Essays on Liquidity in Macroeconomics

by

Guido Lorenzoni

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2001

© Guido Lorenzoni, MMI. All rights reserved.

The author hereby grants to MIT permission to reproduce and
distribute publicly paper and electronic copies of this thesis document
in whole or in part.

Author .......................................................... Guido Lorenzoni

Department of Economics

May 4, 2001

Certified by ...................................................... Ricardo Caballero

Ford International Professor of Economics

Thesis Supervisor

Accepted by ..................................................... Richard S. Eckhaus

Ford International Professor of Economics, Emeritus
Chairman, Departmental Committee on Graduate Studies
Essays on Liquidity in Macroeconomics

by

Guido Lorenzoni

Submitted to the Department of Economics on May 4, 2001, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics

Abstract

This thesis includes four essays on the macroeconomic effects of financial market imperfections. The first essay studies the incentives for banks that participate in an interbank market to keep a sufficient level of reserves. It presents a model where, in presence of imperfect insurance against bank-specific shocks, banks keep an inefficiently low ratio of reserves to deposits. A consequence of this is that the interest rate on the money market will fluctuate too much from a second-best perspective. It discusses the potential benefits and risks associated to central bank intervention, and highlights the complementarity between regulatory reserve requirements and stabilization of the interest rate. The second essay (joint with C. Hellwig) studies the ability of banks to issue liquid liabilities while holding only a fraction of their activities in liquid assets. We study the possibility of self-sustaining equilibria in which banks are prevented from abusing their issuing privilege by the threat of losing it in case of default. The third essay is a contribution to the empirics of precautionary savings and shows evidence of a decreasing relationship between household wealth and the variability of consumption expenditure. The evidence is consistent with the presence of a precautionary motive for wealth accumulation. The fourth essay (joint with F. Broner) shows that the time series of the spreads on emerging market bonds appears consistent with the view that international investors supplying funds to these countries are liquidity constrained at times of large price drops.

Thesis Supervisor: Ricardo Caballero
Title: Ford International Professor of Economics
a Carla
Acknowledgments

I wish to thank my advisors Ricardo Caballero, Daron Acemoglu and Abhijit Banerjee for their precious guidance, their insight and their example. Many enjoyable conversations with Olivier Blanchard and Bengt Holmstrom were of great help at various stages of my research. Some years ago Nobuhiro Kiyotaki’s classes at Pompeu Fabra first prompted me to study the issues that appear in my thesis, I had the fortune of meeting him again during my last year of PhD and I wish to thank him for his advice. I am also very grateful to Withney Newey for his suggestions and encouragement in the writing of Chapter 4.

I wish to thank all the friends with whom I shared my time at MIT, and in particular Roberto, Francesco and Max, Christian B., John and Paolo, Christian H. and Fernando, Augustin, David, Esther, Franz, Jeff, Pol, Thomas and Tobias.

My graduate education began in Turin and Barcelona, and I want to thank my teachers and my friends there for all they taught me and for their generous encouragement.

Finally, I want to thank my parents, Luisa and Fabio, and my sister Livia for their support and their love throughout the years I have spent here.
Contents

1 Introduction

2 Excess Interest Rate Volatility in an Intermediated Economy 19
   2.1 The model ........................................... 24
   2.2 Illiquidity and excess interest rate volatility ............ 32
       2.2.1 The difference between derivatives and intermediaries 40
   2.3 Intermediation and illiquidity ........................... 42
   2.4 Monetary policy and interest rate stabilization .......... 47
       2.4.1 No regional risk ................................ 50
       2.4.2 Liquidity reallocation .............................. 53
       2.4.3 Alternative specifications of monetary policy ....... 56
   2.5 Conclusion .............................................. 58
   2.6 References ............................................. 59
   2.7 Appendix .................................................. 62

3 Fractional Reserves and Liquidity Multiplication 65
   3.1 A simple model of an economy with scarce liquidity .... 69
   3.2 Private notes ............................................. 73
   3.3 Steady state equilibrium with banks ....................... 76
       3.3.1 Households and banks ............................... 77
       3.3.2 Entry ............................................... 80
       3.3.3 No default .......................................... 82
   3.4 Conclusions .............................................. 85
4 Household Wealth and Consumption Variability

4.1 Theoretical motivation ........................................... 93
4.2 Linearization and Identification ............................... 99
4.3 Data and Estimation Strategy ................................. 102
4.4 Results ............................................................. 105
  4.4.1 Some preliminary results ............................... 105
  4.4.2 Euler equation estimates and tests of heteroskedasticity 106
  4.4.3 Non-parametric regressions ........................... 112
4.5 Concluding remarks ............................................. 114
4.6 References ....................................................... 116

5 Supply of Funds, Maturity and The Spreads of Emerging Market Bonds ......................................................... 119

5.1 Volatility of Spreads on Long-Term Bonds .................. 123
5.2 The Model ......................................................... 127
5.3 Equilibrium ...................................................... 131
5.4 Empirical Evidence .............................................. 137
5.5 Concluding Remarks ............................................ 139
5.6 References ....................................................... 141
# List of Figures

2-1 Excess volatility ........................................... 40  
2-2 Intermediation and illiquidity .............................. 46  
2-3 Optimal monetary policy ................................... 55  
4-1 Wealth and variability of consumption (c) .................. 96  
4-2 Wealth and variability of consumption (ln C) ............... 97  
4-3 Wealth and consumption variability (w=0 included, bandwidth=.15) 107  
4-4 Wealth and consumption variability (w > 0, bandwidth=.15) .... 108  
4-5 Wealth and residual variability (w >= 0) ....................... 113  
4-6 Wealth and residual variability (w >= 0) ....................... 114  
4-7 Wealth and residual variability (w > 0) ....................... 115  
4-8 Conditional density of the residual .......................... 116  
5-1 Spread of dollar-denominated Argentine bonds over U.S. bonds. ........................................... 121  
5-2 Forward spreads (or default hazard) at different maturities ................................. 125  
5-3 Return-indices of Argentine sovereign bonds at different maturities. ................................. 138  
5-4 Expected excess returns on Argentine bonds during crises. ........................................... 142  
5-5 Volatility of returns on Argentine bonds during crises. ........................................... 143  
5-6 Excess returns vs. standard deviation on Argentine bonds during crises. Points farther from the origin always correspond to longer-term debt. ........................................... 144  
5-7 Sharpe ratio on Argentine bonds as a function of previous price drops. ........................................... 145  
5-8 Returns, volatility, and Sharpe ratio on Mexican bonds. ........................................... 146
List of Tables

4.1 Log Linearized Model .................................................. 109
4.2 Model with quadratic term correction ................................. 110
Chapter 1

Introduction

This thesis includes four essays on the macroeconomic consequences of liquidity constraints. By liquidity constraints here we mean the limited ability of agents to raise finance for investment or consumption at the times and in the states of the world when they most need it. This limited ability is associated broadly to the fact that agents can sell claims to only part of their future income flows, and is affected by the price at which they are able to sell these claims. The first two chapters deal with theoretical issues and are focused on the theory of banking institutions. They study the ability of private banks to partially solve problems of scarce liquidity by reallocating existing liquidity among agents (chapter 2) and by creating new liquid assets (chapter 3). The third and fourth chapters attempt to evaluate the empirical relevance of liquidity constraints in two different contexts. Chapter 4 focuses on household consumption and is essentially a contribution to the empirics of precautionary savings. Chapter 5 instead derives the implication of the presence of liquidity constrained international investors for the prices of emerging market bonds and shows empirical evidence consistent with the presence of liquidity constraints.

The second chapter presents a model of the interbank market. Banks can raise funds (reserves) in the interbank market borrowing at the current interest rate. In general equilibrium the price at which they do these transfers is determined by the aggregate scarcity of funds. From an ex ante perspective the aggregate scarcity of funds depends on various portfolio choices made by banks. This paper focuses on
two dimensions of the banks' portfolio that affects their exposure to liquidity shocks: the extension of committed source of finance to the private sector (credit lines and deposits) and the accumulation of reserves. If financial markets are complete, that is, when bank specific shocks are insurable, the banks fully internalize the effect of their ex ante portfolio decisions on the scarcity of aggregate liquidity. If financial markets are incomplete, instead, the competitive equilibrium is constrained Pareto inefficient. Banks extend commitments that are too large from a second best perspective and this is associated to excess interest rate volatility in the money market.

In this setup as banks integrate and cover a larger number of consumers their ability to reallocate liquidity increases. Nonetheless, when we compare the banking equilibrium with the second best the relation between the degree of integration of the financial system and the severity of the second best illiquidity problem can be non-monotone. The illiquidity problem can be exacerbated when we move from a less integrated system to a more integrated system, because more integrated banks keep a lower reserves to deposit ratio and make liquidity scarcer in the market. For high levels of integration of the intermediation system, though, the inefficiency tends to disappear, as less liquidity flows through the spot market and more liquidity flows among integrated banks, so that interest rate volatility becomes less important for the allocation of liquidity.

In the presence of a nominal friction the central bank can intervene to stabilize the interest rate. Central bank intervention, though, worsens banks' incentives to keep a high reserve to deposit ratio, and the central bank faces a trade-off between liquidity insurance and nominal stability.

Chapter 3 (jointly written with Christian Hellwig) studies an economy with a scarce liquid asset (trees) in which banks issue liabilities backed by fractional reserves of the liquid asset. The multiplication performed by banks alleviates the problem of scarce liquidity but requires some monitoring of the banks to prevent overissuing of liabilities. Occasional conversion of the liabilities provides a form of decentralized monitoring, but at the same time reduces the total liquidity in the economy. We show which conditions are needed to obtain a self-sustaining equilibrium with private
liquidity creation in which banks have no incentive to overissue liabilities. In a stationary setup with no growth we show that it is impossible to sustain an equilibrium in which banks never default. We show that such an equilibrium is possible if the economy growth rate is large enough. We also characterize the optimal reserve ratio and we show that the equilibrium is fragile near the optimal reserve ratio.

Chapter 4 presents evidence of a decreasing relationship between household wealth and consumption variability using PSID data. This relationship offers evidence in favor of models of precautionary savings vs. models of full insurance and vs. standard models of the permanent income hypothesis. Moreover, the presence of this relationship invalidates the identification conditions behind the log-linearized consumption Euler equation. We construct a specification test of the log-linearized Euler equation based on this finding and we obtain a rejection of the assumption needed for log-linearization.

Chapter 5 (jointly written with Fernando Broner) studies the behavior of spreads on short term and long term bonds issued by emerging countries, in order to characterize the supply of funds faced by these countries. We show that the time series of the bond prices are not consistent with the presence of highly diversified international investors that price default risk in a risk neutral manner. Then we study a simple model of financially constrained international investors and we offer some empirical evidence that supports the view that the price-risk associated with long term bonds is very important in explaining the behavior of international investors, and that drop in prices that negatively affect the balance sheet of these investors are consistently associated with increases in the effective rate of risk aversion observed in the market.
Chapter 2

Excess Interest Rate Volatility in an Intermediated Economy

Central banks appear to care about the effect that the level and the movements of short term interest rates have on the liquidity of the banking system. An unexpected increase in interest rates can plunge a bank with large temporary funding needs into a liquidity crisis, while an interest rate reduction can allow it to survive financial distress. Some observers ascribe the interest rate smoothing behavior displayed by central banks at least in part to an objective of financial market stability. A concern for the stability of the banking system was also reflected in the original purpose of the founders of the Federal Reserve system, which was created partly to "furnish an elastic currency" to the economy. A rich empirical literature has documented how the creation of the Fed virtually eliminated seasonal fluctuations in interest rates and the spikes in interest rates which were associated with financial panics before 1914. The Federal Reserve behavior in the early 80's is indicative of this concern over the financial effects of interest rate volatility: when a money targeting regime was producing wide swings in the short term interest rate, the Fed reverted gradually to an interest rate targeting regime, prompted by —among other things— a preoccupation

---

2Quoted from the preamble to the Federal Reserve Act.
with the health of some banking institutions. Similar concerns over liquidity seem to have influenced the Fed timing of interest rate changes in the aftermath of the 1987 stock market crash and of the East-Asian crisis.

In order to rationalize this concern a useful first step is to consider an economy with financial intermediation where the central bank does not intervene in the interbank market to support the system's liquidity and understand what imperfections might generate excessive movements in the interest rate. Therefore, the main questions this chapter will tackle are: Does an economy where the central bank is absent or cares only about price stability have a tendency to display excessive interest rate variability? How does the level of intermediation in the economy and the type of shocks hitting the economy affect the excess variability problem? Finally, I will study how monetary intervention by a central bank can address the problem, focusing on the trade-off a central bank faces in making such an attempt.

To address these issues I build a model where the level and the variability of the short term interest rate affect the functioning of the payment system, described as a system of intermediaries that trade in a money market to transfer resources among themselves. As usual, "money market" refers to the set of markets where intermediaries and large investors raise short term sources of finance, in particular the Federal funds market and the Repo market in the case of the US. The basic idea is that since intermediaries are in the business of delivering funds on request they need to finance large temporary positions in the money market: whenever there is an unexpected increase in the use of credit lines or in deposits withdrawals, a bank must fund the imbalance by borrowing in the money market. The actual size of positions taken in this market is quite relevant. If we concentrate on US commercial banks at the end of 1999 the total of interbank loans held by these banks exceeded 200 bln (2.2% of GDP). To gauge the size of the Repo market we can look at figures on the positions held by US dealers in government securities, who are large market makers in this market. The average value of repurchase agreements and of reserve repurchase agreements held by dealers at the end of 1999 was 1,320 bln and 1,063
bln, respectively (14.7% and 11.8% of GDP).  

To describe the role of banks as liquidity providers I modify a standard Diamond and Dybvig setup introducing banks located in different regions, exposed to different regional liquidity shocks. Banks offer committed sources of finance to units located in their region. These commitments represent deposit accounts and credit lines from which consumers and firms can withdraw funds on demand. When a bank faces an imbalance in its liquid position it must raise funds in the money market. Financial markets are incomplete because bank-specific liquidity shocks are uninsurable. In this situation the level and the variability of the cost of funds in the money market has non-zero wealth effects ex post which are non-neutral from the welfare point of view. Essentially, borrowing banks are hurt by an interest rate increase while lending banks are favored. In the absence of perfect insurance, the marginal value of a dollar in a borrowing bank will be higher than the marginal value of a dollar in a lending bank. Because of this discrepancy a reduction of the interest rate, by transferring resources from borrowers to lenders, would be welfare improving.

However, the same discrepancy has consequences for the optimal decisions of banks ex ante. When banks decide the level of commitments offered to their customers they will tend to extend commitments that are too large because they will undervalue the social return to a dollar of non committed resources ex post. This happens because an extra dollar available for lending moves the supply of funds to the right and reduces the interest rate with a positive wealth effect in favor of borrowing regions. This tendency to undervalue liquidity ex ante corresponds to the typical result of constrained inefficiency of economies with incomplete markets. In this setup the constrained inefficiency of the equilibrium takes the specific form of excess commitments. Excess commitments in turn generate excess interest rate volatility as the intermediation economy will be excessively exposed to aggregate liquidity shocks.

Constrained inefficiency depends on the manner in which ex post financial markets reallocate liquidity among banks and would not appear if banks could offer a

---

4Source: Federal Reserve Bulletin.
grand intermediation contract to all regions in the economy. Thus, the inefficiency disappears when the economy reaches the maximum level of intermediation. This does not imply, however, that the introduction of banks reduces the inefficiency at lower levels of intermediation. I study economies with different levels of intermediations by allowing banks to cover a variable number of consumers. An interesting possibility that arises is that of a non-monotone relationship between the level of intermediation and the excess volatility in the money market. When the economy is very fragmented, each bank pools the liquidity risk of a small number of consumers and, for precautionary reasons, tends to be more liquid and to supply smaller levels of commitments. Therefore, in such an economy the tendency to be illiquid is weaker. In a more intermediated system, with banks covering more regions, each bank needs less self-insurance because it pools large amounts of idiosyncratic risk and tends to supply more commitments per capita. This in turn generates liquidity shortages at the macro level. In this sense the presence of an intermediation system exacerbates the illiquidity problem at intermediate levels of financial integration.

Monetary policy is introduced in the model by adding a very simple nominal friction — namely, deposits denominated in nominal terms. With this nominal friction, a monetary policy that targets the nominal interest rate affects the real allocation in the economy through two distinct channels. The first channel changes the real value of the deposits by affecting the price level, and in this way allows private contracts to be contingent on the aggregate shocks. The second channel affects the allocation of liquidity in the money market by changing the equilibrium real interest rate. Both channels provide a rationale for interest rate smoothing. The second channel is the central one in the context of this chapter as it is intimately related to the inefficiency discussed above. By exploiting this channel, monetary policy can reallocate funds across banks, increasing efficiency. This reallocation is costly, though, because it worsens the *ex ante* tendency of banks to be illiquid. This tendency corresponds to an excessive extension of money-like liabilities by banks—that is, of liabilities that are payable on demand. In this sense the central bank faces a trade-off between liquidity insurance and control of the monetary aggregates.
This chapter is related to various strands of literature. In the macro literature a few papers have modelled the so-called financial motive for interest rate smoothing. In particular Cuckierman (1991) studies a model of bank balance sheets in which movements of the interest rate can affect banks' profitability. In Cuckierman (1991) an interest rate reduction acts simply as a transfer to the banking system as a whole, and this is due to specific assumptions about the speed of adjustment of active and passive rates offered by banks. Instead, this chapter focuses on the effect of interest rate movements on the allocation of funds across banks, rather than on the effects on the aggregate profits of the banking sector. Woodford (1997) constructs a quasi-representative agent setup with cash-in-advance and a sequential service constraint where the distribution of liquid resources across units hit by different shocks is central. The Woodford model is very much related to the model presented here, but in the quasi-representative-agent setup he adopts wealth effects have no welfare consequences, so that the constrained efficiency issues studied in the present chapter do not arise.

The extension of the setup of Diamond and Dybvig (1983) to a multi-region economy was introduced in Battaharya and Gale (1986). Battaharya and Gale study the problem of the optimal design of an incentive-compatible mechanism to reallocate liquid resources ex post assuming that the central bank offers commercial banks a contingent liquidity contract analogous to that used between consumers and banks. In a sense, in Battaharya and Gale all interbank transactions must be mediated by the central bank, who monitors perfectly the trades made by each individual bank. In this chapter instead an anonymous market mechanism is used to reallocate funds across banks located in different regions, and central bank intervention takes the more conventional form of open market operations. The idea of introducing a competitive financial market to reallocate resources between banks in a multi-region setup has also been introduced in a recent paper by Allen and Gale (2000). That paper discusses the general problem of integrating markets and intermediaries in a model of liquidity provision. The model I study falls into the category of economies that they define as "banking" economies. The crucial difference between their setup and mine is that the
in this chapter I introduce a "liquidation technology" subject to decreasing returns. This makes the interest rate \textit{ex post} depend on the aggregate liquidity shock and on the level of commitments chosen by banks. This assumption makes it possible to address problems of interest rate variability and interest rate stabilization.

The fact that an incomplete markets economy with more than two dates is typically constrained inefficient is a classical result in the literature on general equilibrium with incomplete markets. Some recent articles, in particular Caballero and Krishnamurty (2000) and Holmstrom and Tirole (2000), have shown how this type of constrained inefficiency arises in models of liquidity provision and have explored some of its implications in terms of optimal liquidity management. The excess interest rate volatility presented in this chapter is closely related to these results. Features that distinguish the present chapter from this literature are the presence of an intermediation system (with a variable level of bank sizes) and the introduction of a nominal rigidity to discuss the monetary policy aspects of liquidity management.

The chapter is organized as follows. Section 2 presents the basic assumptions of the model. In section 3 I develop the main results concerning the relationship between the interest rate and the allocation of liquidity and on excess interest rate volatility in the baseline model. In section 4 I show the effects of increased intermediation on excess volatility. In section 5 I discuss monetary policy. At the end of the section I discuss extensions of the model that allow for price rigidities and for a nonzero demand for reserves.

\subsection{The model}

\textit{Preferences and technology}

Consider an economy that lasts three periods, populated by a large number of consumers located in \( n \) different regions. Each region population is normalized to 1. Consumers have preferences represented by the utility function

\[ c_0 + \theta u(c_1) + (1 - \theta) u(c_2) \]
where \( \theta \) is an unobservable preference shock, realized at date 1, which takes the values 0 or 1, as in the standard Diamond and Dibvig setup. Assume that the usual Inada-type conditions hold for \( u \). Consumers have an endowment of \((e_1, e_2)\) of the consumption good in periods 1 and 2. All uncertainty is resolved at date 1, at this date in each region \( i \) a proportion \( \tau^i \) of consumers receives the shock \( \theta = 1 \) (early consumers).

I introduce aggregate and regional risk in the following manner. Assume that at date 1 first a cross sectional distribution \( \{\tau_k\}_{k=1}^n \) of regional shocks is selected from a finite set of possible distributions, then a permutation \( \{k_i\} \) of the regional indices is selected at random and each region \( i \) is assigned a \( \tau^i = \tau_{k_i} \) in a symmetric way. The economy wide proportion of early consumers is denoted by \( \tau = (\sum \tau_k) / n \). The unconditional mean of \( \tau \) will be denoted by \( \mu \). For simplicity, I assume that to any given \( \tau \) corresponds only one cross sectional distribution of regional shocks and from now on I will identify the aggregate state with \( \tau \).

Two special cases are the following. The first is the case of no regional risk, when regional shocks are perfectly correlated: in any state \( \tau \), \( \tau_k = \tau \) for all \( k \). The second is the case of no aggregate risk, when \( \tau \) is constant and equal to \( \mu \). The case of no regional and no aggregate risk will be called a "deterministic economy". Also, it will be convenient for many results to use a linear specification for the shocks

\[
\tau_k = \tau + \beta d_k
\]  

(2.1)

where \( d_k \) is a fixed distribution of shocks, \( \tau \) is a random variable, and \( \beta \) is a parameter that controls the level of regional risk in the economy.

The consumption good can be transformed intertemporally according to the following technology. Investing \( x \) units of the consumption good at time 1 gives \( f(x) \) units at time 2, where \( f \) is a \( C^2 \) concave function. In each region there is one firm per consumer that operates this technology and the firm is owned by the consumer at

---

6Notice that this is a slight misnomer given that idiosyncratic risk is still present at the consumers level.
date 0. The concavity of the function $f$ captures the ability of the economy to react to temporary demand shocks by curtailing investment. We can interpret $x$ as investment in maintenance of the existing equipment and in inventories. Low maintenance and a reduction in inventories will result in a less efficient production and distribution in the next period. The firm faces increasing marginal costs when reducing the investment level $x$ and this is reflected in the concavity of $f$. The possibility to have a non-linear $f$ will be crucial for the analysis of the effects of bank commitments on the interest rate.

**Contracts and markets**

At date 1 there is a financial market where agents can borrow and lend against future production at the gross interest rate $r$. At date 1 all uncertainty is resolved, therefore at date 1 the financial market is complete in a trivial sense. This market will be referred to as the money market or the funds market.

Secondly, we specify the type of intermediation contracts that banks offer. In each region, there is a competitive banking sector. More precisely, there is a large number of bankers with zero endowment and linear preferences in terms of consumption at date 0. At time 0 banks offer contracts of contingent credit supply to consumers located in their region. A contract of contingent credit supply specifies, for every level of the regional shock $\tau_k$ and of the market interest rate $r$, the level of investment $x$ by the firm owned by the consumer and the consumption goods delivered to the consumer in the two periods depending on the revealed individual shock. This is summarized in the following definition\footnote{We can derive the active and passive interest rates $(r_1, r_2)$ implicit in the contract using the conditions: $c_1 = e_1 - x + \frac{e_2 + f(x)}{r_1}$ and $c_2 = r_2 (e_1 - x) + e_2 + f(x)$.}

**Definition 1** A regional banking contract is described by the functions $x(\tau_k, r)$ and $\{c_1^\theta(\tau_k, r), c_2^\theta(\tau_k, r)\}_{\theta=0,1}$.

Consumers choose a contract of contingent credit supply at date 0, and they use it at date 1 to finance their consumption selecting one of the pairs \{c_1^\theta, c_2^\theta\}. Individual
preference shocks are unobservable so the incentive compatibility condition

\[ \theta u(c_1^0) + (1 - \theta)u(c_2^0) \geq \theta u(c_1^{\theta'}) + (1 - \theta)u(c_2^{\theta'}) \text{ for all } \theta, \theta' \]  

(2.2)
is required in order to induce consumers to reveal their preference shock. Given this corner type of preference shocks the incentive compatibility condition will not be binding in the analysis to follow. Moreover, it will be optimal to set \( c_2^1 = 0 \) and \( c_1^0 = 0 \) so from now on I will assume that this is the case and denote \( c_1 = c_1^1 \) and \( c_2 = c_2^0 \). Notice that consumers serviced by a bank contract do not need to participate in the financial market directly, since the banks do that on their behalf. Actually, it is the very fact that banks can limit consumers’ participation that allows them to offer liquidity insurance and to improve upon the financial market allocation. If consumers can do side trades on the financial market the contracts of contingent liquidity supplied by banks become redundant as was first shown by Jacklin (1986).

A bank located in region \( i \) has an advantage in supplying credit to consumers located in region \( i \) which is captured by the following assumption.

**Assumption 1** A bank in region \( i \) can observe (and verify in court) the level of consumption of consumers located in region \( i \). A bank in region \( i \) cannot observe the level of consumption nor the trades on the financial market of consumers and bankers located in other regions.

This assumption makes clear what is the empirical counterpart to the notion of "region" used in the model. A region can be any set of firms and consumers whose balance sheets are easier to monitor for a given bank. Notice that in this type of setup banks can improve upon financial market because they offer a commitment to lend at a rate different from the current market rate when the liquidity shock hits. To successfully implement this type of contract the bank must be able to observe the customer activity to make sure he is not using the commitment simply to channel funds to the money market when the rate differential is favorable irrespectively of his consumption level. That is, the bank must be able to observe the customer balance
sheet and it must be able to impose some penalty on the customer in the case of misuse of the credit line. In actual intermediation contracts, the fact that a bank offers a full set of payment services and corporate finance services, helping a company to issue commercial paper, etc. helps the bank to enforce contingent liquidity arrangements. These considerations underlie the idea that the bank has a limited group of customer whose balance sheet can be monitored.

