The London School of Economics and Political Science

Liquidity, information and coordination in financial markets

Gara Mínguez Afonso

A thesis submitted to the Department of Finance of the London School of Economics for the degree of Doctor of Philosophy, London, May 2008.

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Declaration of Joint Work

I certify that Chapter 3 of this thesis describes joint work with Hyun Song Shin. The work presented in Chapter 3 was conceived together, the theory was developed in somewhat different directions, but the final formulation reflects input from both authors. Examples and simulations were prepared by me. Overall, my net contribution is 60%.

Abstract

No two crises are identical. As we learn from them, they evolve and change. This thesis is an attempt to understand some of their features. We discuss the abandonment of a peg (Chapter 2), full disruption of payments (Chapter 3) and illiquidity in one-sided markets (Chapter 4).

Chapter 2 investigates the consequences of introducing uncertainty about the willingness of a central bank to defend the peg in an economy in which a government runs a persistent deficit. We analyze how not knowing when other arbitrageurs intend to attack a currency affects investors' decision to attack. Specifically, we show how the lack of common knowledge induces arbitrageurs to delay their attack on the currency, which in turn leads to a discrete devaluation of the exchange rate as it is generally observed during currency crises.

In Chapter 3 we examine how financial integration of payment systems creates a feedback channel which might threaten the stability of financial markets. In payment systems, banks rely heavily on incoming transfers to finance outgoing payments requiring a high degree of coordination and synchronization. We study the response of the payment system to disruptions in payments and to changes in the precautionary demand for liquid balances targeted by the different institutions in a payment system. This work aims to shed light on the recent events in credit markets.

The analysis of liquidity in one-sided markets is the focus of Chapter 4. When a market is in distress, liquidity typically vanishes playing a key role in the build-up of one-sided markets. We present an alternative view of market liquidity which results from a tradeoff between market externalities and a congestion effect. When congestion is the dominating effect, as during fire sales, diminishing market frictions can deteriorate liquidity and reduce welfare. Our results provide a rationale for circuit breakers.

Acknowledgements

I am profoundly indebted to Hyun Song Shin for continuing guidance and endless encouragement. By any measure, the best advisor anyone could wish for.

I am sincerely beholden to Dimitri Vayanos and Markus Brunnermeier for their invaluable advice. I would have never survived the 'job market experience' without their help.

For helpful lessons I am grateful to Nobuhiro Kiyotaki. Also, I would like to thank Antoine Faure-Grimaud for enriching discussions. I wish to express my gratitude to the faculty and colleagues at the London School of Economics, Princeton University, Bendheim Center for Finance and Universidad Carlos III. I am especially thankful to Miguel Artola Menéndez for opening the door to a whole new world of possibilities and to Juan José Dolado for his enthusiastic support and to Oriol Aspachs Bracons, Max Bruche, Camille Cornand, Amil Dasgupta, Antoni Espasa, Péter Kondor, Beatriz Mariano, Andrew J. Patton and Jean-Pierre Zigrand.

I owe more than one favour to Runquan Chen. Thank you very much for all your help and for never skipping the Corporate Finance Lectures. I think we will never forget those lectures...

I am very appreciative of the financial support from the Banco de España, Fundación Caja Madrid, the Economic and Social Research Council (research grant RES-156-25-0026), Fondation Banque de France and Fundación Ramón Areces.

I am thankful to my job market "buddie", Julio Cacho Díaz, and his family Claudia, Julio, Paulina and Regina for friendship in time of war. Also, I am enormously grateful to my friends Mari Paz Jr., María Luisa, Angela, Sebastián, Nicolás, Juan, Francis, Viana, Amanda, Mari Paz, Lola, Isa, Eunice, Llanos, Ana, Stéphanie, Gaurav, Arthur, Bernhard, André, Sebas and Juanjo for still being my friends after the job market.

I would like to thank my family for their lifelong love and support. This would have never been possible without your help. I am extremely grateful to Javier and Josefina for their encouragement and understanding and to Cayetano for teaching me the true meaning of teamwork.

From the first school play to graduation, I deeply thank Yeray for her unconditional support. Sis, you are the best! (super-green... well, super-pink if you must.)

I could never repay the debt I owe my mother. Thank you for making me believe I could reach the sky!

Fernando, I could never thank you enough. You are the Maltese falcon.

Gara Mínguez Afonso

I could not finish without sharing with you some of the enlightening advice I have received throughout my PhD. Thank you and enjoy!

How do you write a good paper?

By reading good papers.

Work like crazy.

Stress works.

The job market is the best advertisement for your paper.

Forget about getting a job. You will get one.

We all need luck all the time.

Now the tenure clock is ticking... Can you hear it?

Nobody ever will read your thesis.

To Rosalba and Fernando

In confusion, there is profit.

– OPERATION PETTICOAT (1959)

Contents

CHAPTER 1

Introduction

FINANCIAL CRISES are not a new phenomenon. Historically, they have emerged under a broad variety of financial systems with dissimilar monetary and regulatory regimes. Prominent episodes include the 1720 crisis following the South Sea and Mississippi bubbles[1], the 1763 crisis, the 1928-1933 episode in the US, and more recently, the Scandinavian crises of the early 1990's, the Asian crisis of 1997, the 1998 Russian crisis and the LTCM episode, the 2001-02 Argentinean crisis and even the current turmoil in financial markets[2].

The devastating effects of financial crises[3] have long attracted the interest of academics

[1] A review on the literature on bubbles can be found in Brunnermeier (2008a).

[2] See Kindleberger and Aliber (2005) and Visano (2006). For an account of the 1763 crisis, see Schnabel and Shin (2004).

[3] Bordo et al. (2001) analyzes the costs and frequency of financial crises. See also Boyd et al. (2005).

and policymakers aiming to extract some helpful lessons for future episodes. However, no two crises are identical and as we learn from them, crises evolve and change. As Alan Greenspan puts it in a recent article in the Financial Times[4]:

> In the current crisis, as in past crises, we can learn much, and policy in the future will be informed by these lessons. But we cannot hope to anticipate the specifics of future crises with any degree of confidence.

This constitutes a great challenge for economists seeking an explanation of the onset of financial crises and a considerable disappointment for those aiming to completely avoid them. Traditionally, different theories have highlighted various, sometimes complementary, mechanisms to explain the build-up and contagion of crises. However, in practice, financial crises are very complex phenomena[5]. This thesis presents a theoretical approach to different financial crises in an attempt to better understand some of their distinctive features. We analyze the abandonment of a peg in Chapter 2, full disruption of payments in a payment system (Chapter 3) and, in Chapter 4, the role of liquidity in asset market crashes (and booms).

Chapter 2 focuses on a currency crisis in an economy with a fixed exchange rate and inconsistent government macroeconomic policies that deplete the central bank foreign reserves. Traditional first-generation models of currency crises assume investors are perfectly informed about macroeconomic fundamentals and hence the transition from a fixed to a floating exchange rate occurs without jumps in the exchange rate. The original model is due to Krugman (1979). These models conclude the attack on the currency is accurately predictable and thus implies zero devaluation. However, in practice, large discrete devaluations

[4]'We Will Never Have a Perfect Model of Risk' by Alan Greenspan, *Financial Times*, 16 March 2008 (Greenspan (2008)).

[5]Allen and Gale (2007) presents an excellent discussion of theories of financial crises.

are observed during currency crises.

We investigate the consequences of introducing uncertainty about the willingness of a central bank to defend the peg in an economy in which a government runs a persistent deficit. The main idea relies on the assumption that even if arbitrageurs could precisely know the underlying value of the currency (the shadow price), they would not want to share this information with other investors. We analyze how not knowing when other arbitrageurs intend to attack a currency affects investors' decision to attack. Specifically, we show how the lack of common knowledge induces arbitrageurs to delay their attack on the currency. Not knowing what others know gives arbitrageurs incentives to hold this currency for some time instead of selling it immediately as soon as they realized it is overvalued. They prefer to invest in this currency because it is overvalued, but they would not want to hold it for too long and suffer the capital loss associated with the devaluation of the currency. Arbitrageurs optimally prefer to ride "the bubble" for a while, in Abreu and Brunnermeier (2003)'s spirit, before they attack the currency. Delaying the attack leads to a discrete devaluation of the exchange rate as it is generally observed during currency crises.

Chapter 3 is motivated by the recent events in credit markets[6] and aims to address the question of what would be the consequences to the payment system if one (or few) bank(s) would target more conservative balances. We would like to understand what could happen if one bank were to increase its precautionary demand for liquid balances to conserve cash holdings because some conduits, SIVs or other off-balance sheet vehicles that this bank is sponsoring have drawn on some credit lines.

In the U.S. reserve banking system, banks in aggregate make payments that exceed by a factor of more than 100 their deposits at the Federal Reserve Banks. To achieve such

[6]Greenlaw et al. (2008) reviews the key credit market events since August 2007 and highlights their main policy implications. Also, Fender and Hördahl (2007) presents an overview of the key events over the period from end-May to end-August 2007.

velocities, a high degree of coordination is required. We examine how financial integration of payment systems creates a feedback channel which might threaten the stability of financial markets. The events of September 11, 2001, highlighted how vulnerable the financial system is to disruptions to its payment system[7]. Those disruptions damaged the ability of some financial institutions to execute payments and as a result other entities received fewer payments than expected. Lesser incoming funds impaired their capacity to send out payments causing unexpected shortfalls and liquidity shortages across the entire financial system.

In a payment system, where banks rely heavily on incoming funds to finance outgoing payments, if a bank were to target a more conservative balance, it would then alter the value of payments this bank sends relative to the payments it receives. But changes in outgoing transfers will affect incoming funds and incoming funds changes will then affect outgoing transfers. We use lattice-theoretic methods to solve for the unique fixed point of an equilibrium mapping and conduct comparative statics analysis on changes to the environment. To study the response of the payment system to disruptions in payments and to changes in the precautionary demand for liquidity, we present simulations of a stylized payment system reminiscent of the U.S. Fedwire payment system. We find that changes towards more liquid balances induce an enormous increase in the use of intraday credit provided by the Federal Reserve Banks and, ultimately, full disruption of payments.

Chapter 4 considers dramatic collapses in asset prices when liquidity is scarce. When a market is in distress, liquidity typically vanishes playing a key role in the build-up of one-sided markets. In Chapter 4 we focus on the study of liquidity in one-sided markets to better understand the response of financial systems to the threat of market disruptions such

[7]McAndrews and Potter (2002) gives a detailed account of the consequences of the September 11, 2001, events to the US payment system.

as the 1998 LTCM crisis[8], the September 11, 2001, events[9] and the recent credit market turmoil[10].

In financial markets, we typically think of liquidity as a coordination phenomenon. Investors moving into a market facilitate trade for all investors by reducing the cost of participating in this market. Lower trading costs then attract potential investors. As a result, a market externality is created in which new investors provide market liquidity and market liquidity attracts new investors. However, if investors preferred to concentrate on the same side of the market, trade would become more difficult, decreasing the returns to participating in this market. Lower returns would discourage new investors from entering this market. A negative externality arises. Consider, for instance, a fire sale where investors wish to sell their holdings of an asset. The sell-side of the market for this asset is "congested" and as it gets congested potential buyers may prefer to invest in other markets, making it even more difficult for the sellers to exit this market. The arrival of investors to the congested side of a market "repels" new investors.

We present an alternative view of market liquidity, in which liquidity results from a tradeoff between two effects: the well-known market externality and a congestion effect. When congestion is the dominating effect, for example during fire sales, diminishing market frictions can deteriorate liquidity and reduce welfare. This might seem counterintuitive but it is a very interesting result. It states that fewer frictions in a distressed market can make the market less liquid and investors worse-off. Or equivalently, a policy intended not to deteriorate liquidity or welfare but to enhance it would have to increase frictions in this market, at least temporarily. An example of such policy would be a "circuit breaker"[11]

[8]For an analysis of the events surrounding the market turbulence in autumn 1998, see BIS (1999) and IMF (1998).

[9]Cohen and Remolona (2001) presents a summary of the September 11, 2001 episode in global financial markets.

[10]See Brunnermeier (2008b).

[11]A circuit breaker is a pause in trading at predetermined thresholds during a severe market decline. It is

which could halt trading if market declines beyond trigger levels. Our results thus provide a rationale for circuit breaker trading halts.

References

Abreu, D. and M. K. Brunnermeier (2003). Bubbles and Crashes. *Econometrica 71*(1), 173–204.

Allen, F. and D. Gale (2007). *Understanding Financial Crises.* New York: Oxford University Press.

BIS (1999, October). A Review of Financial Market Events in Autumn 1998. *CGFS Publication* (12). http://www.bis.org/ publ/ cgfs12.htm.

Bordo, M., B. Eichengreen, D. Klingebiel, and M. S. Martínez-Peria (2001, April). Is the Crisis Problem Growing More Severe? *Economic Policy 16*(32), 53–82.

Boyd, J. H., S. Kwak, and B. Smith (2005, December). The Real Output Losses Associated with Modern Banking Crises. *Journal of Money, Credit, and Banking 37*(6), 977–999.

Brunnermeier, M. K. (2008a). Bubbles. In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics. Second Edition.* New York: Palgrave Macmillan.

Brunnermeier, M. K. (2008b). Deciphering the 20078 Liquidity and Credit Crunch. *Mimeo, Princeton University.* http://www.princeton.edu/ ~markus/ index.htm.

Cohen, B. H. and E. M. Remolona (2001, December). Overview: Financial Markets Prove Resilient. *BIS Quarterly Review.* http://www.bis.org/ publ/ r_qt0112a.pdf.

Fender, I. and P. Hördahl (2007). Overview: Credit Retrenchment Triggers Liquidity Squeeze. *BIS Quarterly Review.* http://www.bis.org/ publ/ qtrpdf/ r_qt0709a.pdf.

Greenlaw, D., J. Hatzius, A. K. Kashyap, and H. S. Shin (2008). Leveraged Losses: Lessons from the Mortgage Market Meltdown. *Mimeo, Princeton University.*

intended to reduce volatility and promote investor confidence. Details on these thresholds can be found at http://www.sec.gov/answers/circuit.htm.

http://www.chicagogsb.edu/usmpf/docs/usmpf2008confdraft.pdf.

Greenspan, A. (2008, March 16). We Will Never Have a Perfect Model of Risk. *Financial Times*.

IMF (1998, December). World Economic Outlook and International Capital Markets: Interim Assessment. *Mimeo, McCombs School of Business*. http://www.imf.org/ external/ pubs/ ft/ weo/ weo1298/ index.htm.

Kindleberger, C. P. and R. Aliber (2005). *Manias, Panics, and Crashes: A History of Financial Crises. Fifth Edition.* New Jersey: John Wiley & Sons.

Krugman, P. (1979). A Model of Balance of Payment Crisis. *Journal of Money Credit and Banking 11*, 311–325.

McAndrews, J. J. and S. M. Potter (2002). Liquidity Effects of the Events of September 11, 2001. *Federal Reserve Bank of New York Economic Policy Review*, 59–79.

Schnabel, I. and H. S. Shin (2004). Liquidity and Contagion: the Crisis of 1763. *Journal of the European Economic Association 2*(6), 929–968.

Visano, B. S. (2006). *Financial Crises. Socio-economic Causes and Institutional Context.* New York: Routledge.

CHAPTER 2

Imperfect Common Knowledge in First-Generation Models of Currency Crises

FIRST-GENERATION MODELS assume the level of reserves of a central bank is common knowledge among arbitrageurs, and therefore the timing of the attack on the currency can be correctly anticipated. The collapse of the peg thus leads to no discrete change in the exchange rate. We relax the assumption of perfect information and introduce uncertainty about the willingness of a central bank to defend the peg. In this new setting, there is a unique equilibrium at which the fixed exchange rate is abandoned. The lack of common knowledge will lead to a discrete devaluation once the peg finally collapses.

2.1 Introduction

Traditionally, exchange rates have been explained by macroeconomic fundamentals. However, there seem to be significant deviations from them. The need to understand these unexplained movements has drawn attention on the basic assumptions of the standard models, such as the homogeneity of market consumers or the irrelevance of private information.

First-generation models of currency crises assume that consumers are perfectly informed about macroeconomic fundamentals. The original model is due to Krugman (1979), who drew on the work of Salant and Henderson (1978) on the study of attacks on a government-controlled price of gold[1]. Krugman (1979) presents an economy in which the level of the central bank's foreign reserves is common knowledge among consumers. In this setting, market participants not only know the level of reserves, but also, that the other agents know it too. There is perfect transmission of information and speculators can precisely coordinate the attack on the currency. The model concludes that the attack is therefore accurately predictable and implies zero devaluation. However, during currency crises, large discrete devaluations are normally observed.

In this paper, we incorporate the information structure presented in Abreu and Brunnermeier (2003)[2] to introduce uncertainty about the willingness of the central bank to defend the peg in first-generation models of currency crises. The application of Abreu and Brunnermeier's dynamic model to currency crises presents several advantages over the original setting. Firstly, it provides a reasonable explanation for the assumption that the asset's market price remains constant until the time when the bubble bursts. In our setting, the

[1]For a review on the currency crises literature see Flood and Marion (1999).

[2]Abreu and Brunnermeier (2003) consider an economy in which arbitrageurs learn sequentially, but in random order, that the market price has departed from fundamentals. They assumed this information structure to argue the existence and persistence of asset bubbles despite the presence of rational agents in the economy.

exchange rate does remain invariable until the peg collapses and the cumulative selling pressure will not affect the fixed exchange rate until it exceeds a certain threshold and the central bank can no longer maintain the peg. Secondly, it gives a sound interpretation to the exogenous price paths, which in our model can be intuitively identified with the fixed exchange rate. Thirdly, in Abreu and Brunnermeier's model it is necessary to introduce a final condition which guarantees the burst of the asset bubble even if no arbitrageur sells his shares. In our model, the central bank will be ultimately forced to abandon the peg when the foreign reserves are exhausted (even if no speculator attacks the currency).

We consider an economy with a fixed exchange rate regime and a persistent deficit that reduces the central bank's reserves. In this setting, we suppose a continuum of arbitrageurs who, one by one, cannot affect the exchange rate. They can choose between local and foreign currency. We assume that holding local currency generates higher returns, although there is a risk of devaluation. Arbitrageurs take their investment decisions by evaluating the trade off between these higher profits and the fear of capital losses. At some random time, they become sequentially aware that the shadow price[3] has exceeded the peg, and they have to decide between canceling and maintaining their positions. Arbitrageurs notice the mispricing in random order and they do not know if other arbitrageurs are also aware of it. They would prefer to hold local currency for as long as possible, since it produces higher profits, but they would not want to wait for too long because of the capital losses caused by devaluation.

We suppose that arbitrageurs have financial constraints which limit their individual maximum positions and their impact on the exchange rate. To force the exchange rate off the peg, it will be necessary to coordinate the attack on the currency. They face a synchronization problem and at the same time a competition dilemma: only a fraction of them

[3]The shadow price is the exchange rate that would prevail in the market if the peg were abandoned. This concept was originally developed by Flood and Garber (1984).

can leave the market before the peg collapses, because as soon as a large enough number of arbitrageurs sell out local currency, the central bank will be forced to abandon the fixed exchange rate regime and those who, at that time, still hold local currency will suffer a capital loss.

We prove that there exists a unique equilibrium in which, for moderate heterogeneity among arbitrageurs, the peg is abandoned after enough arbitrageurs have hold overvalued currency for some period of time τ^* since they noticed that the fixed exchange rate lies below the shadow price. Arbitrageurs will hold a maximum long position in local currency until they fear the attack is imminent. At that time, they will sell out causing the fall of the peg. For extreme levels of dispersion of opinion among them, the fixed exchange rate regime collapses when the central bank's foreign reserves are completely exhausted. In either case, the abandonment of the peg implies a discrete devaluation of the currency. In the former, we will prove that arbitrageurs optimally hold overvalued currency for some time $\tau^* > 0$ ("After becoming aware of the bubble, they [arbitrageurs]... optimally choose to ride the bubble over some interval.", in Abreu and Brunnermeier's setting). Hence, the attack will take place strictly after the shadow price exceeds the peg causing a jump in the exchange rate. In the latter, the central bank will defend the peg until the reserves are exhausted. This will necessary occur some time after the shadow price exceeds the peg, which will imply a discrete change in the exchange rate. In this paper we derive a first-generation model of currency crises with imperfect information, which explains the discrete jump in the exchange rate generally observed in the transition from a fixed to a floating exchange rate regime.

This paper is related to the recent literature on currency crises which incorporates private information to accommodate discrete devaluations. Guimarães (2006) introduces uncertainty in a first-generation model by assuming that agents do not know whether they would be able to escape the devaluation. He argues that there is a unique equilibrium in which agents delay their decisions to attack the currency due to these market frictions. A different approach is presented in Pastine (2002). He incorporates a forward-looking rational policy maker that dislikes speculative attacks and it is capable of choosing the timing of the move to a floating exchange rate regime. Pastine (2002) shows that the monetary authority has an incentive to introduce uncertainty into the speculators' decisions in order to avoid predictable attacks on the currency. In this model, the standard perfectly predictable attack is replaced by an extended period of speculation which gradually places increasing pressure on the policy maker.

Broner (2008) relaxes the assumption of perfect information including a new type of investors: the uninformed consumers. The ratio between informed and uninformed consumers is fixed exogenously and determines the resulting equilibria. He concludes that, if the fraction of informed consumers is high enough, market agents face a situation similar to the lack of private information. However, when the private information is high, the model arrives to a large set of equilibria, characterized by possible discrete devaluations, which differ on the informed consumers' propensity to attack the currency. In Broner's model, it is still common knowledge among informed consumers the threshold level of the central bank's reserves and therefore, the exact time when the peg is attacked. In a competitive environment though, rational agents may not be willing to reveal what they know to other participants.

In independent work, Rochon (2006)[4] applies Abreu and Brunnermeier's structure to currency crises. She considers an economy with a fixed exchange rate regime in which a negative shock triggers a gradual depletion of the central bank's foreign reserves. The similarities between Rochon (2006) and our work lie in the application of Abreu and Brunnermeier's information structure to currency crises. However, our paper differs from Rochon (2006)

[4] I am thankful to Fernando Broner for bringing this paper to my attention at the International Economics seminar (London School of Economics, December 2004)

in several ways. We consider a different setting; we assume a first-generation model, as described in Krugman (1979), in which a government runs a persistent deficit which will gradually reduce the central bank's foreign reserves. Rochon (2006) defines a model in which the central bank is committed to defend the fixed exchange rate regime but it is not forced to finance an expansionary monetary policy. Also, we suppose that rational agents have imperfect information about the shadow price, while in Rochon (2006) the key variable is the level of reserves. From there on, the specification and focus of the model, the derivation of the unique equilibrium and the timing of the attack are clearly different.

The paper is organized in the following order. Section 2.2 analyzes the Krugman model. We explain Abreu and Brunnermeier's model in Section 2.3. Section 2.4 illustrates the generalization of the traditional first-generation model of currency crises. We relax the assumption of perfect information and define the new setting. Section 2.5 studies the resulting unique equilibrium. We derive the timing of the attack and analyze the determinants of the period of time τ^* during which arbitrageurs optimally hold overvalued domestic currency. Finally, we present our conclusions in Section 2.6.

2.2 The Krugman Model

The original model of speculative attacks on fixed exchange rates is due to Krugman (1979). He considered an economy with a fixed exchange rate regime where the government runs a budget deficit that will gradually reduce the central bank's reserves. The model concludes that the peg will be abandoned before the reserves are completely exhausted. At that time, there will be a speculative attack that eliminates the lasting foreign exchange reserves and leads to the abandonment of the fixed exchange rate.

In this model, the central bank faces two different tasks. First, it must satisfy the financial needs of the government and second, it has to maintain a fixed exchange rate. The central bank finances the deficit by issuing government debt and defends the peg through direct intervention in the foreign exchange rate market. In this economy, the asset side of the central bank's balance sheet at time t is made up of domestic credit (D_t) and the value in domestic currency of the international reserves (R_t). The balance sheet's liability side consists of the domestic currency in circulation, the money supply (M_t^s). Hence:

$$M_t^s = D_t + R_t$$

The budget deficit grows at a constant rate μ ($\mu > 0$):

$$\mu = \frac{\dot{D}_t}{D_t}$$

Also assume that the purchasing parity holds:

$$S_t = \frac{P_t}{P_t^*}$$

where, at time t, P_t is the domestic price level, S_t the exchange rate of domestic currency for foreign, and P_t^* is the foreign price level. We can fixed $P_t^* = 1$, and therefore the exchange rate can be identified with the domestic price level ($P_t = S_t$).

In this model it is supposed that money is only created through the deficit. As long as the central bank is committed to defend the peg, it will print money to finance the deficit. This will tend to raise the money supply, and hence affect the domestic prices and the exchange rate. Domestic prices will begin to increase bringing about an incipient depreciation of the currency. To maintain the exchange rate fixed, the authorities will reduce the foreign reserves to purchase the domestic currency and foreign reserves will fall as domestic assets continually rise. Ultimately, if the budget is in deficit, pegging the rate becomes impossible, no matter how large the initial reserves were.

However, the model concludes that the attack comes *before* the stock of foreign reserves would have been exhausted in the absence of speculation. Why? In Krugman's model, consumers can correctly anticipate the exhaustion of the reserves, they can only choose between domestic and foreign money and it is also supposed that foreigners do not hold domestic money. Then, the assumption of perfect foresight implies that speculators, anticipating an abandonment of the peg, will attack the exchange rate to acquire the central bank's reserves and to avoid a capital loss.

To determine the timing of the crisis, we introduce the following definition.

Definition 2.1. *(Krugman and Obstfeld (2003)) The shadow floating exchange rate or* **shadow price** *at time t (S_t) is the exchange rate that would prevail at time t if the central bank held no foreign reserves, allowed the currency to float but continued to allow the domestic credit to grow over time.*

In Appendix 2.7.1 we derive an expression for the shadow floating exchange rate. To simplify the analysis, it is convenient to express all magnitudes in logarithms[5]. We present logarithmic versions of the previous equations and describe the monetary equilibrium by the Cagan equation. Then, the logarithm of shadow price is given by:

$$s_t = \gamma + \mu \times t$$

[5]We use the standard notation in which an upper case letter represents a variable in levels and a lower case one its logarithm: $s_t = ln(S_t)$. From now on, exchange rates will be expressed in logarithms. To simplify the reasoning we will still refer to them as fixed exchange rate and shadow price, where, to be precise, it should say fixed *log*-exchange rate and *log*-shadow price.

where γ and μ are constants and μ is the rate of growth of the budget deficit. The time of the attack on the currency T is defined as the date on which the shadow price reaches the peg ($s_t = \bar{s}$):

$$T = \frac{\bar{s} - \gamma}{\mu}$$

In the Krugman model, the level of the central bank's foreign reserves is common knowledge among consumers. Thus, the timing of the attack is accurately predictable and the transition from a fixed to a floating exchange rate regime occurs without discrete jumps in the exchange rate.

2.3 The Abreu and Brunnermeier Model

Abreu and Brunnermeier (2003) present a model in which an asset bubble can survive despite the presence of rational arbitrageurs. They consider an information structure where rational arbitrageurs become sequentially aware that an asset's market price has departed from fundamentals and they do not know if other arbitrageurs have already noticed the mispricing. The model concludes that if the arbitrageurs' opinions are sufficiently dispersed, the asset bubble bursts for exogenous reasons when it reaches its maximum size. And in the case of moderate levels of dispersion of opinion, Abreu and Brunnermeier (2003) prove that endogenous selling pressure advances the bubble collapse. They demonstrate that these equilibria are unique. Also, the model shows how news events can have a disproportionate impact on market prices, since they allow agents to synchronise their exit strategies.

This model considers two types of agents: behavioral traders (influenced by fads, fashions, over-confidence...) and rational arbitrageurs. Initially the stock price p_t grows at the risk-free interest rate r ($p_t = e^{rt}$) and rational arbitrageurs are fully invested in the market.

At $t = 0$, the price starts growing at a faster rate g ($g > r$). Behavioral traders believe that the stock price p_t will grow at a rate g in perpetuity. Hence, whenever the stock price falls below $p_t = e^{gt}$, they are willing to buy any quantity of shares (up to their aggregate absorption capacity κ). Then, at some random time t_0 (exponentially distributed on $[0, \infty)$), rational arbitrageurs become (in random order) sequentially aware that the price is too high. However, the price continues to grow at a rate $g > r$ and hence, only a fraction $(1 - \beta(\cdot))$ of the price is explained by the fundamentals, where $\beta(\cdot)$ represents the "bubble component". Rational agents understand that the market will eventually collapse but still prefer to ride the bubble as it generates higher returns.

In Abreu and Brunnermeier's model, the bubble collapses as soon as the cumulative selling pressure exceeds some threshold κ (the absorption capacity of the behavioral traders) or ultimately at $t = t_0 + \overline{\tau}$ when it reaches its maximum size ($\overline{\beta}$). It is assumed that arbitrageurs, one by one, have limited impact on the price, because of the financial constraints they face. Consequently, large movements in prices require a coordinated attack. They consider that in each instant t, from $t = t_0$ until $t = t_0 + \eta$, a mass of $1/\eta$ arbitrageurs becomes aware of the mispricing, where η can be understood as a measure of the dispersion of opinion among agents concerning the timing of the bubble. Since t_0 is random, they do not know how many other rational arbitrageurs have noticed the mispricing, because they will only become aware of the selling pressure when the bubble finally bursts. Rational arbitrageurs face temporal miscoordination. Then, an arbitrageur who becomes aware of this mispricing at time t_i has the following posterior cumulative distribution for the date (t_0) on which the price departed from its fundamental value, with support $[t_i - \eta, t_i]$:

$$\Phi(t_0|t_i) = \frac{e^{\lambda\eta} - e^{\lambda(t_i - t_0)}}{e^{\lambda\eta} - 1}$$

Therefore, from $t = t_0 + \eta\kappa$ onwards, the mispricing is known to enough arbitrageurs to correct it. Nevertheless, they do not attempt to, since as soon as they coordinate, the selling pressure will burst the bubble. However, there is also a competitive component in the model: only a fraction κ of the arbitrageurs will be able to sell out before the bubble collapses (because it bursts the moment the selling pressure surpasses κ). Thus, arbitrageurs have also an incentive to leave the market.

