nw, \\ (ii) & \frac{\partial L^M}{\partial q_L} < 0 & \Rightarrow & q_L^M = 0. \end{array}$$

Bibliography

[1] Bartolini, Leonardo; Bertola, Giuseppe; Prati, Alessandro (2002), Day-to-Day Monetary Policy and the Volatility of the Federal Funds Interest Rate, Journal of Money, Credit, and Banking, 34 (1), pp. 137-159.

[2] Bartolini, Leonardo; Bertola, Giuseppe; Prati, Alessandro (2001), Banks' Reserve Management, Transaction Costs, and the Timing of Federal Reserve Intervention, Journal of Banking and Finance, 25 (7), pp. 1287-1317.

[3] Beck, Thorsten; Levine, Robert; Loayza, Norman (2000), Finance and the Sources of Growth, Journal of Financial Economics, 58 (1-2), pp. 261-300.

[4] Bencivenga, Valerie; Smith, Bruce (1993), Some Consequences of Credit Rationing in an Endogenous Growth Model, Journal of Economic Dynamics and Control, 17, pp. 97-122.

[5] Besanko, David; Thakor, Anjan (1987a), Collateral and Rationing: Sorting Equilibria in Monopolistic and Competitive Credit Markets, International Economic Review, 28, pp. 671-689.

[6] Besanko, David; Thakor, Anjan (1987b), Competitive Equilibrium in the Credit Market under Asymmetric Information, Journal of Economic Theory, 42 (1), pp. 167-182.

[7] Boyd, John; Prescott, Edward (1986), Financial Intermediary-Coalitions, Journal of Economic Theory, 38, pp. 211-232.

[8] Boyd, John; Smith, Bruce (1993), The Equilibrium Allocation of the Investment Capital in the Presence of Adverse Selection and Costly State Verification, Economic Theory, 3, pp. 427-451.

[9] Boyd, John; Smith, Bruce (1992), Intermediation and the Equilibrium Allocation of Investment Capital: Implications for Economic Development, Journal of Monetary Economics, v. 30 (3), pp. 409-432.

[10] Carpenter, Seth; Demiralp, Selva (2004), The Liquidity Effect in the Federal Funds Market: Evidence from Daily Open Market Operations, Board of Governors of the Federal Reserve System (U.S.), Finance and Economics Discussion Series: 2004-61.

[11] Clouse, James; Dow, James (2002), A Computational Model of Banks' Optimal Reserve Management Policy, Journal of Economic Dynamics and Control, 26 (11), pp. 1787-1814.

[12] Clouse, James; Dow, James (1999), Fixed Costs and the Behavior of the Federal Funds Rate, Journal of Banking and Finance, 23 (7), pp. 1015-1029.

[13] Deidda, Luca; Fattouh, Bassam (2002), Non-linearity between Finance and Growth, Economics Letters, 74 (3), pp. 339-345.

[14] Diamond, Douglas (1984), Financial Intermediation and Delegated Monitoring, Review of Economic Studies, 51, pp. 393-414.

[15] Freeman, Scott (1996), The Payments System, Liquidity, and Rediscounting, American Economic Review, 86 (5), pp. 1126-1138.

[16] Freeman, Scott (1999), Rediscounting Under Aggregate Risk, Journal of Monetary Economics, 43 (1), pp. 197-216.

[17] Fujiki, Hiroshi (2003), A Model of the Federal Reserve Act under the International Gold Standard System, Journal of Monetary Economics, 50 (6), pp. 1333-1350.

[18] Furfine, Craig (2000), Interbank Payments and the Daily Federal Funds Rate, Journal of Monetary Economics, 46 (2), pp. 535-553.

[19] Furfine, Craig (2000), The Interbank Market During a Crisis, European Economic Review, 46 (4-5), pp. 809-820.

[20] Gale, Douglas (1996), Equilibria and Pareto Optima of Markets with Adverse Selection Economic Theory, 7, pp. 207-235.

[21] Gale, Douglas; Hellwig, Martin (1985), Incentive-Compatible Debt Contracts: The One Period Problem, Review of Economic Studies, 52, pp. 647-663.

[22] Gonzalez, Fidel (2002), Robust Control and the Economics of Air Pollution. Working Paper, Center for Applied Research in Economics, The University of Texas at Austin.

[23] Gonzalez, Fidel; Rodriguez, Arnulfo (2004), Robust Control: A Note on the Response of the Control to Changes in the "Free" Parameter Conditional on the Character of Nature, Computational Economics, 24 (3), pp. 223-238.

[24] Goodfriend, Marvin (2002), Interest on Reserves and Monetary Policy, Federal Reserve Bank of New York Economic Policy Review, 8 (1), pp. 77-84.

[25] Goodfriend, Marvin; King, Robert, (1988), Financial Deregulation, Monetary Policy, and Central Banking, Federal Reserve Bank of Richmond Economic Review, 74 (3), pp. 3–22.

[26] Green, Edward (1997), Money and Debt in the Structure of Payments, Federal Reserve Bank of Minneapolis Quarterly Review, Spring 1999, 23 (2), pp. 13-29.

[27] Hamilton, James (1997), Measuring the Liquidity Effect, American Economic Review, 87 (1), pp. 80-97.

[28] Hamilton, James (1996), The Daily Market for Federal Funds, Journal of Political Economy, 104 (1), pp. 26-56.

[29] Hansen, Lars Peter; Sargent, Thomas (2001), Acknowledging Misspecification in Macroeconomic Theory, Review of Economic Dynamics, 4 (3), pp. 519-535.

[30] Hernandez-Verme, Paula (2002), Disruptions of Liquidity, International Financial Crises and Monetary Integration, Macroeconomic Dynamics, forthcoming.