Banks could in principle offer contingent credit lines to banks and consumers located in different regions but assumption 1 makes these contracts redundant. Consider the case of consumers. If a bank were to offer a banking contract to a group of consumers located in a different region the incentive compatibility constraint would take the form

\[ c_1^\theta + \frac{1}{ \tau } c_2^\theta \geq c_1^{\theta'} + \frac{1}{ \tau } c_2^{\theta'} \text{ for all } \theta, \theta'. \]  

(2.3)

It is possible to show that a banking contract that satisfies this condition is redundant, in the sense that consumers achieve the same level of utility offered by a banking contract by just keeping their endowments and trading ex post on the financial market. This is again a consequence of the problem highlighted first by Jacklin's (1986). A similar argument shows that a contract of contingent credit supply between banks located in different regions (where the problem is to induce revelation of the regional shocks \( \tau^i \)) is redundant under assumption 1. In this case the total present value of the transfers between the two banks, computed at the market price \( \tau \), must be equal to zero and these same transfers can be achieved simply by ex post trading between the two banks. Given that all these arrangements are redundant in presence of an open financial market I will simply assume that banks offer contracts of contingent credit supply only to consumers in their own region, preventing them from participating to the financial market, while regional banks interact only through the financial market. In short, consumers obtain liquidity through local banks and local banks obtain liquidity through the money market.

Having described how banks extract information about the unobservable individual shocks in their region it is clear that the regional shocks \( \tau^i \) are private information
of banks located in region $i$. To be more precise, given assumption 1 it is impossible to extract information about the liquidity shocks in different regions in order to support non-zero contingent transfers among regions hit by different liquidity shocks. In different words, all financial instruments that can be designed to induce revelation of the regional shocks must involve contingent transfers between regions hit by different shocks which have zero net present value at the interest rate $r$. These financial instruments are therefore useless for the purpose of insurance against bank-specific shocks. A complete treatment of the endogenous design of financial instruments in a setup with asymmetric information is outside the scope of this chapter. Therefore, to keep matters simple I will just make the following assumption.

**Assumption 2** At date 0 banks can trade securities contingent on the interest rate $r$. There are no securities contingent on the distribution of the regional shocks across banks $\{\tau^i\}$.

In the equilibria I will study the interest rate fully reveals the aggregate shock $\tau$. Given that we assumed that to each $\tau$ corresponds one cross section $\{\tau_k\}$ the interest rate fully reveals it. What the interest rate fails to reveal is the identity of the regions hit by each shock (i.e. the permutation $\{k_i\}$), this is were lies the crucial market incompleteness of the model. Since the shocks are perfectly symmetric and we will study symmetric equilibria all regions will be identical at date 0. Furthermore they are identical at date 1 if we consider the instant immediately after the aggregate shock $\tau$ has been revealed and immediately before the permutation $\{k_i\}$ has been selected. This feature will allow me to study the equilibrium omitting altogether the *ex ante* financial markets and to check at a second time that there is a pricing function at date 0 such that no trade at date 0 is an equilibrium.

The symmetric equilibria I will study are defined below.

**Definition 2** A symmetric competitive equilibrium with banking contracts is defined by a banking contract $\{x(\tau_k, r), \{c^0_1(\tau_k, r), c^0_2(\tau_k, r)\}_{a=0,1}\}$, an *ex post* interest rate function $r(\tau)$, and a date 0 state contingent price function $p(r)$, such that:

(i) the banking contract maximizes the expected utility at date 0 of consumers located
in each region;

(ii) the goods markets and the financial markets at date 0 and 1 clear.

Finally, I impose assume that banking contracts take the form of deposit contracts. More precisely, I assume that for each $\theta c^0(\tau_k; r)$ is restricted to be a constant determined at date 0.

**Assumption 3** Banks are restricted to offer deposit contracts, that is banking contracts with $c^0(\tau_k; r) = c^0$. 

The advantage of introducing this constraint is twofold. First, the contract has a natural interpretation in terms of a commitment. The bank commits to provide finance at pre-specified rate of interest and the customer decides when and whether to use it. The fact that commitments in the form of deposits and credit lines characterize banks' activity has been recently emphasized by Kashyap et al. (1999). Secondly, this specification allows us to introduce nominal contracts and monetary policy in a very simple way, by making the deposit fixed in nominal terms. For the monetary policy to have real effects we need to introduce some form of nominal rigidity, the sequential service constraint allows us to do it with the minimum added complication. The disadvantage of this assumption is that there will be a fundamentally unexplained asymmetry between monetary policy and private contracts, since the former can be conditioned on aggregate shocks while the latter does not allow for contingent values of $c_1$. A similar problem was identified by Wallace in the original Diamond and Dybvig paper, where there is a fundamental asymmetry between private contracts and the contingent taxation scheme that supports deposits insurance. Here, I will live with this asymmetry, and I will stress the fact that interest rate stabilization has two types of effect and that the effect that is central to this chapter (which I will dub the reallocation effect) can be studied independently of this specific asymmetry.

*First best*

This is the place to briefly introduce the first best allocation. In the first best allocation all early consumers are treated identically, independently of the region
where they are located, and the same applies to late consumers. The first best allocation is derived by solving the problem

\[
\begin{align*}
\max & \quad E\tau u(c_1) + (1 - \tau)u(c_2) \\
\text{s.t.} & \quad \tau c_1 = e_1 - x \\
& \quad (1 - \tau)c_1 = e_2 + f(x)
\end{align*}
\]

Given the corner assumption about preference shocks the first best is also incentive compatible, that is, the first best can be achieved by a mechanism under which each consumers truthfully reports his own preference shock to the planner and the planner assigns him \(c_1\) or \(c_2\) accordingly. For future reference it is convenient to write down the first order condition characterizing the first best solution

\[
u'(c_1) = f'(x)u'(c_2)
\]

Some basic results about the first best allocation are collected in the following lemma (proved in the appendix) where \(r\) stands for the shadow interest rate in the problem above and is equal to \(f'(c_1)\).

**Lemma 1** In the first best solution \(c_1\) and \(x\) are decreasing in \(\tau\), while \(c_2\) and \(r\) are increasing in \(\tau\).

Notice that if we eliminated assumption 3 the competitive economy would achieve the first best allocation under *either* of the following conditions: (i) if one bank can cover the whole economy, (ii) if banks could perfectly insure against regional shocks. This clarifies the role of the three imperfections that have been introduced in the model. Assumption 3 creates a rigidity which is clearly suboptimal except in the case of no aggregate risk. If assumption 3 is eliminated, the combination of assumption 1 (regional banking) and assumption 2 (incomplete markets) is keeping the economy away from the first best.
2.2 Illiquidity and excess interest rate volatility

In the previous section we have introduced a decentralized financial system, that is, an economy where each bank is active in a single region and banks located in different regions are connected only through the anonymous interbank market. Given that bank specific shocks are not insurable we do not expect the economy to achieve the first best risk sharing across regions, what is more, since we are in an economy with more than 2 dates the equilibrium will typically be constrained Pareto inefficient. In this section I show under what conditions this inefficiency arises and takes the form of excess illiquidity of banks balance sheets.

I will first characterize the equilibrium at date 1. Consider a bank which has extended a certain level of commitments $c_1$ in an economy where the average level of commitments is $\bar{c}_1$. Since we assumed that the banking sector is competitive in equilibrium banks will offer contracts that maximizes expected consumers utility. Conditional on a given realization of the aggregate shock $\tau$ a bank hit by the shock $\tau_k$ chooses the level of investment $x$ and consumption $c_{2k}$ that solve the problem:

$$\begin{align*}
\max & \quad \tau_k u(c_1) + (1 - \tau_k) u(c_{2k}) \\
\text{s.t.} & \quad \tau_k c_1 + (1 - \tau_k) c_{2k} / r = e_1 - x + (e_2 + f(x)) / r
\end{align*}$$

(2.4)

where given the sequential service constraint $c_1$ has been fixed at date 0.

Every bank will choose the same level of $x$, in order to maximize $f(x)/r - x$ on the right hand side of the budget equation. The equilibrium interest rate $r$ is determined equating total demand of consumption with production at time 1. Therefore for every state $\tau$ we have an equilibrium price $r$ that is determined by the condition

$$r(\tau, \bar{c}_1) = f'(e_1 - \tau \bar{c}_1)$$

(2.5)

Let now consider the decision about deposit commitments $c_1$ made at time 0. It is convenient to introduce the value function $V(c_1, \bar{c}_1; \tau)$. This represents the expected utility of a representative consumer conditional on the aggregate shock $\tau$, before
knowing the individual and the regional idiosyncratic shock. The first argument of $V$ represents the level of deposit commitments contracted by the consumer. The second argument, $\bar{c}_1$, represents the average level of deposits in the economy, which determines the ex post interest rate according to the equilibrium condition (2.5). Assuming that each bank acts as a price taker in the money market implies that each bank takes $\bar{c}_1$ as given when maximizing $EV$ on behalf of consumers. Therefore, a competitive equilibrium with banking contracts is fully characterized by a level of deposits $\bar{c}_1$ such that $\bar{c}_1 = \arg \max_{c_1} EV(c_1, \bar{c}_1; \tau)$. Differentiability of $V$ and concavity in the first argument can be proved easily in this context. Therefore, the level of deposits in a competitive equilibrium is characterized by the first order condition associated to the maximization of $EV$ with respect to its first argument, imposing the equilibrium condition $c_1 = \bar{c}_1$. Applying standard envelope arguments to the problem above, this gives

$$EV_1 = E \left[ \tau_k \left( u'(c_1) - \lambda_k \right) \right] = E \left[ \tau_k \left( u'(c_1) - ru' (c_{2k}) \right) \right] = 0$$

(2.6)

where $\lambda_k$ is the marginal utility of funds for the bank hit by the shock $\tau_k$.

Clearly the competitive equilibrium does not in general achieve the first best. If there are aggregate shocks this follows simply from the presence of the sequential service constraint. If there are only regional shocks the sequential service constraint is not actually binding at the first best (that is a fixed level of $c_1$ is optimal). Nonetheless the fact that only the financial market can be used to reallocate liquidity across banks limits the ability of the economy to achieve the first best allocation. Notice that in the first best allocation $c_{2k}$ does not depend on the regional shock $\tau_2$. In the competitive equilibrium, instead, a fixed level of $c_{2k}$ does not in general satisfy the budget equation except in the special case when $c_{2k} = c_2 = rc_{1}$. This special case corresponds to the first best when the utility function is logarithmic. When this is not the case the financial market will not be able to achieve the first best allocation. This corresponds to the common result in the literature that a grand contract of financial intermediation

\textsuperscript{5}The formal definition of $V$ is in the appendix.

\textsuperscript{9}Notice that strict concavity implies that only symmetric equilibria exist.
is superior to \textit{ex post} financial markets from the point of view of liquidity insurance.

Consider now the problem of a social planner that can only choose the level of deposits $c_1$ committed \textit{ex ante} and then lets the financial market determine the \textit{ex post} allocation after uncertainty is resolved. That is, we assume that after date 0 the social planner is subject to the same constraints faced by the private economy: reallocation across regions can take place only through borrowing and lending (incentive compatibility with non-monitorable side trades) and $c_1$ cannot be made contingent upon realizations of $\tau$ (sequential service constraint). This corresponds to the usual thought experiment introduced in the incomplete markets literature when discussing constrained efficiency.\footnote{Geanakoplos and Polemarchakis. See the discussion of the constrained optimality in Magill and Quinzii.} Under this situation the social planner will maximize $EV(c_1, \bar{c}_1; \tau)$ internalizing the equilibrium condition $c_1 = \bar{c}_1$.

The necessary first order condition for this problem is $E(V_1 + V_2) = 0$. Therefore the difference between the constrained efficient allocation and the competitive equilibrium will depend on the presence of a term $EV_2$ different from zero. Applying standard envelope arguments to problem (2.4) we obtain

$$EV_2 = E \left[ u'(c_{2k})(e_1 - x - \tau k c_1) \frac{dr}{dc_1} \right]$$

(2.7)

The essential difference between the competitive allocation and the second best allocation is that competitive banks do not take into account the effect of their choice on the \textit{ex post} price of funds $r$. When the sign of $EV_2$ is negative the private banking economy is inefficiently illiquid, that is banks issue too many deposits commitments \textit{ex ante}. This is precisely the case we are interested in, because when $c_1$ is large the interest rate is more responsive to aggregate shocks as we can see from condition (2.5). Therefore when $EV_2 < 0$ the real economy displays \textit{excess illiquidity}, that manifests itself in excessive interest spikes as $r$ is too responsive to aggregate shocks. The sign of $EV_2$ will in general depend on the parameters of the model.

The case of a negative $EV_2$ arises under the following additional assumptions, for economies with small amounts of aggregate risk.
**Assumption 4** The utility function displays a coefficient of relative risk aversion greater than one, $-\frac{w''(x)}{w'(x)} x > 1$.

**Assumption 5** The production function satisfies $f'(x) \geq 1$ with $f'(0) > 1$. Defining $\hat{x} = \inf\{x|f'(\hat{x}) = 1\}$ the endowment satisfies the inequality

$$\frac{e_1 - \hat{x}}{\mu} < \frac{e_2 + f(\hat{x})}{1 - \mu}$$

These two assumptions are sufficient conditions to show that for small aggregate shocks the equilibrium allocation will satisfy $c_1 > \frac{c_2k}{r}$ for each $k$. An economic interpretation of these assumptions is the following. We are studying economies where liquidity at date 1 is scarce in the sense that at date 1 the average consumption path would be increasing in absence of regional shocks. Assumption 3 implies that in this situation consumers at date one will require more insurance against early liquidity shocks, and this implies that banks offer them to borrow at a low rate when the liquidity shock hits and to lend at a higher rate when they have excess funds. To put it differently, the net present value of the transfer to early consumers will be higher than the present value of the transfer to late consumers (which corresponds to the inequality $c_1 > \frac{c_2k}{r}$). This corresponds to the notion that banks offer liquidity insurance by charging less than the corresponding market rate (e.g. a commercial paper rate) in states of scarce liquidity.

Under these assumptions we can establish that $c_{2k}$ is lower for regions hit by a higher shock $\tau_k$ which are also the regions that borrow more. Since $r$ is increasing in $c_1$ a larger $c_1$ makes liquidity more expensive, and this hurts borrowing banks more than lending banks, that have larger consumption levels at date 2. Proposition 1 establishes that in this case banks tend to be inefficiently illiquid.

For a given bank his available liquidity depends on the possibility of borrowing ex post against future income, and the terms at which it can do so depends on the price $r$. For the economy as a whole the available liquidity is governed by the intertemporal technology $f$. A single bank tends to be illiquid by extending too many deposits.
because it does not internalize the effect of his decision on the market cost of funds \textit{ex post}. A larger $c_1$ not only makes first period banks balance sheet more sensitive to the shock $\tau_k$ but also makes the economy wide interest rate more sensitive to the aggregate shock $\tau$, making liquidity more expensive for other banks. Competitive banks do not take into account the latter effect and tend to be illiquid from a social point of view.

This effect would be irrelevant if the banks could insure \textit{ex ante} against the shocks $\tau_k$. In this case price fluctuations \textit{ex post} would affect the wealth distribution \textit{ex post} (a high interest rate clearly hurts borrowers and favors lenders) but the ratios of the marginal utility of income of any two banks would be constant across states and equal to the ratios at date 0, so there would be no room for Pareto improvements. In an incomplete market setup instead it is possible that the marginal social value of a dollar transferred from a lender to a borrower is greater \textit{ex post} when interest rates are high, therefore interest rate stabilization in order to relieve borrowing banks is a desirable objective. Also the effect is irrelevant if regional banks shocks are perfectly correlated. In this case there is no trade \textit{ex post} and being $(e_1 - x - \tau_k c_1)$ zero for all $k$'s it follows that $V_2 = 0$.\footnote{It is useful to observe that if we take seriously assumption 1 constrained Pareto inefficiency is a slightly misleading term. The source of the market imperfection is the fact that individual balance sheets cannot be monitored (cannot prevent side trades). $c_1$ is part of a bank's balance sheets, and assuming that the planner can monitor it violates a primitive assumption of the model. Therefore despite the 'inefficiency' the allocation just described could be derived as the second best solution of an appropriately stated planner problem (i.e. a mechanism where competitive side trades can't be prevented).}

The following proposition summarizes the previous discussion and states a result of excess illiquidity for an economy with no aggregate risk. This result can be extended by a continuity argument to economies with small amount of aggregate risk. This shows that under assumptions 4 and 5 economies with small amounts of aggregate risk display excess illiquidity.

\textbf{Proposition 1} \textit{If (2.7) is non-zero the banking economy is constrained Pareto inefficient. In particular if $EV_2 < 0$ banks supply excessive deposits \textit{ex ante} and the interest rate is excessively reactive to aggregate liquidity shocks. Consider an economy with}
no aggregate risk that satisfies Assumptions 4 and 5. In this economy the inefficiency term (2.7) is negative.

**Proof.** It has been shown in the text that the inefficiency depends on the sign of the expression 2.7. Notice that $V_2$ can be rewritten

$$V_2 = E[u'(c_{2k})(e_1 - x - \tau_k c_1)|\tau] \frac{dr}{dc_1}$$

Notice that $E[(e_1 - x - \tau_k c_k)|\tau] = 0$ by market clearing. This follows from the symmetry assumption that guarantees that conditional on $\tau$ the cross sectional distribution of $\tau_k$ corresponds to the distribution faced by an individual bank. Notice also that $c_{2k}$ is decreasing in $\tau_k$ by the bank budget constraint provided that $c_1 > c_{2k}/r$. This inequality holds in an economy with no aggregate risk that satisfies assumptions 4 and 5 (see lemma 2 in the appendix). With a strictly concave utility function and a non degenerate distribution of $\tau_k$, $u'(c_{2k})$ is decreasing with $(e_1 - x - \tau_k c_k)$. This implies $E[u'(c_{2k})(e_1 - x - \tau_k c_1)|\tau] < 0$. Finally $\frac{dr}{de}$ is a positive constant given that $r = f'(e - \tau c_1)$, and $f$ is strictly concave because $x < \bar{x}$ in an economy that satisfies assumption 4 (see lemma 2).

The substantial interest of this result depends on the magnitude of the externality captured in expression (2.7). Therefore, this is the place to discuss under which conditions we expect this term to be sizeable.

- The first condition is that some regional risk has to be present, if there are no region specific shocks $e_1 - x - \tau c_1$ is always zero, all banks have a zero net position on the federal funds market and the economy is constrained efficient. The volume of trades on the money market documented in the introduction indicates that this term is likely to be sizeable in actual economies.

- The second condition is that $u''$ has to be large. If agents are risk neutral (at least in period 2) $u'$ is constant and movements in $c_{2k}$ have no consequences on net welfare. The underlying economic idea here is that when banks balance sheets are hit the marginal value of a dollar in a liquidity needy bank is different
(larger) than the marginal value of a dollar funds for liquidity abundant banks. A possible interpretation of this condition is that it captures, in a reduced form way, some imperfections in the credit market, such that a project that was financed by a bank $i$ cannot be easily transferred to a different bank $j$. In this case even if we replaced consumers with risk neutral entrepreneurs we could derive a profit function which is concave in the amount of funds available from the firms' bank.

- The third condition concerns the term $\frac{dr}{dc_1}$. This terms depends on the curvature of $f$ and captures the ability of the real economy to adjust to temporary movements in demand by e.g. curtailing investment in equipment, running down inventories.

The last result in proposition 1 tells us that in an economy with no aggregate risk it would be desirable to reduce the interest rate to reallocate liquidity. This is a result about an excessive interest rate level. When we introduce aggregate risk an economy characterized by aggregate illiquidity ($EV_2 < 0$) will also be characterized by an interest rate $r$ that reacts too much to liquidity shocks (in particular $\frac{1}{r} \frac{dr}{dr} = c_1$ is too large). This is a first sense in which this economy displays excess interest rate volatility. To be clear, this type of excess interest volatility is simply an external manifestation of an underlying problem of excess illiquidity.

We can however discuss excessive interest rate volatility in a second, more substantial, sense. This second notion of excessive volatility refers to the case in which the distortionary term $V_2$ is monotone increasing in $r$. When this is the case the gains from liquidity reallocation are larger when the interest rate is larger and aggregate liquidity is scarcer. In other words, in this case the welfare gain from reducing the interest rate is larger when the interest rate is higher, and there is a motive for reducing interest rate spikes. It is not easy to establish monotonicity of $V_2$ without the help of specific assumptions on $f$ and $u$, but we can prove a very general result about how the sign of $V_2$ changes with $r$. This result is established in the following proposition, that does not rely on assumptions 4 and 5. The proof is in the appendix.
**Proposition 2** Consider an economy with aggregate and regional risk and a strictly concave $f$. There exists a $\hat{\tau} \in [0,1]$ such that at a competitive equilibrium $V_2 < 0$ and $r > 1$ if $\tau > \hat{\tau}$, $V_2 = 0$ and $r = 1$ if $\tau = \hat{\tau}$, and $V_2 > 0$ and $r < 1$ if $\tau < \hat{\tau}$.

This proposition and proposition 1 are different in spirit, and they highlight two complementary aspects of the constrained inefficiency. Proposition 1 considers an economy that is mostly affected by scarce liquidity, because of assumptions 4 and 5. In that economy the main problem are spikes in the interest rate above 1 in periods of high liquidity needs, and the inefficiency is characterized by excess illiquidity at date 0. Proposition 2 has a more symmetric flavor, it describes an economy where there can be both low interest rates that are excessively low and high interest rates that are excessively high. At the same time, under the more general hypotheses of proposition 2 the sign of $EV_2$ at date 0 is generally ambiguous. Therefore, we have two distinct results, a result of excessive illiquidity for economies that satisfy assumptions 4 and 5 and are subject to small aggregate shocks, and a result about excessive movements of the interest rate above and below the rate of time preference that is valid more generally.

Finally, let me illustrate the monotone relationship between $r$ and $V_2$ in a simple numerical example. Figure ?? plots $V_2$ for an economy with CRRA utility, $u'(c) = c^{-\gamma}$\textsuperscript{12}.

Despite the symmetry of the aggregate shock this economy is excessively illiquid at date 0 in the sense of proposition 1, that is $EV_2 < 0$ in equilibrium. Moreover, the measure of conditional illiquidity $V_2$ is decreasing in $\tau$, that is, high interest rates are associated with a greater upward distortion of the interest rate itself. This picture illustrates a typical case where the interest rate is excessively volatile in the following senses: (i) it is excessively reactive to the aggregate liquidity shocks $\tau$, (ii) there are welfare gains from reducing high rates and increasing low rates, and (iii) the gains are larger in absolute value for extreme values of the interest rate.

\textsuperscript{12}The parameters for Figure 1 are: $\gamma = 2$, $e_1 = e_2 = 1$, $f(x) = 1 - 2x^2$. The aggregate shock $\tau$ is uniformly distributed on $[.4,.6]$. The regional shocks follow the linear model with $d = \{-1,1\}$ with equal probabilities.
2.2.1 The difference between derivatives and intermediaries

Since the inefficiency derived in proposition 1 is related to the excess cost of liquid funds at date 1, one may wonder whether the introduction of more sophisticated financial products could overcome the inefficiency. In the context of this model, though, we can show that the introduction of derivatives is irrelevant. Suppose banks can trade at date 0 all kinds of derivatives based on the price \( r \). A large set of derivatives can essentially reproduce a set of state contingent assets, so we assume that a complete set of securities contingent on each possible realization of \( r \) is available. Let the price of one unit of consumption in the aggregate state \( r \) be \( p(r) \) in terms of consumption at time 0.

The problem of banks at date 0 is:

\[
\begin{align*}
\max \quad & c_0 + E \tau_k u(c_1) + (1 - \tau_k) u(c_{2k}) \\
\text{s.t.} \quad & \tau_k c_1 + q(1 - \tau_k) c_{2k} = e_1 - x + q(e_2 + f(x)) + w(r) \\
& \int w(r)p(r)dG(q) + c_0 = c_0
\end{align*}
\]
Optimality of the portfolio \( w \) requires that the first order condition \( E[\lambda_k|r] = E[r u'(c_k)|r] = p(r) \) is satisfied for every \( k \). However, given the symmetry of the uncertainty setup notice that \( E[r u'(c_k)|r] \) is equal across regions in the equilibrium with no derivatives. Therefore we can take an equilibrium of the economy without derivative securities and let \( p(r) = E[r u'(c_k)|r] \). Then we can show that the prices \( \{r_r, p(r_r)\} \) the original consumption allocation together with \( w(r) = 0 \) for all regional banks \( i \) constitute an equilibrium of the economy with derivatives. Therefore, we have proved the following.

**Remark 1** Suppose a complete set of financial assets with payoffs conditional on the interest rate \( r \) are traded at date 0. The equilibrium of the economy with no date 0 financial markets corresponds to an equilibrium of the economy with date 0 financial markets. The latter equilibrium is characterized by no trade ex ante.

The essential difference between an option contract and a credit line appears clear here. If two regional banks could monitor each other’s balance sheet a bank could commit to use his credit line contingent on its total borrowing. That is, under an intermediation contract an agent can exercise his option to borrow cheaply only when he actually needs the funds, and not just in order to reinvest it on the money market. Suppose that instead a regional bank buys from another bank an option contract. If it has the option to borrow at —say— 5% and the interbank rate is 5.5% the first bank will exercise the option no matter what the regional liquidity needs are. To put it another way, the *ex post* transfer induced by the option cannot be made contingent on individual shocks. Given the symmetry assumption we made the value of this option will be the same for both banks and the option will not be traded in equilibrium.

This distinction between the option contract and the credit line is of practical relevance. Banks know that when offering a loan commitment they incur the risk that the client will just use his line when the money market rates are high and refinance the loan outside when the rates are low\(^{13}\). In dealing with commercial customers banks can use many ways to protect themselves from this misuse of commitments:

\(^{13}\)See Stigum (1990) pp.149-153.
for example they can monitor firms balance sheets and current account activity in order to check that the use of the credit line is not immediately followed by money market operations. In dealing with other banks, though, this monitoring is much harder, given that the other banks business is to be on the money market all the time. It is the very fact that banks are in the same business of providing liquidity that makes harder for them to provide liquidity to each other. The ability to check that credit lines are used only for 'real' needs is what essentially makes an intermediation contract different from an anonymous, market-based instrument of insurance like an option.