In this setting, each arbitrageur can sell all or part of her stock of shares until a certain limit due to some financial constraints. It is possible to buy back shares and to exit and re-enter the market multiple times. The strategy of an agent who became aware of the bubble at time t_i is defined as the selling pressure at time t: $\sigma(\cdot, t_i) = [0, t_i + \overline{\tau}] \mapsto [0, 1]$. The action space is normalised to be the continuum between $[0, 1]$, where 0 indicates a maximum long position and 1 a maximum short position. Then, the aggregate selling pressure of all agents at time $t \geq t_0$ is given by $s(t, t_0) = \int_{t_0}^{min\{t,\ t_0+\eta\}} \sigma(t, t_i) dt_i$ and therefore the bursting time can be expressed as:

$$T^*(t_0) = \inf\{t | s(t, t_0) \geq \kappa \text{ or } t = t_0 + \overline{\tau}\}$$

Given this information structure, arbitrageur t_i's beliefs about the date on which the bubble bursts are described by:

$$\Pi(t_0|t_i) = \int_{T^*(t_0)<t} d\Phi(t_0|t_i)$$

In their analysis, Abreu and Brunnermeier focus on trigger-strategies in which an agent, who sells out at t, continues to attack the bubble at all times thereafter. In this case, solving the optimisation problem of the arbitrageur who notices the bubble at time t_i and sells out at time t yields the following condition:

Lemma 2.1. *(Abreu & Brunnermeier) (sell out condition). If arbitrageur t_i's subjective hazard rate is smaller than the 'cost-benefit ratio', i.e.*

$$h(t|t_i) < \frac{g-r}{\beta(t - T^{*-1}(t))}$$

trader t_i will choose to hold the maximum long position at t. Conversely, if $h(t|t_i) > \frac{g-r}{\beta(t-T^{*-1}(t))}$ she will trade to the maximum short position.

where $h(t|t_i) = \frac{\pi(t|t_i)}{1-\Pi(t|t_i)}$ is the hazard rate that the bubble will burst at time t, $\Pi(t|t_i)$ represents the arbitrageur t_i's beliefs about the bursting date and $\pi(t|t_i)$ denotes the associated conditional density. They conclude that an arbitrageur who becomes aware of the mispricing at time t_i will hold a maximum long position until his subjective hazard rate becomes larger than the cost-benefit ratio. That is, arbitrageur t_i will ride the bubble until his subjective probability that the bubble will burst in the next trading round is high enough. At that time, arbitrageur t_i will trade to the maximum short position to get out of the market.

They consider two different scenarios. When arbitrageurs' opinions are sufficiently dispersed, Abreu and Brunnermeier (2003) prove that the selling pressure does not affect the time when the bubble collapses, because each arbitrageur optimally rides the bubble for so long that, at the end of the horizon ($t = t_0 + \overline{\tau}$), there is not enough pressure to burst the bubble (less than κ will have sold out). They show that there is a unique equilibrium at which the bubble bursts for exogenous reasons at $t = t_0 + \overline{\tau}$. A different conclusion is reached when a moderate level of heterogeneity is assumed. In this case, they demonstrate that there is a unique and symmetric equilibrium in which each arbitrageur sells her shares τ^* periods after becoming aware of the mispricing. The bubble bursts at $t = t_0 + \eta\kappa + \tau^*$ ($< t_0 + \overline{\tau}$, given small values for η).

The model assumes an information structure based on the lack of common knowledge (when an arbitrageur becomes aware of the mispricing, he does not know if others know) and derives that these equilibria are unique. However, in typical applications, the symmetry information game has multiple equilibria. Abreu and Brunnermeier (2003) argue that the fact that arbitrageurs are competitive (since at most a fraction of them can leave the market prior to the crash) leads to a unique equilibrium even under symmetric information.

2.4 Imperfect Common Knowledge

This section presents a first-generation model of currency crises in which the traditional assumption of perfect information is relaxed. We introduce the new setting and we derive the sell out condition that determines the moment when an arbitrageur fears the abandonment of the peg and prefers to attack the currency.

Consider an economy similar to the one described in Krugman (1979) and summarized in Section 2.2. In this setting, the level of the central bank's foreign reserves is common knowledge among arbitrageur and therefore, the peg is attacked whenever it leads to no discrete change in the price level, i.e., as soon as the shadow price reaches the fixed exchange rate ($s_t = \bar{s}$). In our model we incorporate the information structure presented in Abreu and Brunnermeier (2003) (and reviewed in Section 2.3) to introduce uncertainty about the willingness of the central bank to defend the peg. We consider that the level of reserves is no longer common knowledge among arbitrageurs and in this paper we analyze the consequences of this uncertainty on the abandonment of the fixed exchange rate.

2.4.1 The Model

The following process is assumed in our setting. The central bank establishes the fixed exchange rate at a certain level $\bar{s}$ as depicted in Figure 3.1. We denote (the logarithm of) the shadow price[6] by s_t. We assume that $\widetilde{t}$ is exponentially distributed on $[0, \infty)$ with cumulative distribution function $F(\widetilde{t}) = 1 - e^{-\lambda\widetilde{t}}$ $(\lambda > 0,\ \widetilde{t} \geq 0)$. Prior to $t = \widetilde{t}$, the peg lies above the shadow price s_t $(\bar{s} > s_t)$ and the fixed exchange rate cannot collapse, since arbitrageurs would only attack the peg if it profitable for them. If before $t = \widetilde{t}$ the peg is abandoned, the currency would immediately revaluate to reach the shadow price (the local currency would worth more while arbitrageurs hold short positions in domestic currency). Hence, anticipating this capital loss, arbitrageurs will not attack the currency and therefore no speculative attack will occur before $t = \widetilde{t}$. From $t = \widetilde{t}$ onwards, the shadow price exceeds the fixed exchange rate $(\bar{s} < s_t)$ and the peg might be attacked.

Agents and Actions

In our model, there is only one type of agent: the rational arbitrageurs. We assume a continuum of arbitrageurs, with mass equal to one, who one by one cannot affect the exchange rate because of some financial constraints which limit their maximum market positions. In currency crises, however, it may seem more realistic to consider that only a few relevant institutions actively participate in currency markets and that information might be clustered. This assumption would not modify the intuition of our results. In Brunnermeier and Morgan (2006), they prove that the equilibrium delay in such games always exceeds equilibrium delay in the game with a continuum of agents and no information clustering (i.e. in the Abreu and Brunnermeier model). Hence, defining a finite number of

[6]In Appendix 2.7.1 we prove that the logarithm of the shadow price is a linear function of time ($s_t = \gamma + \mu \times t$), or equivalently, that the shadow price grows exponentially.

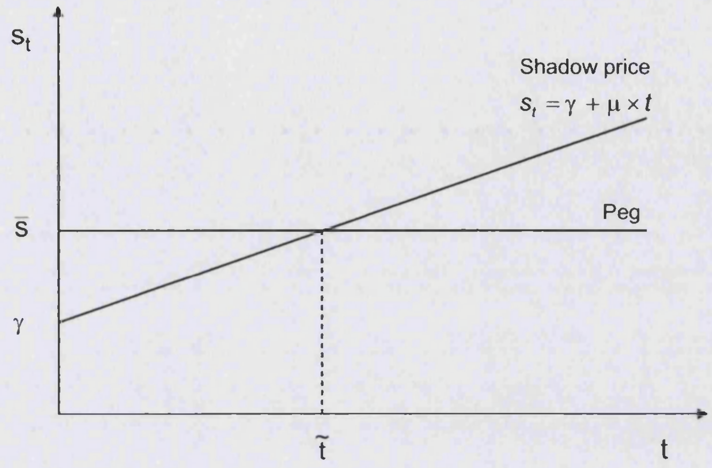

Figure 2.1: FIXED EXCHANGE RATE ($\bar{s}$) AND SHADOW PRICE (s_t) - We represent a fixed exchange rate regime in which the central bank finances a persistent deficit and maintains the exchange rate fixed at a certain level ($\bar{s}$). We plot the random time $t = \tilde{t}$, when (the logarithm of) the shadow price (s_t) reaches the peg ($\bar{s}$).

arbitrageurs and allowing for information clustering increases the optimal waiting time τ^*, and therefore it delays longer the attack on the fixed exchange rate regime, causing a larger discrete devaluation of the home currency.

Let us denote by arbitrageur i the agent who, at time t_i, receives a signal indicating that the shadow price exceeds the fixed exchange rate. Arbitrageur i may take one of two actions. He can hold local currency or buy foreign currency. Investing in domestic currency generates a return equal to r while the foreign investment yields r^*. We impose the following condition:

Assumption 2.1. $r > r^*$

We consider that the local currency pays an interest rate r higher than the foreign currency ($r > r^*$) to guarantee that, initially, arbitrageurs invest in domestic currency. Hence, in our economy, rational arbitrageurs originally prefer to invest in domestic currency because of the higher profits, but they understand that the exchange rate will be attacked and the peg eventually abandoned. Therefore, the only decision is when to exit. Selling too early leads to less profitable outcomes, but if they wait too long and they do not leave the market before the fixed exchange rate collapses, they will incur in capital losses associated with the devaluation.

An individual arbitrageur is limited in the amount of currency he can buy or sell. As in Abreu and Brunnermeier (2003), we can normalise the action space to lie between $[0, 1]$ and define the strategy of arbitrageur i by his selling pressure at time t: $\sigma(\cdot, t_i) = [0, t_i + \overline{\tau}] \mapsto [0, 1]$. A selling pressure equal to zero ($\sigma(t, t_i) = 0$) indicates a maximum long position in local currency and a value equal to one ($\sigma(t, t_i) = 1$) implies that arbitrageur i has sold out all his holdings of domestic currency (maximum pressure). Let $s(t, \widetilde{t})$ denote the aggregate selling pressure of all arbitrageurs at time $t \geq \widetilde{t}$.

Collapse of the Peg

The fixed exchange rate can collapse for one of two reasons. It is abandoned at $t = \widetilde{t} + \tau^* + \eta\kappa$ when the aggregate selling pressure exceeds a certain threshold κ ($s(t, \widetilde{t}) \geq \kappa$) or ultimately at a final time when all foreign reserves are exhausted, say at $t = \widetilde{t} + \overline{\tau}$. Let us denote this collapsing date by $T^*(\widetilde{t}) = \widetilde{t} + \zeta$, where $\zeta = \tau^* + \eta\kappa$ if the abandonment of the peg is caused by arbitrageurs' selling pressure (endogenous collapse) and $\zeta = \overline{\tau}$ if it is due to the exhaustion of reserves (exogenous collapse). Since we have assumed that arbitrageurs have no market power, they will need to coordinate to force the abandonment of the peg. However, only a proportion $\kappa < 1$ of arbitrageurs can exit the market before the peg is abandoned. Therefore, arbitrageurs face both, cooperation and competition.

Information Structure

To simplify the analysis, we assume that at the random time $t = \widetilde{t}$ when the shadow price reaches the peg, arbitrageurs begin to notice this mispricing. They become aware sequentially and in random order and they do not know if they have noticed it early or late compared to others. They cannot know if they are the firsts or the lasts to know. Specifically, at each instant (between $\widetilde{t}$ and $\widetilde{t}+\eta$), a new mass $1/\eta$ of arbitrageurs receives a signal indicating that the shadow price exceeds the fixed exchange rate, where η is a measure of the dispersion of opinion among them. The timing of arbitrageur i's signal is uniformly distributed on $[\widetilde{t}, \widetilde{t}+\eta]$, but since $\widetilde{t}$ is exponentially distributed each arbitrageur does not know how many other others have received the signal before him. Arbitrageur i only knows that at $t = t_i + \eta$ all other arbitrageurs received their signals. Conditioning on $\widetilde{t} \in [t_i - \eta, t_i]$, arbitrageur i's posterior cumulative distribution function for the date $\widetilde{t}$ on which the shadow price reached the peg is $\Phi(\widetilde{t}|t_i) = \frac{e^{\lambda\eta}-e^{\lambda(t_i-\widetilde{t})}}{e^{\lambda\eta}-1}$. Then, arbitrageur i's posterior cumulative distribution function over the collapsing date $T^*(\widetilde{t})$ is

$$\Pi(T^*(\widetilde{t})|t_i) = \frac{e^{\lambda\eta} - e^{\lambda(t_i+\zeta-T^*(\widetilde{t}))}}{e^{\lambda\eta}-1}$$

given that $T^*(\widetilde{t}) \in [t_i + \zeta - \eta, t_i + \zeta]$.

Further Assumptions

We consider the following statements to simplify the analysis and the specification of our setting:

Assumption 2.2. *In equilibrium, an arbitrageur holds either a maximum long position or a maximum short position in local currency:* $\sigma(t, t_i) \in \{0, 1\}\ \forall\ t, t_i$.

We consider that an arbitrageur prefers to invest his whole budget in local currency, since it generates higher returns, until a certain time when he fears that the attack on the currency is imminent and decides to cancel his position by selling all his stock of domestic currency. Hence, his selling pressure is initially equal to 0 (when he is fully invested in domestic currency) and equal to 1 once he sells out. The information structure considered in our model and Assumption 2.2 imply the following result:

Corollary 2.1. *If arbitrageur i holds a maximum short position at time t in local currency ($\sigma(t, t_i) = 1$), then at time t any arbitrageur j ($\forall t_j \leq t_i$) has already sold out his stock of domestic currency ($\sigma(t, t_j) = 1,\ \forall t_j \leq t_i$).*

We assume that once an arbitrageur sells his stock of domestic currency, any arbitrageur that became aware of the mispricing before him, is already out of the market.

Assumption 2.3. *No re-entry.*

To simplify the analysis we suppose that once an arbitrageur gets out of the market, he will not enter again. Intuitively, an arbitrageur sells out when he believes that the attack is close. Then, even if he does not observe the attack during some period of time after leaving the market, he still will not know when the fixed exchange rate will collapse, but certainly it will happen sooner than he thought when he exited the market. Therefore, if he does not change his beliefs, he will not have an incentive to re-enter the market.

The Sell Out Condition

In our economy, an arbitrageur can choose between buying domestic or foreign currency. Initially, they are fully invested in local currency because of the higher returns ($r > r^*$), but there is a risk of devaluation. Hence, each arbitrageur will sell exactly at the moment when the fear of the devaluation of the home currency offsets the excess of return derived from investing in local currency. Ideally, he would like to sell *just* before the exchange rate is abandoned and the domestic currency suffers devaluation or equivalently just before the appreciation of the foreign currency. In Appendix 2.7.2 we define the size of the expected appreciation of the foreign currency perceived by arbitrageur i as $A^i_f(t - t_i) = \left| E\left[\frac{\frac{1}{S_t} - \frac{1}{\overline{S}}}{\frac{1}{\overline{S}}} \middle| t_i \right] \right| = 1 - E[\frac{\overline{S}}{S_t}|t_i] \geq 0$ (if $E[\frac{\overline{S}}{S_t}|t_i] \leq 1$), and we present the optimisation problem which yields the following sell out condition:

Lemma 2.2. (sell out condition). *Arbitrageur i prefers to hold a maximum long position in local currency at time t if his hazard rate is smaller than the 'greed-to-fear ratio', i.e., if*

$$h(t|t_i) < \frac{r - r^*}{1 - E[\frac{\overline{S}}{S_t}|t_i]} = \frac{r - r^*}{A^i_f(t - t_i)}$$

He trades to a maximum short position in local currency, if $h(t|t_i) > \dfrac{r - r^*}{A^i_f(t - t_i)}$.

In our model, an arbitrageur who notices the mispricing at time $t = t_i$ compares his subjective hazard rate ($h(t|t_i)$) with the 'greed-to-fear ratio' ($\frac{r-r^*}{A^i_f(t-t_i)}$) and trades to a maximum short position as soon as he observes that the probability of devaluation given that the peg still holds is larger that the 'greed-to-fear ratio'.

2.5 Equilibrium

The exchange rate collapses as soon as the cumulative selling pressure exceeds a threshold κ or at a final date $t = \widetilde{t} + \overline{\tau}$ when all foreign reserves are exhausted. This statement implies that no fixed exchange rate regime in an economy with persistent deficit can survive in the long term. We will focus our analysis in the first scenario, in which the peg collapses for endogenous reasons.

2.5.1 Endogenous Collapse of the Peg

We have seen that arbitrageurs become aware of the mispricing in random order during an interval $[\widetilde{t},\ \widetilde{t}+\eta]$, where $\widetilde{t}$ is exponentially distributed and represents the time when the shadow price reaches the peg. To simplify the analysis we have supposed that $\widetilde{t}$ is also the moment when the first arbitrageur notices the mispricing. η is a measure of the heterogeneity of the arbitrageurs (a larger η corresponds to a wider dispersion of opinion among them). Since all arbitrageurs are ex-ante identical, we restrict our attention to symmetric equilibria. Then, for moderate values of the parameter η, we will show that there exists a unique symmetric equilibrium in which the peg falls when the aggregate selling pressure surpasses a certain threshold (κ).

Consider the following backward reasoning. If at the final date $t=\widetilde{t}+\overline{\tau}$ the selling pressure has not exceed the threshold κ, we have assumed that the exchange rate collapses because the central bank's foreign reserves are exhausted, and that this final condition is common knowledge among all arbitrageurs[7]. This induces arbitrageur i to sell out at τ_1 periods after he observes the mispricing ($t_i+\tau_1<\widetilde{t}+\overline{\tau}$) in order to avoid a capital loss. Therefore, the currency will come under attack at $t=\widetilde{t}+\tau_1+\eta\kappa<\widetilde{t}+\overline{\tau}$ (since we consider small values of the parameters η and κ). But if the peg is abandoned at $t=\widetilde{t}+\tau_1+\eta\kappa$,

[7]Specifically, $\overline{\tau}$ is common knowledge among arbitrageurs but not $t=t_i+\overline{\tau}$, which will depend on the random time when each arbitrageur notices the mispricing. It is supposed that once an arbitrageur finds out that the shadow price has exceeded the peg, he knows that the central bank's foreign reserves will last *at most* $\overline{\tau}$ periods and he also knows that all other arbitrageurs, who are aware of the mispricing, will know it too. But this arbitrageur does not know if he has learnt this information early or late compared to the other arbitrageurs, and hence he does not know if the attack on the currency will happen before this final date.

arbitrageurs will sell out earlier, let us say $\tau_2 < \tau_1$ periods after they notice that the shadow price exceeds the peg. But given the new timing, arbitrageurs will choose to sell even earlier (τ_3) and so on. As the selling date advances, the cost of devaluation of the domestic currency (or the appreciation of the foreign currency) diminishes and therefore, the benefit from holding local currency rises, that is, the 'greed-to-fear ratio' increases as the selling date advances. In our setting, at each instant, an arbitrageur compares his hazard rate to his 'greed-to-fear ratio'. He will prefer to sell local currency until the time ($t = t_i + \tau^*$) when the 'greed-to-fear ratio' equals the probability of the peg collapsing given that it still holds (the equality defines the switching condition). This guarantees that arbitrageurs will have an incentive to hold "overvalued" local currency for some period of time after they become aware of the mispricing, or equivalently, that the decreasing sequence of periods converges to $\tau^* > 0$. This result is depicted in Figure 2.2.

We can derive an expression for τ^* from previous results. We have assumed that, in equilibrium, an arbitrageur holds either a maximum long position or a maximum short position in local currency, depending on the relation between the probability of the peg collapsing and the profits derived from holding local currency. Hence, we can obtain τ^* from the time ($t = t_i + \tau^*$) when agent i will switch from maximum holding to maximum selling. This is given by:

$$\tau^* = \frac{1}{\mu} \ln \left[\left(1 - \frac{r - r^*}{h} \right)^{-1} \right] - \tau'$$

where $r - r^*$ is the excess of return, $h = h(t|t_i)$ denotes the hazard rate (which we will prove that remains constant over time), μ is a positive constant corresponding to the slope of the linear logarithm of the shadow price (s_t) which represents the rate of growth of the budget deficit and τ' is indicative of the difference between the date t_i at which the arbitrageur

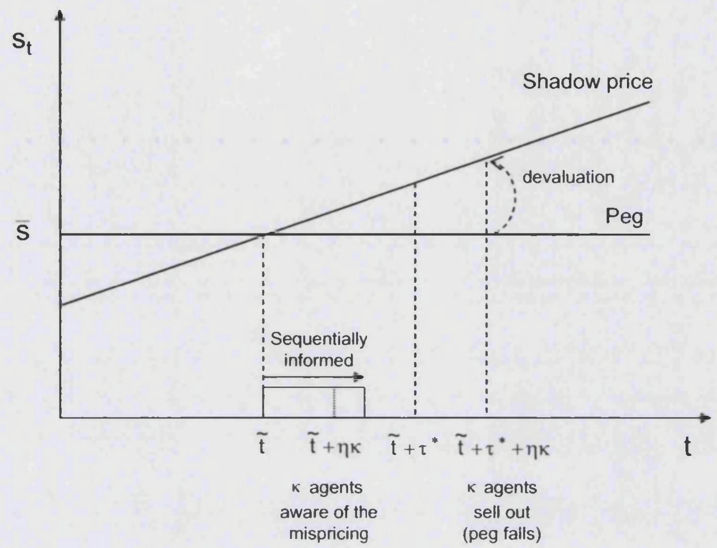

Figure 2.2: COLLAPSE OF THE PEG (MODERATE LEVELS OF DISPERSION OF OPINION)- We plot (the logarithm of) the fixed exchange rate ($\bar{s}$), (the logarithm of) the shadow price (s_t) and relevant moments in time. At $t = \tilde{t}$, the shadow price exceeds the peg and arbitrageurs become sequentially aware of it. At $t = \tilde{t} + \eta\kappa$, enough arbitrageurs have noticed the mispricing but they prefer to wait τ^* periods before selling out. At $t = \tilde{t} + \tau^* + \eta\kappa$, the selling pressure surpasses the threshold κ and the peg is finally abandoned. There is a discrete devaluation of the exchange rate.

receives the signal about the mispricing and the time when he believes the foreign currency begins appreciating.

We can summarize this result in Proposition 2.1:

Proposition 2.1. *There exists a* **unique symmetric equilibrium** *at which each arbitrageur sells out τ^* periods after becoming aware of the mispricing, where:*

$$\tau^* = \frac{1}{\mu} \ln \left[\left(1 - \frac{r - r^*}{h} \right)^{-1} \right] - \tau'$$

Thus, the fixed exchange rate will be abandoned at $t = \tilde{t} + \tau^* + \eta\kappa$.

Proof

Arbitrageur i prefers to invest in local currency for as long as possible, since this strategy generates higher returns than buying foreign currency ($r > r^*$). But, at a certain moment in time ($t = t_i$), he learns that the shadow price exceeds the peg and that there exists a risk of devaluation. We have argued that he still prefers to hold domestic currency (for some period of time τ^*) until he fears that the attack on the currency is imminent and he decides to sell out. This occurs at $t = t_i + \tau^*$. Arbitrageur i sells whenever his 'greed-to-fear ratio' equals his hazard rate, i.e. when $h(t|\widetilde{t}) = \frac{r-r^*}{A_f^i(t-t_i)}$. From this switching condition we will derive the optimal waiting time τ^*, i.e., the period of time when an arbitrageur knows that he is holding overvalued currency.

We will organise the proof in three steps. In the first one, we demonstrate that the 'greed-to-fear ratio' is decreasing in time. Step 2 shows why the hazard rate is constant in time. Finally, in Step 3 we derive the expression for τ^* from the time when the hazard rate equals the 'greed-to-fear ratio' and arbitrageur i changes from a maximum long position to a maximum short position in local currency.

Step 1. *The 'greed-to-fear ratio' decreases in time.*

Proof. The 'greed-to-fear ratio' is defined as:

$$\frac{r - r^*}{A_f^i(t - t_i)}$$

where $r - r^*$ is the excess of return derived from investing in domestic currency and $A_f^i(t - t_i)$ denotes the size of the expected appreciation of the foreign currency feared by agent i:

$$A_f^i(t - t_i) = \left| E\left[\left. \frac{\frac{1}{S_t} - \frac{1}{\bar{S}}}{\frac{1}{\bar{S}}} \right| t_i \right] \right| = 1 - E\left[\left. \frac{\bar{S}}{S_t} \right| t_i \right]$$

and

$$E\left[\frac{\overline{S}}{S_t}\bigg|t_i\right] = \int_{t_i-\eta}^{t_i} e^{-\mu(t-\widetilde{t})}\phi(\widetilde{t}|t_i)d\widetilde{t} = ke^{-\mu(t-t_i)}$$

where $k = \frac{\lambda}{\lambda-\mu}\frac{e^{(\lambda-\mu)\eta}-1}{e^{\lambda\eta}-1}$, $k \in [0,1]$ for $\lambda \geq \mu$. Hence, $A_f^i(t-t_i)$ is an strictly increasing and continuous function of the time elapsed since agent i received his signal indicating that the shadow price exceeded the peg.

Assumption 2.1 establishes that the excess of return $(r-r^*)$ is positive and constant in time. Therefore the 'greed-to-fear ratio', $\frac{r-r^*}{A_f^i(t-t_i)}$, decreases in time. Intuitively, the further in time, the larger the possible appreciation of the foreign currency once the peg collapses, and therefore the smaller the benefits (in relative terms) that a rational agent obtains from holding local currency. Thus, the 'greed-to-fear ratio' will decrease in time. □

Step 2. *The hazard rate is constant in time.*

Proof. The hazard rate is defined as: $h(T^*(\widetilde{t})|t_i) = \frac{\pi(T^*(\widetilde{t})|t_i)}{1-\Pi(T^*(\widetilde{t})|t_i)}$, where $\pi(T^*(\widetilde{t})|t_i)$ is the conditional density function and $\Pi(T^*(\widetilde{t})|t_i)$, the conditional cumulative distribution function of the date on which the peg collapses. The hazard rate represents, at each time, the probability that the peg is abandoned, given that it has survived until $t = t_i$. We have considered that the timing of agent i's signal is uniformly distributed on $[\widetilde{t}, \widetilde{t}+\eta]$ and that $\widetilde{t}$ is exponentially distributed. Then, an arbitrageur, who becomes aware that the shadow price exceeds the peg at $t = t_i$, has a posterior density function of the date on which the peg is abandoned, with support $[t_i+\tau^*+\eta\kappa-\eta, t_i+\tau^*+\eta\kappa]$, given by: $\pi(T^*(\widetilde{t})|t_i) = \frac{\lambda e^{\lambda(t_i+\zeta-T^*(\widetilde{t}))}}{e^{\lambda\eta}-1} = \frac{\lambda e^{\lambda(t_i+\tau^*+\eta\kappa-T^*(\widetilde{t}))}}{e^{\lambda\eta}-1}$, where $\zeta = \tau^*+\eta\kappa$ if the peg collapses for endogenous reasons. This is depicted in Figure 2.3.

At time $t = t_i$, arbitrageur i only knows if the peg has collapsed or not. But if the fixed rate regime has not been attacked, arbitrageur i cannot know when it will happen, since

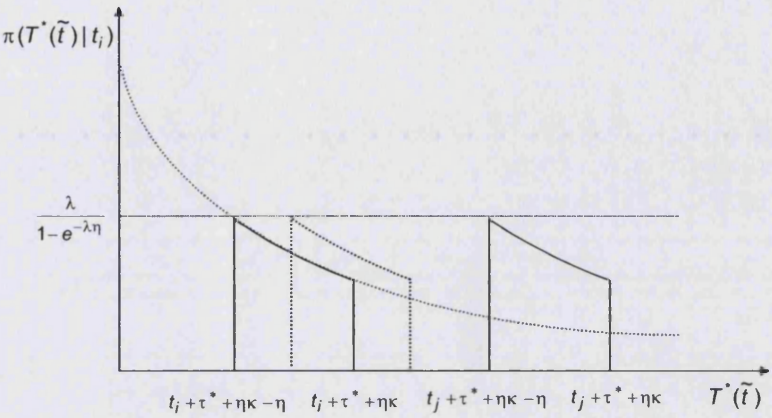

Figure 2.3: POSTERIOR DENSITY FUNCTION FOR ARBITRAGEURS i AND j

he does not know if any other agent became aware of the mispricing before him. At any other time $t = t_j > t_i$, arbitrageur j faces an equivalent scenario (shifted from t_i to time t_j, but with no additional information), i.e., arbitrageurs cannot learn from the process (if the peg has not collapsed, they cannot know when the attack will take place). Therefore, the hazard rate, over the collapsing dates $t = t_i + \tau^*$, is constant in time and it is given by the following expression:

$$h(t_i + \tau^*|t_i) = \frac{\lambda}{1 - e^{-\lambda\eta\kappa}} \equiv h$$

□

Step 3. *The optimal τ^* is given by: $\tau^* = \frac{1}{\mu} \ln\left[\left(1 - \frac{r-r^*}{h}\right)^{-1}\right] - \tau'$*

Proof. We have obtained that the 'greed-to-fear ratio' is a decreasing function of time, while the hazard rate is constant. Therefore, we can derive τ^* from the time $t = t_i + \tau^*$ when arbitrageur i fears that the collapse of the peg is imminent and decides to sell out the local currency. We have proved that he will hold local currency during τ^* periods after becoming aware of the mispricing, i.e., arbitrageur i will hold a long position in local currency until

$t = t_i + \tau^*$, when his 'greed-to-fear ratio' equals his hazard rate:

$$\underbrace{\frac{r - r^*}{A_f^i(t - t_i)}}_{\substack{\text{Greed to Fear} \\ \text{Ratio}}} \Bigg|_{t=t_i+\tau^*} = \underbrace{\frac{\lambda}{1 - e^{-\lambda\eta\kappa}}}_{\text{Hazard Rate}} \quad \Rightarrow \quad \frac{r - r^*}{1 - ke^{-\mu(t_i+\tau^*-t_i)}} = h \quad \Rightarrow$$

$$\Rightarrow \ \tau^* = \frac{1}{\mu}\ln\left[\left(1 - \frac{r - r^*}{h}\right)^{-1}\right] - \tau' \quad \Rightarrow \quad \tau^* = \frac{1}{\mu}\ln\left[\frac{h}{h - (r - r^*)}\right] - \tau'$$

where $\tau' = \frac{1}{\mu}\ln\frac{1}{k} \in [0, \infty)$. The optimal waiting time τ^* has two components. The first one defines the trade-off between the excess of return derived from investing in domestic currency and the capital loss bore by arbitrageur i if the fixed exchange rate is abandoned before he sells out. The second component τ' is due to the information structure and measures the period of time elapsed between the date $t = t_i$ at which arbitrageur i receives the signal and the time when he believes that the foreign currency starts appreciating. To simplify the reasoning suppose that there is no delay, i.e., assume that at $t = t_i$ arbitrageur i receives the signal about the mispricing and believes that the shadow price has "just" reached the peg. In this case, $\tau' = 0$ and each arbitrageur waits $\tau^* = \frac{1}{\mu}\ln\left[\left(1 - \frac{r-r^*}{h}\right)^{-1}\right]$ after receiving the signal. Then, he exits the market.