[31] Jaffee, Dwight; Stiglitz, Joseph (1990), Credit Rationing, in Handbook of Monetary Economics, vol. 2, pp. 837-88, North-Holland, Amsterdam, Oxford and Tokyo.

[32] Kahn, Charles; Roberds, William (2001), Real-time Gross Settlement and the Costs of Immediacy, Journal of Monetary Economics, 47 (2), pp. 299-319.

[33] Keeton, William (1979), Equilibrium Credit Rationing, Garland Press, New York.

[34] Kendrick, David (2002), Stochastic Control for Economic Models, 2nd. Edition, at *http://www.eco.utexas.edu/faculty/Kendrick/papers/kendbook.pdf*

[35] Kendrick, David; Tucci, Marco (2001), A Robust Control Permanent Income Model in the Duali Software: Quadratic Tracking Version, Working Paper, Center for Applied Research in Economics, The University of Texas at Austin.

[36] Kendrick, David (1981), Stochastic Control for Economic Models, McGraw Hill Book Company, New York.

[37] Krieger, Sandra (2002), Recent Trends in Monetary Policy Implementation: A View from the Desk; By C.; Federal Reserve Bank of New York Economic Policy Review, 8 (1), pp. 73-76.

[38] Lacker, Jeffrey (2004), Payment System Disruptions and the Federal Reserve Following September 11, 2001, Journal of Monetary Economics, 51 (5), pp. 935-965.

[39] Lacker, Jeffrey (1997), Clearing, Settlement and Monetary Policy, Journal of Monetary Economics, 40 (2), pp. 347-381.

[40] Madigan, Brian; Nelson, William (2002), Proposed Revision to the Federal Reserve's Discount Window Lending Programs, Federal Reserve Bulletin, 88 (7), pp. 313-319.

[41] Martin, Antoine (2004), Optimal Pricing of Intraday Liquidity, Journal of Monetary Economics, 51 (2), pp. 401-424.

[42] Mills, David (2004), Mechanism Design and the Role of Enforcement in Freeman's Model of Payments, Review of Economic Dynamics, 7 (1), pp. 219-236.

[43] Mitusch, Kay; Nautz, Dieter (2001), Interest Rate and Liquidity Risk Management and the European Money Supply Process, Journal of Banking and Finance, 25 (11), pp. 2089-2101.

[44] Prati, Alessandro; Bartolini, Leonardo; Bertola, Giuseppe (2003), The Overnight Interbank Market: Evidence from the G-7 and the Euro Zone, Journal of Banking and Finance, October 2003, 27 (10), pp. 2045-2083.

[45] Poole, William (1968), Commercial Bank Reserve Management in A Stochastic Model: Implications for Monetary Policy. Journal of Finance, 23 (5), pp. 769-791.

[46] Prescott, Edward; Townsend, Robert (1984), Pareto Optima and Competitive Equilibria with Adverse Selection and Moral Hazard, Econometrica, 52 (1), pp. 21-45.

[47] Rioja, Felix; Valev, Neven (2004), Finance and the Sources of Growth at Various Stages of Economic Development, Economic Inquiry, 42 (1), pp. 127-140.

[48] Rothschild, Michael; Stiglitz, Joseph (1976), Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information, Quarterly Journal of Economics, 90, pp. 629-649.

[49] Rousseau, Peter; Wachtel, Paul (2002), Inflation Thresholds and the Finance-Growth Nexus, Journal of International Money and Finance, 21 (6), pp. 777-793.

[50] Rustem, Berc (1988), A Constrained Min-Max Algorithm for Rival Models, Journal of Economic Dynamics and Control, 12 (1), pp.101-107.

[51] Sellon, Gordon Jr.; Weiner, Stuart (1996), Monetary Policy without Reserve Requirements: Analytical Issues, Federal Reserve Bank of Kansas City Economic Review, 81 (4), pp. 5-24.

[52] Stiglitz, Joseph; Weiss, Andrew (1981), Credit Rationing in Markets with Imperfect Information, American Economic Review, 71 (3), pp. 393-410.

[53] Thornton, Daniel (2001), Identifying the Liquidity Effect at the Daily Frequency, Federal Reserve Bank of St. Louis Review, 83 (4), pp. 59-78.

[54] Wijkander, Hans (1992), Financial Intermediation, Equilibrium Credit Rationing, and Business Cycles, in Contract Economics by Lars Werin and Hans Wijkander, eds., pp. 218-39, Blackwell, Cambridge, Mass. and Oxford.

[55] Williamson, Stephen (1986), Costly Monitoring, Financial Intermediation and Equilibrium Credit Rationing, Journal of Monetary Economics, 18, pp. 159-179.

[56] Walsh, Carl (1998), Monetary Theory and Policy, MIT Press, Cambridge and London.

[57] Zhou, Ruilin. (2000), Understanding Intraday Credit in Large-Value Payment Systems, Federal Reserve Bank of Chicago Economic Perspectives, 24 (3), pp. 29-44.

VITA

Justen Rene Kok Lye Ghwee was born in Malacca, Malaysia on June 25, 1972, the son of Low Besar and Ghwee Kuan Hock. Upon completing his work at the Gajah Berang Secondary School, Malacca, in 1991, he won the Monbusho scholarship to pursue his studies in Japan. In Japan, Justen took an intensive one-year Japanese language education at the Tokyo University of Foreign Studies. He then entered Hitotsubashi University, where he received his Bachelor of Commerce and Master of Arts in Business/Finance in 1997 and 1999 respectively. Justen started his graduate studies in Economics at the University of Texas at Austin in the fall of 1999. He received his Master of Science in Economics in May 2001.

Permanent Address: 79, Jln. Utama 4, Tmn. Bukit Beruang Utama, Malacca 75450,
Malaysia.

This dissertation was typed by the author.