2.3 Intermediation and illiquidity

In the previous section we have seen under what conditions an illiquidity effect arises in the money market. At first glance, the presence of intermediaries in this setup should alleviate the illiquidity problem, given that a bank reallocates liquidity among its customers in a more efficient way. At the same time, though, the presence of intermediaries affects the functioning of the money market because a bank will behave differently from the sum of its customers. In this section I will show that, due to this effect, the presence of an intermediation system may actually exacerbate the illiquidity problem of a decentralized financial system. This does not mean that more financial integration is welfare reducing, rather that the presence of a more integrated financial system while it improves the efficiency of internal liquidity channels can hurt the efficiency of the external liquidity channel (the money market) by making the market price more volatile. This in turns can widen the gap between the constrained efficient economy and the competitive equilibrium.

The objective of this section is to study how the presence of different levels of intermediation affects the illiquidity problem illustrated in the previous section. We know that if a bank could cover all regions there would be no constrained inefficiency: every bank will be affected by the same aggregate shock \( \tau \), there would be no trade on the money market and the expression (2.7) would be identically zero. Therefore, at
high levels of integration the presence of a banking system eliminates the inefficiency. I will show here that at lower levels of integration this is not necessarily the case. It is actually possible that moving from a very disintegrated system to a system with more intermediation the illiquidity problem worsens. That is, the relationship between integration of the banking system and illiquidity can be non-monotone. The essential intuition behind this result is simple. When two banks merge each is able to supply liquidity to the other every time the liquidity shocks of the two banks are asyncronized. This induces them to issue more commitments \textit{ex ante} and worsens the aggregate illiquidity problem. At the same time when the two banks merge they need to use the money market \textit{ex post} less, so the cost of illiquidity and interest rate volatility will be smaller for them. The first effect makes the expression $V_2$ larger, while the second effect will tend to reduce it. A hump-shaped relation between financial integration and illiquidity arises if the first effect dominates at low levels of integration and the second effect dominates for higher levels of integration.

I will study the two effects separately in the following two propositions and then I will show with an example the possible shape of the relation between integration and illiquidity. First of all, though, we need to introduce a metric to give an operational meaning to the notion of "more bank integration". In order to do that consider for a moment the following environment. Suppose that there is a large number $m$ of consumers with independent shocks $\theta \in \{0, 1\}$ where $\theta = 1$ with probability $\mu$. Then, if each bank serves $s$ consumers we will have $n = m/s$ "regions" and they will have bank-specific shocks distributed as a normalized Bernoulli, $\tau_k = \frac{\theta}{s}$. The variance of the shocks $\tau_k$ will be smaller as $s$ increases. Moreover, $\tau_k$ will be approximately distributed as a normal and we can write $\tau_k \simeq \mu + \beta s d_k$ where $d_k$ is a standardized normal distribution and $\beta s$ is decreasing in $s$. Therefore, we can consider the model in the previous section with no aggregate risk and a linear specification of the shocks as an approximation of this environment, when $m$ and $s$ are large numbers. This environment suggests to parametrize the level of integration of the banking system using the coefficient $\beta$ in the equation above. More specifically, I will consider economies where the liquidity shocks are described by the linear model (2.1) and I will say that
an intermediated economy is more integrated if, everything else equal, it is characterized by a smaller \( \beta \). Essentially, when banks can service more consumers they will exploit better the law of large numbers in the internal allocation of liquidity, and their bank-specific shock will be more concentrated around the economy-wide shock.

The first proposition uses the first order condition (2.6) to show that the equilibrium level of commitments is larger for economies characterized by more integration. The proofs of both propositions are presented in the appendix.

**Proposition 3** Consider an economy with no aggregate risk and linear shocks, that satisfies assumptions 4, 5 and \( u''' \geq 0 \). The level of commitments \( c_1 \) is decreasing with \( \beta \).

Essentially, when banks are larger they can better allocate internal liquidity across consumers with offsetting liquidity shocks. This makes the expected marginal utility of income at date 1 less volatile, and induces them to increase the level of commitments. As a consequence they tend to be less liquid and to offer larger deposits \( c_1 \). This in turns, as we saw in the previous section, increases the ex post interest rate determined by (2.5).

Notice, though, that at the same time since banks are more integrated they will resort less to the interbank market for their liquidity needs. This second effect is captured in the following proposition.

**Proposition 4** Consider an economy with no aggregate risk that satisfies assumptions 4 and 5. For a given \( \hat{\beta} \) fix \( \hat{c}_1 \) at the corresponding equilibrium level. The absolute value of the expression \( V_2(\hat{c}_1, \hat{c}_1; \tau) \) is decreasing with \( \beta \) at \( \hat{\beta} \).

This proposition shows that as \( \beta \) increases, keeping fixed the average illiquidity of the economy, bank-specific liquidity shocks become smaller and the welfare effect of an increase in the interest rate is less relevant. The total effect of increased integration on \( V_2 \) is therefore ambiguous as it depends on these two opposing forces: reduced precaution by banks increases aggregate illiquidity (\( c_1 \)) and this tends to increase \( V_2 \).
on the other hand, keeping fixed $c_1$ better integration reduces the need for market channeled reallocation and this tends to reduce $V_2$.

To see how these two effects balance each other at various level of integration I study a simple example. I consider an economy with CRRA utility, $u'(c) = c^{-\gamma}$ a production function which is quadratic for low levels of investment and linear for higher levels, and I allow for two aggregate shocks.\footnote{The parameters for Figure 1 are: $\gamma = 2$, $e_1 = e_2 = 1$, $f(x) = 1 - 2x^2$ for $x < 0$ and $f(x) = x$ for $x > 0$. The aggregate shock is $\tau \in \{5, .52\}$ with probabilities $\{.7, .3\}$. The regional shocks follow the linear model with $d = \{-0.01, .01\}$ with equal probabilities.}

To represent the inefficiency I have plotted in Figure ?? the real interest rate in the state of scarce aggregate liquidity in the competitive case $r^{CE}$ and in the constrained efficient case $r^{Ef}$ for different values of $\beta$. Notice that in this example when the low aggregate liquidity shock hits $r^{CE} = r^{Ef} = 1$. The difference between these two interest rates is one possible measure of the severity of the illiquidity problem in the economy. Remember that a value of $\beta = 0$ corresponds to the case of maximum integration in which a bank can cover all consumers in the economy (i.e. a perfectly representative sample). In that case we see that there is no inefficiency and $r^{CE}_1 = r^{Ef}_1$. Also we see that for $\beta > 0$ the economy is constrained inefficient. But it is interesting to notice that the gap between the two levels of interest rates is larger for intermediate levels of $\beta$. That is, as we approach low levels of financial integration (high levels of $\beta$) the illiquidity problem of the money market tends to be less pronounced. This is essentially due to the prevalence of the first effect, and in particular to the curvature of the utility function which imparts a precautionary behavior to small intermediaries. When internal insurance is not well developed banks precautionary behavior dominates and as banks keep a low level of commitments this keeps $r$ close to 1. When instead banks can pool together the risk of a larger number of customers they tend to be less liquid. Moreover, they tend to increase their commitments faster than what would be required in the second best. Therefore the gap between second best commitments and competitive commitments expands.

Notice that also the constrained efficient solution requires banks to be less liquid
Figure 2-2: Intermediation and illiquidity

for lower levels of $\beta$. As banks can sustain better interest rate shocks and the wealth effects are less relevant, the interest rate is more volatile at the second best solution. Essentially as wealth effects are less relevant from a welfare point of view it is optimal to let the interest rate move more in order to fulfil its intertemporal allocation purpose. Notice that the upper bound on the interest rate corresponds to the shadow interest rate at the first best allocation\(^\text{15}\). What happens in this example is that as $\beta$ is reduced banks tend to increase their commitments faster than what would be required by the constrained efficient allocation. In short interest rate volatility increases with financial integration in both the competitive case and in the constrained efficient solution, but the gap between the constrained efficient allocation and the competitive solution increases first and then declines as both allocations converge to the first best.

The conclusion of this section is that intermediated finance acts as an illiquidity multiplier at intermediate stages of banks evolution. This helps to explain why central banks have typically been introduced after the expansion of intermediated finance,

\(^{15}\)To be precise it corresponds to the shadow interest rate at the first best allocation subject to the use of deposit contracts.
and especially in those systems (as the UK and the US) where the financial market played a central role in reallocating resources among banks. Also this analysis warns us that the expansion of the ability of banks to pool liquidity risk can have destabilizing effects on the money market. What makes policy conclusions particularly difficult to draw is the fact that also the second best solution does involve an increase in interest rate volatility as financial integration proceeds, therefore an increase in volatility following banking deregulation and expansion is not necessarily a sign of inefficiency. Inefficiency is not related to the level of volatility of the interest rate but to the difference between the two levels of volatility: the competitive level and the second best level. The difference between the two is what we call excess volatility. In certain circumstances, it is possible that banks' deregulation and expansion is associated with a greater excess volatility defined in this sense.

2.4 Monetary policy and interest rate stabilization

In this section I introduce nominal contracts and monetary policy. Assume that deposits instead of promising real units of goods promise a fixed amount of dollars $d$. The presence of deposit contracts denominated in units of account gives the central bank some ability to influence the interest rates.

Monetary policy works in the following way. Banks clear their positions through an account at the central bank—the reserve account—, denominated in dollars. This account pays zero interest on positive balances held at the end of a period and requires the payment of a prohibitive interest on negative balances (discount window borrowing). Therefore, provided that the market nominal rate is positive, every bank with excess funds will rather lend them on the interbank market and each bank with a lack of funds will rather borrow in the interbank market. This means that every bank will act so as to keep his balance with the central bank at zero, and equilibrium can be achieved only if the central bank offers zero settlement balances.

With this arrangement the central bank can move the interbank rate by supplying a minimal amount of reserve balances different from zero. For example, suppose the
central bank wants to push up the interest rate \( i \). It can do so by just borrowing a minimal amount \( \epsilon \) on the interbank market. The central bank does a repo operation borrowing \( \epsilon \) from a bank, then the correspondent bank has a negative position of \( \epsilon \) written on its reserve account, and tries to balance it at the end of the day by borrowing \( \epsilon \). Since no bank wants to end up with a negative position in its reserve account, the ultimate effect of this operation is to generate an excess demand for funds on the funds market that pushes up the interest rate\(^{16}\).

This mechanism allows the central bank to peg the interest rate in terms of the unit of account used to denominate the reserve balances. This has effects on the real economy because banks' contract are defined in terms of that unit of account. In this model, for the sake of simplicity, this is the only type of nominal rigidity. We could also introduce nominal rigidities on the production side of the economy. In this way monetary policy would have effects on output and would affect real rates also through this channel. I discuss this extension in subsection 2.4.3. By adopting this simple cashless model I do not need to specify the details of the demand for outside money (i.e. positive reserve balances), monetary control is achieved with zero open market operations in equilibrium, and we do not need to specify the government budget.

Let \( p_t \) be the price level at date \( t \). A monetary policy rule is described by a wicksellian rule, that is by a map \( i(p_t) \). Assume also that date 2 price level \( p_2 \) is fixed at 1 by a government commitment to accept nominal claims for one dollar in exchange for one unit of good. The interest rate rule and the fiscal commitment at date 2 are sufficient to fully characterize the effects of monetary policy in this model by selecting a determinate equilibrium of the real economy.

The equilibrium of the monetary model is characterized by the Fisher equation

\[
i(p_t)p_t = r
\]

\(^{16}\)This appears a quite realistic description of the working of open market operations. A further step towards realism would be to add a small and steep precautionary demand for reserve balances. This addition would not change the essentials of the mechanism. For a discussion of the microstructure of open market operations in a cashless world see Woodford (2000).

48
and by the equilibrium conditions:

\[
E \left[ \frac{\tau_k}{p_1} \left( u'(c_1) - ru' (c_{2k}) \right) \right] = 0
\]  

(2.8)

\[
c_1 = \frac{d}{p_1}
\]

\[
\tau c_1 = e_1 - x
\]

\[
r = f'(x)
\]

\[
\tau_k c_1 + (1 - \tau_k)c_{2k} = e_1 - x + (e_2 + f(x))/r
\]

The first equation corresponds to optimization \textit{ex ante} in terms of nominal commitments \(d\). This condition will turn out to be the crucial constraint for the monetary authority, because a policy that induces inflation of the price level \(p_1\) will induce banks to issue larger nominal value of commitments to increase the value of \(c_1\) and this will limit the ability of the central bank to influence the value of commitments.

I will focus on equilibria where \(p\) fully reveals \(\tau\). In this case the monetary policy can be described equivalently as a map \(c_1 = c(\tau)\) that satisfies the conditions above where the first two equilibrium equations are replaced by the condition:

\[
E \left[ \tau_k c_1 (u'(c_1) - ru' (c_{2k})) \right] = 0
\]  

(2.9)

The optimal monetary policy will maximize the utility of the representative consumer \textit{ex ante} subject to the constraints above.

Consider the value function \(V(c_1, c_1; \tau)\) defined in section 2.2. Notice that now the value of \(c_1\) can be dependent on the shock \(\tau\) thanks to nominal variability. Then the problem of optimal monetary policy can be stated in terms of the function \(c(\tau)\) as

\[
\max_{c(\cdot)} \quad EV(c, c; \tau)
\]

(2.10)

\[
s.t. \quad EcV_1(c, c; \tau) = 0
\]

(2.11)

where the constraint (2.11) corresponds to the condition (2.9) above.
To check that the constraint set defined by this condition is not empty notice that a constant level of $c$ equal to the real equilibrium described in the previous section satisfies (2.11)\textsuperscript{17}. At the same time it is clear that not any map $i(p)$ or $c(\tau)$ results in a feasible policy. Notice also that in this simple setup there is a trivial form of nominal indeterminacy as for any equilibrium with $p(\tau)$ and $d$ there is an equilibrium with $p' = \alpha p$ and $d' = \alpha d$. The indeterminacy is immaterial to the real side and to the study of price and interest rate variability. We eliminate it by letting $d = 1$.

2.4.1 No regional risk

We start considering the case of no regional risk. In this case monetary policy has an easy remedy to the only imperfection in the private economy: the lack of contingent banking contracts. By letting $p$ move the central bank can make $c_1$ contingent on the aggregate state of the economy, and can do so as to achieve the first best allocation. To check that this is the case notice that when $\tau_k = \tau$ the first best consumption $c(\tau)$ satisfies all the constraints above. In particular the first best satisfies the budget equation because —all banks being identical ex post— there is no trade in equilibrium. It satisfies condition (2.8) because the first best allocation satisfies $u'(c_1) - ru'(c_2) = 0$ conditional on any $\tau$. The optimal consumption rule $c(\tau)$ can be implemented by the optimal interest rate rule $i(p) = \frac{1}{p} f'(e_1 - \tau \frac{1}{p})$. This rule involves stabilization of the real interest rate in the following sense, for a given level of deposits monetary policy makes $p$ moves in the same direction as $\tau$ and in this way makes $r$ less reactive to $\tau$. The nominal interest is stabilized even more given that for high levels of $\tau$ both $r$ is reduced and $p$ is increased with respect to the fixed price allocation.

We have thus proved the following result.

**Proposition 5** When there is only aggregate risk optimal monetary policy achieves the first best allocation.

\textsuperscript{17}Notice that this fixed price equilibrium is achievable only in the limit by an interest targeting policy of the type described above, because if $p_1$ is fixed at 1 the central bank has no way of observing $t$. Therefore, we can show that the constraint set is non-empty using the observation above together with a continuity argument.
This proposition is not very surprising. Since nominal contracts together with an appropriate monetary policy can make $c_1$ state contingent, while private contracts can not, a monetary policy with a certain degree of price instability is beneficial. The result is analogous to proposition 2 in Gale and Vives (1999) that apply it to the study of exchange rate policy and dollarization. It is also related to the notion that nominal variability can substitute for the lack of state contingent contracts, an idea that has been developed in various strands of the macro literature with nominal rigidities. The closest case to the one treated here is that of models where the only contract fixed in nominal terms is government debt. Calvo and Guidotti (1993) and more recently Sims (1999) and Woodford (1999) in the context of ‘fiscal theory’ models\textsuperscript{18} have shown that price level variability can generate a source of contingent taxation by affecting the real value of the government liabilities. In this economy unanticipated shocks to the price level act like a capital levy on holders of government nominal liabilities. Variable inflation imparts a state contingent payoff to nominal assets and disposes for the need of state contingent taxation of labor income that would be more distortionary. If the monetary authority has no problems of credibility adjusting the price level \textit{ex post} is a non-distortionary way of adjusting tax revenues to government expenditure by adjusting the \textit{ex post} inflation levy. Clearly the beneficial effect of price level instability appears only because markets are incomplete and there are no assets contingent on government expenditure.

In our model the role of nominal instability is similar. The level of nominal deposits is chosen at time zero and a variable price level allows a contingent transfer from early to late consumers, this can be done without jeopardizing the incentive to keep nominal deposits \textit{ex ante}. That a given level of $d$ remains an optimal choice \textit{ex ante} when the variable price level is taken into account is guaranteed by condition (2.8). This condition plays the same role as the condition that holdings of nominal claims is \textit{ex ante} optimal given expected inflation in models of government finance. The main difference of the present model is that the beneficiary of the contingent transfer is not the government but the banks hit by liquidity shocks. When the high liquidity shock

\textsuperscript{18}It is modeled in detail in section 4 of Christiano and Fitzgerald (2000).
hits and consumers use credit lines and withdraw deposits denominated in nominal terms, an increase in the price level lessens the real burden on the banking system.

In the literature on government debt deflation a crucial issue is credibility and time consistency of the optimal policy. Here, in the case of only aggregate shocks, it is possible to show that the optimal policy is also time consistent. As we will see shortly this ceases to be the cases when idiosyncratic shocks are reintroduced and the government will be tempted to reduce the value of deposits to address the inefficiency problem derived in proposition 1 above.

It is useful to reconsider the mechanism in terms of monetary aggregates. In this model there is only bank money. Assume that the investment expenditures at date 1 are financed by a credit line for $p_1x$ issued at the beginning of period 1, matched by a balancing deposit account. Then total bank money in the economy is $M = px + d$. The total volume of transactions financed at date 1 is $T = px + \tau d$. Therefore money velocity in this economy is equal to

$$\frac{T}{M} = \frac{px + \tau d}{px + d}.$$ 

This means that in this model the aggregate liquidity shock $\tau$ is a source of velocity shocks. Finally we have a quantity-like equation $T = pe_1$ that determines the price level.

With this definitions we can reinterpret the action of the central bank as allowing an expansion of bank money to finance investment that induces a reduction of real consumption on the consumer side. This is related to Wicksell's idea of 'forced savings', generated by a monetary expansion. Bank money expands by $\Delta M = p\Delta x$ to finance firms investment, this increases the price level and hurts the consumption level of holders of nominal commitments. Notice that savings by consumers correspond to $e_1 - \frac{\tau d}{p}$ in real terms, therefore an increase in $p$ increases real savings to match the increased investment by firms. By stabilizing the interest rate the central bank allows commercial banks to collect this form of 'banks seignorage' that supports investment and the consumption of late consumers.
2.4.2 Liquidity reallocation

Consider now the case of aggregate and regional shocks, when there is a non-degenerate distribution of the $\tau_k$. In this case there is an additional gain from interest rate manipulation. The reduction of $r$ achieved through a reduction of the nominal interest rate impacts now on the cost of borrowing, favoring liquidity needy regions. The central bank, though, is constrained in its ability to deflate deposits by the condition that banks be willing to hold a given amount of nominal deposits at date 0.

Going back to the general formulation of optimal monetary policy in (2.10) it is clear that the result of the previous subsection is due to the fact that with no regional shocks the condition $V_2 = 0$ holds at the first best. When this is the case the unconstrained optimum—that satisfies the necessary condition $V_1 + V_2 = 0$—satisfies also the constraint (2.11) automatically. In presence of idiosyncratic shocks, instead, $V_2$ can be non zero, in particular under the conditions of proposition 1 the term is negative reflecting the illiquidity of the banking sector. This generates an additional motive for interest rate stabilization, that I call the liquidity reallocation motive. By reducing the interest rate when liquidity needs are different across regions and $V_2$ is large in absolute terms the central bank reallocates funds from liquidity abundant regions to liquidity scarce regions.

Proposition 2 and the examples presented in section 2.2 suggest that the term $V_2$ will be larger for larger values of $r$. This means that when aggregate liquidity is scarcer the gains from liquidity reallocation are larger. Therefore, the liquidity reallocation effect will tend to make the optimal interest rate flatter in presence of idiosyncratic shocks. The central bank will let deposits deflate, by keeping the interest rate low when $p$ grows, and inflate them when $p$ is small and there is abundant liquidity in the interbank market.

Before discussing further the optimal policy I will first introduce a negative result on what monetary policy can not do in presence of idiosyncratic shocks and on the time inconsistency of optimal monetary policy in this case.

Let me define the constrained efficient contingent allocation as the allocation that
arises in the *ex post* equilibrium when $c(\tau)$ is obtained from the maximization (2.10) omitting the constraint (2.11). This allocation is characterized by the necessary condition $V_1 + V_2 = 0$. Therefore in the case of pure idiosyncratic shocks we can establish the following proposition.

**Proposition 6** Consider an economy with only regional risk that satisfies the assumptions of proposition 1, so that the real equilibrium is constrained inefficient. In this economy monetary policy can not achieve the constrained efficient contingent allocation, at the constrained efficient contingent allocation $V_1(c, c; \tau) > 0$.

Let me discuss two consequences of this proposition. First notice that when only regional shocks are present the central bank will not be able to use monetary policy at all to solve the illiquidity problem. The power of monetary policy is the ability to deflate deposits to increase liquidity in the interbank market. But if there are no aggregate shocks monetary policy will be expected to deflate deposits with probability one and banks will simply adjust up the nominal value of deposits to keep up with the expected inflation. Notice that here I have been implicitly focused on deterministic monetary policies, that is a policy that to any realization of $\tau$ assigns only one value of $c$ (i.e. a policy that assigns to any $p$ a single value of $i$). Therefore, we may imagine that a randomized monetary policy that satisfies constraint (2.11) can be superior to the real equilibrium with $p = 1$. This is a theoretical possibility which I leave aside for the moment. I conjecture that under a plausible parametrization the optimal policy will be deterministic.

A second consequence is that optimal policy will not be time consistent. Once banks choose a certain level of nominal deposits the central bank would like to implement the constrained efficient allocation. But as we saw above at the constrained efficient allocation the condition in (2.11) holds as an inequality. This means that in the case of time inconsistency the equilibrium with nominal deposits unravels. In this case there is no equilibrium with nominal deposits. The model could be amended in various ways in order to analyze the case of no-commitment. One possibility would be to introduce some cost of nominal instability in the central bank objective function,
if other contracts (e.g. prices) are set in nominal terms a more volatile policy will involve misallocation costs at date 1 and a central bank that lacks commitment would have an incentive to limit nominal variability.

Let now go back to the optimal policy with commitment. In presence of aggregate shocks the central bank tries to correct the inefficiency in the following way: deflate deposits at times when liquidity reallocation is particularly beneficial (high $V_2$) and inflate them at times when liquidity reallocation is less necessary. In this way the bank induces the banks to issue the desired amount of nominal deposits ex ante and uses its power to deflate deposits only when liquidity reallocation motive is stronger.

![Figure 2-3: Optimal monetary policy](image)

Figure 2.10 illustrates optimal monetary policy for an example economy\textsuperscript{19}, $r^{CE}$ corresponds to the competitive equilibrium, $r^{OM}$ correspond to the optimal monetary policy, $r^M$ is the equilibrium that arises if monetary policy allows for contingent consumption using the variability of the price level but disregards the liquidity reallocation effect (i.e. the monetary policy that arises by solving $V_1 = 0$ state by state),

\textsuperscript{19}The parameters for Figure 2 are: $\gamma = 5$, $e_1 = e_2 = 1$, $f(x) = 1 - 2x^2$ for $x < 0$ and $f(x) = x$ for $x > 0$. The aggregate shock is $\tau \in \{.5, .52\}$ with probabilities {.7, .3}. The regional shocks follow the linear model with $d = \{-0.01, .01\}$ with equal probabilities.
finally $r^{Ef}$ is the constrained efficient (state contingent) allocation. The reduction of the high interest rate from $r^{CE}$ to $r^M$ corresponds to the effect studied in the previous subsection. The difference between $r^M$ and $r^{Ef}$ shows that the economy is still constrained inefficient, despite the sequential service constraint has been effectively removed. Optimal monetary policy though cannot reach the constrained efficient allocation because it has to satisfy the ex ante constraint on the nominal value of deposits, therefore optimal monetary policy, $r^{OM}$, involves both an interest reduction at high levels of the interest rate and an increase at low levels to induce banks to limit the commitments held.

In terms of monetary aggregates notice that the dilemma of monetary policy can be stated in the following terms. Liquidity insurance reallocates in an efficient way resources across banks. At the same time, though it induces banks to increase $d$. Recalling that velocity in this model is equal to $\frac{px+qd}{px+q}$, liquidity insurance makes the economy more exposed to velocity shocks. This shows in a different way the trade-off between nominal stability and liquidity insurance that arises in this type of models. It will be a topic for future research to study this trade off in more detail introducing explicit costs of nominal instability in the model. Finally, notice that in this model liquidity insurance by a central bank could be highly facilitated by explicit regulations that tend to limit the expansion of monetary aggregates like $d$, such as reserve requirements. That is, with the help of this type of regulations the central bank can achieve the same level of insurance with a lower level of nominal instability. It is also a topic for further research to study the interplay of direct regulation and liquidity insurance channeled through open market operations.

2.4.3 Alternative specifications of monetary policy

This model of monetary policy is extremely simplified, so let me discuss briefly how it could be extended to allow for outside money and for price rigidities.

First, I have assumed no outside money in order to emphasize the fact that the liquidity reallocation effect of monetary policy can be studied abstracting from the role of money as a medium of exchange. When there is a positive demand for money
different considerations arise given that the nominal interest rate represent also the opportunity cost of holding money balances. In this case there are additional motives for (nominal) interest rate stabilization, related to the reduction of the deadweight costs of positive nominal interest rates.

Secondly, in this model the ability of the central bank to affect the real interest rates only passes through the rigidity of the nominal deposits. The model could be enriched introducing a flexible supply of goods $e_1$ and allowing for fixed nominal prices. In this case when the central bank allows banks to expand inside money this would have also an effect on current production. Going back to the quantitative exposition at the end of subsection 2.4.1 if we allow $e_1$ to vary together with nominal demand $M$ this would give the central bank an enhanced ability to influence the interest rate using monetary policy (notice that an increase in $e_1$ reduces $r$ in equation 2.5). In this case if the central bank allows the banking system to increase nominal expenditure $M$, this increase may be satisfied by an increase in $e_1$, and have a small short term impact on $p_1$. The introduction of this type of price rigidities would also introduce a cost of nominal instability that is not present in the present model.