Figure 2.4 represents the decreasing 'greed-to-fear ratio', the hazard rate, τ^* and the time of the attack. It is interesting to note that if the hazard rate is larger than the excess of return ($h > r - r^*$), arbitrageurs wait a finite positive period of time ($\tau^* > 0$) before selling out. At some point in time, they believe that the attack on the fixed exchange rate is imminent and they cancel their positions in home currency. There is an endogenous collapse of the peg. However, if $h = r - r^*$ arbitrageurs sell overvalued currency after waiting for an infinite time ($\tau^* = \infty$) to elapse since receiving the signal. Finally, if $h < r - r^*$, arbitrageurs never sell, i.e., if the probability that the peg collapses (given that it still holds) is lower than the excess of return derived from investing in domestic currency, then arbitrageurs do not fear a devaluation and hence they never leave the market. In this case ($h \leq r - r^*$) the fixed exchange rate collapses for exogenous reasons when the central bank's foreign reserves are exhausted.

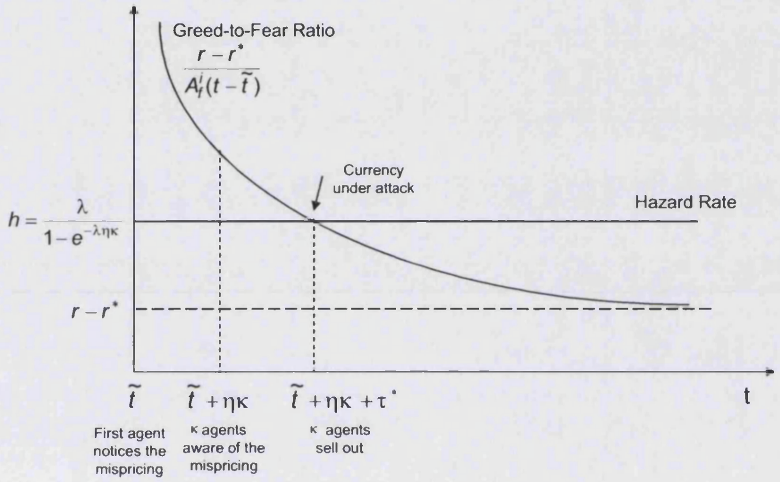

Figure 2.4: TIMING OF THE ATTACK - An arbitrageur holds a maximum long position in local currency until he believes that the attack is imminent, and he prefers to sell out, that is, when his 'greed-to-fear ratio' equals his hazard rate. At $t = \tilde{t} + \tau^* + \eta\kappa$ *enough* arbitrageurs have sold out and the currency comes under attack.

□

2.5.2 Determinants of τ^*

Excess of return. The period of time τ^* during which arbitrageurs optimally hold overvalued local currency depends directly on the excess of return derived from investing in domestic currency. Increasing the spread between returns ($r - r^*$) delays the attack on the peg, since a more attractive local currency will induce arbitrageurs to hold overvalued currency for a longer time. The size of the delay depends on:

$$\frac{\partial \tau^*}{\partial (r - r^*)} = \frac{1}{\mu} \cdot \frac{1}{h - (r - r^*)}$$

This suggests that rising returns will have a small impact on the delay of the attack if the probability of the peg collapsing (given that it still holds) and the slope of the (logarithm of the) shadow price are high.

Hazard rate. τ^* is a decreasing function of the hazard rate ($\frac{\partial \tau^*}{\partial h} < 0$). The hazard rate represents the probability that the fixed exchange rate is abandoned given that the peg is still in place. Assuming the information structure presented in this paper, the hazard rate remains constant over time[8] and equal to:

$$h = \frac{\lambda}{1 - e^{-\lambda \eta \kappa}}$$

where λ characterises the exponential distribution of $\widetilde{t}$ (the time when the shadow price exceeds the peg), η is a measure of the dispersion of opinion among agents and κ defines the threshold level of cumulative selling pressure that triggers the attack on the currency.

A lower heterogeneity among market participants increases the hazard rate and advances the currency crises. In the limit as η tends to zero (no private information), we converge to the Krugman setting in which the fundamentals are common knowledge and the peg is attacked as soon as the shadow price reaches the fixed exchange rate. In this case, $h \longrightarrow \infty$ and $\tau^* = 0$. On the other extreme case where there is large dispersion of opinion among arbitrageurs, the hazard rate tends to zero and the peg is abandoned whenever the central

[8]This is proved in Subsection 2.5.1.

bank's foreign reserves are exhausted.

Also, a higher threshold κ delays the crisis. If the central bank is determined to commit a larger proportion of its foreign reserves to defend the fixed exchange rate, the peg will survive longer.

Slope of the shadow price. The optimal time τ^* depends inversely on the slope of (the logarithm of) the shadow price μ ($\frac{\partial \tau^*}{\partial \mu} < 0$). μ can be seen as the speed of depletion of the central bank's foreign reserves. This suggests that a steeper (logarithm of the) shadow price implies a faster rate of exhaustion of reserves and ultimately that the central bank's reserves will be exhausted earlier. Hence, a higher slope, reduces τ^* and advances the attack on the fixed exchange rate.

μ also represents the rate of growth of the government's budget deficit. Then, the faster the level of government's expenditure, the shorter arbitrageurs will be willing to hold the overvalued local currency, hence, advancing the attack on the currency.

Our analysis suggests that in a first-generation model in which a government runs a persistent deficit, which grows at a constant rate μ, rising the spread between returns to make the local currency more attractive, committing more foreign reserves to defend the peg and inducing dispersion of opinion among arbitrageurs will delay the attack on the currency. However, since the expansionary monetary policy makes a fixed exchange rate ultimately unsustainable, these policy instruments would only increase the size of the devaluation whenever it occurs. The only effective means would be to reduce the rate of growth of the budget deficit (μ). This would delay the speculative attack on the fixed exchange rate and diminish the size of the devaluation when the peg finally collapses.

2.6 Conclusion

In this paper we relax the traditional assumption of perfect information in first-generation models of currency crises. We consider an economy similar to the one described in Krugman (1979) and, to introduce uncertainty about the willingness of a central bank to defend the peg, we incorporate the information structure presented in Abreu and Brunnermeier (2003).

At a random time, the shadow price exceeds the fixed exchange rate and sequentially, but in random order, arbitrageurs become aware of this mispricing. They understand that the currency will be attacked and the peg eventually abandoned, but still prefer to hold local currency during some period of time (τ^*) after they notice the mispricing, since we have assumed that holding domestic currency generates higher returns. We derive an expression for τ^*. The optimal period of time τ^* is the same for all arbitrageurs and it is independent from the time when each arbitrageur notices the mispricing. Increasing the excess of return ($r-r^*$) obtained from investing in domestic currency, reducing the hazard rate or diminishing the level of persistent deficit, would increase the period of time when arbitrageurs prefer to hold overvalued local currency and therefore it would delay the attack on the currency.

In our model, arbitrageurs sequentially know that the peg lies below the shadow price but they do not know if other arbitrageurs have already noticed it, i.e., macroeconomic fundamentals are no longer common knowledge among market participants. Therefore, arbitrageurs in this economy face a synchronisation problem. However they do not have incentives to coordinate, since as soon as they do, the peg collapses. There is also a competitive component in our model: only a fraction κ of arbitrageurs can leave the market before the fixed exchange rate is abandoned. Consequently, in our setting, the currency does not come under attack the moment the shadow price exceeds the peg, as in Krugman's model, but at a later time when a large enough mass of arbitrageurs has waited τ^* periods and decides to leave the market because of fear of an imminent attack.

We conclude that there is a unique equilibrium in which the fixed exchange rate collapses when the selling pressure surpasses a certain threshold (κ) or ultimately at a final date $t = \widetilde{t} + \overline{\tau}$, when the central bank's foreign reserves are exhausted, if dispersion of opinion among arbitrageurs is extremely large. This equilibrium is unique, depends on the heterogeneity among agents and leads to discrete devaluations. This result differs from the zero-devaluation equilibrium in Krugman's model and also from other settings (Broner (2008)) in which lack of perfect information brings multiple equilibria.

2.7 Appendix

2.7.1 The Shadow Price

In this section, we will derive an expression for the shadow price in a first-generation model of currency crises. To simplify the calculations, it is convenient to express magnitudes in logarithms. We will use upper case letters to represent a variable in levels and lower case letters to indicate its logarithm.

We consider an economy with a fixed exchange rate regime ($\bar{s} = ln(\overline{S})$), in which a government runs a persistent deficit. In particular, we assume that it grows at a positive constant rate μ:

$$\mu = \frac{\dot{D}_t}{D_t} = \frac{1}{D_t}\frac{d(D_t)}{dt} = \dot{d}_t \quad \Rightarrow \quad \dot{d}_t = \mu \tag{A.1}$$

where $d_t = ln(D_t)$ is the logarithm of the domestic credit.

The central bank has two main tasks: to finance the government's deficit by issuing debt and to maintain the exchange rate fixed through open market operations. In our economy there are no private banks. Then, from the central bank's balance sheet the money supply at time t, M_t^s, is made up of domestic credit (D_t) and the value in domestic currency of the international reserves (R_t):

$$M_t^s = D_t + R_t \tag{A.2}$$

We assume that the purchasing parity holds. Then, the exchange rate S_t is defined as:

$$S_t = \frac{P_t}{P_t^*} \quad \Rightarrow \quad s_t = p_t - p_t^*$$

We can take the foreign price as the numeraire $(P_t^* = 1 \;\Rightarrow\; p_t^* = 0)$. Then,

$$s_t = p_t \tag{A.3}$$

The monetary equilibrium is described by the Cagan equation:

$$m_t^s - p_t = -\delta \times \dot{p}_t$$

By equation A.3, the Cagan equation can then be written as:

$$m_t^s - s_t = -\delta \times \dot{s}_t \tag{A.4}$$

The shadow price is the exchange rate that would prevail in the market if the peg is abandoned. The central bank will defend the peg until reserves reach a minimum level. To simplify the analysis, assume that the central bank abandons the fixed exchange rate when the reserves are exhausted, i.e., when $R_t = 0$. Then, the money supply (equation A.2) is given by:

$$M_t^s = D_t \quad \Rightarrow \quad m_t^s = d_t \tag{A.5}$$

Equations A.1 and A.5 imply:

$$\dot{m_t^s} = \dot{d_t} = \mu \quad \Rightarrow \quad m_t^s = m_0^s + \mu \times t$$

Hence, substituting this result in equation A.4 we obtain:

$$m_0^s + \mu \times t - s_t = -\delta \times \dot{s_t} \quad \Rightarrow \quad m_0^s + \mu \times t - s_t + \delta \times \dot{s_t} = 0$$

To solve this differential equation, we can try a linear solution:

$$s_t = constant + \mu \times t$$

Then,

$$m_0^s + \mu \times t - constant - \mu \times t + \delta \times \mu = 0 \quad \Rightarrow \quad constant = m_0^s + \delta \times \mu$$

Therefore, (the logarithm of) the shadow price is given by:

$$s_t = \underbrace{m_0^s + \delta \times \mu}_{\gamma} + \mu \times t = \gamma + \mu \times t \quad \Rightarrow \quad s_t = \gamma + \mu \times t \tag{A.6}$$

where γ and μ are constants and μ represents the rate of growth of the budget deficit.

We have proved that (the logarithm of) the shadow is a linear function of time.

2.7.2 Sell Out Condition

In this section we derive the sell out condition stated in Lemma 2.2. In our economy, an arbitrageur can choose between either holding local or foreign currency. Investing in domestic currency generates a return equal to r, while the foreign currency yields r^*. Assumption 2.1 ($r > r^*$) makes the domestic investment more attractive. Therefore, arbitrageurs are initially fully invested in domestic currency. At some random time $t = \tilde{t}$ the shadow price reaches the fixed exchange rate and from then onwards the peg might be attacked.

We want to determine the optimal selling date for the arbitrageur who becomes aware of the mispricing at time $t = t_i$. Each arbitrageur's payoff from selling out depends on the price at which he can sell the domestic currency. At time $t \geq \tilde{t}$ the peg may or may not hold. The price (in local currency) of an asset which yields a constant rate r is e^{rt}. The payoff function is denominated in foreign currency, hence the price of domestic currency if the peg still holds is $p_t = e^{rt}\frac{1}{\overline{S}}$, where $\overline{S}$ is the fixed exchange rate. However, if the peg has been abandoned, the price will be: $E[p_t|t_i] = E[e^{rt}\frac{1}{S_t}|t_i]$, which can be expressed as a fraction of the pre-crisis price as: $E[p_t|t_i] = e^{rt}\frac{1}{\overline{S}}E[\frac{\overline{S}}{S_t}|t_i]$, and $E[\frac{\overline{S}}{S_t}|t_i]$ can be understood as a rate of variation in the exchange rate. Then, arbitrageur i's payoff[9] from selling out at time t is given by:

$$\int_{t_i}^{t} e^{-r^*s}\, e^{rs}\, E\left[\frac{1}{S_s}\middle| t_i\right]\, \pi(s|t_i)\, ds + e^{-r^*t}\, e^{rt}\, \frac{1}{\overline{S}}\, (1 - \Pi(t|t_i))$$

where $\Pi(t|t_i)$ represents agent i's conditional cumulative distribution function of the date on which the peg collapses and $\pi(t|t_i)$ indicates the associated conditional density.

[9] We assume there are no transaction costs to simplify the specification of the payoff function. We could easily incorporate transaction costs in our setting. For example, let us define the transaction cost at time t equal to ce^{r^*t} (as in Abreu and Brunnermeier (2003)). This convenient formulation guarantees that the optimal solution is independent of the size of the transaction costs.

Differentiating the payoff function with respect to t yields:

$$\underbrace{\frac{\pi(t|t_i)}{1-\Pi(t|t_i)}}_{h(t|t_i)} = \frac{r-r^*}{1-E[\frac{\bar{S}}{S_t}|t_i]} \quad \Rightarrow \quad h(t|t_i) = \frac{r-r^*}{A_f^i(t-t_i)}$$

where $A_f^i(t-t_i)$ is the size of the expected appreciation of the foreign currency feared by agent i once the fixed exchange rate is abandoned. $A_f^i(t-t_i)$ is a strictly increasing and continuous function of $t-t_i$, the time elapsed since arbitrageur i becomes aware of the mispricing.

Therefore, arbitrageur i maximises his payoff to selling out at time t when his hazard rate equals the 'greed-to-fear ratio'. Thus, arbitrageur i holds:

- a maximum long position in local currency, if $h(t|t_i) < \dfrac{r-r^*}{A_f^i(t-t_i)}$, or
- a maximum short position in local currency, if $h(t|t_i) > \dfrac{r-r^*}{A_f^i(t-t_i)}$.

References

Abreu, D. and M. K. Brunnermeier (2003). Bubbles and Crashes. *Econometrica 71*(1), 173–204.

Broner, F. A. (2008). Discrete Devaluations and Multiple Equilibria in a First Generation Model of Currency Crises. *Journal of Monetary Economics*, forthcoming.

Brunnermeier, M. K. and J. Morgan (2006). Clock Games: Theory and Experiments. *Mimeo, Princeton University*. http://www.princeton.edu/ ~markus/ research/ papers/ clock_ games.pdf.

Flood, R. P. and P. M. Garber (1984). Collapsing Exchange Rate Regimes: Another Linear Example. *Journal of International Economics 17*(1-2), 1–13.

Flood, R. P. and N. Marion (1999). Perspectives on the Recent Currency Crisis Literature. *International Journal of Finance and Economics 4*(1), 1–26.

Guimarães, B. (2006). Dynamics of Currency Crises with Asset Market Frictions. *Journal of International Economics 68*(1), 141–158.

Krugman, P. (1979). A Model of Balance of Payment Crisis. *Journal of Money Credit and Banking 11*, 311–325.

Krugman, P. R. and M. Obstfeld (2003). *International Economics: Theory and Policy. Sixth Edition*. Boston: Addison Wesley.

Pastine, I. (2002). Speculation and the Decision To Abandon a Fixed Exchange Rate Regime. *Journal of International Economics 57*(1), 197–229.

Rochon, C. (2006). Devaluation without Common Knowledge. *Journal of International Economics 70*(2), 470–489.

Salant, S. W. and D. W. Henderson (1978). Market Anticipation of Government Policies and the Price of Gold. *Journal of Political Economy 86*(4), 627–648.

CHAPTER 3

Systemic Risk and Liquidity in Payment Systems

IN THIS CHAPTER we study liquidity and systemic risk in high-value payment systems. Flows in high-value systems are characterized by high velocity, meaning that the total amount paid and received is high relative to the stock of reserves. In such systems, banks rely heavily on incoming funds to finance outgoing payments, necessitating a high degree of coordination and synchronization. We use lattice-theoretic methods to solve for the unique fixed point of an equilibrium mapping and conduct comparative statics analyses on changes to the environment. We find that banks attempting to conserve liquidity cause an increase in the demand for intraday credit and, ultimately, full disruption of payments.

3.1 Introduction

Since the credit market turmoil began last summer, investment banks and other financial institutions have become seriously preoccupied with their liquidity. Banks have attempted to conserve cash holdings concerned about the possibility that they might face large draws on the standby liquidity facilities and credit enhancements of the special-purpose investment vehicles (SIVs) they sponsored. Moreover, as some of these SIVs were in danger of failing, banks came under raising pressure to rescue them by taking the assets of these off-balance sheet entities onto their own balance sheets. Greenlaw et al. (2008) and Brunnermeier (2008) present a detailed analysis of this recent episode in financial markets.

Banks, concerned about liquidity, have attempted to target more liquid balances as financial tension intensified. However, as banks increase their precautionary demand for liquid balances, they become less willing to lend to others. As a result, interbank funding rates have been showing clear signs of distress since August 2007. This has been highlighted by Fed Chairman Ben S. Bernanke in a speech last January[1]

> ...these developments have prompted banks to become protective of their liquidity and balance sheet capacity and thus to become less willing to provide funding to other market participants, including other banks. As a result, both overnight and term interbank funding markets have periodically come under considerable pressure, with spreads on interbank lending rates over various benchmark rates rising notably.

A shorter, but perhaps an even sharper episode of the systemic implications of the gridlock in payments came in the interbank payment system following the September 11,

[1] 'Financial Markets, the Economic Outlook, and Monetary Policy', speech by Ben S. Bernanke, 10 January 2008 (Bernanke (2008)).

2001 attacks. The interbank payment system processes very large sums of transactions between banks and other financial institutions. Moreover, one of the reasons for the large volumes of flows is due to the two-way flow that could potentially be netted between the set of banks. That is, the large flows leaving bank A is matched by a similarly large flow into bank A over the course of the day. However, the fact that the flows are not exactly synchronized means that payments flow backward and forward in gross terms, generating the large overall volume of flows.

The nettable nature of the flows allows a particular bank to rely heavily on the inflows from other banks to fund its outflows. McAndrews and Potter (2002) notes that banks typically hold only a very small amount of cash and other reserves to fund their payments. The cash and reserve holdings of banks amount to only around 1% of their total daily payment volume. The rest of the funding comes from the inflows from the payments made by the other banks. To put it another way, one dollar held by a particular bank at the beginning of the day changes hands around one hundred times during the course of the day. Such high velocities of circulation have been necessitated by the trend toward tighter liquidity management by banks, as they seek to lend out spare funds to earn income, and to calculate fine tolerance bounds for spare funds.

There is, however, a drawback to such high velocities that come from the fragility of overall payment flows to disruptions to the system, or a small step change in the desired precautionary balances targeted by the banks. After the September 11 attacks, banks attempted to conserve liquidity and raised their precautionary cash balances as a response to the greater uncertainty. Given the high velocity of funds, even a small change in target reserve balances can have a marked effect on overall payment volumes, and this is exactly what happened after September 11. McAndrews and Potter (2002) gives a detailed account of the events in the U.S. Fedwire payment system following the September 11 attacks.

Our paper addresses the issue of liquidity in a flow system. The focus is on the interdependence of the agents in the system, and the manner in which equilibrium payments are determined and how the aggregate outcome changes with shifts in the parameters describing the environment. In keeping with the systemic perspective, we model the interdependence of flows and show how the equilibrium flows correspond to the (unique) fixed point of a well-defined equilibrium mapping. The usefulness of our approach rests on the fact that our model abstracts away from specific institutional details, and rests only of the robust features of system interaction. The comparative statics exercise draws on methods on lattice theory, developed by Topkis (1978) and Milgrom and Roberts (1994), and allows us to analyze the repercussions on the financial system of a change in precautionary demand for liquid balances. Specifically, we aim at better understanding the systemic implications of a shift towards more conservative balance sheets targeted by one or a small set of market participants in a payment system.

We find that a reduction in outgoing payments to conserve cash holdings translates into lesser incoming funds to other banks, but lesser incoming funds will then affect outgoing transfers. Our findings show that if few banks targeted more liquid balances, there will be an increase in the demand for intraday liquidity provided by the Federal Reserve System and it could even lead to a full disruption of payments.

The outline of the paper is as follows. In the next section we introduce a theoretical framework for the role of interlocking claims and obligations in a flow system. An application to the interbank payment system then follows. Section 3.3 briefly reviews the US payment system paying special attention to the Fedwire Funds Service. Section 3.4 presents numerical simulations based on a stylized payment system. Then, Section 3.5 analyzes the response of payment systems to a change in precautionary balances. Miscoordination in payments and a potential policy intended to economize on the use of intraday credit are discussed in

Section 3.6. Section 3.7 concludes.

3.2 The Model

There are n agents in the payment system, whom we will refer to as "banks". Every member of the payment system maintains an account to make payments. This account contains all balances including its credit capacity.

Banks in a payment system rely heavily on incoming funds to make their payments. Let us denote by y_t^i the time t payments bank i sends to other members in the payment system. These payments are increasing in the total funds x_t^i bank i receives from other members during some period of time (from $t-1$ to t). We do not need to impose a specific functional form on this relationship. In particular, we will allow each bank to respond differently to incoming funds. The only condition we impose is that each bank only pays out a proportion of its incoming funds. Formally, it entails that transfers do not decrease as incoming payments rise and that its slope is bounded above by 1 everywhere. Then, outgoing transfers made by bank i at time t are given by:

$$y_t^i = f^i(x_t^i, \theta_t)$$

where $\theta_t = (b_t, c_t)$ and b_t represents the profile of balances b_t^i and c_t is the profile of remaining credit c_t^i. Outgoing payments made by bank i will depend on incoming funds, which in turn depends on all payments sent over the payment system. Then, for every member in the payment system we have:

$$y_t^i = f^i(x_t^i(y_{t-1}), \theta_t) \qquad i = 1, \dots, n$$

This system can be written as:

$$y_t = F(y_{t-1}, \theta_t)$$

where $y_t = [y_t^1, y_t^2, \dots, y_t^n]^\top$ and $F = [f^1, f^2, \dots, f^n]^\top$.

The task of determining payment flows in a financial system thus entails solving for a consistent set of payments - that is, solving a fixed point problem of the mapping F. We will show that our problem has a well-defined solution and that the set of payments can be determined uniquely as a function of the underlying parameters of the payment system. We will organize the proof in two steps. Step 1 shows the existence of at least one fixed point of the mapping F. We will show uniqueness in Step 2.

Step 1. *Existence of a fixed point of the mapping F.*

Lemma 3.1. *(Tarski (1955) Fixed Point Theorem) Let $(Y, \leq)$ be a complete lattice and F be a non-decreasing function on Y. Then there are y^* and y_* such that $F(y^*) = y^*$, $F(y_*) = y_*$, and for any fixed point y, we have $y_* \leq y \leq y^*$.*

A *complete lattice* is a partially ordered set $(Y, \leq)$ which satisfies that every non-empty subset $S \subseteq Y$ has both a least upper bound (join), $\sup(S)$, and a greatest lower bound (meet), $\inf(S)$. In our payments setting, we can define a complete lattice $(Y, \leq)$ as formed by a non-empty set of outgoing payments Y and the binary relation $\leq$. Every subset S of the payment flows Y has a greatest lower bound (flows are non-negative) and a least upper bound which we will denote by $\overline{y}_i$. $\overline{y}_i$ represents the maximum flow of payments bank i can send through the payment system. This condition can be understood as a maximum flow capacity due to some technological limitations of the networks and communication systems used by the banks to receive and process transfer orders. We have:

$$Y = [0, \bar{y}_1] \times [0, \bar{y}_2] \times ... \times [0, \bar{y}_n]$$

The relation $\leq$ formalizes the notion of an ordering of the elements of Y such that $y \leq y'$ when $y_i \leq y_i'$ for all the components i and $y_k < y_k'$ for some component k.

In our payments problem, $(Y, \leq)$ is a complete lattice and since outgoing payments made by bank i do not decrease as incoming funds rise, i.e. f^i is a non-decreasing function, then $F = [f^1, f^2, ... f^n]^\top$ is non-decreasing on Y. Our setting hence satisfies the conditions of the Tarski's Theorem and as a result there exists *at least* one fixed point of the mapping F. Moreover, in Step 2 we will show that the fixed point is unique.

Step 2. *Uniqueness of the fixed point of the mapping F.*

Theorem 3.1. *There exists a unique profile of payments flows y_t that solves $y_t = F(y_{t-1}, \theta_t)$.*

Proof. F is a non-decreasing function on a complete lattice $(Y, \leq)$. Then, by Tarski's Fixed Point Theorem (Lemma 3.1), F has a largest y^* and a smallest y_* fixed point. Let us consider, contrary to Theorem 3.1, that there exist two distinct fixed points such that $y_i^* \geq y_{i*}$ for all components i and $y_k^* > y_{k*}$ for some component k. Denote by x_i^* the payments received by bank i evaluated at y_i^* and by x_{i*} the payments received by bank i evaluated at y_{i*}. By the Mean Value Theorem, for any differentiable function f on $[x_{i*}, x_i^*]$, there exists a point $z \in (x_{i*}, x_i^*)$ such that

$$f(x_i^*) - f(x_{i*}) = f'(z)(x_i^* - x_{i*})$$

We have assumed that the slope of the outgoing payments is bounded above by 1 everywhere ($\frac{\partial f^i}{\partial x_i} < 1$ everywhere). Hence,

$$\begin{cases} y_1^* - y_{1*} = f^1(\overline{y}_1, x_1^*) - f^1(\overline{y}_1, x_{1*}) \leq x_1^* - x_{1*} \\ y_2^* - y_{2*} = f^2(\overline{y}_2, x_2^*) - f^2(\overline{y}_2, x_{2*}) \leq x_2^* - x_{2*} \\ \vdots \\ y_k^* - y_{k*} = f^k(\overline{y}_k, x_k^*) - f^k(\overline{y}_k, x_{k*}) < x_k^* - x_{k*} \\ \vdots \\ y_n^* - y_{n*} = f^n(\overline{y}_n, x_n^*) - f^n(\overline{y}_n, x_{n*}) \leq x_n^* - x_{n*} \end{cases}$$

Re-arranging the previous system of equations we get

$$\begin{cases} x_{1*} - y_{1*} \leq x_1^* - y_1^* \\ x_{2*} - y_{2*} \leq x_2^* - y_2^* \\ \vdots \\ x_{k*} - y_{k*} < x_k^* - y_k^* \\ \vdots \\ x_{n*} - y_{n*} \leq x_n^* - y_n^* \end{cases}$$

Summing across banks we have

$$\sum_{i=1}^{n} x_{i*} - \sum_{i=1}^{n} y_{i*} < \sum_{i=1}^{n} x_i^* - \sum_{i=1}^{n} y_i^*$$

so that the total value of the balances including credit capacity is strictly larger under y^*, which is impossible. Therefore, there cannot exist two distinct fixed points and as a result $y^* = y_*$. □

Although uniqueness is relevant to our analysis of payment systems, our key insights stem from the comparative statics results due to Milgrom and Roberts (1994).

3.2.1 Comparative Statics

Theorem 3.2. *Let $y_t^*(\theta_t)$ be the unique fixed point of the mapping F. If for all $y_t \in Y$, F is increasing in θ_t, then $y_t^*(\theta_t)$ is increasing in θ_t.*

Proof. Let F be monotone non-decreasing and Y a complete lattice. From Tarski's Fixed Point Theorem (Lemma 3.1) and Theorem 3.1 there exists a unique fixed point $y_t^*(\theta_t)$ of the mapping F. For the simplicity of the argument, let us suppress the subscript t. Define the set $S(\theta)$ as

$$S(\theta) = \{y | F(y, \theta) \leq y\}$$

and define $y^*(\theta) = \inf S(\theta)$. Since F is non-decreasing in θ, the set $S(\theta)$ becomes more exclusive as θ increases. Hence, $y^*(\theta)$ is a non-decreasing function of θ. Formally, if F is increasing in θ, then for $\theta' > \theta$, $F(\theta') > F(\theta)$ and

$$S(\theta') = \{y | F(y, \theta') \leq y\} \subset S(\theta)$$

Thus,

$$y^*(\theta') = \inf S(\theta') > \inf S(\theta) = y^*(\theta)$$

Therefore, if F is increasing in θ, the fixed point $y^*(\theta)$ is increasing in θ too. □

3.3 Payment Systems

Payment and securities settlement systems are essential components of the financial systems and vital to the stability of any economy. A key element of the payment system is the interbank payment system that allows funds transfers between entities. Large-value (or wholesale) funds transfer systems are usually distinguished from retail systems. Retail funds systems transfer large volumes of payments of relatively low value while wholesale systems are used to process large-value payments. Interbank funds transfer systems can also be classified according to their settlement process. The settlement of funds can occur on a net basis (net settlement systems) or on a transaction-by-transaction basis (gross settlement systems). The timing of the settlement allows another classification of these systems depending on whether they settle at some pre-specified settlement times (designated-time (or deferred) settlement systems) or on a continuous basis during the processing day (real-time settlement systems).

A central aspect of the design of large-value payment systems is the trade-off between liquidity and settlement risk. Real-time gross settlement systems are in constant need of liquidity to settle payments in real time while net settlement systems are very liquid but vulnerable to settlement failure[2]. In the last twenty years, large-value payments systems have evolved rapidly towards greater control of credit risk[3].

In the United States, the two largest large-value payment systems are the Federal Reserve Funds and Securities Services (Fedwire) and the Clearing House Interbank Payments System (CHIPS). CHIPS, launched in 1970, is a real-time, final payment system for US dollars that

[2]Zhou (2000) discusses the provision of intraday liquidity by a central bank in a real-time gross settlement system and some policy measures to limit the potential credit risk.

[3]Martin (2005) analyzes the recent evolution of large-value payment systems and the compromise between providing liquidity and settlement risk. See also Bech and Hobijn (2006) for a study on the history and determinants of adoption of real-time gross settlement payment systems by central banks across the world.

uses bi-lateral and multi-lateral netting to clear and settle business-to-business transactions. CHIPS is a bank-owned payment system operated by the Clearing House Interbank Payments Company L.L.C. whose members consist of 46 of the world's largest financial institutions. It processes over 300,000 payments on an average day with a gross value of $1.5 trillion.

Fedwire is a large-dollar funds and securities transfer system that links the twelve Banks of the Federal Reserve System[4]. The Fedwire funds transfer system, which we will discuss in more detail below, is a real-time gross settlement system, developed in 1918, that settles transactions individually on an order-by-order basis without netting. The average daily value of transactions exceeded $2 trillion in 2005 with a volume of approximately 527,000 daily payments. Settlement of most US government securities occurs over the Fedwire book-entry security system, a real-time delivery-versus-payment gross settlement system that allows the immediate and simultaneous transfer of securities against payments. More than 9,100 participants hold and transfer US Treasury, US government agency securities and securities issued by international organizations such as the World Bank. In 2005 it processed over 89,000 transfers a day with an average daily value of $1.5 trillion. Figure 3.1 depicts the evolution of the average daily value and volume of transfers sent over CHIPS and Fedwire.

3.3.1 Fedwire Funds Service

Fedwire Funds Service, owned and operated by the Federal Reserve Banks, is an electronic payment system that allows participants to make same-day final payments in central bank money. An institution that maintains an account at a Reserve Bank can generally become a Fedwire participant. Approximately 9,400 participants are able to initiate and receive funds transfers over Fedwire. When using the Fedwire Funds Service, a sender instructs a Federal

[4]See Gilbert et al. (1997) for an overview of the origins and evolution of Fedwire.

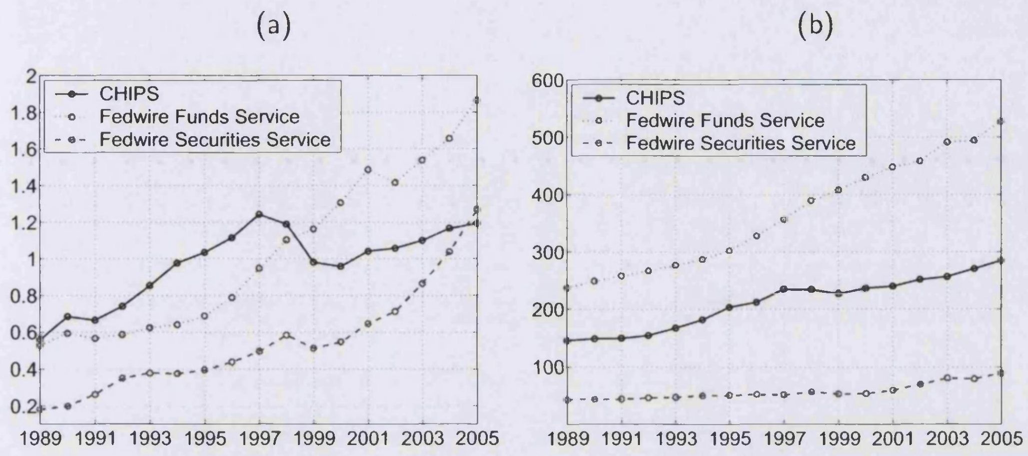

Figure 3.1: Average daily value (a) ($ trillion) and volume (b) (thousands) of transactions over CHIPS, Fedwire Funds Service and Fedwire Securities Service, 1989-2005. Source: The Federal Reserve Board and CHIPS.

Reserve Bank to debit its own Federal Reserve account for the amount of the transfer and to credit the Federal Reserve account of another participant.

The Fedwire Funds Service operates 21.5 hours each business day (Monday through Friday), from 9.00 p.m. Eastern Time (ET) on the preceding calendar day to 6.30 p.m. ET[5]. It was expanded in December 1997 from ten hours to eighteen hours (12:30 a.m. - 6:30 p.m.) and again in May 2004 to accommodate the twenty-one and a half operating hours. This change increased overlap of Fedwire's operating hours with foreign markets and helped reduce foreign exchange settlement risk.

A Fedwire participant sending payments is required to have sufficient funds, either in the form of account balance or overdraft capacity, or the payment order may be rejected. The Federal Reserve imposes a minimum level of reserves, which can be satisfied with vault cash[6] and balances deposited in Federal Reserve accounts, neither of which earn interest. A

[5]A detailed description of Fedwire Funds Service operating hours can be found at www.frbservices.org/Wholesale/FedwireOperatingHours.html.

[6]Vault cash refers to U.S. currency and coin owned and held by a depository institution.

Fedwire participant may also commit itself or be required to hold balances in addition to any reserve balance requirement (clearing balances). Clearing balances earn no explicit interest but implicit credits that may offset the cost of Federal Reserve services. Fedwire participants thus tend to optimize the size of the balances in their Federal Reserve accounts[7].

When an institution has insufficient funds in its Federal Reserve account to cover its debits, the institution runs a negative balance or *daylight overdraft*. Daylight overdrafts result because of a mismatch in timing between incoming funds and outgoing payments (McAndrews and Rajan (2000)). Each Fedwire participant may establish (or is assigned) a maximum amount of daylight overdraft known as *net debit cap*[8]. An institution's net debit cap is a function of its capital measure. Specifically, it is defined as a cap multiple times its capital measure, where the cap multiple is determined by the institution's cap category. An institution's capital measure varies over time while its cap category does not normally change within a one-year period. Each institution's cap category is considered confidential information and hence it is unknown to other Fedwire participants (Federal Reserve (2005), Federal Reserve (2006d)).

In 2000 the Federal Reverse Board's analysis of overdraft levels, liquidity patterns, and payment system developments revealed that although approximately 97 percent of depository institutions with positive net debit caps use less than 50 percent of their daylight overdraft capacity, a small number of institutions found their net debit caps constraining (Federal Reserve (2001)). To provide additional liquidity, the Federal Reserve now allows certain institutions to pledge collateral to gain access to daylight overdraft capacity above their net debit caps. The maximum daylight overdraft capacity is thus defined as the sum

[7]Bennett and Peristiani (2002) find that required reserve balances in Federal Reserve accounts have declined sharply while vault cash applied against reserve requirements has increased. They argue that reserve requirements have become less binding for US commercial banks and depository institutions.

[8]Appendix 3.8.1 briefly reviews the evolution of net debit caps and describes the different cap categories and associated cap multiples.

of the institution's net debit cap and its collateralized capacity.

To control the use of intraday credit, the Federal Reserve began charging daylight overdraft fees in April 1994. The fee was initially set at an annual rate of 24 basis points and it was increased to 36 basis points in 1995[9]. At the end of each Fedwire operating day the end-of-minute account balances are calculated. The average overdraft is obtained by adding all negative end-of-minute balances and dividing this amount by the total number of minutes in an operating day (1291 minutes). An institution's daylight overdraft charge is defined as its average overdraft multiplied by the effective daily rate (minus a deductible). Table 3.4 presents an example of the calculation of a daylight overdraft charge. An institution incurring daylight overdrafts of approximately \$3 million every minute during a Fedwire operating day would face an overdraft charge of \$6.58.

At the end of the operating day, a Fedwire participant with a negative closing balance incurs overnight overdraft. An overnight overdraft is considered an unauthorized extension of credit. The rate charged on overnight overdrafts is generally 400 basis points over the effective federal funds rate. If an overnight overdraft occurs, the institution will be contacted by the Reserve Bank, it will be required to hold extra reserves to make up reserve balance deficiencies and the penalty fee will be increased by 100 basis points if there have been more than three overnight overdraft occurrences in a year. The Reserve Bank will also take other actions to minimize continued overnight overdrafts (Federal Reserve (2006a)).

[9]Fedwire operates 21.5 hours a day, hence the effective annual rate is 32.25 basis points ($36 \times \frac{21.5}{24}$) and the effective daily rate is 0.089 basis points ($32.25 \times \frac{1}{360}$).

3.4 An Example of Payment System

In this section we present numerical simulations of a stylized payment system reminiscent of Fedwire. We first describe the payment system and next we introduce the characteristics of a standard day of transactions in this payment system.

3.4.1 The Payment System

Consider a network of four banks. Each bank sends and receives payments from other members of the payment system. The payment system opens at 9.00 p.m. on the preceding calendar day and closes at 6.30 p.m. Every bank begins the business day with a positive balance at its central bank account and may incur daylight overdrafts to cover negative balances up to its net debit cap. For simplicity we assume initial balances and net debit caps of equal size. The expected value of bank i's outgoing payments equals the expected value of its incoming funds to guarantee that no bank is systematically worse off. Each member of the payment system is subject to idiosyncratic shocks which determine its final payments.

Following McAndrews and Potter (2002) we define outgoing transfers as a linear function of the payments a bank receives from all other banks. Specifically, at every minute of the operational day, bank i pays at most 80 percent of its cumulative receipts and a proportion of its reserves and credit capacity (which we fix at 10 percent of the bank's net debit cap). We assume banks settle obligations whenever they have sufficient funds. When the value of payments exceeds 80 percent of a bank's incoming funds and 10 percent of its net debit cap, payments are placed in queue. Queued payments are settled as soon as sufficient funds become available[10].

When banks use more than 50 percent of their own daylight overdraft capacity[11], they become concern about liquidity shortages and reduce the value of their outgoing transfers. Inspired by McAndrews and Potter's estimates of the slope of the reaction function of banks during the September 11, 2001, events, we assume that banks would then pay at most 20 percent of their incoming funds. Table 3.1 summarizes how banks organize their payments.

Banks pay at most:	
NORMAL CONDITIONS	CONCERNED ABOUT LIQUIDITY
80% of its cumulative receipts and up to 10% of the bank's net debit cap	20% of its cumulative receipts and up to 10% of the bank's net debit cap

Table 3.1: Outgoing payments.

Once bank *i* becomes concerned about a liquidity shortage and reduces the slope of its reaction function, it faces one of two possible scenarios. Its balance may become positive (it has been receiving funds from all other banks according to the 80 percent rule while it has been paying out only 20 percent of its incoming transfers). The "episode" would be over and bank *i* would return to normal conditions. However, it may also be possible that despite reducing the amount of outgoing payments its demand for daylight overdraft continues to rise. Bank *i* would incur negative balances up to its net debit cap. At that time, it would stop using intraday credit to make payments and any incoming funds would be devoted to settle queued payments and to satisfy outgoing transfers at the 20 percent rate per minute.

We first introduce the baseline setting. Then, in Section 3.5, we analyze what happens

[10]To avoid excessive fluctuations we consider that if at any time bank *i*'s use of reserves and credit capacity is below the 10 percent threshold, bank *i* will devote its spare capacity to settle queued payments. Otherwise, payments will remain in queue.

[11]According to a Federal Reserve Board's review, in 2000, 97 percent of depository institutions with positive net debit caps use less than 50 percent of their daylight overdraft capacity (Federal Reserve (2001)).

when a bank attempts to conserve cash holdings. Section 3.6 discusses the potential implications of miscoordination in payments and a policy intended to economize on the use of intraday credit.

3.4.2 Standard Functioning of the Payment System

We consider a payment system as the one just described above and focus on the functioning of the payment system during one business day. The value of payments by time of the day is depicted in Figure 3.2(a). Payments are defined to follow the pattern of the average value of transactions sent over the Fedwire Funds Service[12]. Thus, as in the case of Fedwire, the market opens at 9.30 p.m. on the preceding calendar day, there is almost no payment activity before 8 a.m. and from then on the value of payments increases steadily and it peaks around 4.30 p.m. and again around 5.15 p.m.[13] The market closes at 6.30 p.m.

Each bank starts the operating day with a positive balance in their Federal Reserve accounts, which we assume equal to 10. Figure 3.2(b) plots the balances at the central bank account of each member of the payment system during this business day. Before 8 a.m. all balances remain close to the opening balance because of the low payment activity. Let us focus our attention on banks B and C, for instance. Bank C initially receives more payment orders than transfers. Bank B represents the opposite case. Just after 1 p.m. bank C starts running negative balances and thus incurring daylight overdrafts as illustrated in Figure 3.2(c). Overdrafts peak at 5.10 p.m., shortly after bank C places some payments in queue. The top panel of Figure 3.2(d) presents the payments placed in queue at each minute of the operating day. In this case, queued payments are settled at the next minute.

[12]See McAndrews and Rajan (2000) (Chart 3) and Coleman (2002) (Chart 1).

[13]The average value of Fedwire funds peaks at 4.30 p.m. and at 5.15 p.m. most likely from settlement at the Depositary Trust Company and from institutions funding their end-of-day positions in CHIPS respectively (Coleman (2002)).

After that, bank C begins receiving more payments than payment orders. At 5.25 p.m. it runs a positive balance and ends the day with a positive balance (its closing balance more than doubles its opening balance).

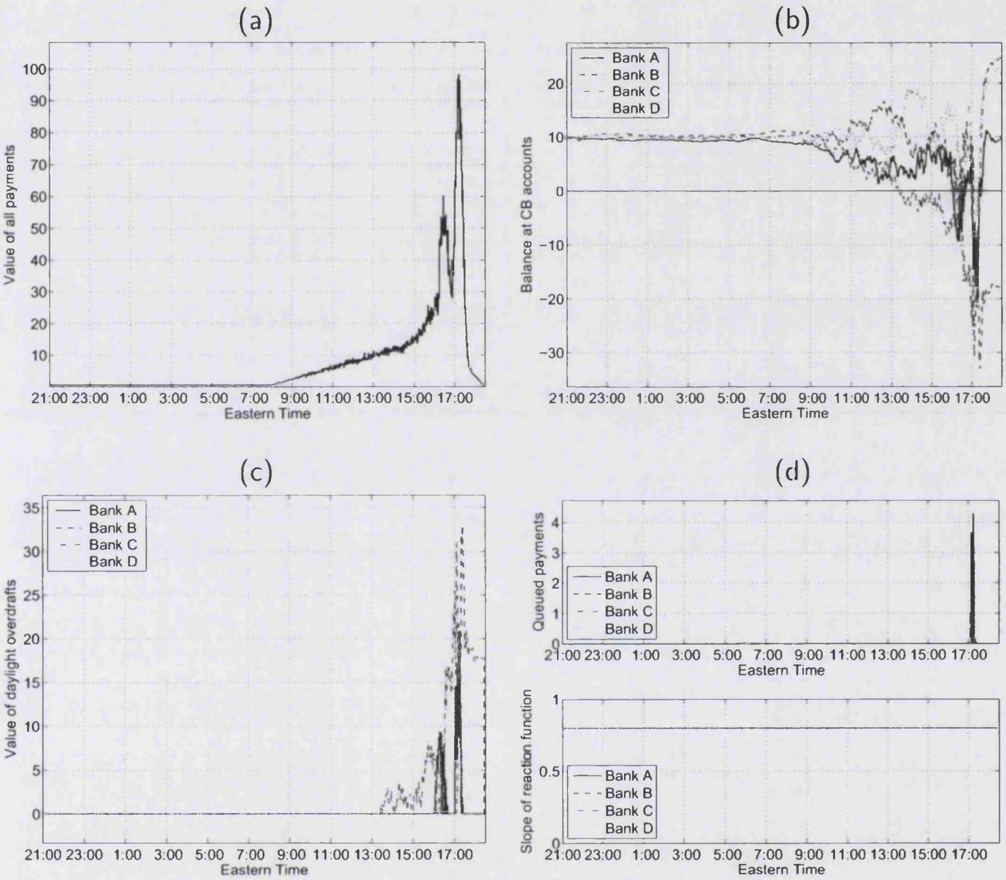

Figure 3.2: STANDARD FUNCTIONING OF THE PAYMENT SYSTEM - Total value of payments sent over the Payment System (a), banks' balances (b), value of daylight overdrafts (c), queued payments and slopes of reaction functions (d) by time of the day.

As shown in Figures 3.2(c) and the top panel of 3.2(d), banks A, B and D also incur daylight overdrafts and delayed payments. Banks A and D reach the end of the operating day with positive balances while B runs a negative closing balance and it will need to "sweep" deposits from another account to its account at the central bank to avoid an overnight overdraft charge (Figure 3.2(b)).

In this exercise, we set net debit caps equal to 100. During this business day, none of the four banks have reached half of their net debit caps (their balances never fall below −50 (50 percent of their cap)) and hence every bank sends out payments according to the 80 percent rule. The slope of their reaction functions is thus 0.8 as depicted in the bottom panel of Figure 3.2(d).

Overall, this example pictures the smooth functioning of the payment system. Let us now introduce a more interesting scenario.

3.5 Increased Precautionary Demand

Consider a member of the payment system becomes suddenly concerned about a liquidity shortage. Suppose, for instance, this bank wants to conserve cash holdings because the conduits, SIVs or other off-balance sheet vehicles that it is sponsoring have drawn on credit lines as experienced in credit markets during the recent market turmoil.

We are interested in the consequences of an increase in the liquid balances targeted by one bank in our payment system. Specifically, we assume bank A is the one concerned about a liquidity shortage. To preserve cash, bank A decides to pay only 20 percent of the funds it receives (and up to 10 percent of its net debit cap per minute). Banks B, C and D initially behave as in the baseline setting, i.e. they send out payments according to the 80 percent rule. As a result, even though there is almost no payment activity before 8 a.m., the size of bank A's balance increases steadily as it receives transfers at the 80 percent rate while paying out at most 20 percent of the funds it receives. The evolution of the balances hold at the central bank accounts as a function of time is depicted in Figure 3.3(b). Comparing Figure 3.3(b) to Figure 3.2(b) clearly shows that both the size and pattern of these balances differ from the standard functioning of the payment system described in Subsection 3.4.2.

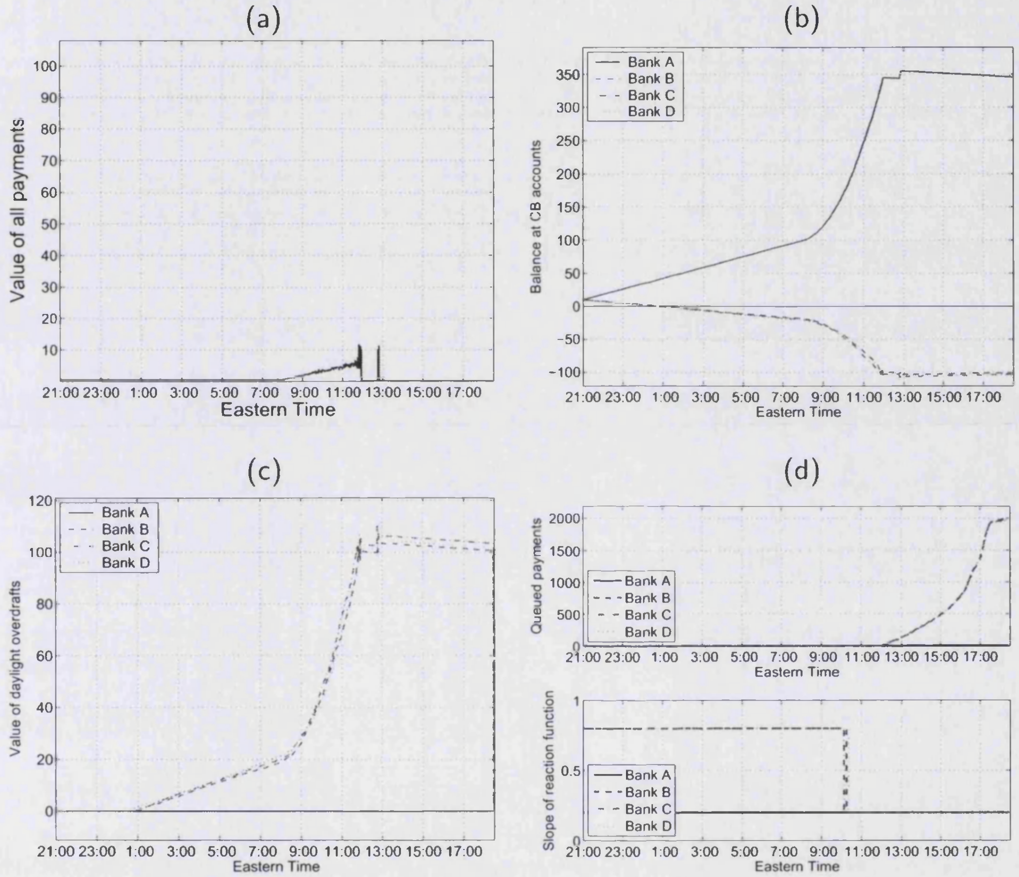

Figure 3.3: INCREASED PRECAUTIONARY DEMAND - Total value of payments sent over the Payment System (a), banks' balances (b), value of daylight overdrafts (c), queued payments and slopes of reaction functions (d) by time of the day.

Just before 1 a.m., banks B, C and D begin running negative balances (Figure 3.3(b)) and incurring daylight overdrafts (Figure 3.3(c)). Around 10.15 a.m. their daylight overdrafts exceed half of their caps (Figure 3.3(b)) and they start paying out at most 20 percent of the funds they receive as illustrated in the bottom panel of Figure 3.3(d). At noon, banks

B, C and D's daylight overdrafts reach their net debit caps (Figure 3.3(c)) and they start placing payments in queue (top panel of Figure 3.3(d)). At that time, total payments are finally disrupted as shown in Figure 3.3(a).

It is important to highlight that a change in preferences of a member of the payment system towards more liquid balances induces the following effects. First, it causes full disruption of payments. Payment activity is disrupted as soon as the other members reach their maximum credit capacity. Second, the size of banks' balances hold at the Federal Reserve increases compared to the standard functioning of the payment system. Thirdly, a raise in precautionary demand leads to an enormous use of intraday credit.

3.6 Miscoordination and Multiple Settlements

In this section we analyze, first, if the payment system is sensitive to timing miscoordination. Secondly, we discuss the possibility of having two synchronization periods (instead of having only one late in the afternoon).

3.6.1 Timing Miscoordination

In the U.S. payment system, banks in aggregate make payments that exceed their deposits at the Federal Reserve Banks by a factor of more than 100[14]. To achieve such velocities a high degree of coordination and synchronization is required. In the standard functioning of the payment system, introduced in Subsection 3.4.2, we assumed banks could synchronize their payment activity perfectly, i.e., we considered the value of the payments made by every bank exhibited exactly the same pattern. In the next example, we examine the response of

[14]See McAndrews and Potter (2002).

the payment system to miscoordination in the timing of payments. Specifically, suppose that banks experience a five-minute delay with respect to each others. Figure 3.4 summarizes our findings.

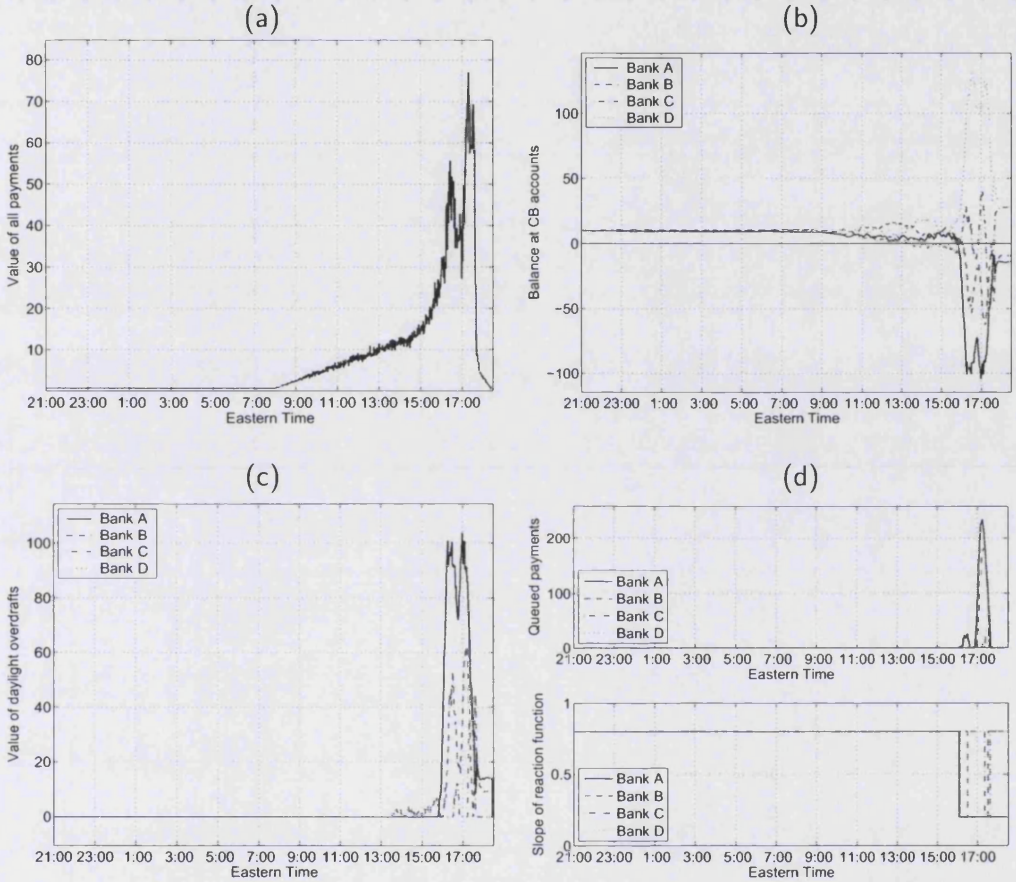

Figure 3.4: TIMING MISCOORDINATION - Total value of payments sent over the Payment System (a), banks' balances (b), value of daylight overdrafts (c), queued payments and slopes of reaction functions (d) by time of the day.

Payments are organized as follows. Bank A starts sending out payments first. B begins five minutes after A, then C after B and D will be the last one. As a result, there will be a mismatch in timing between the settlement of payments owned and the settlement of payments due. Initially, bank A makes more payments that it receives (Figure 3.4(b)) and hence it incurs daylight overdrafts as shown in Figure 3.4(c). The bank that pays first will demand the largest amount of intraday credit. Then, the bank after the first one and so on (Figure 3.4(c)). This pattern persists across business days (simulations). Once a bank has used half of its credit capacity, it starts making payments according to the 20 percent rule. This is depicted in the bottom panel of Figure 3.4(d). The top panel of Figure 3.4(d) reports the payments placed in queue.

Banks A and B end the operating day with a negative balance while banks C and D run positive closing balances. We could think this is a consequence of the time mismatch. However, this is not the case. In this exercise, payments are delayed but the expected value of outgoing funds and incoming payments is still the same. To emphasize this result we present a different business day in Figure 3.5. Now, banks A and B hold a positive closing balance while banks C and D will need to "sweep" deposits to avoid the overnight overdraft penalty rate.

A five-minute miscoordination in payments thus induces an increase in the size of balances at the central bank accounts and a more intense use of the intraday credit compared to the standard functioning of the payment system.

3.6.2 Multiple Settlement Periods

To economize on the use of intraday credit, a potential operational change in settlement systems which is being considered (Federal Reserve (2006b)) is the possibility of developing multiple settlement periods. An example of such policy could be the establishment of two synchronization periods, one late in the morning and then another early in the afternoon peak, as proposed by McAndrews and Rajan (2000).