It is interesting to notice the symmetry between the model presented here and models with nominal price rigidity. Essentially, the constraint (2.11) above plays the role of the equation for optimal price fixing ex ante by the firms in models of price rigidities. In models of price rigidities it has been shown that under certain conditions optimal monetary policy will essentially replicate the flexible price equilibrium of an economy with no rigidities. In the current setup this is correct when there are only aggregate shocks, in that case the optimal monetary policy implements the equilibrium that would arise if $d$ could be made state contingent. When there are regional shocks, instead, the central bank will tend to dampen interest rate fluctuations with respect to the flexible deposits allocation, to correct the inefficiency discussed in section 2.2.
2.5 Conclusion

In this chapter I have developed a simple model of the interbank market with endogenous interest rate volatility. The unobservability of liquidity shocks and the inability of banks to monitor each others' portfolios produce financial market incompleteness. With incomplete financial markets the equilibrium of the economy is constrained Pareto inefficient and under plausible conditions the inefficiency takes the form of excess illiquidity and excess interest rate volatility.

A financial system with a limited degree of intermediation can display a greater measure of excess illiquidity than a system with no intermediation, despite the fact that excess illiquidity eventually disappears in in a system with maximum intermediation. When banks are able to pool together a larger number of consumers under the same intermediation contract, they tend to maintain less liquid balance sheets and higher levels of commitments. This has a destabilizing effect on the interest rate. Eventually, when the system is fully integrated, this effect is irrelevant, as banks completely substitute contractual sources of liquidity for market sources of liquidity. But, in the intermediate range, the market interest rate tends to be excessively volatile in terms of constrained efficiency.

If monetary policy can affect the short term real rate, it can reallocate liquidity across banks and enhance efficiency by reducing its variability. In this sense the model provides a rationale for the observed central banks' tendency to smooth interest rates. The interest rate matters here for the cross-sectional allocation of resources among banks subject to different shocks. A higher interest rate is associated with larger liquidity shocks, and under certain conditions larger liquidity shocks are associated with a larger gain from liquidity reallocation.

At the same time, the model highlights the dilemma faced by central bank engaging in interest rate stabilization. By supplying insurance the central bank worsens banks' incentives to maintain a liquid balance sheet. The problem is a type of market mediated moral hazard: by stabilizing the interest rate the central bank provides insurance against liquidity shocks but at the same time it reduces incentives for banks
to limit nominal commitments. Essentially central bank intervention makes it less costly to issue commitments that are fixed in nominal terms. This in turn exposes the economy to larger velocity shocks and greater nominal instability. In this sense liquidity insurance by the central bank generates a need for regulatory control of the monetary aggregates. This helps to explain why central bank support of the financial system liquidity has been traditionally associated with various regulations of bank balance sheets to control the expansion of money-type liabilities of the banks (e.g. reserve requirements, portfolio restrictions, caps on deposits rates). A study of the link between the provision of liquidity insurance through open market intervention and this type of regulation is a subject for future research.

2.6 References


Allen, Franklin and Douglas Gale, 2000, ”The Role of Banks and Markets in the Optimal Allocation of Liquidity”, unpublished manuscript.


Caballero, Ricardo and Arvind Krishnamurty, 2000, ”International and Domestic Collateral Constraints in a Model of Emerging Markets Crises”, unpublished manuscript.

Caballero, Ricardo and Arvind Krishnamurty, 2000, ”Dollarization of Liabilities”, unpublished manuscript.


Holmstrom, Bengt and Jean Tirole, 2000, "Models of Liquidity (Wicksell lectures)", unpublished manuscript


Sims, Christopher, 1999, "Fiscal Consequences for Mexico of Adopting the Dollar", unpublished manuscript.


Wicksell, Knut, 1936, Lectures in Political Economy, Vol. II

Woodford, Micheal, 1997, "Loan Commitments and Optimal Monetary Policy", Journal of Monetary Economics


2.7 Appendix

Definitions of $V$. The indirect utility function $V$ is defined as:

$$V(c_1, c_1; \tau) = E[\max_{c_2, x} \{\tau_k u(c_1) + (1 - \tau_k)u(c_2);$

$$s.t. \quad \tau_k c_1 + (1 - \tau_k)c_2/r = e_1 - x + (e_2 + f(x))/r$$

$$r = f'(e_1 - \tau c_1)]/\tau]$$

**Lemma 2** An economy with no aggregate risk that satisfies assumptions 4 and 5 has a unique equilibrium and in equilibrium the inequality $c_1 > w$ is satisfied.

**Proof.** Assumption 4 implies that

$$ru'(rw) - u'(w) < 0$$

for every $w$ and every $r > 1$ (the derivative of this expression is $u'(rc) + rcu''(rc) < 0$).

Define $d_k = \tau_k - \mu$. Consider the family of economies indexed by $\beta$ characterized by the distribution of shocks

$$\tau_k = \mu + \beta d_k$$

Consider the equilibrium of the economy with $\beta = 0$. Notice that this economy has a unique equilibrium characterized by $u'(c_1) - ru'(c_2) = 0$. Notice that $u'(\frac{e_1 - x}{\mu}) - f'(x)u'(\frac{e_1 + f(x)}{1 - \mu})$ is an increasing function of $x$ and is positive at $x = \hat{x}$. Therefore the equilibrium involves $x < \hat{x}$ and $r > 1$ and $ru'(rc_1) < u'(c_1) = ru'(c_2)$. This implies $rc_1 > c_2$. Letting $w = \mu c_1 + (1 - \mu)c_2/r$ this also implies that $c_1/w > 1$. Now notice that the equilibrium value of $c_1/w$ varies continuously with $\beta$ and notice that

$$c_{2k} = \frac{1 - \frac{\tau_k}{w} r}{1 - \tau_k} r w$$

If there is a $\beta \in (0, 1]$ such that $c_1/w = 1$ then for that $\beta$, $c_{2k}$ would be equal to a constant $c_2 = r'w' = r'c_1 = r'\frac{e_1 - x}{\mu}$. But notice that the pair $c_1', c_2'$ satisfies all
the conditions for an equilibrium of the economy with $\beta = 0$, and since $c'_2 = r'C'$ this is a contradiction. Therefore $c_1/w$ has to be bounded below 1 for all $\beta$ and the equilibrium with $\beta = 1$ is characterized by $c_{2k} < rw < rc_1$.  

**Lemma 3** If $\phi$ is an increasing function, $d$ is a random variable with $Ed = 0$ and $\beta' > \beta$ then:

$$E\phi(\beta'd)d' \geq E\phi(\beta)d$$

**Proof.** $E\phi(\beta'd)d' = E(\phi(\beta'd) - \phi(0))\beta'd \geq E(\phi(\beta d) - \phi(0))\beta d = E\phi(\beta)d$.  

**Proof of Proposition 2.** Define $w(x) = e_1 - x + (e_2 + f(x))/f'(x)$. We can easily show that concavity of $f$ implies that $w'(x) \geq 0$ and $\frac{d}{dx}[w(x)f'(x)] \leq 0$. Recall that $x = e_1 - \tau c_1$. Then the ratio $c_1/w$ is decreasing in $\tau$. Let $\hat{\tau}$ be such that $w(e_1 - \hat{\tau}c_1) = c_1$. Then $c_{2k}$ will be a decreasing function of $\tau_k$ for $\tau > \hat{\tau}$ and an increasing function for $\tau < \hat{\tau}$. As a consequence the covariance term $E[u'(c_{2k})(e_1 - x - \tau_k c_1)|\tau]$ changes sign for $\tau \leq \hat{\tau}$.

**Proof of Proposition 3.** Notice that $\tau u'(c_1) - E\tau_k ru'(c_{2k})$ is a decreasing function of $c_1$. Then, we need to show that if we fix an equilibrium for $\beta$, the following inequality holds for $\beta' > \beta$: $\tau u'(c_1) - E(\tau + \beta'd_k)ru'(c_{2k}) < 0$, where $c_{2k}$ is derived from the budget constraint when we substitute $\beta'$ for $\beta$. This can be established using the following inequality: $\tau r E u'(c_{2k}) + rE\beta'd_ku'(c_{2k}) > \tau r E u'(c_{2k}) + rE\beta d_ku'(c_{2k})$. This inequality can be proved in two parts. $E u'(c_{2k}) > E u'(c_{2k})$ is derived using the convexity of $u'$ and the fact that the function $\frac{1-\tau_k}{1-\tau_k}w$ is concave decreasing. $E\beta'd_ku'(c_{2k}) > E\beta d_k u'(c_{2k})$ follows immediately by lemma 3 since assumption 3 guarantees that $u'(c_{2k})$ is an increasing function of $\beta d_k$.

**Proof of Proposition 4.** Observe that $E u'(c_{2k})(e_1 - x - \tau_k c_1)\frac{df}{dc_1} = -E u'(c_{2k})\beta d_k \tau f''(x)$ thus we can immediately apply lemma 3.

**Proof of Proposition 6.** At the competitive allocation $c^* V_1 = 0$ and $V_2 < 0$ hold; at the constrained efficient allocation $c^* V_1 + V_2 = 0$ holds. If $V_2(c^*, c^*, \tau) > 0$,
by continuity of $V_1$ and $V_2$ and by the uniqueness of the equilibrium there must be a $c'$ in the interval between $c^e$ and $c^*$, such that $V_1 < 0$ and $V_2 = 0$ at $c'$. But $V_2(c', c'; \tau) = 0$ implies that the corresponding values for $c'_{2k}$ are constant, which is only compatible with $c'_1 = \frac{\omega}{r} = w$, and this contradicts $V_1 < 0$, because in this case $V_1 = \tau(u'(w) - ru'(rw)) > 0$ by virtue of Assumption 4.
Chapter 3

Fractional Reserves and Liquidity Multiplication

Joint with Christian Hellwig

This chapter studies an economy in which private banks multiply the liquid resources of an economy by issuing convertible claims against reserves of a real asset. The convertibility of the banks' claims plays a double role. First, if only a fraction of the claims are converted every period the bank does not need to hold 100% of his assets in liquid reserves and can invest the rest in more profitable private projects, therefore the bank operates as a liquidity multiplier. Secondly, the fact that every period a fraction of the claims are converted acts as a form of decentralized monitoring of the banks balance sheet and prevents banks from running a Ponzi scheme. There is a basic tension in this economy: increasing the level of monitoring guarantees the reliability of the banking system but reduces the total liquidity of the economy.

This chapter analyzes the role of banks in financial development from the point of view of liquidity models. In our setup banks transform a non pledgeable return into a liquid claim and they face the threat of termination of their banking privilege if they overexpand the amount of claims circulating. The chapter studies the relationship between the value of the bank franchise, the entry of banks and the incentives to overexpand credit and default.
We study an economy in which agents have access to a private investment technology that generates a stream of profits that cannot be pledged. Moreover, there is a real asset, trees, which is in fixed supply and pays a fixed unit return which is fully pledgeable. In this economy credit is limited by the amount of wealth accumulated in the form of trees: when an agent need resources to invest in its own private technology he can sell the trees accumulated, and buy trees when he receives the return on his investment. If the existing stock of trees is scarce the economy displays underinvestment in the private technology and imperfect consumption smoothing across time.

Then we allow some agents to setup a bank and finance their investment and consumption by issuing private claims. In the absence of a centralized system of accounts this economy has to solve two problems: (1) monitoring the amount of claims issued by a bank, and (2) punishing a bank that issues an excessive amount of claims. In this chapter we ask whether a reserve mechanism can solve the two problems above, to what extent and under what conditions. The reserve mechanism is the following: banks keep reserves of the tree-asset, issue private claims that are convertible in the underlying asset, if they fail to convert their claims they lose their ability to issue claims in the future. In terms of monitoring this mechanism can work under some conditions. First, if we want to maintain an atomistic setup, some friction must prevent banks from rolling over smoothly any amount of liabilities. Otherwise a Ponzi scheme (by an atomistic agent) would go completely undetected. Secondly, we must prevent a bank from accumulating infinite amounts of reserves. Otherwise the bank could still expand his liabilities to infinity without ever violating his reserve requirement, and infinite dilution would go undetected. We describe an environment in which these two requirements are satisfied and we show that the reserve requirement is eventually translated into a borrowing constraint.

The reserve mechanism faces more difficulties on the punishment side. We study the conditions for a steady state competitive equilibrium with banks and we show that the steady state value of the bank franchise is actually negative in a stationary environment. This is due to the fact that in an environment with scarce liquidity
a young bank that is allowed a large borrowing limit tends to exhaust her franchise value early in her life by smoothing consumption and over-investing. When the bank reaches her long run steady state she has exhausted the benefits of her license to borrow, so much so that she would actually be better off defaulting on her debt even though she will not be able to ever borrow again. Therefore, we are led to study an equilibrium with endemic bank failures, in which there is a constant creation of new banks that have a high franchise value and a constant outflow of defaulting banks. Notice that an equilibrium with less than 100be superior to an equilibrium with 100failures. This result in a sense questions the conventional wisdom about the success of the Canadian free banking experience versus the US experience. Fast redemption and no bank failures can simply be associated to an economy with low liquidity needs. In such an economy the banking system is very safe but at the same time it adds very little to the liquidity present in the economy. On the other hand a banking system with a slow redemption of notes and frequent bank failures can be the efficient response to a situation of scarce liquidity.

We also discuss modifications of the original setup that would allow for a no-default steady-state. In particular we study the introduction of growth and the introduction of a harsher punishment imposed by a central authority in the event of a default.

This chapter is related to some recent papers in the literature that have also described the role of banks as liquidity creators\(^1\). In particular, Kiyotaki and Moore (2001) have studied a model of scarce aggregate liquidity in which banks operate a costly technology that essentially transforms non-pledgeable returns into pledgeable returns. Diamond and Rajan (2001) have studied a model in which banks increase the pledgeable portion of a firm returns, by setting up a contract subject to a sequential service constraint, therefore exposing themselves to runs that reduce their incentive to renegotiate. In our model also banks can create liquidity, and they do so by virtue of a reputational mechanism. Essentially if they fail to satisfy their promises they are punished by losing the ability to issue claims in the future. In that dimension

\(^1\)As opposed to the role of banks as institutions that reallocate existing liquidity across agents, as in Diamond and Dybvig (1983) and in Holmstrom and Tirole (1999)
this chapter is also related to the literature on sovereign debt and on credit markets with default and limited enforcement (e.g. Levine and Kehoe (1998)). We know from Bulow and Rogoff (1989) that, in a decentralized market, if an agent can unilaterally default on his debts it is impossible to sustain an equilibrium with positive debt if the only threat he faces is the refusal of future loans. Here we take a course that is close to Bulow and Rogoff, in that we consider a very decentralized market setup, essentially assuming that there is not a central record of the total creditor and debtor positions of each agent, so that dilution of claims is a very serious problem. Nonetheless, we assume that a minimal legal structure exists so that if somebody issue claims against a reserve of the tree asset then he cannot hide his reserve holdings when he defaults. This slight deviation from the Bulow and Rogoff setup allows debt to be issued in equilibrium. Clearly, if debt had to be 100% backed by trees this would be essentially irrelevant for the economy. Therefore we need to establish the possibility of an equilibrium with less than 100% reserves. A paper closely related to ours is Cavalcanti et al. (1999). They also study a redemption mechanism as a monitoring device. It is interesting to compare our results, obtained in a semi-walrasian setup, with their results, obtained in a search framework. We will highlight how the search frictions play a similar role to the frictions introduced in our model and we will discuss in more detail the difference between the two models when we introduce our impossibility result, which seems to contradict their central proposition.

This chapter is also related to Bolton and Von Thadden (1999). They study the optimal design of assets with different liquidity from the corporate finance perspective and analyze some of its general equilibrium effects on the liquidity premium. Here we take a rather simplistic approach to the issue of asset design, namely we restrict attention to debt partially collateralized. The incentives in this setup are associated to the rate of conversion of debt notes into the underlying asset. A problem we want to analyze here is that even though the presence of dispersed debtors acts as a commitment device for banks, if the debt claims can circulate in the economy the incentive to redeem them may not be strong enough. Essentially, if the bank note is a perfectly good substitute for trees there is no need to redeem it before expiration.
But this allow the bank to roll over the debt notes without ever running into trouble, and a Ponzi scheme cannot be prevented. Therefore a positive rate of redemption is necessary to support an equilibrium with valued notes.

3.1 A simple model of an economy with scarce liquidity

To begin, we consider a simple variant of the simple environment studied by Woodford (1991) in which a lack of intertemporal synchronization between investment opportunities generates a demand for a liquid asset. Consider an infinite-horizon, discrete-time economy populated by a unit mass of small, infinitely lived households. There are two goods in this economy, a non-storable consumption good, which we call "corn", and a durable good, "trees". Household preferences are represented by the utility function $\sum_{t=0}^{\infty} \beta^t u(c_t)$. Households must invest $k$ units of corn in period $t$, in order to produce $f(k)$ units of corn in period $t+1$. The production technology $f$ is not transferable. After production, households must wait at least one period in order to invest again. For technical purposes, we assume that $u(.)$ is strictly increasing, strictly concave, twice differentiable, bounded above by $\bar{U} < \infty$, and $u(0) = 0$, $\lim_{c \to 0} u'(c) = +\infty$. $f(.)$ is strictly increasing, strictly concave, twice differentiable, and with marginal returns converging to 0.

This lack of coincidence between production and investment means that households must find a way of transferring wealth from one period to the next, and within periods, transfer funds from households who produce to households who want to invest, if ever they want to get production off the ground. The second good, "trees", serves this purpose. We assume that trees come in a fixed total supply of $\bar{z}$. In each period, a Walrasian market opens, in which households can trade trees for corn, before they consume. At the end of each period, a tree pays a dividend of $r$ units of fruits, fruits and corn are perfect substitutes in household consumption.

We can easily solve for the (symmetric) Pareto-optimal allocations, as well as the
steady state competitive equilibrium of this economy. The Pareto optimum solves the problem

$$\max_{\{k_t\}_{t=0}^{\infty}} W(k_0) = \sum_{t=0}^{\infty} \beta^t u\left(\frac{1}{2} (f(k_{t-1}) - k_t) + r \bar{z}_t\right)$$

(3.1)

The standard solution to this optimal growth problem shows that for any positive initial investment level, investments converge to a unique optimal steady-state level of investment, whose value is given by

$$\beta f'(k^*) = 1$$

Let us first assume that there is no way of identifying agents over time, that is, if an agent signs a borrowing contract it will be impossible to check the authenticity of his signature the following period. In this case credit contracts to finance $k$ are not present in this economy and the only way for a consumer to transfer resources from the harvesting period to the next investment period is to accumulate trees and sell them. In a competitive steady-state equilibrium of this economy, household demand is given by the solution to the following dynamic programming problem, a complete discussion of which is given in Appendix 1.\(^2\)

$$V(z_t) = \max_{c_t, \bar{c}_{t+1}, \bar{z}_{t+1}, \bar{z}_{t+2}, k_t} \left\{ u(c_t) + \beta u(\bar{c}_{t+1}) + \beta^2 V(\bar{z}_{t+2}) \right\}$$

subject to:

$$\begin{align*}
    p \bar{z}_t &= c_t + k_t + (p_t - r) \bar{z}_{t+1} \\
    p \bar{z}_{t+1} &= \bar{c}_{t+1} - f(k_t) + (p - r) \bar{z}_{t+2} \\
    \bar{z}_{t+1} &\geq 0 \\
    \bar{z}_{t+2} &\geq 0
\end{align*}$$

(3.2)  (3.3)  (3.4)  (3.5)

\(^2\)As a convention throughout the chapter, we use upperbar for consumption and stocks at the end of periods where households are sellers, and lowerbar to denote consumption and stocks in periods when households are buyers.
In addition, the following market-clearing and steady-state conditions must hold:

\[
\begin{align*}
\frac{1}{2} (\bar{z}_t + \bar{z}_t) & = \bar{z} \\
\bar{c}_t + \bar{c}_t & = \frac{1}{2} (f(k_{t-1}) - k_t) + r \bar{z} \\
\bar{z}_t & = \bar{z}_{t+H} = \bar{z}; \bar{z}_{t+1} = \bar{z}
\end{align*}
\]  

(3.6)  
(3.7)  
(3.8)

\(\bar{z}_t\) and \(\bar{z}_t\) denote the stock of trees held by investing and harvesting households, respectively, while \(p\) denotes the price of trees (taking corn as the numeraire). The following proposition characterizes the steady state equilibrium of this economy.

**Proposition 7** There is a \(z^*\) such that if the supply of trees is \(\bar{z} \geq z^*\) the unique steady-state equilibrium achieves first-best efficiency. Prices and allocations are given by:

\[
\begin{align*}
1 & = \beta f'(k^*) \\
p^* & = \frac{r}{1 - \beta} \\
\bar{c} & = \bar{c} = \frac{1}{2} (f(k^*) - k^*) + r \bar{z}
\end{align*}
\]

If \(z < z^*\) the liquidity constraint is binding, the unique steady-state is given by the solution to the following equations:

\[
\begin{align*}
\frac{u'(\bar{c})}{u'(\bar{c})} & = \beta f'(k) \\
\frac{u'(\bar{c})}{u'(\bar{c})} & = \frac{1}{\beta} \frac{p - r}{p} \\
\bar{c} & = 2p\bar{z} - k \\
\bar{c} & = f(k) - 2(p - r) \bar{z}
\end{align*}
\]

Notice that when liquidity is abundant the price of trees equals the present discounted value of dividends at the rate of intertemporal discount. In this case transfers of trees allow full consumption smoothing and this competitive equilibrium achieves Pareto efficiency.
Plugging equilibrium consumption into the household’s per period budget constraint and simplifying yields

\[ \bar{z} - z = \frac{1}{2} \frac{1 - \beta f(k^*) + k^*}{r} \]

Using the market-clearing condition (3.7), we can now rewrite the liquidity constraint (3.4) as:

\[ \hat{z} \geq z^* = \frac{1}{2} \frac{1 - \beta f(k^*) + k^*}{1 + \beta} \]

This condition is best understood as a condition relating the real value of the liquidity base (the stock of trees) to the transactions demand:

\[ \hat{z} p^* \geq \frac{1}{2} \frac{f(k^*) + k^*}{1 + \beta} \]

(3.10)

The left hand side of this inequality is the real base, the right hand side gives the aggregate monetary demand for transactions, multiplying the proportion of buyers \( \frac{1}{2} \) with the demand from each buyer: \( \frac{f(k^*) - k^*}{1 + \beta} \) for consumption and \( \frac{2k^*}{1 + \beta} \) for investment.\(^3\)

If \( \hat{z} < z^* \), the competitive steady-state equilibrium cannot implement Pareto-optimal allocations, and the environment exhibits all the symptoms of an economy with scarce liquidity: consumers do not achieve full consumption smoothing (\( \bar{c} > c \)) and there is underinvestment (\( \beta f'(k) > 1 \)). The price of the liquid asset is strictly greater than \( p^* \) and incorporates a liquidity premium.

This minimal environment has the essential properties of an economy with insufficient liquidity, in the tradition of Bewley (1980) and Sheinkman and Weiss (1986). As consumers cannot borrow against the private investment \( k \) they need to use trees to transfer wealth. Their demand for liquidity drives up \( p \) and reduces the risk-free rate of return below the individual rate of time preference. In the following sections of this chapter, we will discuss, how financial institutions and the payment system may enable the economy to create additional liquidity, in order to reduce these distortions.

---

\(^3\)The correction by \( \frac{1}{1+\beta} \) instead of \( \frac{1}{2} \) occurs because of the dividend on trees.
3.2 Private notes

We now introduce the possibility for some agents to issue private liabilities to finance consumption and investment. At any point in time, a household may decide to pay a fixed cost $\gamma$ and set up a 'bank'. A bank has a 'signature' and has a printing press that is used to print bills that cannot be counterfeited. These bills are simple IOU's that promise the bearer of the bill one unit of trees at the beginning of the coming period. The cost parameter $\gamma$ is distributed across the population according to some distribution $G(.)$ on $\mathbb{R}_+$.

We assume that these contracts can be traded (and enforced by third parties), i.e. in the language of Kiyotaki and Moore (2000), they are "saleable". In particular, they may serve as a means of payment for corn between buyers and sellers. In order to complete these simple contracts, we need to explain what happens if a bank is not able to deliver trees on the presentation of a note, or if the bank strategically decides to default on the notes it has issued. Here, we assume that a "bankruptcy procedure" takes over and (exogenously) determines the continuation value for the defaulting banker. In what follows, we will consider two different bankruptcy procedures, and their effect on liquidity creation through the banking sector: Bankruptcy law A is one in which a defaulted bank merely losess its banking privilege for a certain number of periods, but continues to participate in the market as a non-bank household. A second type of bankruptcy law, bankruptcy law B, will punish the defaulted banker much harsher: in particular we assume that for a certain number of periods, the household is excluded from the market altogether. In both cases, we assume that the court can seize any liquid assets that the defaulting banker holds at the time of default (though any resources of corn, or future output of corn, which we have assumed non-contractible, cannot be seized).

We also make the following assumptions in order to introduce a simple transactional friction in the economy$^4$ Each period is split into two subperiods. In the morning, all holders of notes issued in the previous period present these to the banker

$^4$The conclusions from the previous section are not affected by these assumptions.
who issued them. The banker gives them a choice, whether they want to "roll them over" for one period, or whether they want to withdraw the notes and receive trees instead. If a household demands to be paid out in trees, we assume that the banker cannot go and present some other note to another banker or borrow and promise to repay later, but he must meet the pay-outs in trees out of his own stock of trees (we will relax this assumption, when we go on to discuss the role of interbank markets). During the afternoon, a Walrasian market opens, in which buyers can use notes or trees to buy corn. Notice that in the absence of a transactional friction of this type it would be essentially impossible to detect a Ponzi scheme, in an environment with non-exclusivity, since an agent will always be able to satisfy his liabilities by issuing new ones and buying trees with the receipts.

In this environment, the capacity of banks to create liquidity will depend on the amount of reserves a bank is required to hold in order to satisfy conversion needs. Formally, we will find that the maximum amount of notes that a bank can circulate is a multiple of its reserve holdings, where the multiplier is the inverse of the proportion of withdrawals that occur in each settlement phase. The capacity of banks to create liquidity is directly linked to the use of these notes in transactions: If no notes were used in transactions, all notes would be withdrawn for trees in each period, and banks would have to back up the total circulation of notes by trees, thus, the net increase in liquidity would be 0. On the other hand, if every household is perfectly happy to roll over his stock of notes in each period, banks can create any amount of additional liquidity by issuing private notes.