Assume there is an additional synchronization period around noon such that the value of

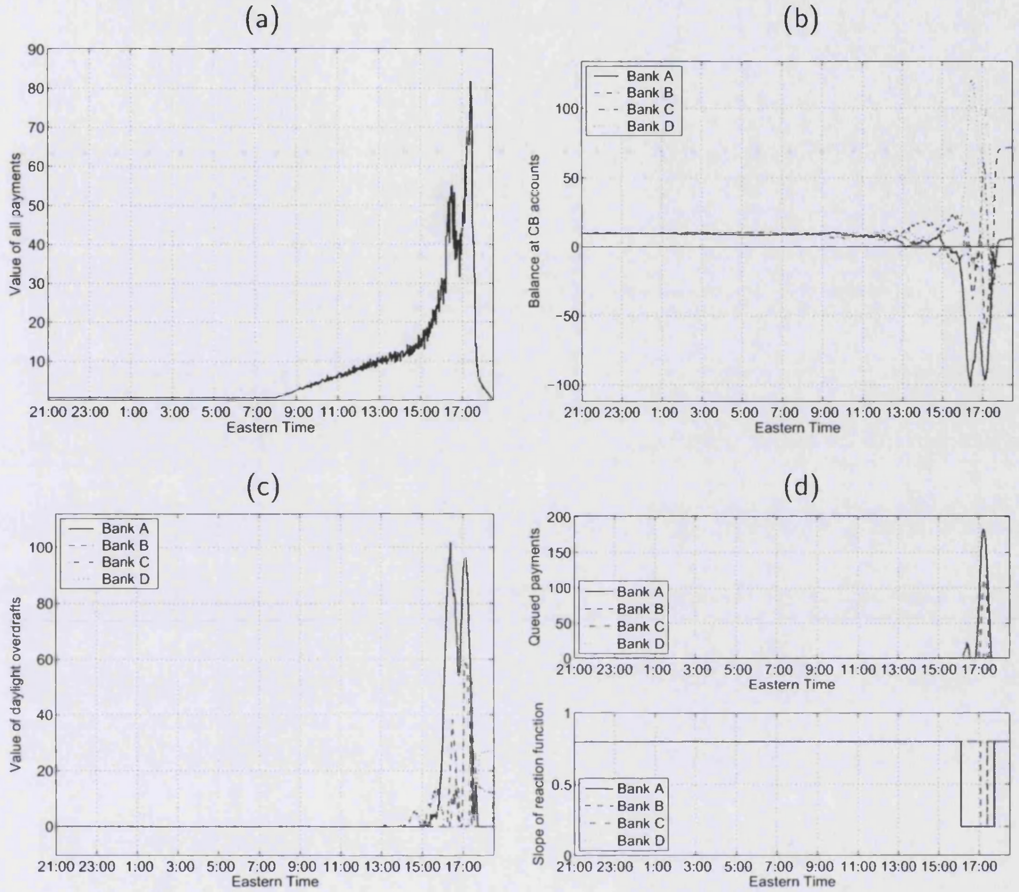

Figure 3.5: TIMING MISCOORDINATION (Different business day)

payments sent over the payment system follows the pattern in Figure 3.6(a). Let us discuss the response of the payment system to such policy. Our results are reported in Figure 3.6.

Relative to the standard functioning, we find that introducing multiple synchronization periods does not alter significantly the size of banks' balances at their central bank accounts or the ratio between outgoing and incoming funds. This is depicted in Figures 3.2(b) and 3.6(b) and at the bottom panel of Figures 3.2(d) and 3.6(d). On the contrary, it reduces the use of daylight overdraft (Figures 3.2(c) and 3.6(c)) and the amount of payments in

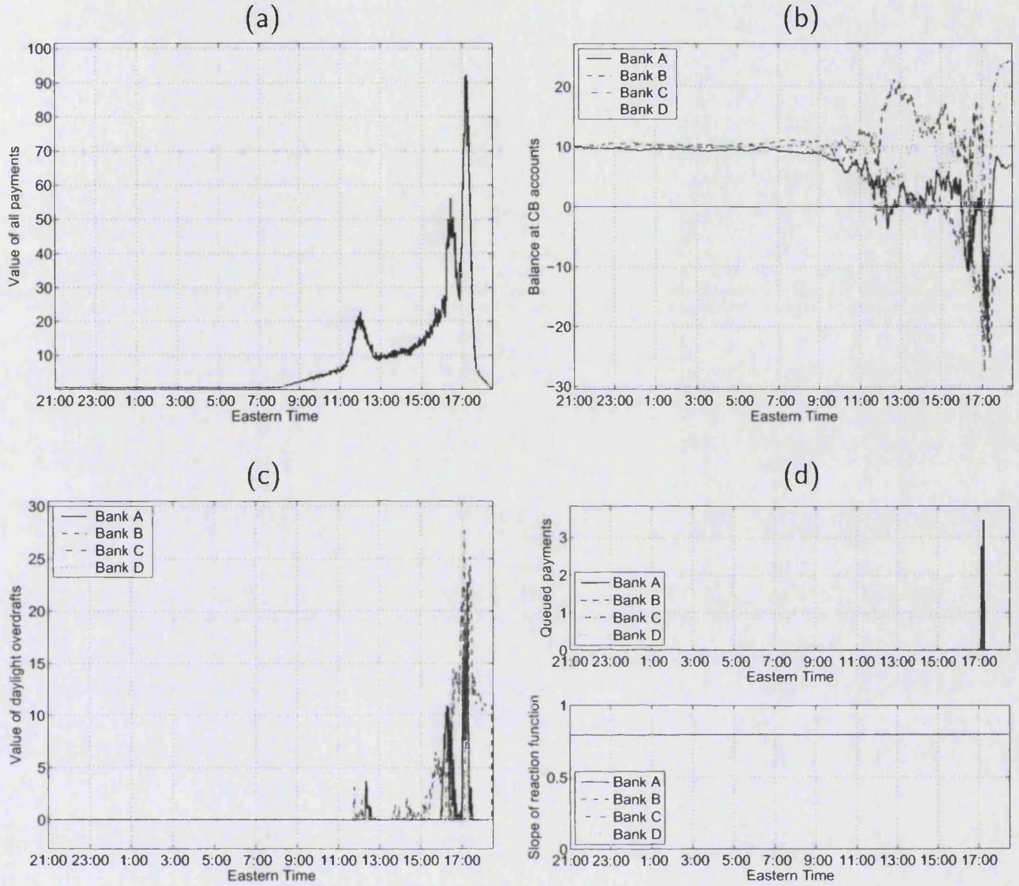

Figure 3.6: MULTIPLE SYNCHRONIZATION PERIODS - Total value of payments sent over the Payment System (a), banks' balances (b), value of daylight overdrafts (c), queued payments and slopes of reaction functions (d) by time of the day.

queue (bottom panel of Figures 3.2(d) and 3.6(d)).

We conclude that having two synchronization periods does economize on the use of intraday credit.

3.7 Concluding Remarks

The focus of the paper is on the role of liquidity in a flow system. We argue for the importance of the interdependence of the flows in high-value payment systems. High-value payment systems such as the interbank payment systems that constitute the backbone of the modern financial system, link banks and other financial institutions together into a tightly knit system. Financial institutions rely heavily on incoming funds to make their payments and as such, their ability to execute payments will affect other participants' capability to send out funds. Changes in outgoing transfers will affect incoming funds and incoming funds changes will affect outgoing transfers. The loop thus created may generate amplified responses to any shocks to the high-value payment system.

We draw from the literature on lattice-theoretic methods to solve for the unique fixed point of an equilibrium mapping in high-value payment systems. Using numerical simulations based on simple decision rules which replicate the observed data on the Fedwire payment system in the U.S., we then perform comparative statics analysis on changes to the environment of this payment system. We find that changes in preferences towards more conservative balances by one bank in the payment system leads to a full disruption of payments, increased balances at the Federal Reserve accounts and an immense use of intraday credit.

Our framework also allows simulations of counterfactual "what if" scenarios of disturbances that may lead to gridlock and systemic breakdown, as well as the consequences of potential policies such as the possibility of multiple settlement periods. We show that introducing a second synchronization period late in the morning economizes on the demand for intraday credit.

3.8 Appendix

3.8.1 Net Debit Caps

In 1985, the Federal Reserve Board developed a payment system risk policy on risks in large-dollar wire transfer systems. The policy introduced four categories of limits (net debit caps) on the maximum amount of daylight overdraft credit that the Reserve Banks extended to depository institutions: *high*, *above average*, *average* and *zero*. In 1987 a new net debit cap (*de minimis*) was approved. It was intended for depository institutions that incur relatively small overdrafts. The Board incorporated a sixth cap class (*exempt-from-filing*) and modified the existing *de minimis* cap multiple in 1990. The *de minimis* cap multiple was then increased in 1994 when daylight overdraft fees were introduced[15]. A brief summary[16] of the actual cap categories and their associated cap multiples for maximum overdrafts on any day (single-day cap) and for the daily maximum level averaged over a two-week period (two-week average cap) are presented in Tables 3.2 and 3.3.

3.8.2 Example Daylight Overdraft Charge Calculation

Table 3.4 contains an example of the calculation of a daylight overdraft charge.

[15] For a comprehensive study of the history of Federal Reserve daylight credit see Coleman (2002). See also Federal Reserve (2005).

[16] For a detailed reference, see Federal Reserve (2006c).

Table 3.2: Brief definition of cap categories.

Cap Category	Chosen by institutions that	Requirements
High *Above Average* *Average*	Regularly incur daylight overdrafts in excess of 40 percent of their capital. They are referred to as "self-assessed".	Self-assessment of own creditworthiness, intraday funds management, customer credit and operating controls and contingency procedures. Each institution's board of directors must review the self-assessment and recommend a cap category at least once in each twelve-month period.
De minimis	Incur relatively small daylight overdrafts.	Board-of-directors resolution approving use of daylight credit up to de minimis cap at least once in each 12-month period.
Exempt-from-filing	Only rarely incur daylight overdrafts.	Exempt from performing self-assessments and filing board-of-directors resolutions.
Zero	Do not want to incur daylight overdrafts and associated fees. A Reserve Bank may assign a zero cap to institutions that may pose special risks.	

Table 3.3: Net debit cap multiples of capital measure.

Cap Category	Single Day	Two-week Average
High	2.25	1.50
Above Average	1.875	1.125
Average	1.125	0.75
De minimis	0.4	0.4
Exempt-from-filing[a]	min{\$10 million,0.2}	min{\$10 million,0.2}
Zero	0.0	0.0

[a]The net debit cap for the *exempt-from-filing* category is equal to the lesser of \$10 million or 0.20 multiplied by a capital measure.

Table 3.4: Daylight Overdraft Charge

Example of Daylight Overdraft Charge Calculation[a]
Policy parameters
Official Fedwire day = 21.5 hours
Deductible percentage of capital = 10%
Rate charged for overdrafts = 36 basis points (annual rate)
Institution's parameters
Risk-based capital = $50 million
Sum of end-of-minute overdrafts for one day = $4 billion
Daily Charge calculation
Effective daily rate = .0036 x (21.5/24) x (1/360) = .0000089
Average overdraft = $4,000,000,000 / 1291 minutes = $3,098,373
Gross overdraft charge = $3,098,373 x .0000089 = $27.58
Effective daily rate for deductible = .0036 x (10/24) x (1/360) = .0000042
Value of the deductible = .10 x $50,000,000 x .0000042 = $21.00
Overdraft charge = 27.58 - 21.00 = **$6.58**

[a] Federal Reserve (2006d).

References

Bech, M. L. and B. Hobijn (2006, September). Technology Diffusion within Central Banking: The Case of Real-Time Gross Settlement. *Federal Reserve Bank of New York Staff Reports* (260).

Bennett, P. and S. Peristiani (2002, May). Are U.S. Reserve Requirements Still Binding? *Economic Policy Review 8*(1), 1–16.

Bernanke, B. S. (2008, 10 January). Financial Markets, the Economic Outlook, and Monetary Policy. http://www.federalreserve.gov/ newsevents/ speech/ bernanke20080110a.htm.

Brunnermeier, M. K. (2008). Deciphering the 20078 Liquidity and Credit Crunch. *Mimeo, Princeton University*. http://www.princeton.edu/ ~markus/ index.htm.

Coleman, S. P. (2002). The Evolution of the Federal Reserve's Intraday Credit Policy. *Federal Reserve Bulletin*, 67–84.

Federal Reserve (2001, May). Policy Statement on Payments System Risk (Docket No. R-1107). http://www.federalreserve.gov/ boarddocs/ press/ BoardActs/ 2001/ 20010530/ attachment.pdf.

Federal Reserve (2005). Overview of The Federal Reserve's Payments System Risk Policy on Daylight Credit. http://www.federalreserve.gov/ PaymentSystems/ PSR/ overview.pdf.

Federal Reserve (2006a). Account Management Guide.

Federal Reserve (2006b). Consultation Paper on Intraday Liquidity Management and Payment System Risk Policy (Docket No. OP-1257). http://www.federalreserve.gov/ boardDocs/ press/ other/ 2006/ 200606152/ attachment.pdf.

Federal Reserve (2006c). Federal Reserve Policy on Payments System Risk. http://www.federalreserve.gov/ PaymentSystems/ PSR/ policy.pdf.

Federal Reserve (2006d). Guide to the Federal Reserve's Payments System Risk Policy on Daylight Credit. http://www.federalreserve.gov/ paymentsystems/ psr/ guide.pdf.

Gilbert, A. M., D. Hunt, and K. C. Winch (1997). Creating an integrated payment system: The evolution of Fedwire. *Federal Reserve Bank of New York Economic Policy Review*, 1–7.

Greenlaw, D., J. Hatzius, A. K. Kashyap, and H. S. Shin (2008). Leveraged Losses: Lessons from the Mortgage Market Meltdown. *Mimeo, Princeton University*. http://www.chicagogsb.edu/usmpf/docs/usmpf2008confdraft.pdf.

Martin, A. (2005, First Quarter). Recent Evolution of Large-Value Payment Systems: Balancing Liquidity and Risk. *Federal Reserve Bank of Kansas City Economic Review*, 33–57.

McAndrews, J. J. and S. M. Potter (2002). Liquidity Effects of the Events of September 11, 2001. *Federal Reserve Bank of New York Economic Policy Review*, 59–79.

McAndrews, J. J. and S. Rajan (2000). The Timing and Funding of Fedwire Funds Transfers. *Federal Reserve Bank of New York Economic Policy Review*, 17–32.

Milgrom, P. and J. Roberts (1994). Comparing Equilibria. *The American Economic Review 84*(3), 441–459.

Tarski, A. (1955). A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics 5*(2), 285–309.

Topkis, D. M. (1978, March - April). Minimizing a Submodular Function on a Lattice. *Operations Research 26*(2), 305–321.

Zhou, R. (2000, Third Quarter). Understanding intraday credit in large-value payment systems. *Federal Reserve Bank of Chicago Economic Perspectives 24*, 29–44.

CHAPTER 4

Liquidity and Congestion

THIS CHAPTER studies the relationship between the arrival of potential investors and market liquidity in a search-based model of asset trading. The entry of investors into a specific market causes two contradictory effects. First, it reduces trading costs, which then attracts new investors (thick market externality effect). But secondly, as investors concentrate on one side of the market, the market becomes "congested", decreasing the returns to participating in this market and discouraging new investors from entering (*congestion effect*). The equilibrium level of market liquidity depends on which of the two effects dominates. When congestion is the leading effect, some interesting results arise. In particular, we find that diminishing trading costs in our market can deteriorate liquidity and reduce welfare.

4.1 Introduction

Liquidity is sometimes defined as a coordination phenomenon. In financial markets, as investors move into a specific market they facilitate trade for all investors by reducing the cost of participating in this market. At the same time, easier trade and lower trading costs attract potential investors. There is a thick market externality where new investors provide market liquidity and market liquidity attracts new investors. However, as investor prefer to join one side of a market, i.e. as they become buyers or sellers, this side of the market becomes "congested", hindering trade. Congestion then discourages investors from entering this market.

One-sided markets arise during financial booms and, more drastically, during market crashes. When a market is in distress, liquidity typically vanishes playing a key role in the build-up of one-sided markets. The study of liquidity in one-sided markets is thus vital to understand the response of financial systems to the threat of market disruptions. Recent episodes of market distress include the LTCM crisis[1] in 1998, the September 11, 2001, events[2] and the turbulence in credit markets[3] during the summer of 2007.

In this paper we present an alternative view of market liquidity. The main difference with the previous literature is that we consider not only a thick market externality but also a *congestion effect*. In our model, the arrival of new investors causes two opposite effects. First, it diminishes transaction costs and eases trade, which attracts potential investors. But

[1] For an analysis of the events surrounding the market turbulence in autumn 1998, see BIS (1999) and IMF (1998).

[2] Cohen and Remolona (2001) presents a summary of the September 11, 2001 episode in global financial markets. Also, McAndrews and Potter (2002) gives a detailed account of the consequences of the September 11, 2001, events on the US payment system and of the actions of the Federal Reserve System to provide liquidity to the financial system.

[3] See Fender and Hördahl (2007) for an overview of the key events over the period from end-May to end-August 2007. Greenlaw et al. (2008) analyzes relevant credit market events since August 2007 with special emphasis on their main policy implications. See also Brunnermeier (2008).

secondly, as investors concentrate on one side of the market, trade becomes more difficult, reducing the returns to participating in this market and discouraging potential investors from entering. Market liquidity thus results from the tradeoff between thick market externalities and a congestion effect.

We assume an infinite-horizon steady-state market where agents can invest in one asset which can be traded only bilaterally. In this market, investors cannot trade instantaneously but it takes some time to find a trading partner resulting in opportunity and other costs. Once an investor buys the asset, he holds it until his preference for the ownership change and he prefers to liquidate the investment and exit the market. To model the search process we adopt the framework introduced in Vayanos and Wang (2007). In our setting though, investors are heterogeneous in their investment opportunities in the sense that some investors have access to better investment options than others.

We compute explicitly the unique equilibrium allocations and the price at which investors trade with each other and show how they depend on the flow of new investors entering the market. Prices negotiated between investors are higher in the flow of potential investors. However, investors' entry decision is endogenous and thus depends on market, asset and investors characteristics. A change in investors' search abilities, for instance, affects both the rate of meetings between trading partners and the flow of investors entering the market, which then determines the distribution of potential partners with whom they can meet.

Moreover, the equilibrium flow of investors arises from a tradeoff between thick market complementarities and a congestion effect. When congestion is the dominating effect some interesting results come to light. First, reducing market frictions can *decrease* market attractiveness. Under some cases, one-sided markets can develop. A regulatory reform or the introduction of a technological advance, such as a new electronic trading system, can induce an adverse effect on the distribution of investors during upswings. Specifically, it would allow the few sellers present in the market to exit at faster rates leading to an even more unbalanced distribution of investors. Congestion then intensifies as the market becomes more one-sided, discouraging potential investors and thus dampening down the attractiveness of this market.

Second, *diminishing market frictions can deteriorate market liquidity and reduce welfare.* The reason for this counterintuitive result is the following. In a one-sided market with more sellers than buyers, for example during a fire sale, introducing a measure that improves the efficiency of the search process makes it easier for one of the few buyers present in the market to acquire the asset. But when the buyer purchases the asset (and a seller exits), the proportion of buyers to sellers falls further and the market becomes more one-sided. As investors cluster on the sell-side of this market, buyers gain a more favorable position in the bargaining process and try to lower the price they pay to acquire the asset. Reducing market frictions in a distressed market thus magnifies the effect of congestion and results in a lower asset price (a higher price discount) and ultimately in a less liquid market. Investors who hold this asset and those trying to sell it are clearly worse-off as the market becomes more one-sided, leading to a decrease in overall welfare. From this point of view, this paper provides an example of the Theory of the Second Best. Improving the efficiency of the search process, when there are other imperfections in the market such as the ones arising from the congestion effect, is not necessarily welfare enhancing.

Third, market illiquidity measured by the price discount can increase while trading volume rises. Reducing search frictions during downswings amplifies the effect of congestion, resulting in a higher price discount and in a less liquid market. But a more efficient search process also increases the frequency of meetings between the investors already present in the market. More frequent meetings then translates into a higher trading volume. Thereby, a measure intended to shorten the waiting times needed to locate a trading partner in a market experiencing distressed selling can cause both higher price discount and higher trading volume. This third result joins the discussion on the measurement of the effect of liquidity on asset prices and shows how alternative measures capture different dimensions of market liquidity.

The outline of the paper is as follows. In the next section, we discuss the related literature. We introduce a theoretical framework to examine the relationship between market liquidity and the arrival of potential investors to this market in Section 4.3. Section 4.4 determines the population of investors, their expected utilities and the price of the asset, taking as given investors' decision to enter the market. Then, Section 4.5 endogenizes the entering rule and characterizes the study of the unique market equilibrium. Market liquidity and welfare are discussed in Section 4.6. Finally, Section 4.7 concludes. Some proofs and additional results are presented in the appendices.

4.2 Related Literature

The notion of thick market complementarity is clearly captured in Diamond (1982a). He considers an economy where islanders face production opportunities and decide whether to remain unemployed or to climb a palm tree and retrieve coconuts. Trees differ in their heights (the cost of production). Islanders only climb trees shorter than a certain height and they cannot consume the coconuts they pick. They need to search for a trade to swap the coconuts. The likelihood of meeting a trading partner in this economy increases in the number of potential traders available. This key feature constitutes the basis of the strategic complementarity in Diamond's model. This is highlighted in Cooper and John (1988), where they discuss the economic relevance of strategic complementarities in agents' payoffs and explain how they can lead to coordination failures. A related argument is presented in

Milgrom and Roberts (1990). They show the Diamond-type search model is a supermodular[4] game, where more production or participation activity by some islanders raises the returns to increased levels of activity by others.

Building on strategic complementarities Brunnermeier and Pedersen (2007) and Gromb and Vayanos (2002) analyze the link between capital and market liquidity. Also, Pagano (1989) focuses on the feedback loop between trading volume and liquidity to study concentration and fragmentation of trade across markets. In Dow (2004), multiple equilibria with di-fferent degrees of market liquidity result from informational asymmetries. Plantin (2004) assumes investors can learn privately about an issuer's credit quality by holding an asset. This "learning by trading" also creates a thick market externality. From a broad perspective, this literature studies liquidity as a self-fulfilling phenomenon where both liquid and illiquid market equilibria may arise. Illiquid markets are thus a consequence of a coordination failure.

Our paper is also related to the search literature. The economics of search have their roots in Phelps (1972). Search-theoretic models such as the frameworks introduced in labour markets[5] by Diamond (1982a), Diamond (1982b), Mortensen (1982) and Pissarides (1985) have been broadly used in different areas of economics. In asset pricing[6], Duffie, Gârleanu and Pedersen introduce search and bargaining in models of asset market equilibrium to study the impact of these sources of illiquidity on asset prices. This paper is related to Duffie et al. (2005), which presents a theory of asset pricing and marketmaking in over-the-counter markets with search-based inefficiencies. They conclude that risk neutral investors receive narrower bid-ask spreads if they have easier access to other investors and marketmakers. Similarly to Duffie et al. (2005) we consider risk-neutral agents who can only invest in

[4]In the unidimensional case, a supermodular game is a game exhibiting strategic complementarities in which each agent's strategy set is partially ordered. See Topkis (1979) and Cooper (1999) for a formal definition.

[5]See Pissarides (2001) for a review of the literature on search in labour markets.

[6]For an excellent review on liquidity and asset prices, see Amihud et al. (2005).

one asset. In our model though, investors can only trade with other investors and our focus, rather than on liquidity and marketmaking, lies on the endogenous relationship between market liquidity and the arrival of potential investors to this market.

Duffie et al. (2007) extends their setting to incorporate risk aversion and risk limits and finds that, under certain conditions, search frictions as well as risk aversion, volatility and hedging demand increase the illiquidity discount. Lagos and Rocheteau (2007) also generalizes Duffie et al. (2005) to allow for general preferences, unrestricted long positions, idiosyncratic and aggregate uncertainty and entry of dealers. Our paper shares with theirs the existence of strategic complementarities and an endogenous entry decision. To define the entry of dealers, Lagos and Rocheteau (2007) specify that the contact rate between investors and dealers increases sublinearly in the number of dealers. In our framework, entry is the result of a decision problem where investors compare the benefits of this market to their best investment opportunities.

Weill (2007) and Vayanos and Wang (2007) extend the framework of Duffie, Gârleanu and Pedersen to allow investors to trade multiple assets[7]. They show that search frictions lead to cross-sectional variation in asset returns due to illiquidity differences. In Vayanos and Wang (2007) investors are heterogeneous in their trading horizons while in Weill (2007) investors are homogeneous, but there are differences in the assets' number of tradable shares. From a methodological point of view, our paper is closely related to Vayanos and Wang (2007). The main difference with their work is that we consider only one asset and focus on the analysis of the liquidity in the market for this asset rather than on the liquidity across two assets.

Our paper is close in spirit to Huang and Wang (2007). They also find that decreasing

[7]See also Vayanos and Weill (2007) for an application to the on-the-run phenomenon, by which recently issued bonds have higher prices than older ones with the same cash flows. They develop a multi-asset model where both the spot market and the repo market operate through search.

market frictions can diminish the level of market liquidity. However, their framework and the general mechanism that yields this result clearly differ from ours. Rather than a search-based model, they consider a centralized market where exogenous transaction costs take the form of participation costs. Agents can pay an ex-ante cost to trade constantly (and become market makers) or pay a spot cost to trade after observing their trading needs. Huang and Wang (2007) argues that, when there is insufficient supply of liquidity, lowering the cost to enter on the spot can decrease welfare because it reduces investors' incentives to become market makers. In our model, market liquidity results from a tradeoff between thick market externalities and congestion effects. We show that, when the congestion effect dominates, the market becomes one-sided and improving the efficiency of the search process can diminish market liquidity because it discourages agents from investing into our market.

This paper also relates to the literature on asset pricing with exogenous trading costs studied in Amihud and Mendelson (1986), Vayanos (1998) and Acharya and Pedersen (2005), among others. We complement this literature by endogenizing transaction costs.

4.3 The Model

Time is continuous and goes from zero to infinity. There is only one asset traded in the market with a total supply S. This asset pays a dividend flow d.

Consider risk-neutral agents, whom we will refer to as investors. By assuming risk neutrality, we aim to study the concentration of liquidity in a specific market without reference to investors' shifts in their attitudes towards risk. Investors are infinitely lived and have time preferences determined by a constant discount rate equal to $r > 0$. At some random time, investors decide to enter the market and aim to buy one unit of the asset. They become *buyers-to-be.* Once they purchase the asset, buyers-to-be become *non-searcher investors.* Non-searchers hold the asset and enjoy the full value d of its dividend flow until they receive a liquidity shock which makes them want to liquidate their portfolio and leave the market. We assume liquidity shocks arrive with a Poisson rate γ and reduce investors' valuation to a lower level $d - x$ of flow utility, where $x > 0$ captures the notion of a liquidity shock to the investors, for example, a sudden need for cash or the arrival of a good investment opportunity in another market. x could also be understood as the holding cost borne by the investor who receives a liquidity shock and is aiming to exit the market. At that time, non-searcher investors become *sellers-to-be* and seek to sell[8]. Upon selling, investors exit the market and join the initial group of outside investors.

The flow of investors entering the economy is defined by a function f. Investors are heterogeneous in their investment opportunities κ, i.e. we consider they differ on their outside options as some investors enjoy better investment possibilities than others. We assume f is a continuous and strictly positive function of the investor's investment opportunity class κ, such that the total flow of investors entering the economy is given by $\int_{\underline{\kappa}}^{\overline{\kappa}} f(\kappa)d\kappa$, where $[\underline{\kappa}, \overline{\kappa}]$ is the support of $f(\kappa)$. Only a fraction $\nu(\kappa)$ of the flow of investors entering the economy decides to invest in this market. At any point in time there is a non-negative flow of every class of investor from the outside investors' group into the market, and hence the total flow of investors entering the market is defined by $g = \int_{\underline{\kappa}}^{\overline{\kappa}} \nu(\kappa)f(\kappa)d\kappa$.

We assume markets operate through search, with buyers and sellers matched randomly over time in pairs. Search is characteristic of over-the-counter markets where investors need to locate trading partners and then bargain over prices. There is a cost associated to this search process. In a market where it is more likely to find a counterpart in a short time,

[8] Investors are risk neutral and thus have linear utility over the dividend flow d. Consequently, they optimally prefer to hold a maximum long position in the asset (which we can normalize to 1) or zero units of the asset (once they seek to exit the market).

the search cost is smaller and liquidity, measured by search costs, is higher. But we could think of a broader interpretation of the search friction. In a centralized market, it represents the cost of being forced to trade with an outside investor who does not understand the full value of the asset and requires an additional compensation for trading. These investors only buy the asset at a discount and sell it at a premium. This transaction cost decreases in the abundance of investors. In the market of a frequently traded asset, it is less likely that it is necessary to trade with an outside investor who "mis-values" the asset and hence the transaction cost linked to this asset is smaller and its liquidity higher. In this paper, we use the first intuition because of its more transparent interpretation.

We adopt the search framework presented in Vayanos and Wang (2007). To define the search process, we first need to describe the rate at which investors willing to buy meet those willing to sell and once they meet we need to specify how the asset price is determined. The ease in finding a trading partner depends on the availability of potential partners. Let us consider that an investor seeking to buy or sell meets other investors according to a Poisson process with a fixed intensity. Thus, for each investor the arrival of a trading partner occurs at a Poisson rate proportional to the measure of the partner's group. Denote by η_b the measure of buyers-to-be and by η_s the measure of investors seeking to sell (sellers-to-be). Then, a buyer-to-be meets sellers-to-be with a Poisson intensity $\lambda\eta_s$ and a seller-to-be meets buyers-to-be at a rate $\lambda\eta_b$, where λ measures the efficiency of the search and a high λ represents an efficient search process. The overall flow of meetings[9] between trading partners is then given by $\lambda\eta_b\eta_s$.

Once investors meet they bargain over the price p of the asset. These meetings always result in trade as Proposition 4.5 shows. For simplicity we assume that either the investor

[9]See Duffie and Sun (2007) for a formal proof of this result. This application of the exact law of large numbers for random search and matching has previously been used in Duffie et al. (2005), Duffie et al. (2007) and Vayanos and Wang (2007) among others.

willing to buy or the one willing to sell is chosen randomly to make a take-it-or-leave-it offer to his trading partner. Denoting by $\frac{z}{1+z}$ the probability of the buyer-to-be being selected to make the offer and thus by $\frac{1}{1+z}$ the probability that the seller-to-be makes the offer, $z \in (0, \infty)$ captures the buyer's-to-be bargaining power.

Figure 4.1 describes our market, specifying the different types of investors and the flows between types. η_0 denotes the measure of non-searcher investors.

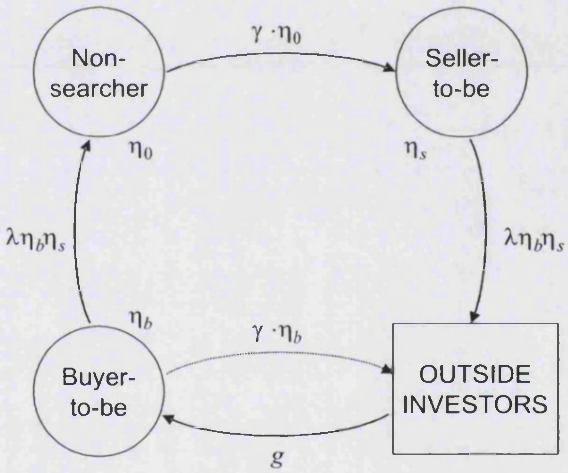

Figure 4.1: An outside investor enters the market and becomes a buyer-to-be aiming to meet a seller-to-be. If he suffers a liquidity shock before meeting a trading partner, he exits the market. On the contrary, if he meets a seller-to-be, he bargains over the price, buys the asset (pays p) and becomes a non-searcher. He holds the asset until he receives a liquidity shock. At that time, he becomes a seller-to-be seeking a buyer-to-be. When he meets a buyer-to-be, he bargains over the price, sells the asset (receives p) and exits the market returning to the group of outside investors.

4.4 Analysis

In this section we first solve for the steady-state measure of every type of investor in the market. Next, in Subsection 4.4.2, we describe investors' flow utilities. We show in Subsection 4.4.3 that every meeting between trading partners results in trade and we discuss thick market externalities in Subsection 4.4.4.

4.4.1 Measure of Investors

In this subsection we determine the measure of buyers-to-be (η_b), non-searcher investors (η_0) and sellers-to-be (η_s). Although investors are heterogeneous in their investment opportunities κ, once they enter the market their class does not alter their behavior in this market. Investors develop sudden needs for cash at the same Poisson rate γ, independently of their outside investment opportunities κ. In consequence, we do not need to consider the distribution of investment opportunities within each population but the aggregate measure of buyers-to-be, non-searcher investors and sellers-to-be. This assumption could be generalized by considering γ a function of the outside option κ. The analysis would be similar but the notation more complicated, as we would need to take into account the distribution of investment opportunities κ within each group of investors rather than the aggregate measures[10].

In equilibrium, the market needs to clear and thus the supply of the asset equals the measure of investors holding the asset, each of whom holds one unit of the asset. Specifically, the sum of the measures of non-searchers and sellers-to-be is equal to the total supply of

[10]See Section 3 in Vayanos and Wang (2007) for a particular case.

the asset:

$$\eta_0 + \eta_s = S \quad \Rightarrow \quad \eta_s = S - \eta_0 \tag{4.1}$$

In a steady state, the inflow of investors joining a group matches the outflow such that the rate of change of the group's population is zero. The inflow and outflow of the different types of investors are summarized in Figure 4.1. Let us first consider the non-searcher investors. In this case, inflows are given by the buyers-to-be who meet a trading partner and buy the asset ($\lambda\eta_b\eta_s$), while non-searchers receiving a liquidity shock ($\gamma\eta_0$) constitute the outflow. Setting inflow equal to outflow and using equation (4.1) yields:

$$\eta_b = \frac{\gamma}{\lambda}\frac{\eta_0}{S - \eta_0} \tag{4.2}$$

We now analyze the population of buyers-to-be. The flows of investors coming from the outside group are defined by g. The outflow is comprised of the buyers-to-be who receive a liquidity shock before meeting a trading partner ($\gamma\eta_b$) and of those who meet sellers-to-be and buy the asset ($\lambda\eta_b\eta_s$). Then,

$$g = \gamma\eta_b + \lambda\eta_b\eta_s$$

Using equations (4.1) and (4.2) we can rewrite the previous equation as:

$$g = \gamma\left(1 + \frac{\gamma}{\lambda}\frac{1}{S - \eta_0}\right)\eta_0 \tag{4.3}$$

Equation (4.3) determines η_0 as a function of g. Then, substituting η_0 in equations (4.1) and (4.2) specifies η_s and η_b respectively. Let us first assume the flow of investors entering the market g is constant. We generalize our results in Subsection 4.5.1.

Proposition 4.1. *Given g constant, there is a unique solution to the system (4.1) - (4.3) given by:*

$$\eta_0 = \frac{1}{2\gamma}A \tag{4.4}$$

$$\eta_s = S - \frac{1}{2\gamma}A \tag{4.5}$$

$$\eta_b = \frac{\gamma}{\lambda}\frac{A}{2\gamma S - A} \tag{4.6}$$

where $A = (g + \gamma S + \frac{\gamma^2}{\lambda}) - \sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}$.

The proof is presented in Appendix 4.8.1. It is interesting to note how the different measures of investors respond to changes in the parameters of our model. For instance, as the flow of investors g entering the market rises, the measure of investors willing to buy (buyers-to-be) and of those passively holding the asset (non-searchers) increase. However, given that there are more investors seeking to buy the asset, it is now easier for a seller-to-be to find a trading partner and hence the measure of investors seeking to sell falls. This is summarized in Proposition 4.2 and proven in Appendix 4.8.1.

Proposition 4.2. *The measure of buyers-to-be and non-searcher investors is increasing in g* ($\frac{\partial \eta_b}{\partial g}, \frac{\partial \eta_0}{\partial g} > 0$) *while the measure of sellers-to-be decreases in g* ($\frac{\partial \eta_s}{\partial g} < 0$).

Given the measures of investors η_b seeking to buy and those η_s seeking to sell, the efficiency of the search process λ defines the overall flow of meetings (and transactions, according to Proposition 4.5) in our market. However, the measures of the different types of investors also depend on the efficiency of the search process. In particular, for the same level of investors entering the market, if the search process is more efficient, there will be a lower measure of investors "waiting" to meet a potential seller ($\frac{\partial \eta_b}{\partial \lambda} < 0$). Thus, outside investors, who enter the market, meet a trading partner and become non-searcher investors at a faster rate if the search process is more efficient ($\frac{\partial \eta_0}{\partial \lambda} > 0$). A proportion of non-searcher investors then joins the pool of sellers-to-be and hence there is a higher flow of investors coming from the non-searchers to the group of sellers-to-be. And, although there are more inflows of investors and less investors seeking to buy, if the search process is more efficient, the measure of sellers-to-be "waiting" to sell is reduced ($\frac{\partial \eta_s}{\partial \lambda} < 0$). Proposition 4.3 presents these results:

Proposition 4.3. *Given g constant, the measure of buyers-to-be and sellers-to-be is decreasing in* λ ($\frac{\partial \eta_b}{\partial \lambda}, \frac{\partial \eta_s}{\partial \lambda} < 0$) *while the measure of non-searcher investors increases in* λ ($\frac{\partial \eta_0}{\partial \lambda} > 0$).

The proof is in Appendix 4.8.1.

4.4.2 Expected Utilities and Price

We now determine the expected utility of the buyers-to-be (v_b), the non-searcher investors (v_0) and the sellers-to-be (v_s), as well as the price p. Investors exit this market because of a need for cash. We assume that the expected utility of outside investors is zero. Once they are out of the market, investors have different investment opportunities and decide where to invest next. They could even choose to re-enter this market again.

To derive the expected utility of every type of investor we analyze the possible transitions between types. For example, a buyer-to-be can leave the market if he receives a liquidity shock, remain a potential buyer or meet a seller-to-be and become a non-searcher. This is summarized in Figure 4.2:

The utility flow rv_b of buyers-to-be is thus equal to the expected flow of exiting the

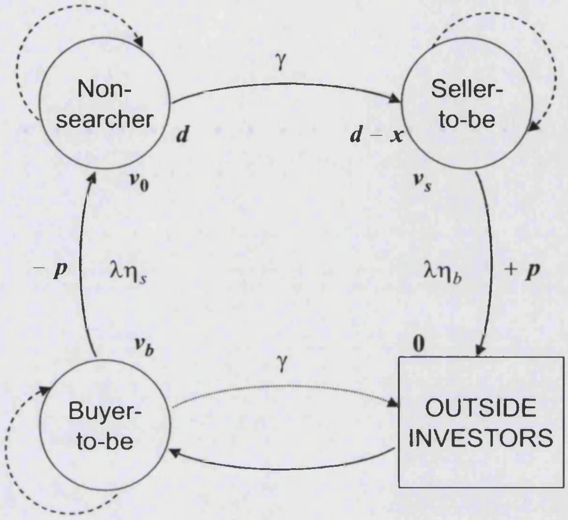

Figure 4.2: Groups of investors and transitions between groups.

market $((0 - v_b)\gamma)$ and becoming an outside investor plus the expected flow derived from meeting a trading partner seeking to sell (which occurs at rate $\lambda\eta_s$), buying the asset (paying p) and becoming a non-searcher investor $(\lambda\eta_s(v_0 - v_b - p))$. Then,

$$rv_b = -\gamma v_b + \lambda\eta_s(v_0 - v_b - p) \tag{4.7}$$

Non-searcher investors can either remain non-searchers enjoying the full value d of the asset's dividend flow or receive a liquidity shock with probability γ and become a seller-to-be. In this case, the flow of utility of being a non-searcher is

$$rv_0 = d + \gamma(v_s - v_0) \tag{4.8}$$

Sellers-to-be exit the market as soon as they meet a trading partner, i.e., with intensity

$\lambda\eta_b$ they sell the asset (receiving p) and become outside investors with zero expected utility. Meanwhile, they enjoy a low level $d - x$ of utility. Thus,

$$rv_s = (d - x) + \lambda\eta_b(p + 0 - v_s) \tag{4.9}$$

The asset price is determined by bilateral bargaining between a buyer-to-be and a seller-to-be. We have assumed that with probability $\frac{z}{1+z}$ the buyer-to-be makes a take-it-or-leave-it offer to his trading partner and offers him his reservation value v_s. With probability $\frac{1}{1+z}$, the seller-to-be is chosen to offer the buyer-to-be his reservation value $v_0 - v_b$. As a result,

$$p = \frac{z}{1+z}v_s + \frac{1}{1+z}(v_0 - v_b) \tag{4.10}$$

where z measures the buyer's-to-be bargaining power which we treat as exogenous. Proposition 4.4 summarizes this subsection's main result. The proof is in Appendix 4.8.1.

Proposition 4.4. *Given g constant, the system of equations (4.7)-(4.10) has a unique solution given by:*

$$\begin{aligned}
v_b &= k\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\frac{\lambda\eta_s z}{r+\gamma} && (4.11)\\
v_0 &= \frac{d}{r} - k\Big(\frac{x}{r} + \frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\Big)\frac{\gamma}{r+\gamma} && (4.12)\\
v_s &= \frac{d}{r} - k\Big(\frac{x}{r} + \frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\Big) && (4.13)\\
p &= \frac{d}{r} - k\frac{x}{r} && (4.14)
\end{aligned}$$

where $k = \dfrac{(r+\gamma+\lambda\eta_s)z+\gamma}{(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)}$.

The price of the asset as given by equation (4.14) is thus equal to the present value of all future dividend flows d, discounted at the rate r, minus a price discount due to illiquidity. The second term is the product of present value of the holding cost x borne by investors seeking to exit the market and a function k. $k \in (0,1)$ measures the severity or intensity of the illiquidity discount[11].

It is interesting to highlight that the asset price will be higher when fundamentals are stronger (i.e. if the asset pays a higher dividend flow d) and whenever the demand for the asset increases ($\frac{\partial p}{\partial d}, \frac{\partial p}{\partial \eta_b} > 0$). On the contrary, the price decreases with investors trying to sell the asset and in the buyer's-to-be bargaining power ($\frac{\partial p}{\partial z}, \frac{\partial p}{\partial \eta_s} < 0$). If during the bargaining process the buyer-to-be holds a more favorable position, he would try to lower the price paid to acquire the asset. The proof of this set of comparative statics is presented in Appendix 4.8.2.

4.4.3 Trade among Investors

In this subsection we prove a result we have assumed so far in our analysis:

Proposition 4.5. *All meetings between buyers-to-be and sellers-to-be result in trade.*

Proof. Trade between buyers-to-be and sellers-to-be occurs if the gain from trade is strictly positive, i.e., if the buyers'-to-be reservation value $v_0 - v_b$ exceeds the sellers'-to-be reservation value v_s. Let us see if $(v_0 - v_b) - v_s > 0$. Subtracting equations (4.13) and (4.11) from (4.12), we get:

$$(v_0 - v_b) - v_s = \frac{x(1+z)}{(r+\gamma)(1+z) + \lambda\eta_s z + \lambda\eta_b}$$

[11] See Section 4.6 for a discussion of market liquidity.

which is always strictly greater than zero since $x, r, \gamma, \lambda, \eta_s, \eta_b, z > 0$.

Therefore, once investors meet, trade among partners always occurs. □

4.4.4 Thick Market Externality

In financial markets, thick market externalities arise when the gains from investing in a market depend on the number of investors who decide to come to the market. In this case, the more traders move into a market, the easier become the transactions and as a result the bigger is the gain derived from participating in this market. In our framework, the price of the asset is higher as the flow of investors moving into the market increases[12] ($\frac{\partial p}{\partial g} > 0$). As investors arrive to this market, the costs of search are reduced and hence the illiquidity discount is diminished. This increases the returns to investing in this market, making it more attractive to new investors. To understand how higher participation may encourage further participation we need to endogenize investors' entry decisions.

4.5 Equilibrium

In our setting, market equilibrium is determined by the fraction of investors entering the market, a measure of each group of investors, their expected utilities and the price of the asset. We center our study on the steady-state analysis. In the previous section we take as given investors' decision to enter the market, and we now endogenize the entering rule in Subsection 4.5.1. A formal definition of the market equilibrium is then presented in Subsection 4.5.2. Subsection 4.5.3 introduces the congestion effect.

[12] The proof is presented in Appendix 4.8.2.

4.5.1 Entering Rule

In this subsection we endogenize the entering rule. In our framework, outside investors can choose between entering the market, which we will refer to as *our market*, and investing in an alternative market. Investors are heterogeneous in their outside investment opportunities κ, i.e. each class of investor has access to different investment opportunities. However, once they enter our market, their type no longer influences their decisions in the sense that every buyer-to-be, for instance, enjoys the same expected utility independently of his original outside opportunity. Interestingly, a buyer's-to-be expected utility does depend on the flow of investors who entered this market before him.

Let us refer to the investor who is deciding between moving or not into our market as the marginal investor. And, let us denote by κ' and by $v_{alt}(\kappa')$, respectively, the best outside investment opportunity of the marginal investor and his expected utility from investing in that alternative market. For simplicity, we assume $v_{alt}(\kappa') = \kappa'$, such that an investor with a better outside option (higher κ) enjoys a higher level of expected utility.

When an investor faces the decision to choose a market, he prefers to enter our market if the expected utility v_b of being a buyer-to-be in this market is higher than the expected utility v_{alt} derived from his best outside option. Then, if our market represents the best opportunity for the marginal investor, it is also preferred by any other investor with a worse investment opportunity, i.e. any investor whose type $\kappa < \kappa'$ moves into our market too. As a result, when our market is chosen by a marginal investor with a high type, a high flow of investors enters our market. A high flow of investors implies an increase in the measure of buyers-to-be, which then affects the expected utility of being a buyer-to-be. Thus, even though each investor's type does not alter his expected utility, the type of the last investor who enters does. The type of this last investor defines the total flow who prefers our market and hence determines how concentrated the population of buyers-to-be is.

Let us define the fraction $\nu(\kappa)$ of investors with outside investment opportunity κ who enters the market as follows:

$$\nu(\kappa) = \begin{cases} 0 & \text{if} \quad \kappa > \kappa' \\ [0,1] & \text{if} \quad \kappa = \kappa' \\ 1 & \text{if} \quad \kappa < \kappa' \end{cases}$$

where $1 - \nu(\kappa)$ represents the fraction of investors with outside option κ who invests in alternative markets. The total flow of investors moving into our market is thus given by: $g(\kappa') = \int_{\underline{\kappa}}^{\overline{\kappa}} \nu(\kappa) f(\kappa) d\kappa$, where f defines the total flow of investors entering the economy. In equilibrium, as we discuss in more detail in the next subsection, the total flow g^* depends on the equilibrium fraction of investors ν^* entering our market. But the equilibrium fraction of investors is determined by the marginal investor who is indifferent between our market and his best outside option. We refer to this investor as the indifferent investor. For the indifferent investor, the expected utility of being a buyer-to-be equals the expected utility of his best outside option:

$$v_b\left(g^* = \int_{\underline{\kappa}}^{\overline{\kappa}} \nu^*(\kappa) f(\kappa) d\kappa\right) = v_{alt}(\kappa^*) \tag{4.15}$$

Before we proceed, let us introduce the formal definition of market equilibrium.

4.5.2 Equilibrium Definition and Characterization

Definition 4.1. *A* **market equilibrium** *consists of a fraction $\nu(\kappa)$ of investors entering the market, measures (η_s, η_b, η_0) of investors and expected utilities and prices (v_b, v_0, v_s, p) such that:*

- *$(\eta_s^*, \eta_b^*, \eta_0^*)$ solve the market-clearing condition and inflow-outflow equations given by the system (4.1) - (4.3),*
- *$(v_b^*, v_0^*, v_s^*, p^*)$ solve the flow-value equations for the expected utilities and the pricing condition given by the system (4.7) - (4.10),*
- *$\nu^*(\kappa)$ solves the entering condition given by the system (4.15).*

To analyze the equilibria in this market, we need to solve for the fixed points of the system of equations (4.1) - (4.3), (4.7) - (4.10) and (4.15). There are two types of possible scenarios depending on the behavior of the expected utility v_{alt} of investing in an alternative market and the expected utility v_b of being a buyer-to-be in our market. There is an equilibrium where all investors clearly prefer one market (either all enter or no one enters) or an equilibrium where a fraction of investors is better off by investing in our market while others prefer not to enter. Theorem 4.1 summarizes a key result:

Theorem 4.1. *There is a unique market equilibrium.*

The proof is in Appendix 4.8.3. To gain some intuition for this result, let us introduce Figure 4.3. Figure 4.3 represents the expected utility v_{alt} of investing in an alternative market and the expected utility of being a buyer-to-be of the marginal investor, i.e. the one deciding whether or not to enter our market. Consider, for example, the marginal investor with outside investment opportunity κ_1'. He compares the utility of his outside option, $v_{alt}(\kappa_1') = \kappa_1'$, to the utility of being a buyer-to-be, $v_b(g(\kappa_1'))$, given that investors with outside opportunities $\kappa < \kappa_1'$ have already entered our market. He enters since $v_b(g(\kappa_1')) > v_{alt}(\kappa_1')$, as shown in Figure 4.3. Now, let us focus on the marginal investor with investment opportunity κ_2'. The expected utility of being a buyer-to-be has decreased because now all investors with $\kappa < \kappa_2'$ are in the market. Still he is better-off by moving into our market. Suppose marginal investor κ^* is now facing the entry decision. For him, $v_b(g(\kappa^*)) = v_{alt}(\kappa^*)$ and he is indifferent between markets. Any investor with a better outside opportunity prefers not to enter.

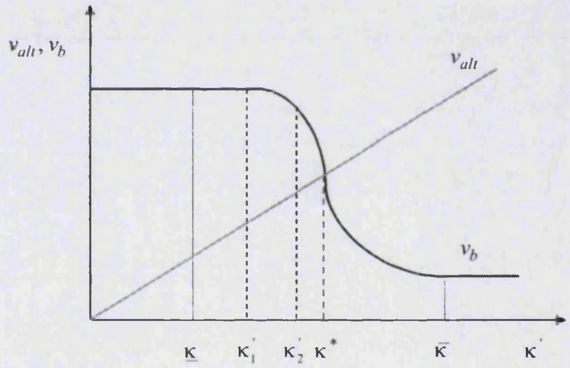

Figure 4.3: UNIQUE MARKET EQUILIBRIUM - Investors compare expected utilities v_b and v_{alt} and decide to move into our market if $v_b > v_{alt}$. κ^* defines the outside investment opportunity which makes investors indifferent between entering or not our market. $\underline{\kappa}$ and $\overline{\kappa}$ determine the support of the flow of investors who enter the economy.

Let us see why the equilibrium is unique. Given non-negative expected utilities, if $v_b(\kappa' = 0) > v_{alt}(\kappa' = 0)$ and v_b decreases in κ' while v_{alt} is strictly increasing, then by continuity there exists a unique threshold κ^* such that expected utilities are equal and investors indifferent between markets. A unique threshold κ^* then defines a unique flow of investors $g^* \equiv g(\kappa' = \kappa^*)$ entering our market. And given a unique flow of investors g^*, steady-state measures, expected utilities and the asset price can be determined uniquely as stated in Propositions 4.1 and 4.4. Consequently, market equilibrium is unique. It is interesting to note that the expected utility of buyers-to-be decreases as more investors enter our market. We discuss this result in the following subsection.

4.5.3 Market Congestion

Why is the expected utility of a buyer-to-be reduced as the flow of investors entering the market rises? Because buyers-to-be suffer from a *congestion effect*. In our market, an increase in the flow of investors g affects differently the steady-state measures of investors. Every investor who enters our market becomes a buyer-to-be first. Then, only a proportion of buyers-to-be meets a trading partner, purchases the asset and becomes a non-searcher. Only a fraction of non-searcher receives a liquidity shock becoming a seller-to-be. But, given that the measure of buyers-to-be has increased, it is now easier for a seller-to-be to meet a trading partner and hence the steady-state measure of investors seeking to sell is reduced as the flow of investors g increases[13]. As a result, in our framework buyers-to-be are worse off when g rises because it is now more difficult for them to meet a seller-to-be and purchase the asset. There is a congestion effect as investors move into our market in the sense that buyers-to-be face a crowded market where there is increasing competition among buyers-to-be for the fewer sellers-to-be.

[13] Comparative statics of the steady-state measures of investors in the market were introduced in Section 4.4.1 (See Proposition 4.2).

4.6 Liquidity, Market Efficiency and Welfare

In this section we first discuss the relationship between market liquidity and the equilibrium flow of investors who move into our market. In our model, the equilibrium flow is endogenously determined and depends on the characteristics defining the market, the asset and the investors. To analyze, for instance, the consequences on market liquidity of a change in market efficiency we need to understand both the direct effect of this change on the asset price, and hence on liquidity, and also the indirect effect through the equilibrium flow of investors. Subsection 4.6.2 examines the introduction of a new electronic system in our market to provide some intuition for the interaction between search costs and the equilibrium flow of investors and thus to better understand this indirect effect. The general relationship between the flow of investors and the parameters of the model, including search efficiency, is derived in Subsection 4.6.3. Finally, in Subsection 4.6.4 we introduce welfare and study the implications on welfare and market liquidity of an improvement in the efficiency of the search process when our market experiences a fire sale.

4.6.1 Market Liquidity

In our model, an investor willing to buy or sell needs to find a trading partner and bargain over the asset price before the transaction takes place. Investors cannot trade instantaneously but there is a time delay due to this search process. This search cost can be identified with the expected time required to locate a trading partner and, as a result, liquidity can be viewed as inversely related to this time delay. In Subsection 4.4.2 we define illiquidity as measured by the illiquidity discount

$$k\frac{x}{r}$$

where $\frac{x}{r}$ is the present value of the holding cost x and $k = \frac{(r+\gamma+\lambda\eta_s)z+\gamma}{(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)}$. Let us denote by $\tau^s \equiv \frac{1}{\lambda\eta_b}$ the expected time required to locate a buyer-to-be and by $\tau^b \equiv \frac{1}{\lambda\eta_s}$ the expected time it takes for a buyer-to-be to meet a seller-to-be. The function k is increasing in τ^s and decreasing in τ^b. Then, as the time a seller-to-be needs to wait before he can leave the market (τ^s) increases, k rises and the effect of the illiquidity discount is more severe. In contrast, if a buyer-to-be needs to wait longer to locate a seller-to-be, the effect of illiquidity discount is diminished. The equilibrium level of market liquidity thus rises in η_b^* but diminishes in η_s^*. In our market, an increase in the equilibrium measure of buyers-to-be and a reduction in the equilibrium measure of sellers-to-be occurs whenever the equilibrium flow of investors, g^*, moving into our market increases[14]. We formalize this result in the following proposition, which we prove in Appendix 4.8.4:

Proposition 4.6. *Liquidity increases in the flow of investors entering the market.*

Understanding the relationship between market liquidity and investors' decision to enter a market constitutes one of the main motivations of our analysis. In our model, there is a trading externality as the arrival of new investors facilitates the search process for every investor in the market. If the flow of potential traders increases, trade becomes easier and liquidity rises.

However, as the flow of investors moving into a market increases, the congestion effect makes it more difficult for a buyer-to-be to locate a trading partner. Consequently, as the market gets crowded, it becomes less attractive to investors. This translates into a lower flow of investors entering the market and as a result into a less liquid market.

In equilibrium, the flow of investors and hence the level of market liquidity result from a tradeoff between thick market complementarities and congestion effects. The equilibrium

[14]See Proposition 4.2.

flow of investors though is determined endogenously in our framework and depends on the market, investors and asset characteristics such as search efficiency, frequency of the liquidity shocks and dividend flow, among others. If we were interested in the consequences on market liquidity of a change in any of these characteristics, we would need to consider two different type of implications. Assume, for instance, an improvement in the efficiency of the search process. It would not only increase the rate at which investors meet but it would also affect the flow of investors who enter our market. Specifically, it raises the frequency of meetings between buyers-to-be and sellers-to-be, favoring market liquidity, and induces two opposite effects on the equilibrium flow of investors who move into our market. First, trading externalities attract potential investors, increasing the flow. But, secondly, congestion deters investors from entering our market, diminishing the flow. The overall level of market liquidity thus depends on this tradeoff and on the effect of the improvement in efficiency on the trading frequency. We discuss the aggregate effect on market liquidity in Subsection 4.6.4, but we first introduce the following example to better understand the interaction between trading costs and flow of investors.

4.6.2 An Example of a Technological Innovation

We consider a search-based market of asset trading as the one described in the previous sections. We are interested in understanding the consequences of a technological innovation intended to increase the efficiency of the search process, such as the introduction of a new electronic trading system. The efficiency of the search process in our model is defined by the parameter λ. A high value of λ represents an efficient search process and corresponds to a market where the rate at which investors meet trading partners is high and hence the friction introduced by the search process and its associated cost are low.

We assume the flow of investors f entering the economy is uniformly distributed[15] with support $[0, 5]$, where $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$. Investors have time preferences with discount rate equal to 1% ($r = 0.01$). The asset pays a dividend flow $d = 2$ and is in total supply $S = 2$. The holding cost is defined as a 40% of the dividend flow to indicate that once an investor receives a liquidity shock his valuation of the asset drops to a 60% of the initial value. Liquidity shocks arrive at a Poisson rate $\gamma = 0.2$ and hence the expected time between shocks is 5. The value of z is chosen such that buyers-to-be and sellers-to-be have the same bargaining power, i.e. $z = 1$. We refer to this example as the baseline setting.

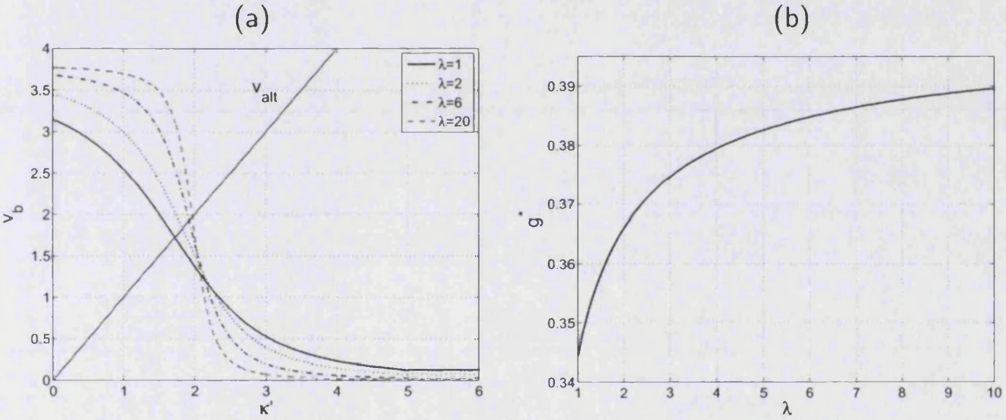

Figure 4.4: BASELINE SETTING - Improving efficiency (higher value of λ) attracts more investors to our market (higher g^*). The value of the model parameters is set at the following: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.2$, $a = 1$, $b = 1$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

Figure 4.4(a) represents the expected utility v_{alt} of investing in an alternative market and the expected utility v_b of being a buyer-to-be as a function of the marginal investor's

[15]Formally, we assume a beta distribution defined on the interval $[0, 5]$ with shape parameters $a = 1$ and $b = 1$, which is identical to a uniform distribution with support $[0, 5]$. The beta distribution is a flexible class of distributions defined on the unit interval $[0, 1]$, whose density function may take on different shapes depending on the choice of the two parameters. These include the uniform density function and hump-shaped densities (See Evans et al. (1993)). We introduce the beta distribution to facilitate the comparison between settings when we later discuss the second example.

outside investment opportunity κ'. The expected utility v_b of buyers-to-be is plotted for four different values of the market efficiency λ, where a higher λ indicates a more efficient search process. The intersection between v_b and v_{alt} gives, for each level of search efficiency, the threshold κ^* that defines the indifferent investor. κ^* is hence the solution to our fixed point problem. In equilibrium, investors whose best outside investment opportunity κ' is below the threshold value κ^* enter our market, while those with $\kappa' > \kappa^*$ prefer the alternative market. Figure 4.4(a) shows that an improvement in the efficiency of the search process (higher value of λ) makes our market attractive to more investors (higher κ^*). A higher threshold κ^* then corresponds to an increase in the flow of investors g^* who prefer our market. Figure 4.4(b) depicts the equilibrium flow of investors g^* entering our market, which is strictly increasing in the efficiency of the search process.