So far, in this set-up, there is no reason why households would want to convert notes into trees. To model this choice we add a simple geographical structure. There are two locations, with half the population living in each location. In each period, households are either buyers or sellers. Sellers (those whose previous investment has just come to fruition) are always happy to meet their consumption needs out of the corn they have harvested. Buyers (those who are about to invest, and who have wealth to spend) have a random taste shock at the beginning of each period. With some probability \( \tau \), they will travel during the period and buy their corn for consumption.
and investment at the other location. With the complementary probability, they buy their corn at home. Consumers learn where they will be buying at the beginning of each period, before they engage in note withdrawals. We assume that the consumers living in region 1 can recognize liabilities issued by banks in region 1 (e.g. they can check the authenticity of a bank note), but they do not recognize liabilities issued by banks in region 2. As a consequence region 1 consumers traveling to region 2 need to pay with trees, because their notes are worthless to region 2 sellers. Thus, travelers will need to redeem their notes into trees at the beginning of each period. Moreover, the afternoon Walrasian markets in the two locations are separated, so that conversion of notes into trees has to happen at the beginning of the period, before travelling occurs.

In the following sections, we study how deposit contracts, and various other improvements on the payment system and the flow of funds help to create additional liquidity. In addition to the usual competitive equilibrium prices, however, an equilibrium with a stable banking system, without defaults must satisfy additional incentive constraints on issuing notes. Here, we have not introduced any centralized mechanisms that serve to create a "collective memory" of the notes of each individual bank in circulation. Banks thus have the possibility, if they want to, to issue any number of notes at a given date, and default during the next clearing market. As we will show,

---

5 This geographical structure is a simple way to introduce a transactional demand for trees into the model. This transactions demand may come from other sources. To this day, the smallest change transactions are most efficiently dealt with by using money, and not checks, credit cards and other forms of private liquidity. Geography may matter as well: Over large distances, contracting issues, asymmetric information and the difficulty to verify information about an unknown bank or its client on the spot imply that private deposits secured against a bank account rarely circulate far beyond a bank's region of operation. It remains open to see how more and more efficient networks for the use of private deposits as liquidity, as well as improved verification and contract enforcement technology, impact on the "technological" transactions demand for base money. Aside from these technological considerations, a final advantage of base money over deposits is its anonymity. This generates an already much discussed "criminal" transactions demand for money (oftentimes also linked to tax evasion). The advantage of anonymity also prevails, however for less criminal reasons, in environments where the loss of access to financial markets and/or institutions is used as a disciplining device against bankruptcy, or where past creditors may enforce repayment of defaulted debt by seizing any bank deposits, thus leaving base money as the only way in which an individual can participate in transactions after a default.

In this model, we are content to just assume the existence of such a transactions demand, and summarize its importance by the probability of travelling.
in equilibrium, an incentive constraint must prevent bankers from misbehaving in this way. Essentially we are assuming that a bank balance sheet cannot be monitored by her creditors (depositors), in the language of general equilibrium with asymmetric information we are studying a case of non-exclusive borrowing contracts.

With non-exclusive borrowing contracts two types of misbehavior are possible on the bank's side. First, as we just noticed, a bank can dilute borrowing claims issued at a point in time. Secondly, a bank can run a Ponzi game that would go undetected as soon as the bank remains an infinitesimal agents. In each period, a bank can adjust its portfolio holdings so that its stock of trees matches the proportion of notes that are not renewed, and at the same time proportionally increases its stock of notes (i.e. a bank running ever larger deficits). Under the assumptions made so far, a bank could presumably misbehave in this way without ever being detected, despite the presence of redemptions. Here, we will rule out this type of behavior, simply by assuming that there is an upper limit to the amount of trees any agent can hold, \( z \). We will return to this problem in our discussion of interbank markets.\(^6\)

3.3 Steady state equilibrium with banks

In this section, we formally set up the optimization problems that banks and households are facing within this economy, and define a steady-state equilibrium. We want to study the possibility of having a steady state in which banks never default and banks’ bills and trees are perfect substitutes. In characterizing the equilibrium, we will proceed in 3 steps. First we characterize and discuss the individual optimization problem, and we characterize the steady-state equilibrium for a given fixed fraction of participating banks. Then we determine the equilibrium size of the banking sector introducing an entry condition. Finally, we analyze whether the equilibrium described is consistent with the no-default incentive constraint of banks.

\(^6\)This problem also arises in Cavalcanti et al. in a search-theoretic set-up. There, the authors recognize the problem in footnote 8 where they discuss the technical problem of imposing an upper bound on reserves holdings, and have to appeal to the positive death rate. In their model, a bound to the unlimited printing of notes accompanied by accumulation of reserves is imposed by the fact that banks die with some probability, but can accumulate reserves only gradually over time.
We will see that the last requirement is impossible to satisfy if the only punishment imposed on defaulting banks is the loss of the banking privilege. Therefore, we will discuss which modifications need to be introduced in order to study an equilibrium with valued bank bills.

### 3.3.1 Households and banks

Given a steady-state equilibrium price \( p \) for trees and \( q \) for notes, the household's optimization problem is given by:

\[
V(\bar{z}_t, \bar{d}_t) = \max_{\bar{z}, \bar{c}, \bar{z}_{t+2}, \bar{k}, \bar{d}_{t+2}} \{ u(\bar{c}) + \beta u(\bar{c}) + \beta^2 V(\bar{z}_{t+2}; \bar{d}_{t+2}) \}
\]  \hspace{1cm} (3.11)

subject to:

\[
p(\bar{z}_t + \bar{d}_t) = \bar{c} + k + (p - r) \bar{z} + q \bar{d}
\]  \hspace{1cm} (3.12)

\[
p(\bar{z} + \bar{d}) = \bar{c} - f(k) + (p - r) \bar{z}_{t+2} + q \bar{d}_{t+2}
\]  \hspace{1cm} (3.13)

\[
\bar{z} \geq 0; \bar{d} \geq 0
\]  \hspace{1cm} (3.14)

\[
\bar{z}_{t+2} \geq 0; \bar{d}_{t+2} \geq 0
\]  \hspace{1cm} (3.15)

Where \( d \) is used to denote the stock of notes a household holds. If \( q = p - r \) and banks never default households will be indifferent between holding notes and trees. Since we are treating notes and trees as perfect substitutes from the perspective of net wealth and transactions, this will also be the equilibrium price. Substituting \( w = z + d \) as net wealth, we see that the previous problem is analogous to the problem in section 3.1 with \( w \) in the place of \( z \). Thus, the analysis in appendix 1 applies.

Now consider a bank. Let \( y \) denote a bank outstanding notes, \( z \) denotes reserves of the real asset and \( x \) holdings of other banks' notes. Again, if in equilibrium banks never fail banks claims are traded at par, at the same price \( p \) of the real asset. Suppose that every period every bank and every consumer convert a random fraction \( \tau \) of their notes holdings. If bank notes are issued in small denomination \( \tau \) will also
be the proportion of notes issued by a single bank that are redeemed in a period. Notice that in this case the bank budget equations will take the following form

\[ c_t + k_t = p(x_t + z_t - y_t) - (p - r)(\hat{x}_{t+1} + \hat{z}_{t+1} - \hat{y}_{t+1}) \]

\[ \overline{c}_{t+1} = f(k_t) + p(x_{t+1} + z_{t+1} - y_{t+1}) - (p - r)(\hat{x}_{t+2} + \hat{z}_{t+2} - \hat{y}_{t+2}) \]

where hat variables denote assets holdings in the morning of period \( t + 1 \), and the redemption mechanism imposes the following constraints

\[ z_{t+1} = \hat{z}_{t+1} - \tau \hat{y}_{t+1}, \quad y_{t+1} = (1 - \tau) \hat{y}_{t+1}, \quad z_{t+1} = \hat{z}_{t+1} - \tau x_{t+1} \]

\[ z_t, \hat{z}_{t+1}, \hat{z}_{t+1} \leq s \]

We first assume that banks hold enough reserves so as to never fail, that is they always satisfy the reserve constraint \( \hat{z}_{t+1} \geq \tau \hat{y}_{t+1} \). Define the total asset position of the banks \( w_t = (x_t + z_t - y_t) \). At this point we can derive an effective borrowing constraint on the banks by putting together the physical requirement \( x_t \geq 0 \), the convertibility condition \( \hat{z}_t \geq \tau \hat{y}_t \) and the condition of limited storage \( \hat{z}_t \leq s \). These three inequality together imply

\[ w_t \geq -\left(\frac{1}{\tau} - 1\right)s \]

From this expression it is clear that at the level of the individual bank a reduction in the conversion rate \( \tau \) has the effect of relaxing the borrowing constraint.

Define the implicit borrowing constraint enforced by the reserve system as

\[ B = \left(\frac{1}{\tau} - 1\right)s \]

\[ ^7 \text{Notice the difference with Cavalcanti et al. In their search setup with indivisible money a law of large numbers approximation does not make sense if we let the number of banks and consumers grow to infinity at the same rate. Here instead we can just exploit the divisibility of the bank notes, and achieve a correct LLN approximation just by letting the denomination of the notes go zero. A more extensive discussion of the randomness of } \tau \text{ appear in section.} \]
Therefore we can compactly state the conditions characterizing steady state consumption and investment of banks and non-banks:

\[
\begin{align*}
\bar{c}^b &= f(k^b) - (p - r) w^b - pb \\
\bar{x}^b + k^b &= pw^b + (p - r) b \\
u'(\bar{c}^b) &= \beta \frac{p}{p - r} u'(\bar{x}^b) \\
1 &= \beta^2 \frac{p}{p - r} f'(k^b)
\end{align*}
\]  

(3.16)  
(3.17)  
(3.18)  
(3.19)

where the non-banks have a borrowing constraint \( b = 0 \) and the banks \( b = B \). We can now state the following proposition.

**Proposition 8** Fix a given proportion \( \alpha \) of banks in the economy and a given conversion rate \( \tau \). There is a \( z' \) such that if the supply of trees is \( \hat{z} \geq z' \) the unique steady-state equilibrium achieves first-best efficiency. If \( \hat{z} < z' \) there is a unique steady-state equilibrium which does not achieve first-best efficiency and is characterized by the equations 3.16 to 3.19 for \( b = 0, B \) together with the market clearing condition

\[
\frac{\alpha}{2} \left[ w^B - \left( \frac{1}{\tau} - 1 \right) s \right] + \frac{(1 - \alpha)}{2} w^0 = \hat{z}
\]

The market clearing condition is derived noticing that the cross holdings of notes in the hands of each bank and consumers must cancel in the aggregate so that the aggregate net asset position of all agents has to sum up to the supply of the outside liquid asset.

Notice that the proposition above excludes the possibility of a steady state in which banks are not constrained while consumers are. In the present setup this is impossible. Notice if the liquidity constraint is not binding for banks, then in a steady state we would have \( \beta^2 \frac{p}{p - r} = 1 \) and also consumers would achieve full intertemporal smoothing and efficient investment. Therefore, in a steady state either liquidity is abundant for all agents in the economy or they will all be constrained in alternating periods.

Notice that in this lemma we have assumed the presence of a fixed number \( \alpha \) of
bank that never default. As we noticed above, in order to complete the description of a candidate steady state equilibrium we must check two things: first, that the marginal bank is indifferent between entry and no-entry, second, that banks never default by issuing more notes than \( B \) and facing termination the next period.

### 3.3.2 Entry

In order to have equilibrium entry in steady state it must be the case that there is a cutoff cost of entry \( \gamma \) such that

\[
\alpha = G(\gamma)
\]

and \( \gamma \) is equal to the utility gain that a consumer will obtain by acquiring the ability of issuing notes. Define this utility gain as \( \Delta U(R, w) \). This utility gain is fully determined by the steady state interest rate and by the level of assets available to a constrained consumer. It is clear that the maximum benefit would apply to a consumer in the investment stage so let us focus on his decision problem. As the economy is in steady state, this consumer has a stock \( w^0 \) of claims. If he acquires a printing press and issues bills he will be able to invest up to \( pw^0 + (p - r)B \). Notice that this is more than what an existing bank is able to raise in steady state, because in a steady state \( w^0 > w^B \). Therefore a consumer can use less than his maximum borrowing ability at early stages of his banking career. More precisely we can state the following.

**Lemma 4** Consider a consumer that starts a bank in his investment period with a level accumulated wealth of \( w^0 \) and plans to never default in the future. Let \( \{c_{t+j}\}_{j=0}^\infty \) and \( \{k_{t+2j}\}_{j=0}^\infty \) be his optimal consumption and investment plan. There is a number of periods \( J \) such that before \( t + J \) the consumer is not liquidity constrained and consumption satisfies

\[
u'(c_{t+j}) = \beta R u'(c_{t+j+1})
\]

and investment satisfies

\[
f'(k_{t+j}) = R
\]
in even periods. For \( j \geq J \) the liquidity constraint is binding \((w_t = -B)\) at all even periods and consumption and investment sequences converge to their steady state values \((c^B, \bar{c}^B, k)\).

Therefore a consumer who just acquired banking privileges will front-load his consumption path, and gradually exhausts his borrowing ability. At some point he hits the borrowing constraint and from then on he will converge towards the steady state. Notice that in their early stage banks invest more than consumers and old banks. Actually young banks overinvest with respect to the first best, since they face a return \( R \) that is lower than the individual rate of time preference \( \frac{1}{\beta} \). In the later stages of his life the bank will converge towards the steady state, and will go back to underinvestment. Notice that if we are in an equilibrium with abundant liquidity the consumer behavior is unchanged by his ability to issue bills, since his borrowing constraint was not binding to start with. Therefore we can state that \( \Delta U (1/\beta, w) = 0 \). If \( R\beta < 1 \) instead the gain from creating a new bank will be positive, since it corresponds to the gain from the relaxation of the borrowing constraint for a consumer with a given level of initial wealth.

In conclusion the no-entry condition for a steady state equilibrium with banks takes the form

\[
\alpha = G(\Delta U (R, w^0)) \tag{3.20}
\]

If the cost of creating a bank is essentially positive we can state the following result. Let \( z^* \) be the cutoff defined in section 3.1.

**Proposition 9** Assume bank creation is costly \((G(0) = 0)\). If \( z \geq z^* \) liquidity is abundant and we have a steady state equilibrium with no banks as described in Proposition 3.1. If \( z < z^* \) we can have a steady state equilibrium with bank entry. A steady state equilibrium with bank entry, though, is always characterized by \( \beta R < 1 \) and cannot achieve first best allocation.

To conclude this subsection, we have seen that it is possible to have a steady state equilibrium in which there is a positive utility gain from creating new banks and
the equilibrium proportion of banks is positive and determined by the indifference condition 3.20.

3.3.3 No default

Up to now we have characterized the banking equilibrium under the assumption that banks never overissue bills. Now we have to check under what condition such assumption is warranted. Clearly, a minimal assumption that we need in order to prevent overissuing is to assume that \( u \) and \( f \) are bounded, otherwise a one period deviation with infinite consumption or investment will always be a profitable deviation\(^8\). Even with this assumption in place, though, we show in the following proposition that it is not easy to support an equilibrium with no default. In particular, if the only punishment associated to default is the loss of the banking privilege the no-default condition is incompatible with a steady state equilibrium.

**Proposition 10** Suppose that the only punishment imposed by the bankruptcy code is the loss of the banking privilege. In a steady state equilibrium a consumer running a bank in his investment period finds always profitable to deviate by overissuing bills in the current period and defaulting in the next period.

**Proof.** Aggregating the two intertemporal budget constraints in steady state we obtain two equations characterizing the steady state consumption levels of an agent with borrowing limit \( b \)

\[
\begin{align*}
    u'(\bar{c}^b) &= \beta Ru'(\xi^b) \\
    Rc^b + \xi^b &= Rf(k^b) - k^b + (1 - R^2)(p - r)b 
\end{align*}
\]

(3.21)

Notice that \( k^b \) is independent of the borrowing limit \( b \) therefore the only difference between a bank and a consumer is in the term \((1 - R^2)(p - r)b\), which is negative for \( b = B \) and zero for \( b = 0 \). We deduce that \( \bar{c}^0 > \bar{c}^B \) and \( \xi^0 > \xi^B \). But a bank that

\(^8\)A positive cost of bill issuing could also serve this purpose.
defaults will be able to start next period with $k > k^0$ so he will always be at least as well off as a steady-state non-bank, that has zero accumulated wealth and $k^0$ invested in the private project.

Basically, a new banks in a liquidity constrained environment exploits his borrowing ability early in his life, and ends up spending the rest of his life in a cycle in which he borrows up to the limit every two periods. In this situation the gain from reducing his debt in a low period by defaulting is always larger than the loss given by his reduced ability to borrow. One could wonder whether this result is specific to the simple environment studied in this chapter. It turns out that the underlying argument is much stronger. In a related paper we consider a general setup with income uncertainty and borrowing constraints and show that if consumption follows a stationary process it is impossible to implement debt repayment with certainty using only the threat of tightening the borrowing constraint. This result is clearly reminiscent of the Theorem 1 in Bulow and Rogoff (1989), but it is not a special case of it. The crucial difference between this setup and the sovereign debt model of Bulow and Rogoff is that we have ruled out hidden asset accumulation and we have imposed a bound on the return of the private investment technology. Basically our banks cannot accumulate trees and keep them when they default on their debt obligations, therefore the arbitrage argument of Bulow and Rogoff does not apply here. Here, if a bank has some marketable assets they will all be seized upon default.

Therefore Proposition 10 is a stronger impossibility result: even in presence of a minimal enforcement mechanism that makes an agent assets observable at times of default, the mere threat of losing the ability to borrow is not enough to enforce repayment by a consumer that faces a stationary consumption process. In order to escape this impossibility result and establish the possibility of a steady state equilibrium with banks we need to modify our setup in some direction. In the remain of this section we will discuss some of the possible directions.

First, we can introduce an additional punishment that is imposed on a bank upon default. That is, we can turn to the harsher bankruptcy law B. With this type of punishment we can show that an equilibrium with no-default is sustainable. Notice
that a consumer that is banned from the market in our economy will be restricted to zero investment and zero consumption, since he has no means of transferring resources intertemporally. Therefore, it is clear that if consumers utility at zero is low enough an equilibrium of this type is easy to sustain.

Another possible modification is to assume that bills do not trade one for one against trees and there is a positive probability of default. That is we can study an equilibrium in which in every period new banks enter and some old banks default. The characterization of banks behavior in lemma 4 shows that for certain parameters young banks will not default in equilibrium in order not to squander prematurely their bank franchise. At the same time, it is clear that in this setup banks will deterministically default after they reach a certain age. Therefore, if the bank’s age is observable we cannot support a banking equilibrium of this type (by a backward induction argument banks bills will never be accepted as liquidity). If instead we assume that the bank age is not observable we can support an equilibrium with banks liquidity. Such an equilibrium will be characterized by a constant flow of defaults, that is by endemic bank failures. Nonetheless the equilibrium would be superior to an equilibrium with no banks or with a 100% reserve requirement, because the equilibrium $R$ will be closer to $1/\beta$ thanks to the supply of bills by young banks.

A third possible modification is to introduce growth in the model. When this is the case the negative term in equation (3.21) will be replaced (after normalizing variables in terms of total production) by the term

$$\left(1 - \left(\frac{R}{1+g}\right)^2\right)b$$

where $g$ is the growth rate of the economy. Therefore in this case if $R < 1 + g$ banks steady state utility will be higher than non-banks steady state utility and it is possible to obtain an equilibrium in which bank bills trade at par with trees and default never occurs.

A fourth approach would be to study equilibria in which default by a single bank generates a confidence crisis in a portion of the banking sector of positive mass. If this
is the case, a bank's loss from default cannot be computed by looking at a given steady state. Instead we need to compare the original steady state with a given number of banks to a situation in which a fraction of banks bills are no longer accepted and there is a temporary destruction of aggregate liquidity and the liquidity premium increases for a certain period of time. This aggregate destruction of liquidity may be harsh enough to induce the bank not to default. This would lead us to interpret the value of the claims of the banking system as a whole as a sustained by a collective reputation.

3.4 Conclusions

In this chapter we have used a simple dynamic model a la Bewley to study how banks can expand the available liquidity in the economy by issuing claims that are backed by fractional reserves of a liquid real asset. In the benchmark model credit is absent and consumers rely only on the accumulation and decumulation of a real asset (trees) in order to smooth consumption. We have introduced banks in this model by assuming that some consumers (banks) can, at some cost, issue bills that cannot be counterfeited. An economy with circulating bills has to solve two type of problems: (1) monitoring the total amount of bills issued by a given bank, and (2) punishing a bank that abuses her ability to issue bills.

The first problem can be solved provided that two conditions are satisfied. The banks must face some friction in rolling over their existing claims, so that they cannot smoothly run a Ponzi scheme. Second, the banks must not be able to hoard an unlimited amount of the underlying asset. We have introduced transactional frictions along these lines and we have shown that then the conversion of bills into reserves essentially imposes a borrowing constraint on banks. A low rate of conversion of bills is translated in a looser borrowing constraint for banks, and this is generally beneficial for the whole economy since it increases the aggregate supply of liquidity.

The problem of punishing a bank that defaults on her claims is more difficult to solve. First, we have supposed that banks are punished only by the threat of losing

85
their banking privilege. Therefore the punishment essentially consists in a tightening of the borrowing constraint. We have shown that if this is the only punishment faced by failing banks then a steady state with fractional reserves and no default is impossible. This impossibility result is akin to Bulow and Rogoff (1989), even though in our setup we allow for some legal enforcement of debt contracts, since banks cannot hide their accumulated assets in a situation of default.

This impossibility results points us to work in a two directions. First, we need to study equilibria in which bank default is possible in some states of the world. Secondly, we can modify our initial setup so as to sustain a no-default equilibrium. Two modifications we consider are the introduction of positive growth in the economy, and allowing for harsher punishment of defaulting banks.

3.5 References


### 3.6 Appendix

Throughout this chapter, we are concerned with the analysis of steady-state equilibria, in which households face the following dynamic optimisation problem:

\[
V_i(z) = \max_{c_1, c_2, k, z_1, z'} \{u(c_1) + \beta u(c_2) + \beta^2 V_i(z')\}
\]

where:

\[
c_1 = pz - (p - r) z_1 - k
\]

\[
c_2 = f(k) + pz_1 - (p - r) z'
\]

\[
z_1 \geq \tilde{z}_i; z' \geq \tilde{z}_i
\]

for some distribution of $\tilde{z}_i \leq 0$ across the population. In addition, a steady-state equilibrium will require market-clearing, thus determining $p$.

We first observe that of the two inequality constraints, only the one on $z_1$ will ever be binding. Substituting $R = \frac{p}{p - r}$, the first-order conditions for this problem are:

For $k$:

\[
\frac{u'(c_1)}{u'(c_2)} = \beta f'(k^*)
\]

For $z_1$:

\[
\frac{u'(c_1)}{u'(c_2)} \geq \beta R
\]
With equality, if the liquidity constraint doesn’t bind.

For \( z' \):

\[
\frac{u'(c_2)}{u'(c_3)} = \beta R
\]

We can distinguish three cases:

(i) \( \beta R > 1 \): In that case, the first-order conditions imply:

\[
c_1 < c_2 < c_3 < c_4...
\]

Consumption must be growing for all households, which is impossible in steady-state.

(ii) \( \beta R = 1 \): Now, the first-order conditions imply:

\[
c_1 \leq c_2 = c_3 \leq c_4...
\]

with equality, only if the liquidity constraint doesn’t bind. \( \beta R = 1 \) can be sustained in a steady-state equilibrium, only if no household is liquidity-constrained.

(iii) \( \beta R < 1 \): In this case, the first-order conditions imply

\[
c_1 > c_2 > c_3 > c_4...
\]

With equality, as long as the liquidity constraint doesn’t bind. For liquidity-constrained households, it is possible that \( c_1 = c_3 \), i.e. the household converges to a two-period consumption cycle. In general, \( \beta R < 1 \) can be sustained as a steady-state equilibrium, if and only if (i) households consuming positive amounts are liquidity constrained and (ii) non-constrained households consume 0.

Note that, if the liquidity constraint is weak enough, the household may be able to perfectly smooth consumption without ever hitting the constraint. In this case, steady-state consumption is 0. This is possible, whenever \( \tilde{z}_i \leq z^*_i \). Setting steady-state consumption to 0, and solving for \( z^*_i \), we obtain that households are not liquidity-
constrained, whenever

\[ \tilde{z}_i \leq z^*_i = -\frac{RF(k) - k}{r(R + 1)} \]

where investment \( k \) satisfies \( f'(k) = R \). Solving for the steady-state demand for assets \( z^* \) in the unconstrained period, we obtain

\[ z^* = \frac{-f(k) - Rk}{r(R + 1)} \]

In a steady-state the aggregate demand for wealth by unconstrained households is then given by

\[ (z^*_i + z^*) = \frac{-f(k) - k}{r} \]

which is increasing in \( R \). One observes that the unconstrained will overinvest relative to the first-best investment level. If \( \tilde{z}_i > z^*_i \), households will not be able to fully smooth consumption. In this case, for sufficiently high wealth level, they will be unconstrained in the early periods of their life, but eventually, the wealth constraint will become binding, and they converge to a two-period consumption cycle. The steady-state wealth demands of this cycle solve:

\[ c_1 = p \left( z^* - \frac{1}{R} \tilde{z}_i \right) - k \]
\[ c_2 = p \left( \tilde{z}_i - \frac{1}{R} z^* \right) + f(k) \]
\[ \frac{u'(c_1)}{u'(c_2)} = \frac{1}{\beta R} = \beta f'(k) \]

In this case, \( f'(k) = \frac{1}{\beta R} \), so there will be under-investment. Solving the first two equations yields

\[ rz^* = rR\tilde{z}_i + f(k) - c_2 \]
\[ c_1 + Rc_2 = r\tilde{z}_i (R + 1) + Rf(k) - k \]
The equilibrium demand for wealth will be given by:

\[ r(z^* + \tilde{z}_i) = r\tilde{z}_i \frac{R + 1}{R} + \frac{R - 1}{R} (c_1 + k) \]

Since \( \tilde{z}_i \leq 0 \), the first term is increasing in \( R \). \( k \) increases with \( R \), and, with CRRA preferences, so does \( c_1 \). It follows that the equilibrium demand for wealth is increasing in \( R \) for all liquidity-constrained households.
Chapter 4

Household Wealth and Consumption Variability

This chapter presents evidence of a decreasing relation between household wealth at a given date and the variability of consumption expenditure in the following year. This evidence bears upon two different issues: first, it provides empirical support to models of precautionary savings, second, it provides a specification test for the log-linearized version of the Euler equation.

Models of precautionary wealth accumulation rely either on prudent preferences or on liquidity constraints or both. A typical feature of these of models is that consumers use accumulated financial assets to self insure against income shocks. An empirical prediction of these models is that household with larger levels of accumulated wealth will face a lower level of consumption variability because they are more willing to use accumulated wealth as a buffer stock against income shocks. This happens because consumers with larger levels of accumulated wealth are either less concerned about hitting the liquidity constraint in the near future or less risk averse as they reached higher levels of consumption. This prediction distinguishes models of precautionary behavior both from quadratic utility models of the permanent income hypothesis and from models of full insurance. The first objective of this chapter is to test this prediction by looking at the variability of consumption as a function of household wealth.
The results obtained can be interpreted as a complement to the test of liquidity constraints by Zeldes (1989): he was concentrating on the conditional first moment of the Euler equation residual, which—in presence of liquidity constraints—displays a discontinuous jump at wealth levels at which the constraint is binding, here instead we study the conditional second moment of the Euler equation residual which—in presence of liquidity constraints and/or prudent preferences—is smoothly decreasing with wealth over an extended range. An advantage is that we do not need any assumption about the maximum level of borrowing allowed. Our results can also be interpreted as a further test of the full insurance model in the vein of Townsend (1990) with the twist that we are testing full insurance against a well defined alternative hypothesis, namely against self-insurance through asset accumulation.

In terms of the consumption Euler equation the conditional variability of consumption growth is associated with the conditional second moment of the Euler equation residual. The log-linearized version of the Euler equation is correctly identified only under the implicit assumption that this second moment is uncorrelated with the instruments used. This happens because the log-linearized identification condition is only an approximation of the true identification condition, and the accuracy of the approximation depends on the higher conditional moments of the residual. Current wealth is clearly correlated with past measures of income and consumption, which are usually included in the instruments set. Thus, a relationship between wealth and the second moment of the residuals hints at a possible misspecification of the log-linearized Euler equation. We setup a specification test along these lines and we obtain a rejection of the hypothesis of a second moment uncorrelated with the instruments. Our results constitute a further warning about the use of the log-linearized Euler equation in empirical work and they complement well the discussion by Carroll (1997) which was based on simulation results. While sharing Carroll’s concerns on the log-linearized approach, we maintain a more optimistic view on the general possibility of obtaining useful information in a Euler equation framework. Thus, we try some possible fixes to the identification problem. In presence of heteroskedasticity one can either go back to the nonlinear setup or try and improve the approximation in the
traditional setup adding terms to the Taylor expansion. This chapter tries the second alternative because, under some additional assumptions, it let us keep linearity in the individual effects, and this can be quite useful from the point of view of panel data estimation.

Section 2 contains a theoretical motivation to interpret the evidence presented in terms of models of precautionary savings. We show simulated series for consumption and wealth for a simple model of optimal consumption with incomplete insurance and borrowing constraints and show that the conditional variance of consumption at time \( t \) is decreasing with the level of liquid asset holdings at time \( t - 1 \). Under full insurance or under quadratic preferences the same model displays instead a conditional variance that is constant across wealth levels. In section 3 we discuss the relation between the orthogonality condition derived from consumption theory and the log-linearized version of it often used in the empirical literature and discuss the identification problems that arise in the log-linearized model in presence of heteroskedasticity. In section 4 we describe the data used and the estimation strategy for the log-linearized Euler equation. The empirical results based on the PSID data are reported and discussed in section 5.

### 4.1 Theoretical motivation

In this section I use a standard model of consumption behavior with uninsurable income risk to show, with some simple simulations, that the presence of prudence and/or of borrowing constraints generates a negative relation between household wealth and the variability of consumption.

Consider the problem of an infinitely lived consumer that receives a random labor income \( Y_{it} \), can accumulate a risk free asset, and faces a borrowing constraint \( A_{it} \geq A \).

\[
\begin{align*}
\max \quad & \quad E_t \sum_{s=t}^{\infty} \beta^s u(C_{is}) \\
\text{s.t.} \quad & \quad C_{is} + A_{is} = (1 + r)A_{is-1} + Y_{is}
\end{align*}
\]
\[ A_{is} \geq A \]

Assume that \( Y_{it} \) has a permanent and a transitory component, \( X_{it} \) and \( \eta_{it} \). The income process is described by the two equations

\[
Y_{it} = X_{it} \eta_{it} \\
X_{it} = (1 + g) X_{it-1} \epsilon_{it}
\]

where \( \eta_{it} \) and \( \epsilon_{it} \) are i.i.d. log-normal shocks. Define \( c_t = C_t/X_t \) and \( a_t = A_t/X_t \).

We distinguish three different cases: (1) full insurance, (2) PIH: only a risk free asset is available, utility is quadratic and \( A = -\infty \) (3) precautionary savings with CARA utility and borrowing constraints: only a risk free asset is available, utility is CARA and \( A > -\infty \). We are interested in characterizing the conditional variance of consumption in each case, in particular we will look at \( Var_{t-1}(c_t) \) and at \( Var_{t-1}({\ln C_t}) \).

The case of full insurance requires to embed the consumer’s problem above in a general equilibrium framework. We assume that there is no aggregate uncertainty so that the realized cross sectional averages of \( \epsilon_{it} \) and \( \eta_{it} \) are equal to the mean of the corresponding random variables\(^1\). In this case for any strictly concave utility function full insurance allows any consumer \( i \) to achieve a constant level of consumption across states of the world. In this case both \( Var_{t-1}(c_t) \) and \( Var_{t-1}({\ln C_t}) \) are zero. It is worth noticing that full insurance has a stronger implication: even in presence of aggregate uncertainty, the variability of consumption conditional on the aggregate shock would be zero.

In the case of quadratic utility we need to assume \( \beta \frac{1+r}{1+g} = 1 \) in order to obtain a

---

\(^1\)This assumption is consisten with the assumption of a fixed \( r \) in cases (2) and (3). If we derive \( r \) in a general equilibrium framework \( r \) can be constant if we reach a stationary distribution of the risk-free asset and there is no aggregate risk (see e.g. Huggett (1993)).
stationary process for $a_t$. In this case the optimal consumption policy is

$$C_t = r \left[ (1 + r) A_{t-1} + \sum_{j=0}^{\infty} (1 + r)^{-1} E_t Y_{t+j} \right]$$

and we can show that both $Var_t(c_t)$ and $Var_t(\ln C_t)$ are constant and do not depend either on $a_t$ or on $A_t$.

In case (3) we need to resort to simulations to characterize the stochastic properties of consumption. The budget constraint can be written as: $c_t + a_t = \frac{1+r}{1+g}a_{t-1}/\epsilon_t + \eta_t$. Defining $z_t = \frac{1+r}{1+g}a_{t-1}/\epsilon_t + \eta_t$ we can show that optimal consumption will be characterized by the equation

$$C_t = h(z_t)X_t$$

where the function $h$ satisfies the functional equation

$$u'(h(z)) \geq (1 + r)\beta Ez u' \left( h \left( \frac{1+r}{1+g} (z - h(z)) \exp(-\epsilon) + \exp(\eta) \right) \right)$$

with $h(z) = z$ in case of inequality.

We calibrate the process for income to match the process for individual income estimated by MaCurdy (1982) using PSID data, (in this we follow Deaton (1991)). We use the following parameters: $r = 2\%$, $\beta = .95$, $\gamma = 2$, $Var(\ln \epsilon) = .1$ and $Var(\ln \eta) = .07$. Many studies have characterized in detail the optimal asset accumulation policy in problems of this type: an agent accumulates the asset in good times and runs down his stock in bad times, there is a stationary transition for asset holdings and a long run stationary distribution of asset holdings with bounded support. To compute the optimal consumption policy we follow a simple iteration method on the Euler equation.

Figures 1 and 2 illustrate the simulated relation between $a_{t-1}$ and $Var(c_t|a_{t-1})$ and the simulated relation between $a_{t-1}$ and $Var(\ln C_t|a_{t-1})$. There is a strong negative relation between lagged wealth holdings and the variability of consumption reflecting

---

2Again, this seems a useful requirement if we want to embody the consumer problem in an equilibrium framework and obtain a constant $r$.  

95
the fact that a low variability of the residual is a sign of better ability to insure against adverse income shocks.

Figure 4-1: Wealth and variability of consumption (c)

Notice that the relationship above is strictly related to the concavity of the function \( h \) (see Carroll and Kimball (1996), and Parker (2001)). Notice that for a given \( a \) we have the approximate relation

\[
\text{Var}(c_t|a) \approx (h')^2(a\sigma_{\tilde{y}}^2 + \sigma_{\eta}^2)
\]

Two effects are present here, on the one hand there is a simple scale effect, according to which when financial wealth is larger \( A/X \) is more volatile, on the other hand the concavity of \( h \) makes the term \((h')^2\) larger for low levels of wealth. The figures above show that the second effect seems to dominate for realistic parameter values.

In quantitative terms the effect is sizeable and non-linear in the simulated model. An increase in financial wealth from 0 to 20% of permanent income reduces the standard deviation of consumption from 6.2% to 3.2% percentage points. Further increases in the wealth to permanent income ratio have a much smaller effect on the volatility of consumption. Notice that we have experimented with various values for \( \gamma \)
and the results are very similar: the main effect of increases in $\gamma$ is that the stationary distribution of wealth levels moves to the right but the wealth-variability frontier does not change much.

Summing up, models (1), (2) and (3) differ markedly for their prediction regarding the conditional variability of consumption. In models of full insurance and in quadratic models of the permanent income hypothesis the conditional variability of consumption does not depend on accumulated wealth, in simulated models of precautionary savings models instead household wealth reduces consumption variability especially for low levels of wealth to permanent income ratios. In the empirical section I will use the PSID data on consumption to estimate a nonparametric regression of $\text{Var}_{t-1}(\ln C_t)$ on some measures of lagged asset holdings, and I will compare it with the negative relationship depicted in figure 2.

Let me now briefly compare this test with other similar tests designed to test the full insurance hypothesis and the precautionary saving model separately. The test is close in spirit to the test of liquidity constraints performed by Zeldes (1989). In that article a low level of wealth is associated with a higher probability of facing a liquidity
constraint. The presence of binding liquidity constraints makes the Euler equation invalid for constrained consumers and therefore Zeldes tested whether the conditional moment $E(u_t|a_{t-1})$ was dependent on $a_{t-1}$. Here we are essentially studying the behavior of $E(u_t^2|a_{t-1})$ as a function of $a_{t-1}$. The latter conditional moment is decreasing over the whole wealth range and it does so also in cases in which the consumers are facing a binding constraint very rarely. In these cases a direct test of the liquidity constraint may be inconclusive due to the small number of observations where the constraint is actually binding.

This test is also related to tests of the full insurance model based on the first order properties of the residual as the tests performed by Townsend (1990), Cochrane (1991) and Mace (1991). On the one hand, the present test has more power because it tests the full insurance case against a specific alternative. On the other hand it may be less convincing because it relies on the additional assumptions that taste shocks and measurement error are homoskedastic conditional on lagged values of wealth holdings and income shocks.

We can also compare this test with the a wide class of tests that have been used to study the precautionary motive looking at the relation between individual income variability and wealth holdings, on the grounds that the expectation of higher income variability induces individual to accumulate more wealth ex ante (Browning and Lusardi, Lusardi (1997)). These tests assume some (exogenous) heterogeneity in income variability across individuals, and include this variability in a reduced form savings function. Apart from the use of an explicit saving function these tests rely basically on exogenous cross-sectional heterogeneity in income variability, while the test proposed here exploits the fact that the ex-ante identical consumers may be differently exposed to residual variability depending on their asset holdings. In a sense their approach relies on individual heterogeneity to obtain identification, while the approach taken here relies on having a sufficiently homogeneous population of consumers, so that one can look at the cross sectional dimension to have information about the time series path of income, wealth and consumption.
4.2 Linearization and Identification

The original form of the Euler equation is a conditional moment restrictions and can be written in the form

\[ E[e^u|z] = 1 \]  \hspace{1cm} (4.1)

where \( z \) is any variable that is in the individual information set at time \( t-1 \) —in particular, it may be a lagged value of wealth holdings or income—and \( u \) is the traditional log-linearized residual at time \( t \). The expression for \( u \) I will use in the following is

\[ u^i_{t+1} = \log \beta^i + \log(1 + r^i_{t+1}) + b \Delta d^i_{t+1} + \Delta (\log U'(c^i_{t+1})) + \eta^i_{t+1} \]

where \( \beta \) is the discount factor, \( r \) is the real interest rate, \( d \) is a vector of household characteristics (i.e. observable taste shocks), \( U \) is the instantaneous utility function, \( c \) is consumption and \( \eta \) captures the effect of unobservable taste shocks.

The usual estimation strategy is to use the linearized version of (4.1), that is \( E[u|z] = 0 \).

This approximation may be more or less accurate in different cases. Consider first the simple case in which \( z \) and \( u_t \) are jointly normally distributed. In this case the condition above implies \( E(exp(u+z)) = E(exp(z)) \) which can be transformed into an expression involving only the moments of \( u_t \) and \( z \), that is

\[ Eu + \frac{1}{2}(Var(u) + 2Cov(u,z) + Var(z)) = \frac{1}{2} Var(z) \]

Using this expression when \( z \) is a constant we derive the restriction \( Eu+1/2Var(u) = 0 \) that in turn gives us \( Cov(u,z) = 0 \). The latter equality is exactly the identification assumption required to estimate the traditional linearized version of the Euler equation, thus under joint normality the traditional approach is fully justified.\(^3\) Under normality absence of linear correlation is equivalent to independence, thus the condition

\(^3\)Under normality absence of linear correlation is equivalent to independence, thus the condition
fortunately, not only the joint normal case is a special case, but is also particularly ill-suited in this context, because it embeds homoskedasticity by construction.

In the general case we can take a Taylor expansion of $e^u - 1$ around $u = 0$. Rearranging terms we have

$$E(u|z) = -E(1/2u^2 + 1/6u^3 + ...|z)$$  \hspace{1cm} (4.2)

and if the right hand side of this condition is constant in $z$ we can derive the identification condition $E(u - Eu|z) = 0$.

Many researchers have encountered some version of equation (4.2), that is they have recognized that the mean of $u$ will not be zero and that a variance term should appear in the residual of the regression expression. Usually they have dealt with the problem just redefining appropriately the residual and the constant term in the regression, that is assuming –more or less explicitly– that $E(1/2u^2 + 1/6u^3 + ...|z)$ is constant in $z$.

This last assumption is crucial for the identification of the linearized model. Notice that the first term appearing in the expectation is $E(1/2u^2|z)$, and this term reflects the heteroskedasticity of $u$ conditional on the instrument $z$. Zeldes recognizes correctly this problem: "the conditional variance of the forecast error could be a function of wealth or disposable income. For example, when household assets are especially low, uncertainty about the growth rate of consumption could be higher. (...) In each of these cases the estimation scheme presented below will be inconsistent." 4

Therefore when we do our test of homoskedasticity we are also testing the accuracy of the linearized expression, and if we reject $E(u^2|z) = 0$ this implies that our identification assumption is incorrect. In this setup the presence of heteroskedasticity has negative implications for the consistence of our estimator and not only for its efficiency, due to the form of our original identification assumption and the fact that we are using an approximation of it. After rejecting the homoskedastic hypothesis we cannot rely anymore on the estimates obtained by using $E(u|z) = 0$. At this point

---

1 $E[u|z] = 0$ follows as well.

2 Zeldes (1989), p.319
we can either go back to the non-linear specification or find some fix for the linear identification condition. From the point of view of matching theoretical predictions with identification assumptions the ideal would be to work directly with the nonlinear form, as Hansen and Singleton originally did on aggregate data, and as Runkles did on individual data. But when using panel data there are substantive gains by keeping linearity in the individual effects, so some type of linearization seems useful.

One possibility is to add a term of the Taylor expansion and use as the identification condition

$$E(u + 1/2u^2|z) = 0$$  \hspace{1cm} (4.3)

Clearly to obtain linearity in the individual effects from (4.3) we need to make additional assumptions. The nature of the problem is the following. We don’t actually observe $u^i_t$, instead we observe $v^i_t = u^i_t - \alpha^i - \eta^i_t$. Here $\alpha_i$ is an individual effect constant over time, in our model it corresponds to a different discount factor, that is to $\log \beta^i$ in (4.2). $\eta^i_t$ is a time-varying effect, in our model it correspond to the unobserved taste shock part, but it may also include measurement error in consumption. If we rewrite (4.3) in terms of observables and parameters (the parameters are implicit in $v^i_t$), we have:

$$E[v^i_t + 1/2(v^i_t)^2 - \alpha^i - \eta^i_t|z] - K(z) = 0$$  \hspace{1cm} (4.4)

Where $K(z) = E[(\alpha^i + \eta^i_t)^2 + u^i_t(\alpha^i + \eta^i_t)|z]$

If we are willing to make the additional assumption that $K(z)$ is constant we have an identification assumption which is linear in the individual effects and we can carry out a richer analysis using most of the tools of linear panel data models. Actually, in this chapter I simply analyze a random effect specification (that is I will assume $E(\alpha^i|z) = 0$), but I plan to extend the analysis and do the appropriate specification tests in future work.
4.3 Data and Estimation Strategy

The estimates are computed using a sample of 2,350 families from the Panel Study of Income Dynamics (PSID) covering the years 1976 to 1985. The advantages and disadvantages of using this data set for studying intertemporal consumption have been discussed at length in the literature\(^5\). One specific disadvantage in our context is the recognized presence of large measurement error in the consumption data. Apart from the additional assumptions that need to be done in presence of measurement error to get identification, its presence here means also that we have little information about the total variability of the residual. For this reason we are limited in our ability to make statements about the importance of wealth accumulation in reducing variability in proportional terms, even if we have good estimates of its absolute effect.

The variables used are: food consumption, the real rate of interest, household disposable income, family wealth and two demographic variables.

The food consumption variable is constructed adding the value of food stamps to the food at home variable and then adding food at home and food out of home, each deflated by the appropriate consumption price index. The rate of interest is computed by applying the household marginal tax rate to the rate of return on 3-months Treasury bills, and the real rate of interest is obtained by subtracting the rate of inflation for consumption goods. Disposable income is computed subtracting taxes and social security payments from the total income of the family unit, disposable income is deflated using the price index.

The first measure of family wealth available is housing wealth, and is computed subtracting the value of the outstanding mortgage from the value of the house. The second measure is actually a flow measure: the income from interests and dividends. An approximate stock measure can be obtained dividing this measure by an appropriate rate of return. In the estimation of the euler equation wealth levels are used only as instruments and in that case I have preferred to use the two wealth variables sep-

\(^5\)Studies on consumption using the PSID include: Hall and Mishkin (1982), Zeldes (1989), Keane and Runkle (1992), Gruber and Dynarski (1997).
arately, keeping the second one as a flow variable. For the nonparametric regressions of \( u^2 \) instead I have computed an aggregate measure of wealth, converting the second measure to a flow using the rate of return on treasury bills. \(^6\) The demographic variables used are the age of the head of the household and the size of the family unit.

We can use the definition of the observable residual \( v \) in section 2, assume CRRA utility \( \left( \frac{1}{1-\gamma} \right) \) and rewrite the expression for the linearized residual in (4.2) as:

\[
v_{t+1}^i = b_0 + \log R_{t+1}^i + b \Delta d_{t+1}^i - \gamma \Delta (\log c_{t+1}^i)
\]  

(4.5)

To estimate the parameters of this expression I use the GMM estimator based on the condition \( E(Z_t'v_t) = 0 \) where \( Z_t \) is a vector of instrumental variables that are uncorrelated with \( v_t \). The demographic variables used are: the age of the head, the age of the head squared, and the change in family size. To take into account aggregate forecast error I have added time dummies to the right hand side of (4.5). \( Z_t \) includes lagged values of the wealth variables, lagged values of disposable income (2 lags), and the lagged value of the marginal tax rate.

In principle one could use a different set of instruments for every time period (more lags are available for larger \( t \)), and one could make efficient use of the period-by-period distinct conditions \( \text{plim}_{N \to \infty} (Z_t'v_t)/N = 0 \) for \( t = 2, 3, ..., T \), where \( N \) is the number of families and \( T \) the number of time periods. Here, for simplicity, I have used the same set of instruments for each period and I have used the time aggregate condition \( \text{plim}_{N \to \infty} \sum_t (Z_t'v_t)/NT = 0 \), that is the weighting of the different time period orthogonality conditions is not efficient.\(^7\)

Therefore, my estimation is reduced to a simple efficient instrumental variable regression on the model \( \Delta \log c_t^i = bx_t^i + \epsilon_t^i \), where each \((i, t)\) is treated as a separate

---

\(^6\) In the construction of the variables above I have followed closely Zeldes (1989), therefore I refer to that article's appendix for a careful discussion of this construction.

\(^7\) The efficient use of period-by-period orthogonality conditions separately is computationally more demanding but would allow one to construct a sequence of GMM estimators with asymptotic variances approaching the information bound as the set of instruments is expanded, as shown in Chamberlain (1992).
observation, \( x_t^i \) includes all variables on the r.h.s. of (4.5) except consumption, and 
\[ e_t^i = -(1/\gamma)v_t^i. \]

I am not differencing the \( v_t \) to eliminate the individual effect, that is I am considering a random effects estimator. In making this choice I rely on results by Keane and Runkle (1992). They adapt the Hausman and Taylor (1981) specification test to the case of not strictly exogenous instruments and do not reject the hypothesis of individual effects uncorrelated with the instruments. Therefore, in the following I use the random effects specification.

Still, the panel dimension was taken account of in computing a consistent estimate of the variance matrix \( \Omega = E(Z'\epsilon \epsilon'Z) \) needed for efficient weighting. In doing that I have considered the presence of the individual component in \( v_t^i \) and I have allowed for non-zero values of \( E(\epsilon_t^r \epsilon_t^s) \) for all \( r, s \). The expression below is the appropriate extension of White’s (1980) heteroskedasticity consistent covariance estimator, assuming \( E(\epsilon_t^r \epsilon_t^s) = 0 \) for any \( r, s, i, j \) such that \( i \neq j \).\(^8\)

\[
\hat{\Omega}/N = (1/N) \sum_{t=1}^{T} \sum_{i=1}^{N} (\tilde{\epsilon}_{i,t}^2) z_{i,t}^2 + 2(1/N) \sum_{t=1}^{T} \sum_{i=1}^{N} \tilde{\epsilon}_{i,t} \tilde{\epsilon}_{i,t-l} z_{i,t} z_{i,t-l} (4.6)
\]

The heteroskedasticity test is simply based on the regression of \( \epsilon_{i,t}^2 \) on the vector of instrumental variables in \( Z \). In that regression I have also tested the joint significance of the four variables immediately related to availability of liquid assets, namely the two measures of wealth and the two lagged values of disposable income.

A second set of parameter estimates is obtained from the following orthogonality condition, adjusted by a quadratic term,

\[ E(Z_t'(\epsilon_t - 1/2\gamma e_t^2)) = 0. \] (4.7)

which is derived from (4.4) and, as argued in section 2, is more reliable in case of heteroskedasticity of the residual.

\(^8\) Compare it with (3.12) in Holtz-Eakin, Newey and Rosen and notice that here period-by-period orthogonality conditions are aggregated in a fixed (inefficient) way.
The GMM estimation strategy is analogous to that for the linearized model. Letting \( h = Z'(\epsilon + \gamma 1 / 2\epsilon^2) \), I have first obtained an estimate of the parameters, minimizing \( h'Wh \) with \( W = I \). Then, using the residuals \( \hat{\epsilon} \), I have computed the efficient weighting matrix \( W = \Omega^{-1} \) using an expression analogous to (4.6) with the adjusted residual \( \hat{\epsilon} - \hat{\gamma} 1 / 2\epsilon^2 \) replacing \( \hat{\epsilon} \). Using this efficient weighting matrix I have obtained the parameter estimates.

4.4 Results

4.4.1 Some preliminary results

In order to estimate \( \text{Var}(\ln C_{t+1} | w_{it}) \) we need to make assumptions about individual heterogeneity and about aggregate risk. If aggregate risk is absent, as in the model simulated in section 4.1, then the cross sectional variability of \( \ln C_{it} \) provides information about the perceived variability of consumption for a single consumer. Let us assume, for the moment that aggregate risk has negligible effects, and let us focus on the problem of individual heterogeneity. Denote with \( \theta_{it} \) all sources of individual heterogeneity except the wealth level, in the model in section 4.1 the only element of heterogeneity, aside from the wealth to permanent income ratio, was the level of permanent income and \( \theta_{it} = X_{it} \). Then we can write

\[
\text{Var}_i(\ln C_{it+1} | w_{it}) = E_i(\text{Var}(\ln C_{it+1} | \theta_{it}) | w_{it}) + \text{Var}_i(E(\ln C_{it+1} | w_{it}, \theta_{it}) | w_{it})
\]

Where with \( \text{Var}_i \) we denote a cross sectional variance. Clearly, we are interested in characterizing only the first term on the right hand side. In the model in section 4.1 \( \text{Var}(\ln C_{it+1} | w_{it}, \theta_{it}) \) is actually constant in \( \theta_{it} \), and equal to \( \text{Var}(\ln C_{it+1} | w_{it}) \), therefore if we could isolate this first element on the right hand side we could immediately test the predicted decreasing shape of \( \text{Var}(\ln C_{it+1} | w_{it}) \). Isolating this effect requires a number of identification assumptions. As a preliminary exercise we adopt a rough approach to eliminate the 'level' effect generated by the second term, by just taking
differences. If we compute

\[ \text{Var}_i(\Delta \ln C_{it+1}|w_{it}) = E_i(\text{Var}(\Delta \ln C_{it+1}|w_{it}, \theta_{it})|w_{it}) + \text{Var}_i(E(\Delta \ln C_{it+1}|w_{it}, \theta_{it})|w_{it}) \]

and include \( C_{it} \) in the list of the \( \theta_{it} \) variables, we can exploit the fact that \( E(\Delta \ln C_{it+1}|w_{it}, \theta_{it}) \) is approximately zero if consumption is close to a random walk. In the context of the simulated model in section 4.1 this procedure is correct for any \( w_{it} \) greater than zero if an approximate Euler equation holds.