An increase in the equilibrium flow of investors g^* causes a rise in the equilibrium measures of buyers-to-be η_b^* and non-searchers η_0^* and a reduction in the equilibrium measure of sellers-to-be[16] η_s^*. But the equilibrium measures of investors in our market also depend on the efficiency of the search process[17]. In particular, as the search process becomes more efficient (higher value of λ), the measures of investors "waiting" to buy or sell (η_b^* and η_s^*) decrease while the measure of non-searchers rises. The overall effect on the equilibrium measures is presented in the top panel of Figure 4.5(a):

More interestingly, the bottom panel of Figure 4.5(a) illustrates the ratio between buyers-to-be and sellers-to-be as a function of the efficiency of the search process. This ratio captures the notion of congestion in our market. A high value of the proportion of buyers-to-be to sellers-to-be ($>> 1$) describes a market where there is strong competition among buyers-to-be for the few sellers-to-be. There is congestion on the buyer-side in this market. On the contrary, a very low value of this ratio corresponds to a market where there is

[16]See Proposition 4.2.
[17]See Proposition 4.3.

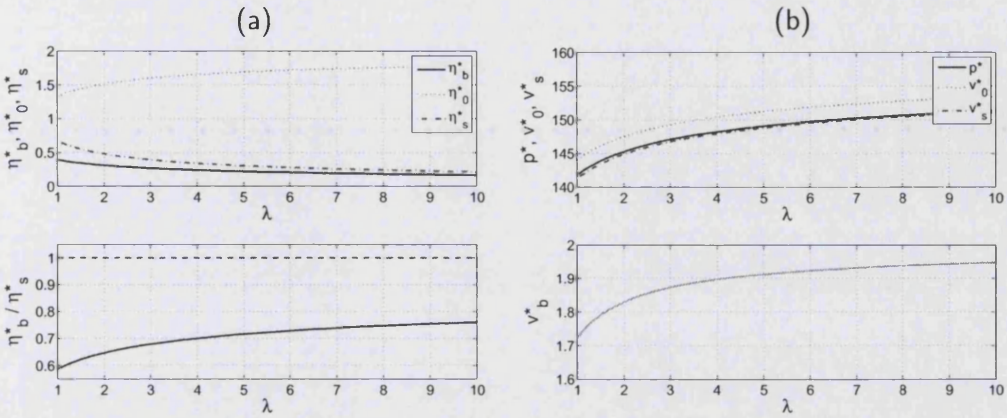

Figure 4.5: BASELINE SETTING - Equilibrium measures of investors in our market and ratio of buyers-to-be to sellers-to-be (a), expected utilities and price (b) as a function of the market efficiency λ. Other parameters are set at the following values: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.2$, $a = 1$, $b = 1$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

congestion on the sell-side (more sellers-to-be than buyers-to-be). The effect of congestion gets attenuated as the ratio between buyers-to-be and sellers-to-be tends to 1 as in our baseline setting.

Figure 4.5(b) depicts equilibrium price and expected utility of sellers-to-be and non-searchers (top panel) and buyers-to-be (bottom panel) as a function of λ. Price and expected utilities increase in the efficiency of the search process.

In this baseline setting a new electronic trading system, which improves search efficiency, enhances the attractiveness of our market. But this is not always the case. Let us introduce the following example.

Market Boom or Market Crash in an Outside Market

Assume a scenario similar to the one we have just discussed in the baseline setting and let us now consider a severe adverse shock which affects investors' outside investment opportunities. The worsening of investors' outside options could correspond to a boom in our market or to a market crash in another market[18] and would affect the distribution of investors f entering the economy as a function of their outside investment opportunities κ. It would lead to a shift to the left of the mass of the distribution of investors f. In particular, we consider a beta distribution with support $[0, 5]$ and parameters $a = 2$ and $b = 15$, which is a right-skewed hump-shaped density function.

We present our results in Figures 4.6 and 4.7, where the value of all parameters (but the distribution parameters) remains as in the baseline setting, i.e., $r = 0.01$, $d = 2$, $S = 2$, $x = 0.4d$, $\gamma = 0.2$ and $z = 1$.

Figure 4.6(a) illustrates the expected utility v_{alt} of investors' outside options and the expected utility v_b of being a buyer-to-be in our market for the same four values of market efficiency considered in the baseline setting. It is interesting to highlight that the equilibrium threshold κ^* now decreases in the search efficiency, such that to a market with a more efficient search process corresponds a lower cutoff value κ^* of the outside option, which then defines a lower equilibrium flow of investors g^* entering the market. As Figure 4.6(b) clearly shows, the equilibrium flow of investors entering our market strictly decreases in the search efficiency.

Why is the equilibrium flow of investors decreasing as the search process becomes more efficient? Let us see why this is the case. Our market is now attractive to more investors

[18] In either case, market conditions improve significantly in our market compared to those in alternative markets. For the ease of exposition, we consider the market crash interpretation.

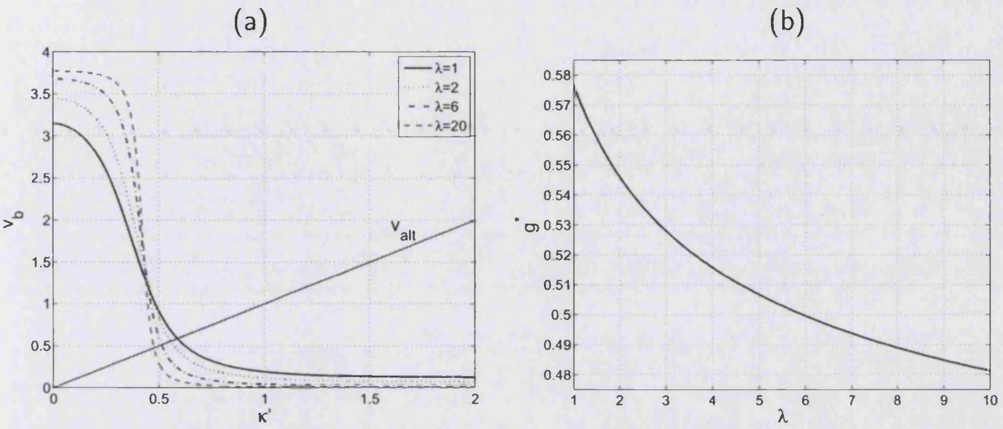

Figure 4.6: MARKET CRASH SETTING - Improving efficiency (higher value of λ) discourages investors from entering our market (lower equilibrium flow of investors g^*). Model parameters are set at the following values: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.2$, $a = 2$, $b = 15$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

because of the worsening of conditions in another market. This is indicated in Figures 4.4(b) and 4.6(b), which show that for any given level of market efficiency (fixing λ), the equilibrium flow of investors now entering our market is higher than in the baseline setting. Then, from the buyers'-to-be perspective, our market has become crowded in the sense that there are too many buyers-to-be for each investor seeking to sell and hence it is now more difficult to meet a trading partner and purchase the asset. If search frictions were then reduced in this market (higher values of λ), the effect of congestion would be amplified. Investors would meet at faster rates, which reduces the measures of buyers-to-be and sellers-to-be as shown in the top panel of Figure 4.7(a) but, most importantly, it would allow sellers-to-be to exit faster leading to an even more unbalanced distribution of investors (bottom panel of Figure 4.7(a)).

As the bottom panel of Figure 4.7(b) illustrates, buyers-to-be are worse-off as the search process becomes more efficient and congestion intensifies. This discourages potential in-

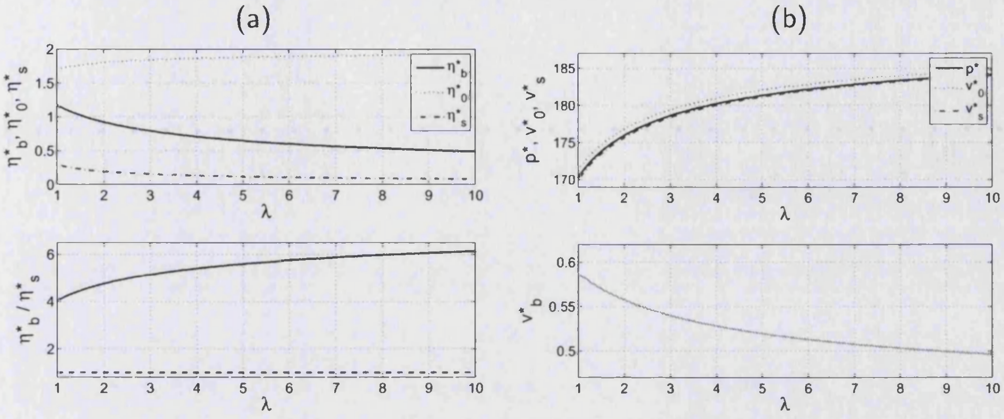

Figure 4.7: MARKET CRASH SETTING - Equilibrium measures of investors in our market and ratio of buyers-to-be to sellers-to-be (a), expected utilities and price (b) as a function of the market efficiency λ. Other parameters are set at the following values: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.2$, $a = 2$, $b = 15$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

vestors from moving into our market, reducing the equilibrium flow of investors g^*.

The reason for this counterintuitive result is that lower trading frictions in a one-sided market magnify the effect of congestion, discouraging investors from entering this market. In this case, congestion dominates thick market externalities and hence the introduction of a measure intended to improve market efficiency results in a less attractive market.

4.6.3 Flow of Investors and Market Efficiency

In this subsection we determine the general relationship between the equilibrium flow of investors g^* entering the market and the efficiency λ of the search process. To simplify the analysis we first derive the equilibrium measure of sellers-to-be (η_s^*) as a function of the market efficiency λ and the other nine parameters of the model (γ, r, S, x, z, a, b, $\underline{\kappa}$ and $\overline{\kappa}$). There is a one-to-one relationship between g^* and η_s^*. Hence, once we compute η_s^*, we can then determine the equilibrium flow of investors g^* who enter our market.

In our setting, market equilibrium is the solution to the system of equations (4.1)-(4.3), (4.7)-(4.10) and (4.15). We thus need to solve for the fixed point of this system, which is reduced to solving the indifference condition that defines investors' entry rule. Investors, in our framework, compare the expected utility v_{alt} of investing in an alternative market to the expected utility v_b derived from being a buyer-to-be in our market and they decide to move in whenever $v_b > \kappa' \equiv v_{alt}$. To present this indifference condition ($v_b = \kappa'$) as a function of the measure of sellers-to-be (η_s), let us first redefine the measure of buyers-to-be η_b as a function of η_s. Using equations (4.5) and (4.6) we find:

$$\eta_b = \frac{\gamma}{\lambda}\frac{A}{2\gamma S - A} = \frac{\gamma}{\lambda}\frac{2\gamma(S-\eta_s)}{2\gamma S - 2\gamma(S-\eta_s)} \qquad \Rightarrow \qquad \eta_b = \frac{\gamma}{\lambda}\frac{S-\eta_s}{\eta_s} \tag{4.16}$$

We can now express the expected utility v_b of buyers-to-be as a function of η_s by substituting equation (4.16) into equation (4.11):

$$v_b = \frac{x}{r+\gamma}\frac{\lambda z \eta_s^2}{\lambda z \eta_s^2 + [(r+\gamma)(1+z)-\gamma]\,\eta_s + \gamma S} \tag{4.17}$$

Next we write κ' as a function of η_s. In this model, the flow of investors g who move into our market is determined by the proportion of the total flow of investors f whose expected utility v_b of being a buyer-to-be exceeds their best outside option κ'. We assume the flow of investors f follows a beta distribution with support $[\underline{\kappa}, \overline{\kappa}]$ and shape parameters[19] a and

[19]The probability density function of the beta distribution defined over the interval $[0,1]$ with shape parameters a and b is:

$$f_{beta}(y;a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}y^{a-1}(1-y)^{b-1}$$

where $a, b > 0$ and $\Gamma(\cdot)$ is the gamma function. For integer values of a and b, the cumulative distribution

b. For notational convenience we omit reference to the shape parameters. Then,

$$g(\kappa') = \int_{\underline{\kappa}}^{\kappa'} f_{beta}(\kappa)d\kappa = F_{beta}(\kappa') \qquad \Rightarrow \qquad \kappa' = F_{beta}^{-1}(g) \tag{4.18}$$

where f_{beta} and F_{beta} denote respectively the probability density function (pdf) and the cumulative distribution function (cdf) of a beta distribution. F_{beta}^{-1} is the inverse cumulative distribution function. Using equation (4.5) and the definition of A in Page 102 we can express the flow of investors g as a function of the measure of sellers-to-be η_s:

$$g = \gamma\left(1 + \frac{\gamma}{\lambda\eta_s}\right)(S - \eta_s) \tag{4.19}$$

Substituting equation (4.19) in equation (4.18) yields:

$$\kappa' = F_{beta}^{-1}\left(\gamma\left(1 + \frac{\gamma}{\lambda\eta_s}\right)(S - \eta_s)\right) \tag{4.20}$$

The indifference condition results from equating the expected utility v_b of buyers-to-be (equation (4.17)) to the marginal investor outside option κ' (equation (4.20)):

$$\frac{x}{r+\gamma}\frac{\lambda z\eta_s^2}{\lambda z\eta_s^2 + [(r+\gamma)(1+z) - \gamma]\eta_s + \gamma S} = F_{beta}^{-1}\left(\gamma\left(1 + \frac{\gamma}{\lambda\eta_s}\right)(S - \eta_s)\right)$$

function of the beta distribution is given by:

$$F_{beta}(y; a, b) = \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} y^j (1-y)^{a+b-1-j}$$

where $\binom{a+b-1}{j} = \frac{(a+b-1)!}{j!(a+b-1-j)!}$.

Rearranging, we get

$$\gamma\left(1+\frac{\gamma}{\lambda\eta_s}\right)(S-\eta_s)=F_{beta}\left(\frac{x}{r+\gamma}\frac{\lambda z\eta_s^2}{\lambda z\eta_s^2+[(r+\gamma)(1+z)-\gamma]\eta_s+\gamma S}\right)$$

Then,

$$\gamma\left(1+\frac{\gamma}{\lambda\eta_s}\right)(S-\eta_s)=\sum_{j=a}^{a+b-1}\binom{a+b-1}{j}\left(\frac{x}{r+\gamma}\frac{\lambda z\eta_s^2}{\lambda z\eta_s^2+[(r+\gamma)(1+z)-\gamma]\eta_s+\gamma S}\right)^j$$
$$\left(1-\frac{x}{r+\gamma}\frac{\lambda z\eta_s^2}{\lambda z\eta_s^2+[(r+\gamma)(1+z)-\gamma]\eta_s+\gamma S}\right)^{a+b-1-j} \tag{4.21}$$

Equation (4.21) is a polynomial of degree $2(a+b)$ in the measure of sellers-to-be[20]. To solve for η_s^* we use the bisection method[21]. Once we compute η_s^*, we can derive g^*:

$$g^*=g^*(\lambda,\gamma,r,S,x,z,a,b,\underline{\kappa},\overline{\kappa})$$

[20] In the simple case of shape parameters of the beta distribution both equal to 1 ($a=1=b$), which corresponds to a uniform distribution with support $[\underline{\kappa},\overline{\kappa}]$, the indifference condition ($v_b=\kappa'$) is:

$$\frac{x}{r+\gamma}\frac{\lambda z\eta_s^2}{\lambda z\eta_s^2+[(r+\gamma)(1+z)-\gamma]\eta_s+\gamma S}=\underline{\kappa}+(\overline{\kappa}-\underline{\kappa})\gamma\left(1+\frac{\gamma}{\lambda\eta_s}\right)(S-\eta_s)$$

Reorganizing terms yields the following polynomial of degree four in the measure of sellers-to-be η_s:

$$\lambda^2 z(\overline{\kappa}-\underline{\kappa})\gamma\eta_s^4 \;+\; \left[\lambda(\overline{\kappa}-\underline{\kappa})\gamma C-\lambda zD+\frac{x}{r+\gamma}\lambda^2 z\right]\eta_s^3+\left[\lambda(\overline{\kappa}-\underline{\kappa})\gamma^2 S(1-z)-CD\right]\eta_s^2-$$
$$-\;\left[\gamma SD+(\overline{\kappa}-\underline{\kappa})\gamma^2 SC\right]\eta_s-(\overline{\kappa}-\underline{\kappa})\gamma^3 S^2=0$$

where $C=(r+\gamma)(1+z)-\gamma$ and $D=\lambda\underline{\kappa}+\lambda(\overline{\kappa}-\underline{\kappa})\gamma S-(\overline{\kappa}-\underline{\kappa})\gamma^2$. There exists closed-form solution to this equation. In particular, there are at most four solutions but only one, η_s^*, (as proved in Subsection 4.5.2) lies in the interval $(0,S)$, the set of possible values of the measure of sellers-to-be. Unfortunately, the solution is intractable. We use the bisection method over the interval $[0,S]$ to determine the zero of this equation.

[21] The bisection algorithm is a numerical method for finding the root of a function. It recursively divides an interval in half and selects the subinterval containing the root, until the interval is sufficiently small. Burden and Faires (1993) presents a clear description of this algorithm as well as other numerical methods for solving root-finding problems.

where g^* is a function of the efficiency of the search process λ, the rate γ at which investors receive liquidity shocks, the discount rate r, the supply of the asset S, the holding cost x, buyer's-to-be bargaining power z, the shape parameters a and b of the beta distribution and the support $[\underline{\kappa}, \overline{\kappa}]$ of the flow of investors f entering the economy. To gain some intuition for how the model parameters affect the equilibrium flow of investors g^*, we set the value of those defining the distribution of f and vary the other parameters of the model. We assume $a = 2$, $b = 15$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$ as in the market crash setting in Subsection 4.6.2. The first set of results is depicted in Figure 4.8:

Figure 4.8 represents g^* as a function of the efficiency of the search process λ, where g^* is plotted for four different values of the discount rate r (a), the supply of the asset S (b), the holding cost x (c) and the buyers'-to-be bargaining power z (d). The distribution of parameters underlying these graphs corresponds to a one-sided market scenario discussed in Subsection 4.6.2. Then, in all four cases, increasing market efficiency (higher values of λ) translates into a lower equilibrium flow of investors entering the market. Also, for a given level of market efficiency (fixed λ), more investors move into our market as we increase the total supply of the asset, the holding cost or the buyers'-to-be bargaining power. The equilibrium flow of investors decreases as they become more impatient (higher r).

More interesting is the interaction between market efficiency λ and the arrival rate of liquidity shocks γ. Figure 4.9(a) demonstrates how the equilibrium flow of investors g^*, who enter our market, varies with the market efficiency λ and the frequency of liquidity shocks γ. Contours are depicted in Figure 4.9(b). Now, the relationship between g^* and λ is non-monotonic. It is first decreasing in market efficiency, corresponding to a one-sided market scenario, but then it becomes increasing in λ for higher values of the liquidity shock rate γ.

If liquidity shocks arrive at very low rates (low values of γ), investors hold the asset,

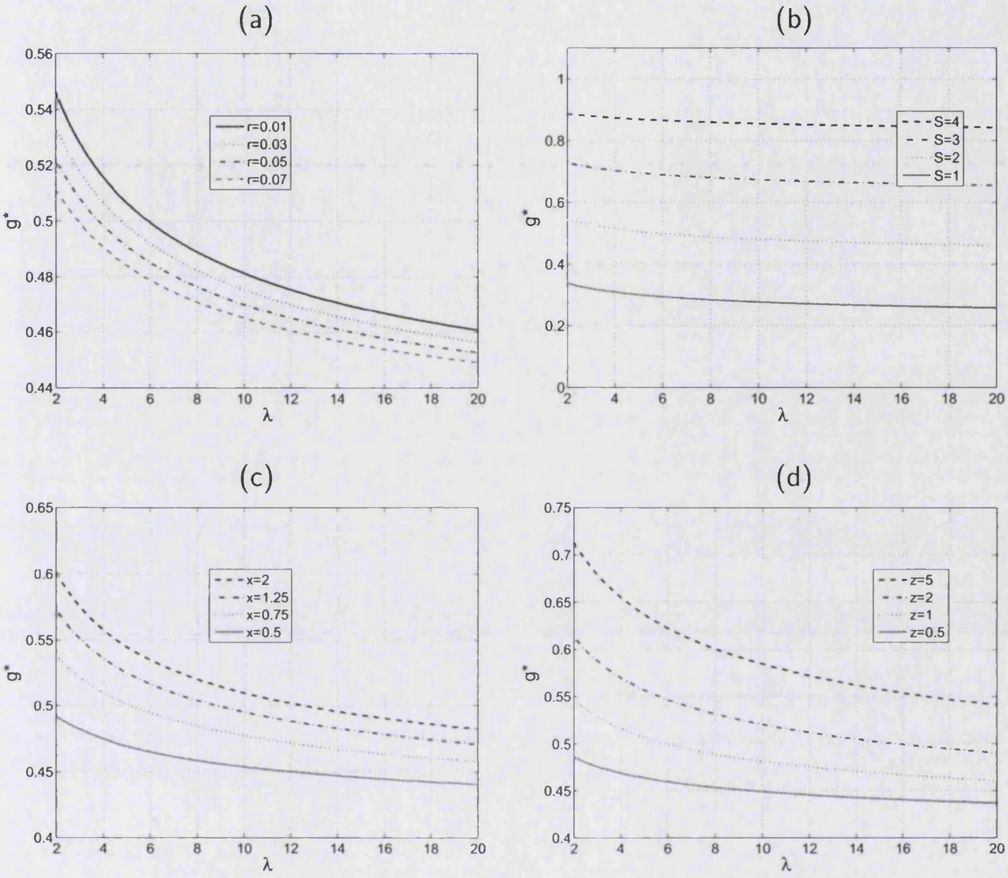

Figure 4.8: Equilibrium flow of investors g^* entering our market as a function of the market efficiency λ for different values of r (a), S (b), x (c) and z (d). Other parameters are set at the following values: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.2$, $a = 2$, $b = 15$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

on average, for a long time. As a result, there are few investors trying to sell and exit the market. Increasing the efficiency of this market (raising λ) attracts new investors, amplifying the effect of congestion. The market becomes one-sided because there are more buyers-to-be and few sellers-to-be. In this case, reducing market frictions diminishes the flow of investors. This is shown in Figure 4.9(c) for values of $\gamma \leq 0.3$. This phenomenon is attenuated as investors need to exit at a faster rate. Then, for intermediate values of γ, there are enough sellers-to-be in our market and improving market efficiency attracts new investors ($\gamma = 0.4$

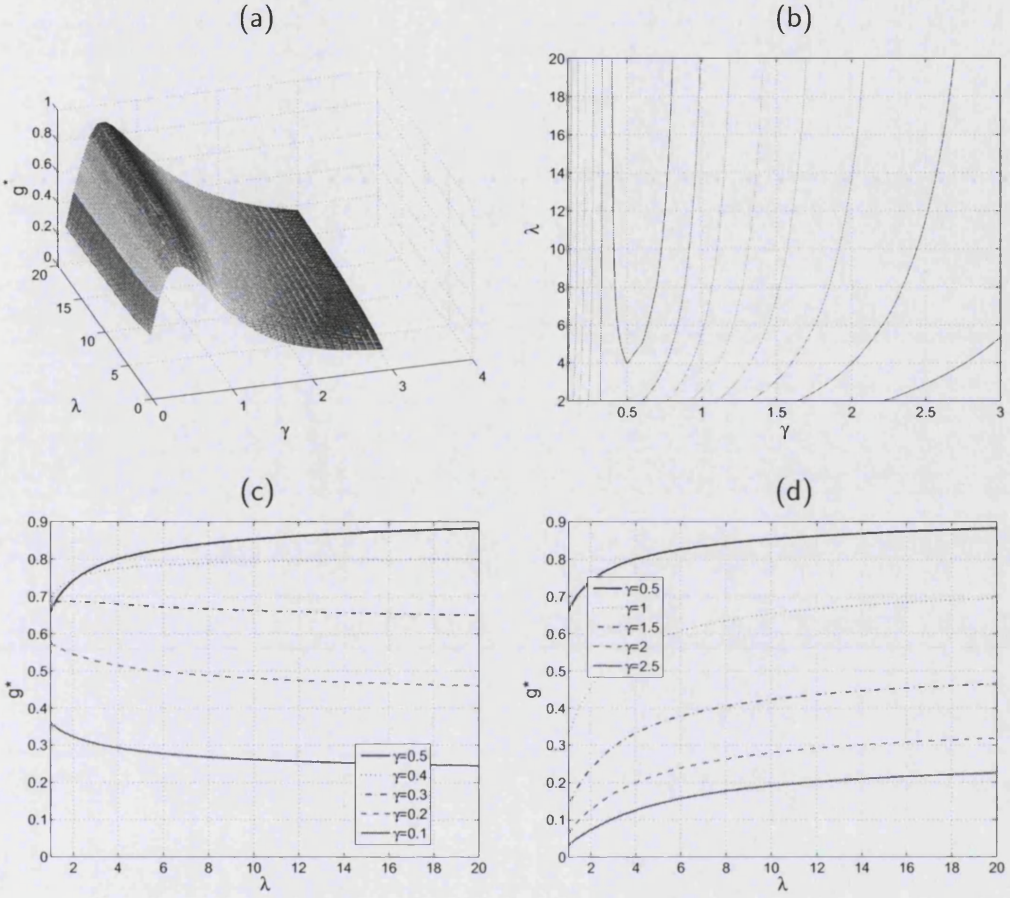

Figure 4.9: Equilibrium flow of investors g^* entering our market as a function of the market efficiency λ and the frequency of liquidity shocks γ. The values of other parameters of the model are set at the following: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $a = 2$, $b = 15$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

and $\gamma = 0.5$ in Figure 4.9(c)). Thick market externalities dominate congestion. Also, as Figure 4.9(d) indicates, if investors need for cash is very frequent (values of γ above 0.5), they prefer not to invest and the flow of investors g^* who enter our market is reduced. Still, for a given frequency of the liquidity shocks γ, diminishing search frictions improves the attractiveness of our market.

4.6.4 Liquidity and Welfare

In this subsection we discuss market liquidity and present the welfare analysis. In particular, we are interested in the implications of potential policies designed to improve the efficiency of the search process and to thus reduce market frictions.

We measure welfare by the weighted sum of investors' expected utilities. Weights are determined by the measure of every type of investors in our economy, including the outside investors. Then, our measure of welfare is:

$$W = \eta_b v_b + \eta_0 v_0 + \eta_s v_s + \int_{\kappa^*}^{\overline{\kappa}} \kappa f(\kappa) d\kappa \qquad (4.22)$$

where the first three terms represent the welfare of the investors who prefer to enter our market ($W_{\text{inside investors}}$) and the last term reflects the welfare of outside investors ($W_{\text{outside investors}}$). Outside investors (those with investment opportunities above the threshold value κ^*) enjoy the expected utility derived from investing in an alternative market v_{alt}, which for simplicity we assume equal to κ', the outside investment opportunity. Substituting equations (4.11)-(4.13) and the *pdf* of a beta distribution into equation (4.22), we get:

$$W_{\text{inside investors}} = \frac{d}{r}S - kx\frac{1}{r+\gamma}\frac{\frac{\gamma}{r}S\left[(\gamma+\lambda\eta_s)z+(r+\gamma)\right]+\eta_s\left[(r+2\gamma+\lambda\eta_s)z+(r+\gamma)\right]}{(r+\gamma+\lambda\eta_s)z+\gamma}$$

$$W_{\text{outside investors}} = \frac{a}{a+b}\left[1-F_{beta}(\kappa^*;a+1,b)\right] = \frac{a}{a+b}\left[1-\sum_{j=a+1}^{a+b}\binom{a+b}{j}(\kappa^*)^j(1-\kappa^*)^{a+b-j}\right]$$

To gain some intuition for how changes in market efficiency affect welfare we introduce the last example.

Fire Sales in our Market

Consider a search-based market similar to the baseline setting described in Subsection 4.6.2 and assume investors need for cash is now more frequent. Specifically, we assume liquidity shocks arrive at a Poisson rate $\gamma = 0.4$. The value of all other parameters remains as in the baseline case: $r = 0.01$, $d = 2$, $S = 2$, $x = 0.4d$, $z = 1$, $a = 1$, $b = 1$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

Investors now prefer to hold the asset, on average, for a shorter period of time and they are willing to sell and exit our market at faster rates. Then, for any given value of the search efficiency, the equilibrium measure of sellers-to-be has increased significantly (top panel of Figure 4.10(a)) compared to the baseline setting (top panel of Figure 4.5(a)), while the equilibrium measure of non-searchers has decreased. Given that there are now more sellers-to-be in our market, it is easier for an investor seeking to purchase the asset to meet a trading partner. As a result, the equilibrium measure of buyers-to-be has diminished compared to the baseline case. Most importantly, the proportion of buyers-to-be to sellers-to-be has fallen drastically. This is depicted in the bottom panel of Figure 4.10(a). Our market is now one-sided and there is severe congestion on the sell-side of the market. This scenario could correspond to a market experiencing a fire sale.