Therefore as a preliminary result I have estimated the regression of \( (\Delta \log(c_{t+1}))^2 \) on the wealth level \( w_{it} \), computed summing the two measures of housing and financial wealth. That is, I have estimated the statistical relation between lagged wealth holdings and variability of consumption. The kernel estimate of this regression appear in figure 3 and 4. \(^9\) The picture seems to support the idea that larger wealth holdings are associated with lower consumption variability. The change in log consumption is an imperfect estimate of the Euler equation residual, for example it does not take into account that changes in demographics may make a given consumption change have different impact on the household marginal utility of income. Therefore we turn now to the estimation of the parameters of a fully specified Euler equation in order to recover the residual from this estimation. After that we will go back to the nonparametric analysis of the conditional behavior of the residual.

4.4.2 Euler equation estimates and tests of heteroskedasticity

The results of the estimation of the linearized model are reported in the first half of table 1. Column II displays the coefficients for an alternative specification that tests for excess sensitivity by adding current income to the group of explanatory variables.

The coefficients of the demographic variables are all significantly different from zero. The coefficient on the interest rate is considerably larger than that usually

---

\(^9\)A simple quadratic kernel was used and various bandwidth values where tried. Figure 3 and 4 correspond to a bandwidth value of 0.15.
found in the literature (usually smaller than 1 and often around 0.5 corresponding to an estimated value of \( \gamma \) around 2). Actually the estimates of this coefficient have pretty large standard errors in all the specifications considered, maybe more efficient estimation procedures (as those discussed in footnote 9) would give results in line with the literature. Observing the coefficient on disposable income in column II we see no evidence of excess sensitivity in the data, in accordance with the findings of Keane and Runkle and of other studies that allow for a certain richness in the specification of the demographic part (see Attanasio and Browning (1995)). The table reports also the Sargan test of overidentifying restrictions, in both cases the test does not reject the specification used.

The second half of table 1 reports the results of the heteroskedasticity regression, the value of the \( \chi^2 \) statistic is reported for the general test of homoskedasticity (all coefficients in the regression equal zero), and for the specific test of homoskedasticity conditional on past income and wealth levels (the relative degrees of freedom are reported in parenthesis). Both tests strongly reject the hypothesis of homoskedasticity.

We can take this as \textit{prima facie} evidence in favor of the insurance role of accu-
mulated wealth. Given the correlation of the income and wealth variables and the fact that they are all included to account for the same factor (availability of liquid assets at time t + 1) we do not expect to get negative and significant coefficients on all variables. The lagged income term carries most of the explanatory power and it is the only one which has a coefficient significantly different from zero. The coefficient on lagged income is negative and — according to our interpretation — it implies that a 10% increase in disposable income in period t allows the consumer to reduce residual consumption variability in the next period by 2.6%. This elasticity is computed using the mean value of the measure of consumption variability, which is 0.1253. Here we encounter the problem mentioned above regarding measurement error. Measurement error affects total observed variability and therefore inflates the value of the average residual variability. At the same time measurement error should not affect our coefficient estimates (at least we hope so, otherwise our identification assumption would be incorrect). As a consequence elasticities of the type just reported will be in general underestimated. Under heteroskedastic errors we know that the identification condition is inaccurate, so I have used the identification condition (4.7) to get new parameter estimates. The results are reported in Table 2. The heteroskedasticity
Table 4.1: Log Linearized Model

<table>
<thead>
<tr>
<th>Indep. Variable</th>
<th>Coeff. I</th>
<th>S.E.</th>
<th>Coeff. II</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.01430</td>
<td>0.00166</td>
<td>0.01458</td>
<td>0.00171</td>
</tr>
<tr>
<td>Age$^2$</td>
<td>-0.00017</td>
<td>0.00002</td>
<td>-0.00017</td>
<td>0.00002</td>
</tr>
<tr>
<td>Δ(Family size)</td>
<td>0.11487</td>
<td>0.00537</td>
<td>0.11511</td>
<td>0.00538</td>
</tr>
<tr>
<td>Real interest rate</td>
<td>1.65004</td>
<td>0.21259</td>
<td>1.29191</td>
<td>0.54788</td>
</tr>
<tr>
<td>Disposable income</td>
<td>-0.00647</td>
<td>0.00917</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overidentif. Restr.</td>
<td>0.93720</td>
<td>(4)</td>
<td>0.42004</td>
<td>(3)</td>
</tr>
</tbody>
</table>

Regression of squared residuals

<table>
<thead>
<tr>
<th>Indep. Variable</th>
<th>Coeff. I</th>
<th>S.E.</th>
<th>Coeff. II</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagged disp. income</td>
<td>-0.03284</td>
<td>0.00482</td>
<td>-0.03301</td>
<td>0.00482</td>
</tr>
<tr>
<td>2 Lagged disp. income</td>
<td>-0.00740</td>
<td>0.00392</td>
<td>-0.00747</td>
<td>0.00392</td>
</tr>
<tr>
<td>Housing wealth</td>
<td>-0.00341</td>
<td>0.00312</td>
<td>-0.00343</td>
<td>0.00312</td>
</tr>
<tr>
<td>Non-housing wealth</td>
<td>0.00186</td>
<td>0.00237</td>
<td>0.00187</td>
<td>0.00237</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>490.610</td>
<td>(16)</td>
<td>492.288</td>
<td>(16)</td>
</tr>
<tr>
<td>$\chi^2$ income/wealth vars.</td>
<td>106.610</td>
<td>(4)</td>
<td>107.816</td>
<td>(4)</td>
</tr>
</tbody>
</table>

regression has been performed also for this model, because of its economic interpretation. The first four coefficients do not show striking differences with the simple linearized specification.

The coefficient of $\hat{\epsilon}^2$ is an additional parameter which is directly estimated in this case and corresponds to $\gamma$ (see (4.7)). Also this estimated $\gamma$ is much smaller than the estimates obtained in most empirical studies and it is actually not significantly different from zero. In this specification we have also another available estimate for $\gamma$, that is obtained as the inverse of the coefficient of the interest rate, which corresponds to 0.7468. We can test the difference of the two by using a delta method to derive the variance of the difference. Even though both estimates have high standard errors we reject the hypothesis that they are equal\(^\text{10}\)

The result that the variability term has a coefficient close to zero is puzzling. This result bears some resemblance with a result obtained by Dynan (1993). She uses a different approximamte expression for the Euler equation that allows for a general form of the utility function, and she obtains a condition analogous to (4.7) with a variabili-

---
\(^{10}\)The difference is 0.5088, and it is asymptotically normal with standard error 0.2145.
Table 4.2: Model with quadratic term correction

<table>
<thead>
<tr>
<th>Indep. Variable</th>
<th>Coeff.</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.01578</td>
<td>0.00254</td>
</tr>
<tr>
<td>Age^2</td>
<td>-0.00019</td>
<td>0.00003</td>
</tr>
<tr>
<td>Δ(Family size)</td>
<td>0.11555</td>
<td>0.00550</td>
</tr>
<tr>
<td>Real interest rate</td>
<td>1.33907</td>
<td>0.47128</td>
</tr>
<tr>
<td>1/2ε^2</td>
<td>0.23798</td>
<td>0.15664</td>
</tr>
<tr>
<td>Overidentif. Restri.</td>
<td>0.39969</td>
<td>(4)</td>
</tr>
</tbody>
</table>

N. obs. 18880
N. Indep vars 12
N. instr. 16

Regression of squared residuals

<table>
<thead>
<tr>
<th>Indep. Variable</th>
<th>Coeff.</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lagged disp. income</td>
<td>-0.03284</td>
<td>0.00483</td>
</tr>
<tr>
<td>2 Lagged disp. income</td>
<td>-0.00759</td>
<td>0.00392</td>
</tr>
<tr>
<td>Housing wealth</td>
<td>-0.00359</td>
<td>0.00313</td>
</tr>
<tr>
<td>Non-housing wealth</td>
<td>0.00185</td>
<td>0.00237</td>
</tr>
<tr>
<td>χ^2</td>
<td>493.83874</td>
<td>(16)</td>
</tr>
<tr>
<td>χ^2 income/wealth vars.</td>
<td>107.75742</td>
<td>(4)</td>
</tr>
</tbody>
</table>

ity term represented by (Δlogc)^2 instead of ε^2. She shows that the coefficient of the variability term is approximately equal to cu''/u'', a measure of consumer 'prudence'. Notice that the CRRA specification imposes cu''/u' = cu'''/u'', so our test of equality between the two estimated γ can be considered —in Dynan setup— as a test of the CRRA specification, and our result would imply a rejection of this specification. Secondly the fact that the coefficient on the squared residual is not significantly different from zero would imply a rejection of the prudence assumption. Starting from the approximate Euler equation Dynan follows a different strategy than ours: first aggregates across time periods and then uses the individual observations obtained in this way to test the significance of the prudence coefficient. She reaches a similar result as that obtained here, namely that the coefficient on the variability term is not significantly different from zero. The approach of aggregating along the time dimension and then using the individual cross section of aggregate data can be criticised because the identification of the Euler equation is basically a time series property, and the identification assumptions needed for cross-sectional analysis may be unwarranted.
Carrol (1997) (pp. 22-23) contains a detailed critique of Dynan's test along these lines. In the present setup we are exploiting the time-series dimension therefore the result seems more disturbing. At the same time, the time series dimension of the panel is so short that would be ridiculous to claim asymptotic validity in that direction. It is outside the scope of this chapter to analyze the delicate question of the use of a time-series identifying assumption with a 'short' panel dataset\(^{11}\), let me just notice that an implicit assumption that justifies the standard approach is the assumption that the shocks realized in the population reflect the shocks experienced by a typical individual across his lifetime. This implicit assumption is perfectly valid for example in the simple model simulated in section 2 where only idiosyncratic shocks are present. In presence of aggregate shocks the question is if these aggregate shocks are 'small' enough and if the time dimension available is large enough to average them away. In this sense it would be interesting to redo the estimation above exploiting the full length of the PSID that covers now 25 years.

Incidentally, it is interesting to observe that a recent article by Attanasio and Browning (1995) seems to bring some indirect comfort to the estimation approach based on condition (4.7). They use cohoort data, and in this way they have a much longer time dimension to exploit. Thus, we hope that they are better protected from Chamberlain’s critique. In their IV regressions they include the growth of squared consumption among the explanatory variables and the coefficient they obtain is always significant and positive. This is not the same as including the squared residual but the two variables are likely to be highly collinear. Actually, their interpretation of that additional term is quite different from the one given here: they stick to the standard linearized Euler equation, but they use a generalized version of the CRRA utility function which, in the end, leads them to write the squared term among the explanatory variables. It would be very interesting to try the estimation strategy outlined above on that dataset, and see which conclusion we obtain regarding consumers’ prudence.

\(^{11}\)The point was first raised by Chamberlain (1984). See also Hayashi (1992) and the reply by Keane and Runkle (1992).
4.4.3 Non-parametric regressions

We now return to the attempt to estimate non-parametrically $\text{Var}(\ln C_{it+1}|w_{it}, \theta_{it})$.
Now we will use the decomposition

$$\text{Var}_i(\Delta \ln C_{it+1}|w_{it}) = E_i(\text{Var}(u_{it+1}|w_{it}, \theta_{it})|w_{it}) + \text{Var}_i(E(u_{it}|w_{it}, \theta_{it})|w_{it}) \quad (4.8)$$

This decomposition is very convenient, because if the approximate Euler equation holds, then the second term will have a negligible effect. Unfortunately, the results in the previous section warn us that this cannot be the case, actually this cannot be the case exactly when the decreasing relation between $w$ and the variability of consumption are correlated. So in a sense our exercise is self defeating, because is leads to a specification test that invalidates our estimation. On the other hand the Dynan result, obtained also in the previous section, could be taken to indicate that, the second order terms are quantitatively small, so that $\text{Var}_i(\Delta \ln C_{it+1}|w_{it})$ mostly capture the behavior of $E_i(\text{Var}(u_{it+1}|w_{it}, \theta_{it})|w_{it})$. Here we take this approach and we simply look at estimates of the left hand side of 4.8. The alternative would be to fully spell out a set of identification assumptions that would allows us to estimate the first term in isolation.

Notice that the results reported below are robust to the use of residual from the standard linearized specification and using the residuals from the adjusted specification.

We report the results for the last specification for brevity and because we still hope it provides a better approximation to the behavior of the actual residual. Figures 5 to 7 show the results of a kernel estimate of the nonlinear regression of the squared residual on $w$. For this estimates we have used a simple quadratic kernel. Given the skewness of the distribution of $w$ we found more convenient to transform the wealth variable according to $w' = \log(1 + w)$ and to use the transformed variable as the explanatory variable in computing the nonparametric regressions\textsuperscript{12}. Notice though

\textsuperscript{12}I also computed some regression without transforming the wealth variable, the results were not very different but the choice of the bandwidth was difficult: increasing the bandwidth implied the
that the figures display the estimated conditional expectation as a function of the original wealth variable. Wealth is expressed in terms of the average income.

Figure 4-5: Wealth and residual variability ($w \geq 0$)

Figure 5 confirms the preliminary result obtained in figure 3. Even after controlling appropriately for demographic variables and the real interest rate the residual variability is clearly influenced by the wealth level of the household. From the picture it appears that most of the reduction in variability is achieved at the lower levels of wealth holdings. The analysis is made more difficult because the distribution of $w$ is clearly not continuous, since 30% of the data correspond to zero wealth holdings. Moreover at zero wealth holdings the conditional variability spikes at 0.1651. This accounts for the steep portion of the regression for low values of wealth, and for the fact that decreasing the value of the bandwidth the steep part shifts left. At this point it is interesting to reestimate the regression admitting a discontinuity at wealth zero. Table 6 displays the results of an estimation using only data with $w > 0$. To get an idea of the jump at zero recall that the mean variability at $w = 0$ is 0.1651.

---

use of a lot of observations for low levels of wealth and and decreasing it implied the use of too few observations at high levels of wealth.
In any event the general result is quite clear. There is a very large reduction in variability passing from zero wealth to some limited amount of wealth holdings, for higher values of wealth holdings the reduction is slower but, interestingly enough, it continues all over the wealth range (the 98% percentile of the distribution is 2.26). Another interesting piece of evidence is contained in the joint distribution of the residual and the wealth variable. The picture of the estimated joint density does not seem very informative, but if we compute the conditional density of the residual at different levels of wealth we can obtain the plot in Figure 8 that confirms the results above, at higher wealth levels corresponds a distribution of the residual that is more concentrated around zero.

4.5 Concluding remarks

In this chapter I have analysed the relationship between household wealth and consumption variability. The relationship has been analysed for its bearing on two different questions: (1) the comparison of a precautionary savings model against models of full insurance or quadratic models of the permanent income hypothesis, and (2)
the econometric use of the approximate consumer Euler equation. Regarding the first question the results support the precautionary model as data display a strong negative relationship between wealth holdings and residual variability of consumption growth. The relationship is non-linear with a sharp decrease in variability associated with levels of wealth close to 0, and a weaker response associated to higher levels of wealth to permanent income ratios. This nonlinearity is also consistent with the simulated model in section 2.

From the econometric point of view the results show that the use of the traditional linearized Euler equation may lead to inconsistent estimates, being the identification condition an inaccurate approximation of the original condition. We tried to improve upon the linear identification adding a second order term. This leads to puzzling results similar to those of Dynan (1993). These results may be due to a classical weakness of the Euler equation approach with short panel data, that is, the unwarranted use of a time series identification condition with a small number of time periods. Or they may be due to flaws in the specification of consumer preferences.

When the approximate Euler equation holds we can estimate in a simple way the
non-parametric relation between wealth and consumption variability, even in presence of individual heterogeneity. When the approximate Euler equation fails, though, this direct estimates are no longer reliable and we need to explicitly state the identification assumptions needed to estimate the relation between wealth and consumption variability. Here, we have basically assumed that unobserved individual heterogeneity may affect the slope of the consumption path but not its variability. In future work we plan to extend the model to allow for more general forms of individual heterogeneity.

4.6 References


Chapter 5

Supply of Funds, Maturity and The Spreads of Emerging Market Bonds

Joint with Fernando Broner

During the 1990's a large number of balance-of-payments (BOP) crises took place in the developing world. These crises were blamed, to varying degrees, on a mismatch in the maturity of assets and liabilities in the affected countries. In particular, observers have argued that these countries were relying too heavily on short-term borrowing to finance long-term projects. Empirical studies have found that the ratio of short-term liabilities over liquid assets (e.g. international reserves) can go a long way in explaining why some countries suffered crises while others did not. In Russia in 1998 and Brazil in 1998, governments had accumulated large amounts of short-term debt. In Indonesia, Korea, and Thailand in 1997 domestic financial institutions were responsible for large amounts of short-term borrowing. In Mexico in 1995, both the government and domestic banks played an important role in the crisis, as a large amount of short-term government debt was held by domestic banks which had borrowed short-term abroad.¹ There is by now significant consensus that

¹Although a large part of short-term government bonds were held by foreign investors, these
countries can decrease their vulnerability to capital-flow reversals by lengthening the maturity structure of their liabilities.\(^2\)

However, there is less agreement on why countries sometimes resort to short-term financing, and on what the trade-offs are in the maturity management of sovereign debt. The literature has emphasized the time consistency issues that arise when countries issue large amounts of long-term liabilities. For example, Rodrik and Velasco (1999) and Jeanne (1999) show that opportunistic governments have less incentives to willingly default on their debts and more incentives to carry out revenue-raising reforms when they have to meet early debt repayments. Namely, short-term debt can serve as a commitment device.\(^3\)

This chapter introduces another important aspect to debt management, by focusing on the "supply side" of funds and its effects on the cost of borrowing at different maturities.

Consider first the case in which debt holders are risk-neutral deep-pockets investors. In this case the expected discounted value of debt repayments necessarily equals total borrowing. As a result, spreads on sovereign bonds should only reflect default probabilities and, in the case of local currency bonds, expectations about future inflation. However, it is difficult to reconcile the observed behavior of spreads with this view. In particular, spreads on long-term bonds are too volatile. For example, consider the behavior of the spread (over U.S. bonds) of Argentine dollar-denominated debt of different maturity. The spread on these bonds should be approximately equal to the average of default probabilities throughout the lifetime of the bond. As a result, in periods of financial turmoil, when the probability of default increases, the spread on long-term bonds should increase much less than that on short-term bonds, as crisis periods are very short compared to the lifetime of long-term bonds.

\(^2\)Giavazzi and Pagano (1990) and Alesina, Prati, and Tabellini (1990) formalize this idea.

\(^3\)Missaie and Blanchard (1994) and Calvo (1988) focus on governments' incentives to lower the real value of debt by creating inflation. They show that this incentive is higher when debt is non-indexed, in domestic currency, and long-term. However, this argument is probably more relevant for OECD countries, since developing countries seldom issue long-term debt in domestic currency. (Probably in part for this reason.)

bonds were involved in swap operations with domestic banks (Garber, 1998).
This is not what casual observation of the data suggests. As figure 5-1 illustrates, the spread on long-term bonds increased almost as much as that on short-term bonds during the Asian crisis. To a varying extent, this behavior can also be observed in other crisis episodes.

![Argentine Spreads at Different Maturities During Asian Crisis](image)

Figure 5-1: Spread of dollar-denominated Argentine bonds over U.S. bonds.

In this chapter, we argue that this observation is more easily reconcilable with a world in which borrowing constraints and liquidity shocks play an important role in investors' behavior. We present a simple model of sovereign debt, in which a government must borrow in order to finance a "project". Investors are assumed to be "specialists" who are risk-averse, financially constrained, and subject to liquidity shocks.\(^4\)

In this context, we show that spreads on long-term bonds should substantially rise in periods of financial turmoil, as investors demand a larger risk-premium to compensate for the price risk associated with holding long-term debt. This result is due to the fact that crises are times in which information is likely to be revealed about the future prospects of the country and, in turn, the probability that long-

\(^4\)Since we want to focus on the lender side, we assume that the government is risk-neutral. If this were not the case, the debt maturity structure would also have implications for risk sharing between investors and borrower countries.
term obligations will be met. Since investors face liquidity shocks, and may need to liquidate their positions early, they effectively become more risk-averse and require a high expected return to hold long-term bonds. Short-term debt, on the other hand, carries less price risk and its return is less dependent on the realization of the “signal”.

Governments thus have incentives to issue short-term debt because it is “cheaper”, in the sense that expected debt repayments are lower than when issuing long-term debt. This is due to the fact that short-term debt allows investors to diversify away the idiosyncratic component of the liquidity shocks. Investors without liquidity needs provide funds to buy new short-term, and constrained investors are repaid at a price that does not depend on the realization of the signal.

The model highlights an important trade-off in the choice between short and long-term debt. By issuing short-term debt, countries can reduce expected debt repayments by shaping their debt maturity structure according to investors’ liquidity needs. On the other hand, the larger the proportion of short-term debt, the larger the losses associated with inefficient liquidation of the project.

The chapter is organized as follows. Section 2 presents some evidence supporting the view that spreads on long-term bonds are too volatile to be consistent with investors being risk-neutral and not financially constrained. Section 3 presents a simple model that highlights the concept of price risk. Section 4 analyzes the model. Section 5 provides further evidence for the importance of specialists' balance sheets in determining spreads on long-term bonds. In particular, by studying realized returns and volatilities we show that the Sharpe ratio of holding emerging market bonds increases after price drops. Assuming that a drop in bond prices decreases the wealth of specialists, this is consistent with investors' effective risk aversion increasing when

---

5 In this chapter, we assume that information is revealed about the probability of success of the project. Alternatively, we could assume that the long-term effects of the crisis are unknown, and that investors learn about them as the crisis progresses.

6 We cannot say whether issuing short-term debt is cheaper than issuing long-term debt just by looking at spreads at different maturities, since spreads must themselves be endogenously determined by, among other factors, the debt maturity structure. For example, it is possible that short-term spreads be lower than long-term spreads because long-term debt holders are residual claimants, without implying that issuing short-term debt is cheaper. This is the case in Rodrik and Velasco (1999).
investors are closer to their borrowing constraints. Section 6 concludes.

5.1 Volatility of Spreads on Long-Term Bonds

In this section we show that the behavior of spreads on developing countries’ sovereign bonds does not seem consistent with an environment in which investors are risk-neutral and spreads equal expected default losses. In particular we will show that spreads on long-term bonds move too much given the time-series properties of default hazard rates.\(^7\)

Let \(s_{t,j}\) denote the log-spread on a zero-coupon bond of maturity \(j\) at time \(t\) and let \(s_t \equiv s_{t,1}\) denote the corresponding current variable. Also, let \(f_{t,j}^s\) denote the forward spread, which satisfies

\[
f_{t,j}^s = (j + 1)s_{t,j+1} - js_{t,j}\tag{5.1}
\]

or, equivalently,

\[
s_{t,j} = \frac{1}{j} \sum_{i=1}^{j} f_{t,i}^s.\tag{5.2}
\]

In the rest of this section we assume that forward spreads arise as a result of expected default probabilities. However, the results we report could also be interpreted more generally as a failure of the expectations hypothesis in terms of forward spreads. Let \(m_{t,j}\) denote the default hazard rate, or expected (as of time \(t\)) probability of default at time \(t + j\) conditional on no default before time \(t + j\). Let \(m_t \equiv m_{t,1}\) denote the corresponding current variable. We assume that bonds are fully repaid if no default occurs, but that when default takes place no further payments are ever made.\(^8\) Under risk-neutrality, the forward spreads must account for default hazard rates and, as a result,

---

7 The methods used in this section are similar to those of Campbell, Lo, and MacKinlay (1997) chapter 10.

8 This assumption is made to simplify the algebra. All repayment schedules would give the same results, even if long-term bonds have seniority over short-term bonds.
\[ m_{t,j} = f_{t,j}^s, \quad (5.3) \]

and

\[ s_{t,j} = \frac{1}{j} \sum_{s=1}^{j} m_{t,s}. \quad (5.4) \]

We compute forward spreads using data on spreads on bonds of different maturities, and equations 5.1 and 5.3. We start by estimating the spread curve at time \( t \) by fitting a smooth curve through the existing \( s_{t,j} \).\(^9\) We report the results obtained using a quadratic function, but other functional forms (such as exponentials) led to similar results.

We then use the continuous time version of equation 5.1 and 5.3 to calculate \( m(t,j) \) from the estimated spread curve:

\[ m(t,j) = j \frac{\partial s(t,j)}{\partial j} + s(t,j). \]

Figure 5-2 shows the behavior of the implied default hazard rates at short (1 year), medium (3 years), and long (6 years) maturities as a function of time \( t \), for both Mexico and Argentina. Figure 5-2 shows that during the Asian and Russian crises, the default hazard rates were very volatile at every maturity for both Mexico and Argentina.

According to the \( j \)-period expectations hypothesis (EH), forward spreads should satisfy

\[ f_{t,j}^s = E_t \left[ f_{t+j}^s \right] + \gamma_j, \quad (5.5) \]

where \( \gamma_j \) is a constant risk-premium. Rather than focusing on the \( j \)-period EH, in what follows we test the \( j \)-period pure expectations hypothesis (PEH) by assuming

\(^9\)Data on bond spreads was obtained from Datastream.

\(^{10}\)Since there are no discount bonds for the countries we study, we proxy \( s_{t,j} \) with the spread of a coupon bond of duration \( j \) at time \( t \). Campbell et al show that this is in fact a good approximation.
\( \gamma_j = 0 \) for all \( j \). In fact, both hypothesis impose the same restrictions on the data, since we are only interested on spread changes rather than levels. We thus test PEH for simplicity of exposition. Under the \( j \)-period PEH, the default hazard rates must satisfy:

\[
m_{t,j} = E_t [m_{t+j}].
\]

We assume that the current default hazard \( m_t \) follows the \( AR(1) \) process

\[
(m_{t+1} - \mu) = \rho(m_t - \mu) + \varepsilon_{t+1},
\]
where $\mu$ is the average default hazard rate.\footnote{We are assuming that $m_t$ is stationary. Although the fact that $m_t \in [0, 1]$ implies stationarity, it is still the case that there might be more permanent shocks that we do not observe in our sample. In a next draft, we plan to estimate how likely this "peso problem" must be in order to make the observed behavior of spreads consistent with the expectations hypothesis.} If condition 5.6 holds, this process implies that the default hazard rates must satisfy

$$
(m_{t,j} - \mu) = \rho^{j - 1}(m_t - \mu).
$$

(5.8)

Since we are interested in studying the volatility of spreads rather than the spread curve at a particular point in time, we take first differences in equation 5.8 to obtain\footnote{This also allows us to interpret the results in terms of the EH, since when 5.5 is satisfied, the process 5.7 implies a similar condition:}

$$
(m_{t+1,j} - m_{t,j}) = \rho^{j - 1}(m_{t+1} - m_t).
$$

(5.9)

Thus, there are two ways of estimating the persistence of shocks $\rho$: from the time series properties of $m_t$, and from the relative size of innovations to $m_{t,j}$ and $m_t$. For the former, we estimate the simple $AR(1)$ of equation 5.7. For the latter, we run the regression

$$
(m_{t+1,3} - m_{t,3}) = \beta(m_{t+1,1} - m_{t,1}) + \varepsilon_{t+1}.
$$

(5.10)

To estimate the $AR(1)$ process we used weekly data from January 1995 through January 2000. Since we have short time series for short-term bonds, we used default hazard rates of 2-year maturity.\footnote{Under the null, default hazard rates at all maturities should have the same $\rho$.} To estimate regression 5.10 we used weekly data from January 1997 through January 2000, and default hazard rates for 1-year and 3-year maturities.

Table 1 summarizes the results of both estimations. To make the interpretation of the coefficients more clear, it presents the "half-life" of innovations, rather than $\rho$. Table 1 shows that the half-life of innovations implied by the time-series behavior of
the default hazard rates is significantly shorter than the one implied by the relative movements of hazard rates for different maturities.

<table>
<thead>
<tr>
<th></th>
<th>Mexico</th>
<th>Argentina</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time series</td>
<td>3.4</td>
<td>3.7</td>
</tr>
<tr>
<td>Relative</td>
<td>34.0</td>
<td>22.9</td>
</tr>
<tr>
<td>volatility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95% conf. int.</td>
<td>2.0–12.6</td>
<td>2.4–7.9</td>
</tr>
<tr>
<td>24.2–49.5</td>
<td>16.4–31.5</td>
<td></td>
</tr>
<tr>
<td>N. obs.</td>
<td>202</td>
<td>215</td>
</tr>
<tr>
<td></td>
<td>153</td>
<td>165</td>
</tr>
</tbody>
</table>

Notes: Half-lives are given in months.

The simple empirical analysis presented in this section suggests that the behavior of spreads on emerging market sovereign debt is not consistent with spreads being driven solely by default probabilities. Rather, it seems that the excess risk-premium investors require to hold long-term bonds instead of short-term bonds is negatively correlated with changes in bond prices.

This behavior of spreads can be accounted for if bond holders are specialists who are financially constrained and subject to liquidity shocks. As a result, when bond prices fall, specialists’ wealth decreases bringing them closer to their constrains and increasing their “effective risk-aversion.” This increases the demanded risk-premium, especially on long-term bonds, as long-term bonds carry price-risk and are thus more risky than short-term bonds. In the next section, we present a simple model that incorporates these ingredients.

5.2 The Model

The model is composed of a government, which has to borrow in order to finance maturing debt, and a set of international investors of mass 1. There are three periods.

Debt structure and default

In period 0, the government must borrow $D_0$ in order to finance old debt coming to maturity. The government can sell either short-term (1-period) or long-term (2-period) bonds. In period 1, the government pays the short-term bonds issued in
period 0 by issuing new short-term debt and by generating revenue. In period 2, the
government generates revenue, pays maturing long and short term bonds, and “con-
sumes” the rest. We abstract from strategic default by assuming that the government
repays its debts whenever it is feasible.\footnote{We are implicitly assuming the existence of costs of default. These costs can be reputational, or involve direct interference by creditors on debtors’ transactions in international goods and capital markets. (See Bulow and Rogoff (1989) for a discussion of the later.)}

The government’s budget constraint in period 0 is

\[ D_0 = q_S D_S + q_L D_L \]

where \( D_S \) and \( D_L \) are the amount of short-term and long-term bonds issued in pe-
riod 0, and \( q_S \) and \( q_L \) are their respective prices.

In period 1, the government has to roll over an amount \( D_S \) of short-term bonds.
The government’s budget constraint in period 1 is

\[ D_S = q_{S,1} D_{S,1} + X \]

where \( D_{S,1} \) is the amount of short-term bonds issued in period 1, \( q_{S,1} \) is their price,
and \( X \) are government revenues in period 1. In order to generate an amount \( X > 0 \)
of revenues in period 1, the government has to resort to “emergency” finance which
entails a cost \( X + c \). The cost \( c \geq 0 \) incorporates the inefficiencies associated with
raising resources “too soon”, and can be thought of as arising from the premature
liquidation of long-term projects (for example, through excessive taxation).\footnote{The choice of \( X = 0 \) as the maximum amount of revenues that can be generated without incurring the cost \( c \) is just a simple normalization.} It is
assumed that the cost \( X + c \) affects the country’s welfare, but does not affect the
availability of resources in period 2.\footnote{By assuming the existence of a fixed cost \( c \) that does not decrease future resources we can solve
the model very easily. Alternatively, we could assume that the “liquidation cost” is proportional
to \( X \), and that it reduces resources in period 2. In both cases, liquidity crises arise when period 2
resources are expected to be too low to allow the government to issue a sufficient amount of short-term
debt in period 1 to repay all maturing short-term debt. Under the second set of assumptions, the
liquidation costs would also resemble a fixed cost, as a small reduction in expected future resources
can make the equilibrium in which investors accept to roll over the short-term bonds disappear,
generating a sizeable liquidation cost. (Namely, liquidation costs are discontinuous in the sense that}
In period 2, the government revenue is $Y$, where $Y$ is a random variable which takes the value $\bar{Y}$ in the good state and 0 otherwise. The extreme case of zero realization in the bad state, and the fact that it will always be the case that in the good state no default takes place (i.e. $\bar{Y} \geq D_{S,1} + D_L$) considerably simplifies the analysis. This is because, in equilibrium, there is never partial default. In order to avoid the possibility of dilution of long-term debt by issuing short-term bonds in period 1, we assume that long-term bonds are senior.

As of period 0, the probability of $Y = \bar{Y}$ is $p_0$. We assume that between periods 0 and 1, information is revealed about the likelihood of success and the probability is updated to $p$. As of period 0, $p$ is a random variable distributed with distribution $F$, which by construction satisfies $p_0 = \int p \, dF(p)$.\(^{17}\)

The government maximizes the objective function

$$W = E_0 \left[ \max \left\{ \bar{Y} - D_L - D_{S,1}, 0 \right\} - I_{(X > 0)} (X + c) \right]$$

where the first term accounts for the resources that can be consumed by the country's residents in period 2 (i.e. output minus debt payments) minus the costs incurred in order to raise revenue in period 1.

**Investors**

Investors preferences are described by

$$E_0 [c_0 + u(c_1) + c_2]$$

where $u$ is a concave function. Investors' budget constraint is

$$b_0 + q_Sd_S + q_Ld_L + c_0 = w_0$$

(they are either zero or large.)

\(^{17}\)Alternatively, we could assume that in period 1 there is a shock that affects the repayment capacity of the country in period 2. This would make the analysis more difficult if it involves the possibility of partial default.
\[ b_1 + q_{S,1} d_{S,1} + q_{L,1} d_{L,1} + c_1 = w_1 + b_0 + d_S + q_{L,1} d_L \]

\[ c_2 = I_{(y = \bar{y})} (d_{S,1} + d_L) + b_1 \]

\[ c_i, b_i, d_{ij} \geq 0 \]

where the \( d \)'s denote holdings of the country's bonds, the \( q \)'s denote bond prices, and the \( b \)'s denote holdings of a risk-free international short-term asset which is offered at exogenous price 1 (e.g. US treasury bills).

Investors' income in period 1, \( w_1 \), is a random variable distributed with density \( g \). We assume pure idiosyncratic uncertainty and, as a result, \( g \) corresponds to the ex post cross sectional distribution of \( w_1 \).\(^{18}\)

This setup captures the idea of specialized investors with limited wealth and subject to liquidity shocks.\(^{19}\) We can think of liquidity shocks as arising from low cash flows in other activities or because of high returns in alternative investment opportunities. Let us define \( \bar{c} \) such that

\[ u'(\bar{c}) = 1. \]

The higher an investor's initial wealth, the more she would behave as risk-neutral. In particular, if she could achieve period 1 consumption larger than or equal to \( \bar{c} \) with probability 1, she would act as totally risk-neutral. Otherwise, when \( w_1 \) is low enough such as the period 1 borrowing constraint binds, she liquidates all her portfolio. This corresponds to \( c_1 = w_1 + b_0 + d_S + q_{L,1} d_L < \bar{c} \), as portfolio liquidation entails selling the whole portfolio at current prices.

An essential feature of the model is that the period-1 price of long-term debt \( q_{L,1} \) is a random variable affected by news on revenue prospects for the government. As a result, holding long-term bonds between periods 0 and 1 is more risky than holding short-term debt.

---

\(^{18}\) In a future draft we will study the case with aggregate uncertainty.

\(^{19}\) Holmstrom and Tirole (1998) derive similar reduced form preferences from primitive assumptions about alternative investment opportunities.
5.3 Equilibrium

We solve the model in two steps. First, we take the maturity structure as given (namely the choice of \(D_L\)) and find the equilibrium bond prices and investment decisions. Then, we choose the maturity structure such as to maximize welfare. We divide the analysis into three special cases. In case 1, there is no information revealed in between periods 0 and 1. As a result, in period 1 investors are able to sell their assets at the same price at which they bought them and, similarly, the government can refinance its debt at constant terms. The debt structure is thus irrelevant. In case 2, there is some information revelation in period 1, but the ex post probability of default is always small enough such that there is no problem rolling over short-term debt in period 1, even in the case in which the all liabilities are short term debt. In this case, since investors dislike the price risk associated with long term debt, full short term financing is optimal. In case 3, the ex post probability of default can be large enough such that the government has trouble rolling-over the short-term debt in period 1. In this case, the trade off between optimal risk sharing with investors and illiquidity risk implies an optimal debt structure with both short and long term liabilities.

Case 1: No price risk

No information is revealed between periods 0 and 1; namely, \(F\) is degenerate and \(p = p_0\). We make the following two assumptions:

**Assumption 1** The government is ex-ante solvent:

\[ p_0\bar{Y} > D_0 \]

**Assumption 2** Investors expected resources satisfy

\[ \int_{\bar{c} - w_0}^{\infty} (w_1 + w_0 - \bar{c})g(w_1)dw_1 \geq D_0 \]

where \(u'(\bar{c}) = 1\).
Under assumptions 1 and 2, for any \( D_S \in [0, D_0] \) there is an equilibrium in which bond prices satisfy

\[
q_S = 1 \quad (5.11)
\]

\[
q_L = q_{L,1} = q_{S,1} = p_0.
\]

The government can issue any combination of short and long-term bonds such that \( p_0D_L + D_S = D_0 \) and the short-term bonds are rolled over in period 1 by issuing an amount \( D_{S,1} = \frac{1}{p_0}D_S \) of new short-term bonds.

In period 1, investors whose income is such that \( w_0 + w_1 < \bar{c} \) liquidate all their assets and consume \( w_0 + w_1 \). Those for whom \( w_0 + w_1 \geq \bar{c} \) consume \( \bar{c} \) and buy the outstanding long-term bonds and the new short-term bonds.

Assumption 1 guarantees that the government can always roll over all short-term debt in period 1. Assumption 2 guarantees that unconstrained investors have enough resources to buy all outstanding long-term bonds and the new short-term bonds issues. Since unconstrained investors have access to the risk-free asset the risk neutral pricing equations 5.11 have to hold.

Since bonds carry no risk-premium it is trivial to prove the following proposition.

**Proposition 11** If no information is revealed between periods 0 and 1, and if assumptions 1 and 2 hold, the government is indifferent among all debt structures that satisfy \( p_0D_L + D_S = D_0 \). For all debt structures the government payoff is constant and equal to \( W_1 = p_0\bar{Y} - D_0 \).

**Case 2: Pure price risk**

Consider now the case in which information is revealed between periods 1 and 2 on the probability of the good state. In this case, the price of long-term bonds in period 1 depends on the realization of \( p \). Since investors who receive low income in period 1 need to liquidate their investments, they need to be compensated \textit{ex ante} for the price-risk associated with holding long-term bonds.
We assume first that short-term debt can always be rolled-over in period 1, with no need of emergency finance.

**Assumption 3** The probability of the good state as of period 1 is always high-enough such that the short-term debt can be rolled over. There is some $\varepsilon > 0$ such that 

$$F\left(\frac{D_0}{Y} - \varepsilon\right) = 0.$$ 

**Assumption 4** Investors' wealth satisfies 

$$G(\bar{c} - w_0) > 0$$

**Proposition 12** Under assumptions 1 through 4, there is a $D < D_0$ such that for any $D_S$ in $[D, D_0]$, there is an equilibrium with $X \equiv 0$ in which bond prices satisfy 

$$q_S = 1$$

$$q_L = Q(D_L)$$

$$q_{S,1} = q_{L,1} = p$$

$Q(\cdot)$ satisfies $Q(D_L) \leq p_0$ with equality if and only if $D_L = 0$.

*Proof:* Assumption 2 guarantees that unconstrained investors have enough resources in period 1 to buy all outstanding long-term bonds and enough new short-term bonds to repay the period-0 short-term bonds. Assumption 3 guarantees that there is never partial default provided that $D_L$ is small enough. Since unconstrained investors have access to the risk-free asset, ex post prices must satisfy $q_{S,1} = q_{S,1} = p$.

The ex ante asset pricing condition for long-term bonds is $q_L = \frac{E_0[u'(c_t)p]}{E_0[u'(c_t)]}$. If $D_L = 0$, investors do not hold any long term bonds and their period-2 consumption is independent of $p$. Therefore we obtain $q_L = \frac{E_0[u'(c_t)p]}{E_0[u'(c_t)]} = E_0[p]$ which implies $Q(0) = p_0$. If instead they hold a positive amount of long-term debt, and assumption 4
holds, period-2 consumption will have a positive covariance with \( p \) and since \( u' \) is decreasing we get \( q_L < \frac{E_0[u'(c_1)]E_0[p]}{E_0[u'(c_1)]} = E_0[p] \) which implies \( Q(D_L) < p_0 \) if \( D_L > 0 \).

In the equilibrium described above the government issues an amount \( D_0 - q_LD_L \) of short term debt. Therefore, in period 1 the government is able to roll over the short-term bonds if \( p\bar{Y} - (p - q_L)D_L \geq D_0 \). For \( D_L \) small enough this condition is always met. However, if \( D_L \) is large, it can be the case the condition is not met and the government is forced to raise revenue in period 1 when \( p \) is low. In this case there is no equilibrium with \( X \equiv 0 \).

As a result, for low levels of long-term debt the government’s payoff in equilibrium is equal to

\[
W = p_0\bar{Y} - D_0 - (p_0 - q_L)D_L, \tag{5.12}
\]

while for high levels of long-term debt the government’s payoff is lower than that.

The following proposition follows immediately.

**Proposition 13** Under assumptions 1 through 4, the optimal debt structure consists of issuing only short-term debt. In this case, the government’s payoff is equal to the one in case 1, namely \( W_1 = p_0\bar{Y} - D_0 \).

This simple example illustrates in which type of environments the price-risk associated with holding long-term bonds makes long-term borrowing for the government expensive. First, investors must expect that in the near future information will be revealed about the long-term prospects of the country. And second, investors must be close to their borrowing constraints, and be subject to liquidity shocks. Both of these features are present in episodes of international financial markets turmoil. On the one hand, crises are times in which a large amount of information is revealed about emerging-market fundamentals.\(^{20}\) On the other hand, investors specialized on emerging-market assets are in danger of becoming constrained, either by withdrawals from their own investors or by further loses.

\(^{20}\) Alternatively, we could just assume that countries are subject to shocks that reduce their long-term prospects, especially during crises.
However, as previously argued, issuing too much short-term debt increases countries’ vulnerability to reversals in capital flows. The trade-off is made clear in the case we consider next.

Case 3: Liquidity crises

We now consider the case in which assumption 3 does not hold. As a result, for any debt structure there is a positive probability that the government cannot issue enough short-term bonds in period 1 in order to repay maturing debt. Since the maximum amount of funds the government can raise in period 1 is \( p(\bar{Y} - D_L) \), the government is forced to set \( X > 0 \) if and only if \( p(\bar{Y} - D_L) < D_S \). In such a case, the government sets \( X = D_S - p(Y - D_L) \), and incurs the fixed welfare cost \( c \).

For a given debt structure, let \( \hat{p} \) be the minimum \( p \) such that the government is solvent in period 1,

\[
\hat{p} = \frac{D_0 - q_L D_L}{\bar{Y} - D_L}.
\]

From investors’ point of view nothing changes since short term debt in repaid with probability 1 in period 1. As a result, prices are determined as in proposition 12. The government’s objective function, on the other hand, now includes a term for the costly roll-over and takes the form

\[
W = p_0 \bar{Y} - D_0 - (p_0 - q_L)D_L - cG(\hat{p}).
\]

In this case the optimal debt structure may involve the use of both short and long term debt. The first order condition for government maximization is

\[
-(p_0 - q_L - Q'D_L) - cg(\hat{p}) \frac{d\hat{p}}{dD_L} = 0.
\]

The first term is positive and reflects the fact that investors value long term debt less than \( p_0 \) because of its associated price-risk. The second term reflects the marginal

135
gain in terms of reduced risk of a costly roll over, and is typically positive.\footnote{It is easy to show that $\frac{\partial Q}{\partial D_L} < 0$ if the elasticity of $Q$ is not too large.}

To illustrate the nature of the optimal solution and to provide some simple comparative statistics we present an example. We assume that $u$ is quadratic

$$u(c_1) = Ac_1 - Bc_1^2$$

and that $w_1$ is a binary random variable. In this case, we can obtain an explicit form for $Q$, and derive an explicit expression for the government’s objective function. Using the pricing equation in proposition 12 the ex ante price of long-term bonds is

$$Q(D_L) = p_0 - \frac{1}{2BD_L} \left( (1 + A) - \sqrt{(1 + A)^2 - 4B^2D_L^2\sigma_p^2} \right).$$

Note that the second term is zero if any of the following applies: consumers are risk neutral ($B = 0$), no information is revealed between periods 0 and 1 ($\sigma_p^2 = 0$), or no long term debt is issued ($D_L = 0$). As shown above, in all three cases the spreads on long-term bonds only reflect the probability of default. If instead neither condition applies, the bond price will be strictly smaller than $p_0$ reflecting the price risk of long-term bonds.

The expression for the government’s payoff is

$$W = p_0\bar{Y} - D_0 - (p_0 - q_L)D_L - cG(\bar{p})$$

$$= p_0\bar{Y} - D_0 - \frac{1}{2B} \left( (1 + A) - \sqrt{(1 + A)^2 - 4B^2D_L^2\sigma_p^2} \right) +
\frac{cG}{\bar{Y} - D_L} \left( \frac{D_0 - p_0D_L + \frac{1}{2B} \left( (1 + A) - \sqrt{(1 + A)^2 - 4B^2D_L^2\sigma_p^2} \right)}{\bar{Y} - D_L} \right)$$

When $\sigma_p^2 = 0$, the expression above is equal to $p_0\bar{Y} - D_0$ for all $D_L$ and we obtain the irrelevance result of case 1. When $c = 0$ the last term disappears and $W$ is strictly decreasing in $D_L$. Therefore, if the government can raise emergency finance
with no distortionary effects it will choose to issue only short-term debt. Finally, under risk neutrality and \( c > 0 \), the third term disappears and \( W \) is unambiguously non-decreasing in \( D_L \). As a result, it is optimal to issue a large amount of long term debt at price \( p_0 \).²²

### 5.4 Empirical Evidence

This section provides preliminary empirical evidence consistent with the main assumptions and predictions of the model presented above. We study the ex-post excess returns on emerging-market debt, after drops in bond prices, concentrating on Argentine and Mexican sovereign bonds. We show that: (i) the expected returns on emerging-market bonds increase substantially during crises; (ii) the increase is much more pronounced for long-term bonds; (iii) there is no appreciable increase in volatility that could account for the increase in expected returns; (iv) the returns on long-term bonds are more volatile; (v) the rise in Sharpe ratios is increasing in the size of previous price drops; and (vi) the rise in Sharpe ratios is increasing in the holding period (for holding periods between 4 and 16 weeks).²³

To determine how the behavior of bond returns depends on their maturity, we construct return indices for different maturities. Given data limitations, we estimated return indices for three maturities: short (residual life of less than 2 years), medium (residual life between 2 and 10 years) and long (residual life of more than 10 years).²⁴ For each point in time, we estimate the return for different maturities by averaging the return on bonds whose maturities fall within the specified range. Figure 5-3

---

²²We have also computed the optimal debt structure for different parameters in the case of \( c > 0 \) and \( p \) uniformly distributed over \( [p_0 - \varepsilon, p_0 + \varepsilon] \). As one would expect, a larger value of \( c \) implies that more long term debt is optimal, while an increase in investors risk aversion implies a shorter optimal maturity. An increase in price variability increases the probability of a costly roll-over, while at the same time lowers the ex-ante price of long-term bonds. Although this seems to imply an ambiguous effect on the optimal debt structure, in the examples we simulated the price risk effect always dominates and the optimal amount of short-term bonds increases with \( \sigma_F^2 \).

²³The last result was unexpected and seems at odds with the spirit of the model.

²⁴Datastream provides data for all outstanding sovereign bonds. As a result, we do not have data on bonds that already expired, which limits the length of our return series, especially for short maturities.
shows the return indices for Argentina. The figure illustrates the high price-volatility of long-term bonds.

The main prediction of the model is that investors should be more sensitive to the high volatility associated with holding long-term bonds when they are closer to their borrowing constraints. Since we cannot observe investors’ balance sheets, we use the past performance of bond prices as a proxy for investors’ capital level. Specifically, we look at the return during the previous 4 weeks on the JPMorgan EMBI index corresponding to the country under study.

The following investment strategies are considered. If the JPMorgan index fell by more than $D$ in the previous 4 weeks, invest in Argentine bonds of maturity $M$ for $H$ weeks, where $D \in \{0\%,3\%,5\%,10\\%\}$, $M \in \{\text{short, medium, long}\}$, and $H \in \{4, 8, 12, 16\}$. We estimate the expected excess returns (over U.S. bonds) and riskiness of each strategy by taking the average and standard deviation of the ex-post returns over all episodes in which the general index fell by more than $D$.

![Return Indices](image)

Figure 5-3: Return-indices of Argentine sovereign bonds at different maturities.

Figure 5-4 shows the estimated expected excess returns for bonds of different maturities, for crises of different magnitude. Each panel corresponds to different holding periods. It is clear that returns of longer maturities have higher expected excess returns and that, for each maturity, the expected returns increase with the
depth of the crisis.

However, we need to account for the possibility that the increase in excess returns be a response to an increase in volatility. Figure 5-5 plots the estimated standard deviations for each strategy. Although the figure shows that the return on long-term bonds during crises is indeed larger than that of short-term bonds, the volatility of returns does not seem to increase appreciably with the magnitude of the crisis.

Figure 5-6 illustrates how the relationship between expected excess returns and volatility changes during crises. The figure highlights a number of important points. First, even though long-term bonds have higher excess returns, this can be explained by their higher associated volatility. In other words, the Sharpe ratio (i.e. the ratio of expected excess return over its standard deviation) of long-term bonds is similar to that of short term bonds. In addition, the Sharpe ratio is increasing in the magnitude of the crisis. This is consistent with investors "effective" risk-aversion increasing as they get closer to their borrowing constraints. Figure 5-7 summarizes this result by plotting the average Sharpe ratio (over different maturities) as a function of the size of the previous price drop. This figure also shows a result which we cannot explain: the Sharpe ratio increases with the holding period, at least for the range of holding periods considered here.

5.5 Concluding Remarks

In this chapter we show that in order to understand the behavior of spreads on developing country sovereign debt it is necessary to take into account the supply side of funds; in particular, the fact that spreads present a large (and highly volatile) risk-premium component.

We further show that the behavior of spreads is consistent with an environment in which bonds are held by specialists subject to liquidity shocks. After drops in bond prices (which presumably bring investors closer to their borrowing constraints),

\[25\] Namely, the points for different maturities fall approximately on a straight line through the origin.
the risk-premium on emerging market bonds rises sharply. Moreover, the rise in the
risk premium seems to reflect an increase in the effective risk aversion of investors as
changes in the volatility of returns, which remain almost constant, cannot account
for the increase in demanded excess returns.

The chapter also shows that the behavior of spreads has important implications for
debt management. Since the price of long-term bonds is very volatile, investors require
a higher risk-premium on them than on short-term bonds. As a result, governments
can save substantially on financing costs by adapting the debt maturity structure
to the liquidity needs of investors. In particular, in periods in which investors are
subject to liquidity shocks, such as in times of financial turmoil, governments should
try to shorten the maturity of debt. However, there is a trade-off in the choice of
debt maturity, since the probability that governments will face difficulties in rolling
over their debt increases with the amount of short-term bonds issued.

Further work is needed in a number of areas. First, we need to obtain longer time
series for short-term bonds in order to make sure that the empirical results we present
are robust. Second, we need to take into account the "peso problem," namely the fact
that the spreads observed in our sample might reflect the possibility of a large crisis
that is not present in our sample.\textsuperscript{26} One way of addressing this issue is to estimate
how likely (and how large) such a crisis would need to be to explain the observed
spreads.\textsuperscript{27} Third, we need to extend the model to account for the possibility of
aggregate liquidity risks. And fourth, the model should have more than three periods
to account for more than two debt maturities. The reason why this is important is
that the roll-over difficulties associated with short-term debt are likely to fall rapidly
as the maturity increases, while the price-risk associated with long-term debt is likely
to fall more slowly. As a result, it seems reasonable that the optimal debt policy in
episodes of crises be to issue medium-term bonds (around 4 or 5 years) rather than
very short or very long-term ones.

\textsuperscript{26}For example, there are no episodes of default in our sample. Benczur (2000) addresses this issue
by using spread data covering a much longer period.
\textsuperscript{27}However, it seems unlikely that a peso problem could account for the difference in the behavior
of short and long-term bonds, as long-term bonds are in general senior.
5.6 References


Figure 5-4: Expected excess returns on Argentine bonds during crises.
Figure 5-5: Volatility of returns on Argentine bonds during crises.
Figure 5-6: Excess returns vs. standard deviation on Argentine bonds during crises. Points farther from the origin always correspond to longer-term debt.
Figure 5-7: Sharpe ratio on Argentine bonds as a function of previous price drops.
Figure 5-8: Returns, volatility, and Sharpe ratio on Mexican bonds.