Increasing the efficiency of the search process (higher value of λ) in this market causes two effects. First, it raises the flow of investors who enter our market as plotted in Figure 4.10(c). Secondly, investors meet at faster rates reducing the equilibrium measure of buyers-to-be and sellers-to-be (top panel of Figure 4.10(a)). The overall effect on the ratio of buyers-to-be to sellers-to-be is presented in the bottom panel of Figure 4.10(a). As the market becomes more efficient, the proportion of buyers-to-be to sellers-to-be falls further and from the sellers'-to-be perspective the market gets even more crowded. Congestion intensifies as it is now more difficult to meet a buyer-to-be and exit the market. Hence,

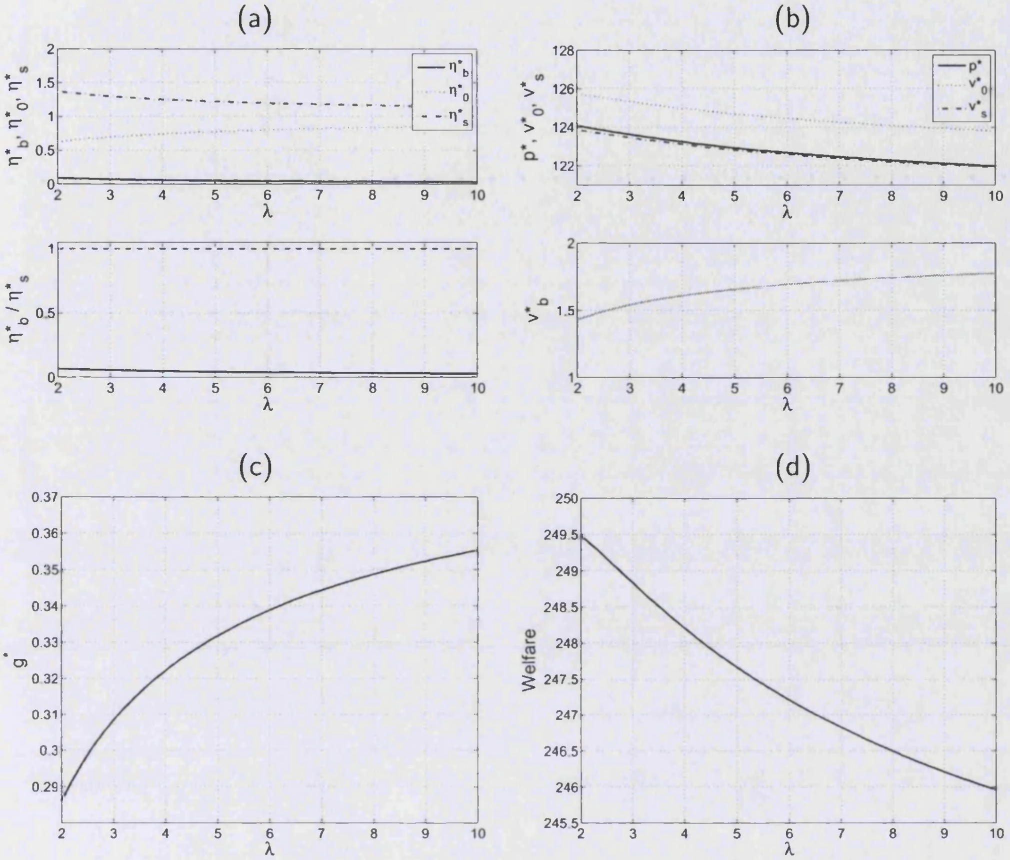

Figure 4.10: FIRE SALES SETTING - Equilibrium measures of investors in our market and ratio of buyers-to-be to sellers-to-be (a), expected utilities and price (b), flow of investors g^* (c) and welfare (d) as a function of the market efficiency λ. Other parameters are set at the following values: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.4$, $a = 1$, $b = 1$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

as the top panel of Figure 4.10(b) illustrates, sellers-to-be and non-searchers (who become sellers-to-be at rate γ) are worse-off as efficiency rises. The expected utility of buyers-to-be increases in λ because they can now acquire the asset at faster rates (bottom panel of Figure 4.10(b)).

A very interesting result is presented in Figure 4.10(d). We find that as search frictions are reduced, welfare decreases. In this market, improving the efficiency of the search process amplifies the effect of congestion. There are then fewer buyers-to-be per each seller-to-be and the expected utilities of investors holding the asset fall. This induces an adverse effect on welfare.

Figure 4.11(a) represents our measure of illiquidity as a function of the efficiency of the search process λ, where illiquidity is defined as the price discount. As market efficiency increases and the population of investors gets saturated with sellers-to-be, the price of the asset falls as shown in the top panel of Figure 4.10(b). This leads to the rise in illiquidity depicted in Figure 4.11(a). Intuitively, given that there are few buyers-to-be compared to sellers-to-be, the price of the asset behaves as if buyers-to-be would hold a more favorable position in the bargaining process. The effect is equivalent to an increase in the buyers'-to-be bargaining power z, which is exogenous in our model. If we were to endogenize z, the effect on the price (and hence on market liquidity) would be amplified.

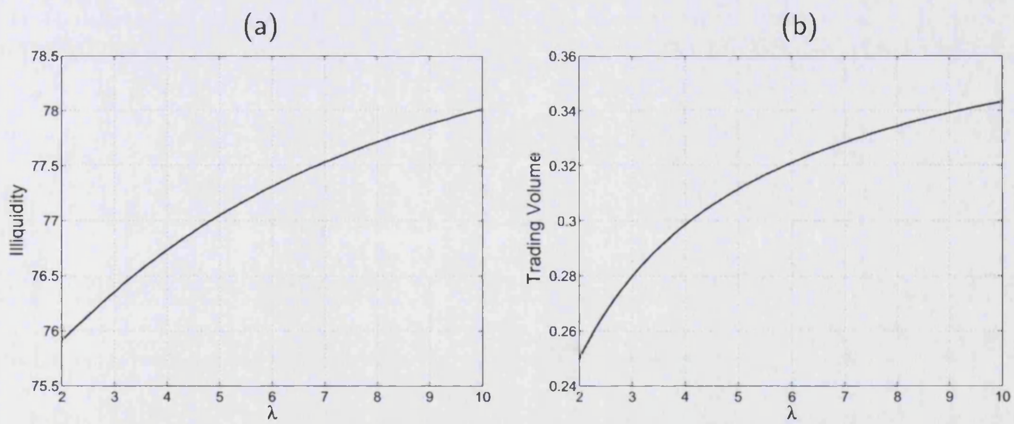

Figure 4.11: FIRE SALES SETTING - Illiquidity measured by price discount (a) and trading volume (b) as a function of the market efficiency λ. The value of the model parameters is set at the following: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.4$, $a = 1$, $b = 1$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

Our market becomes less liquid as the search frictions are reduced. However, as Figure 4.11(b) indicates, trading volume increases. The reason for this counterintuitive result is the following. Facilitating search in our market has two consequences. First, it magnifies the effect of congestion. There are fewer buyers-to-be relative to the measure of investors trying to exit. Buyers-to-be prefer to pay less to purchase the asset, which translates into a lower price and hence into a less liquid market (higher price discount). Second, it raises the frequency of meeting between trading partners. Investors in our market now meet at a faster rate, increasing the trading volume. Consequently, even though our market is less liquid, investors meet faster and trading volume increases.

4.7 Conclusions

This paper proposes a search-based model to study the relationship between market liquidity and the endogenous arrival of potential investors to a specific market. As investors enter a market, they make trade easier, attracting new investors. This gives rise to a thick market externality. Interestingly, as investors get attracted to this market, the market becomes crowded and congestion reduces the returns to investing. This paper aims to complement the literature on self-fulfilling liquidity by incorporating a second effect: the congestion effect.

In our market traders can invest in one asset which can be traded only when a pair of investors meet and bargain over the terms of trade. Finding a trading partner takes time and introduces opportunity and other costs. Investors' ability to trade thus affects the illiquidity discount and ultimately, the equilibrium price. We present a numerical example of an advance in trading technology to illustrate the link between the flow of new investors and market liquidity, and to discuss the implications of search frictions on market liquidity.

We then derive the general relationship between the equilibrium flow of investors moving into a market and the efficiency of the search process and highlight the tradeoff between the thick market complementarity and the congestion effect. The equilibrium outcome depends on which of these two effects dominates. In particular, we find that diminishing trade frictions in a market with many buyers and too few sellers leads to a lower equilibrium flow of investors into this market. Less search frictions would allow sellers to exit faster amplifying the effect of congestion (even more buyers per seller) and further discouraging investors from entering this market. We also show that reducing market frictions, in a "congested" market experiencing a fire sale, induces an adverse effect on both market liquidity and welfare. Improving search efficiency (to facilitate coordination and enhance liquidity), magnifies the effect of congestion (less buyers per seller trying to exit) to the detriment of the overall level of market liquidity and social welfare. From this perspective, this paper presents an example of the Theory of the Second Best, where eliminating one but not all market imperfections does not necessary increase efficiency as it may amplify the effect of the remaining distortions.

4.8 Appendix

4.8.1 Proofs of Propositions 4.1 - 4.4

Proof of Proposition 4.1

Proof. Rearranging equation (4.3), we get

$$h(\eta_0) \equiv \gamma\eta_0^2 - \left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)\eta_0 + Sg = 0$$

where $\eta_0 \in \mathbb{R}_+$. This quadratic function takes positive values as $\eta_0 \to \infty$, is non-negative at $\eta_0 = 0$ and negative at $\eta_0 = S$. Then, by continuity, the polynomial equation has a root in the interval $[0, S)$ and another one in the interval (S, ∞). The two solutions $\eta_0^{(1)}$ and $\eta_0^{(2)}$ are given by:

$$\begin{aligned}
\eta_0^{(1)} &= \frac{1}{2\gamma}\left[(g+\gamma S+\frac{\gamma^2}{\lambda}) - \sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2 - 4\gamma g S}\right] \\
\eta_0^{(2)} &= \frac{1}{2\gamma}\left[(g+\gamma S+\frac{\gamma^2}{\lambda}) + \sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2 - 4\gamma g S}\right]
\end{aligned}$$

where $0 \leq \eta_0^{(1)} < S < \eta_0^{(2)} < \infty$. $\eta_0^{(2)}$ is thus not a valid solution since the total supply of the asset is held either by the non-searchers or by the sellers-to-be and as a result the measure of non-searchers cannot exceed the supply of the asset. Then, there is unique solution to equation (4.3) given by:

$$\eta_0 = \frac{1}{2\gamma}A \tag{A.1}$$

where $A = (g + \gamma S + \frac{\gamma^2}{\lambda}) - \sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2 - 4\gamma g S}$. Plugging equation (A.1) into equations (4.1) and (4.2), we find

$$\begin{aligned}
\eta_s &= S - \frac{1}{2\gamma}A \\
\eta_b &= \frac{\gamma}{\lambda}\frac{A}{2\gamma S - A}
\end{aligned}$$

which proves Proposition 4.1. □

Proof of Proposition 4.2

Proof. Let us compute the partial derivatives of the measures given by the system of equations (4.4) - (4.6) with respect to g:

$$\frac{\partial \eta_0}{\partial g} = \frac{\partial \eta_0}{\partial A}\frac{\partial A}{\partial g} = \frac{1}{2\gamma}\frac{\partial A}{\partial g} \tag{A.2}$$

$$\frac{\partial \eta_s}{\partial g} = \frac{\partial \eta_s}{\partial A}\frac{\partial A}{\partial g} = -\frac{1}{2\gamma}\frac{\partial A}{\partial g} \tag{A.3}$$

$$\frac{\partial \eta_b}{\partial g} = \frac{\partial \eta_b}{\partial A}\frac{\partial A}{\partial g} = \frac{2\gamma^2}{\lambda}\frac{S}{(2\gamma S - A)^2}\frac{\partial A}{\partial g} \tag{A.4}$$

where

$$\frac{\partial A}{\partial g} = 1 - \frac{(g + \gamma S + \frac{\gamma^2}{\lambda}) - 2\gamma S}{\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}} \tag{A.5}$$

To determine the sign of $\frac{\partial A}{\partial g}$, we check if the second term on the right-hand-side of equation (A.5) is greater than 1:

$$\frac{(g + \gamma S + \frac{\gamma^2}{\lambda}) - 2\gamma S}{\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}} > 1 \quad ;$$

$$(g + \gamma S + \frac{\gamma^2}{\lambda}) - 2\gamma S > \sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S} \quad ; \tag{A.6}$$

where the right-hand-side of equation (A.6) is strictly positive since

$$\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S} = \sqrt{(g - \gamma S)^2 + 2(g + \gamma S)\frac{\gamma^2}{\lambda} + \frac{\gamma^4}{\lambda^2}} > 0 \tag{A.7}$$

We analyze two cases. If $(g + \gamma S + \frac{\gamma^2}{\lambda}) - 2\gamma S \leq 0$, then equation (A.6) is not satisfied.

On the contrary, if $(g+\gamma S+\frac{\gamma^2}{\lambda})-2\gamma S>0$,

$$\begin{aligned}\left[(g+\gamma S+\frac{\gamma^2}{\lambda})-2\gamma S\right]^2 &> \left[\sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2-4\gamma gS}\right]^2;\\ (g-\gamma S+\frac{\gamma^2}{\lambda})^2 &> (g+\gamma S+\frac{\gamma^2}{\lambda})^2-4\gamma gS;\end{aligned}$$

Simplifying we arrive to:

$$4\frac{\gamma^3}{\lambda}S<0$$

a contradiction, since γ, λ and $S>0$. Therefore, the second term in equation (A.5) is strictly lower than 1 and as a result:

$$\frac{\partial A}{\partial g}>0 \tag{A.8}$$

Thus, substituting the previous equation into equations (A.2) - (A.4) yields:

$$\begin{aligned}\frac{\partial \eta_0}{\partial g} &> 0\\ \frac{\partial \eta_s}{\partial g} &< 0\\ \frac{\partial \eta_b}{\partial g} &> 0\end{aligned}$$

since γ, λ and $S>0$. □

Proof of Proposition 4.3

Proof. Using equations (4.4) - (4.6) we can compute the partial derivatives of the measures of every type of investor with respect to the efficiency of the search process λ:

$$\frac{\partial \eta_0}{\partial \lambda} = \frac{\partial \eta_0}{\partial A}\frac{\partial A}{\partial \lambda} = \frac{1}{2\gamma}\frac{\partial A}{\partial \lambda} \tag{A.9}$$

$$\frac{\partial \eta_s}{\partial \lambda} = \frac{\partial \eta_s}{\partial A}\frac{\partial A}{\partial \lambda} = -\frac{1}{2\gamma}\frac{\partial A}{\partial \lambda} \tag{A.10}$$

$$\frac{\partial \eta_b}{\partial \lambda} = \frac{\gamma}{\lambda}\frac{1}{2\gamma S - A}\left[\frac{2\gamma S}{2\gamma S - A}\frac{\partial A}{\partial \lambda} - \frac{1}{\lambda}A\right] \tag{A.11}$$

where

$$\frac{\partial A}{\partial \lambda} = -\frac{\gamma^2}{\lambda^2}\left(1 - \frac{g + \gamma S + \frac{\gamma^2}{\lambda}}{\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}}\right) \tag{A.12}$$

We verify whether the second term in the expression in parenthesis is greater than 1 to determine the sign of $\frac{\partial A}{\partial \lambda}$,

$$\begin{aligned}
\frac{(g + \gamma S + \frac{\gamma^2}{\lambda})}{\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}} &> 1; \\
\left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 &> \left[\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}\right]^2; \\
\left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 &> \left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 - 4\gamma g S; \qquad \text{(A.13)}
\end{aligned}$$

where we can square both sides of the expression because, using equation (A.7) and g, γ, S and $\lambda > 0$, the numerator and denominator are strictly positive. Rearranging equation (A.13) we get:

$$4\gamma g S > 0$$

which is true since γ, g and $S > 0$. As a result, the second term in the expression in parenthesis in equation (A.12) is strictly greater than 1 and

$$\frac{\partial A}{\partial \lambda} > 0 \tag{A.14}$$

Thus, plugging the previous equation into equations (A.9) - (A.10) we find:

$$\begin{aligned} \frac{\partial \eta_0}{\partial \lambda} &> 0 \\ \frac{\partial \eta_s}{\partial \lambda} &< 0 \end{aligned}$$

The proof that $\frac{\partial \eta_b}{\partial \lambda} < 0$ is not so straightforward. Let us first rearrange equation (A.12) as follows

$$\frac{\partial A}{\partial \lambda} = \frac{\gamma^2}{\lambda^2} \frac{A}{\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}} \tag{A.15}$$

Now, substituting equation (A.15) in equation (A.11) we get:

$$\frac{\partial \eta_b}{\partial \lambda} = \frac{\gamma}{\lambda^2} \frac{A}{2\gamma S - A} \left[\frac{2\gamma S}{2\gamma S - A} \frac{\gamma^2}{\lambda} \frac{1}{\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}} - 1 \right] \tag{A.16}$$

where we need to derive the sign of the expression in brackets to determine the sign of $\frac{\partial \eta_b}{\partial \lambda}$. Let us then verify if the first term of the expression in brackets in equation (A.16) is strictly lower than 1:

$$
\begin{aligned}
\frac{2\gamma S}{2\gamma S - A}\frac{\gamma^2}{\lambda}\frac{1}{\sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2-4\gamma g S}} &< 1;\\
(2\gamma S - A)\lambda\sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2-4\gamma g S}-2\gamma^3 S &> 0;\\
\lambda\Big[(2\gamma S - A)\sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2-4\gamma g S}-\frac{2\gamma^3 S}{\lambda}\Big] &> 0;
\end{aligned}
$$

Given that $\lambda > 0$ and $A = (g+\gamma S+\frac{\gamma^2}{\lambda}) - \sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2-4\gamma g S}$, then

$$
\begin{aligned}
\Big[2\gamma S - (g+\gamma S+\frac{\gamma^2}{\lambda})\Big]\sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2-4\gamma g S}+\Big[\sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2-4\gamma g S}\Big]^2-\frac{2\gamma^3 S}{\lambda} &> 0;\\
\Big(g-\gamma S+\frac{\gamma^2}{\lambda}\Big)^2-\Big(g-\gamma S+\frac{\gamma^2}{\lambda}\Big)\sqrt{(g-\gamma S+\frac{\gamma^2}{\lambda})^2+\frac{4\gamma^3 S}{\lambda}}+\frac{2\gamma^3 S}{\lambda} &> 0;
\end{aligned}
$$

To simplify the exposition of the proof, let us define $D \equiv g - \gamma S + \frac{\gamma^2}{\lambda}$. Therefore,

$$
D^2 - D\sqrt{D^2+\frac{4\gamma^3 S}{\lambda}}+\frac{2\gamma^3 S}{\lambda} > 0 \tag{A.17}
$$

We consider two possible scenarios. If $D \leq 0$, then equation (A.17) is satisfied since λ, γ and $S > 0$. On the contrary, if $D > 0$, then we need to prove that

$$
D^2+\frac{2\gamma^3 S}{\lambda} > D\sqrt{D^2+\frac{4\gamma^3 S}{\lambda}}
$$

Squaring both sides and rearranging, we find

$$
D^4+\frac{4\gamma^3 S}{\lambda}D^2+\frac{4\gamma^6 S^2}{\lambda^2} > D^2\Big(D^2+\frac{4\gamma^3 S}{\lambda}\Big)
$$

Simplifying,

$$\frac{4\gamma^6 S^2}{\lambda^2} > 0$$

and this is always satisfied. Then, we have shown that the first term in the expression in brackets in equation (A.16) is strictly lower than 1 and as a result

$$\frac{\partial \eta_b}{\partial \lambda} < 0$$

which completes the proof of Proposition 4.3. □

Proof of Proposition 4.4

Proof. Using equation (4.10), we can rewrite equations (4.7) and (4.9) as:

$$\begin{aligned} rv_b &= -\gamma v_b + \lambda \eta_s \frac{z}{1+z}(v_0 - v_b - v_s) \qquad (A.18) \\ rv_s &= d - x + \lambda \eta_b \frac{1}{1+z}(v_0 - v_b - v_s) \qquad (A.19) \end{aligned}$$

Subtracting equation (A.18) from equation (4.8) yields:

$$\begin{aligned} r(v_0 - v_b) &= d + \gamma(v_s - v_0) - \Big[-\gamma v_b + \lambda \eta_s \frac{z}{1+z}(v_0 - v_b - v_s) \Big] = \\ &= d + \gamma(v_s - v_0 + v_b) - \lambda \eta_s \frac{z}{1+z}(v_0 - v_b - v_s) \quad \Rightarrow \end{aligned}$$

$$\Rightarrow \quad v_0 - v_b = \frac{d + (\gamma + \lambda \eta_s \frac{z}{1+z}) v_s}{r + \gamma + \lambda \eta_s \frac{z}{1+z}} \qquad (A.20)$$

We can solve for v_s by plugging equation (A.20) into equation (A.19):

$$\begin{aligned}
rv_s &= d - x + \lambda\eta_b \frac{1}{1+z}\left[\frac{d + \left(\gamma + \lambda\eta_s \frac{z}{1+z}\right)v_s}{r + \gamma + \lambda\eta_s \frac{z}{1+z}} - v_s\right] = \\
&= d - x + \lambda\eta_b \frac{d - rv_s}{(r+\gamma+\lambda\eta_s)z + r + \gamma} \quad \Rightarrow \\
&\Rightarrow \quad \left[1 + \frac{\lambda\eta_b}{(r+\gamma+\lambda\eta_s)z + r+\gamma}\right] rv_s = \left[1 + \frac{\lambda\eta_b}{(r+\gamma+\lambda\eta_s)z + r+\gamma}\right] d - x \quad \Rightarrow
\end{aligned}$$

$$\Rightarrow \quad v_s = \frac{d}{r} - k\frac{x}{r} - k\frac{x}{(r+\gamma+\lambda\eta_s)z + \gamma} \tag{A.21}$$

where

$$k \equiv \frac{(r+\gamma+\lambda\eta_s)z + \gamma}{(r+\gamma+\lambda\eta_s)z + (r+\gamma+\lambda\eta_b)}$$

Given v_s, we can determined v_0, v_b and p uniquely from equations (4.8), (A.18) and (4.10) respectively. Let us compute them. We can solve for v_0 by plugging equation (A.21) into equation (4.8):

$$rv_0 = d + \gamma\left[\frac{d}{r} - k\frac{x}{r} - k\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\right] - \gamma v_0 \quad \Rightarrow$$

$$\Rightarrow \quad v_0 = \frac{d}{r} - k\frac{x}{r}\frac{\gamma}{r+\gamma} - k\frac{\gamma}{r+\gamma}\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma} \tag{A.22}$$

We now compute v_b by substituting equations (A.21) and (A.22) into equation (A.18):

$$
\begin{aligned}
rv_b &= -\gamma v_b + \lambda\eta_s \frac{z}{1+z}\Big[\frac{d}{r} - k\frac{x}{r}\frac{\gamma}{r+\gamma} - k\frac{\gamma}{r+\gamma}\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma} - \\
&- v_b - \frac{d}{r} + k\frac{x}{r} + k\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\Big] \quad \Rightarrow \\
&\Rightarrow \quad \left[r+\gamma+\lambda\eta_s\frac{z}{1+z}\right] v_b = \lambda\eta_s\frac{z}{1+z}\left[k\frac{x}{r+\gamma} + k\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\frac{r}{r+\gamma}\right] \quad \Rightarrow
\end{aligned}
$$

$$
\Rightarrow \quad v_b = k\frac{x}{r+\gamma}\frac{\lambda\eta_s z}{(r+\gamma+\lambda\eta_s)z+\gamma} \tag{A.23}
$$

We now solve for the price. Plugging equations (A.21) - (A.23) into equation (4.10) we get:

$$
\begin{aligned}
p &= \frac{1}{1+z}\Bigg[\left(\frac{d}{r} - k\frac{x}{r} - k\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\right)z + \\
&+ \frac{d}{r} - k\frac{x}{r}\frac{\gamma}{r+\gamma} - k\frac{\gamma}{r+\gamma}\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma} - k\frac{x}{r+\gamma}\frac{\lambda\eta_s z}{(r+\gamma+\lambda\eta_s)z+\gamma}\Bigg] = \\
&= \frac{d}{r} - \frac{1}{1+z}\left[k\frac{x}{r}\left(\frac{\gamma}{r+\gamma} + z\right) + k\frac{x}{r+\gamma}\right] \quad \Rightarrow
\end{aligned}
$$

$$
\Rightarrow \quad p = \frac{d}{r} - k\frac{x}{r} \tag{A.24}
$$

This concludes the proof of Proposition 4.4. □

4.8.2 Additional Proofs

Proof of $\frac{\partial p}{\partial d}, \frac{\partial p}{\partial \eta_b} > 0$ and $\frac{\partial p}{\partial z}, \frac{\partial p}{\partial \eta_s} < 0$

Proof. Using equation (4.14), the partial derivative of the price with respect to the dividend flow d is

$$\frac{\partial p}{\partial d} = \frac{1}{r} > 0 \quad \Rightarrow \quad \frac{\partial p}{\partial d} > 0 \quad \forall d$$

Let us now compute the partial derivative of the price with respect to the measure of buyers-to-be η_b:

$$\frac{\partial p}{\partial \eta_b} = \frac{\partial p}{\partial k}\frac{\partial k}{\partial \eta_b} = -\frac{x}{r}\frac{\partial k}{\partial \eta_b}$$

where:

$$\frac{\partial k}{\partial \eta_b} = -\lambda \frac{(r+\gamma+\lambda\eta_s)z+\gamma}{[(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)]^2}$$

which is strictly lower than zero since $r, \gamma, \lambda, \eta_s, z > 0$. Therefore,

$$\frac{\partial p}{\partial \eta_b} > 0 \quad \forall \eta_b$$

Next, we obtain the partial derivative of the price with respect to the buyer's-to-be bargaining power z:

$$\frac{\partial p}{\partial z} = \frac{\partial p}{\partial k}\frac{\partial k}{\partial z} = -\frac{x}{r}\frac{\partial k}{\partial z}$$

where:

$$\frac{\partial k}{\partial z} = \frac{r\lambda\eta_s\eta_b}{[(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)]^2}$$

which is strictly greater than zero since $r, \lambda, \eta_s, \eta_b > 0$. Then,

$$\frac{\partial p}{\partial z} < 0 \quad \forall z$$

To complete the proof, we calculate the partial derivative of the asset price with respect to the measure of sellers-to-be:

$$\frac{\partial p}{\partial \eta_s} = \frac{\partial p}{\partial k}\frac{\partial k}{\partial \eta_s} = -\frac{x}{r}\frac{\partial k}{\partial \eta_s}$$

where:

$$\frac{\partial k}{\partial \eta_s} = \frac{\lambda z(r+\lambda\eta_b)}{[(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)]^2}$$

which is strictly greater than zero since $r, \lambda, \eta_b, z > 0$. Thus,

$$\frac{\partial p}{\partial \eta_s} < 0 \quad \forall \eta_s$$

□

Proof of $\frac{\partial p}{\partial g} > 0$

Proof. Using equation (4.14), the partial derivative of the asset price with respect to the flow of investors g entering the market is

$$\frac{\partial p}{\partial g} = -\frac{x}{r}\frac{\partial k}{\partial g}$$

where $k = \frac{(r+\gamma+\lambda\eta_s)z+\gamma}{(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)}$. Let us derive the partial derivative of k with respect to the flow of investors g:

$$\frac{\partial k}{\partial g} = \frac{1}{[(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)]^2}\left\{\left(r+\lambda\eta_b\right)\lambda z\frac{\partial \eta_s}{\partial g} - \left[(r+\gamma+\lambda\eta_s)z+\gamma\right]\lambda\frac{\partial \eta_b}{\partial g}\right\}$$

which is strictly lower than zero since $r, \gamma, \lambda, \eta_s, \eta_b, z > 0$ and, as shown in Proposition 4.2, $\frac{\partial \eta_s}{\partial g} < 0$ and $\frac{\partial \eta_b}{\partial g} > 0$. Then,

$$\frac{\partial p}{\partial g} = -\frac{x}{r}\frac{\partial k}{\partial g} > 0$$

which proves the price increases in the flow of investors entering the market. □

4.8.3 Proof of Theorem 4.1

Proof. In our framework, the marginal investor decides whether to enter or not after comparing the expected utility v_{alt} of investing in an alternative market to the expected utility v_b of a buyer-to-be in our market. The expected utility of the marginal investor $v_{alt} = \kappa'$ is a non-negative and strictly increasing function of his outside investment opportunity κ'. Also, $v_b(\kappa' = 0) > v_{alt}(\kappa' = 0) = 0$. Hence, if v_b were decreasing in the outside investment opportunity of the marginal investor, κ', then there would be a unique threshold κ^* satisfying the indifference condition $v_b(g(\kappa^*)) = v_{alt}(\kappa^*)$. Let us show this is the case.

The expected utility v_b of a buyer-to-be is a function of the flow of investors g entering the market. Let us compute the partial derivative of v_b, defined in equation (4.11), with respect to g:

$$\frac{\partial v_b}{\partial g} = \frac{\lambda z x}{r+\gamma}\frac{1}{[(1+z)(r+\gamma)+\lambda(z\eta_s+\eta_b)]^2}\left\{\left[(1+z)(r+\gamma)+\lambda\eta_b\right]\frac{\partial \eta_s}{\partial g} - \lambda\eta_s\frac{\partial \eta_b}{\partial g}\right\}$$

which is strictly negative since $r, \gamma, x, z, \lambda, \eta_b, \eta_s > 0$ and, as shown in Proposition 4.2,

$\frac{\partial \eta_s}{\partial g} < 0$ and $\frac{\partial \eta_b}{\partial g} > 0$. Hence, the expected utility v_b of a buyer-to-be strictly decreases in the flow of investors g entering the market. However g, as given by $g(\kappa') = \int_{\underline{\kappa}}^{\overline{\kappa}} \nu(\kappa) f(\kappa) d\kappa = \int_{\underline{\kappa}}^{\kappa'} f(\kappa) d\kappa$, is increasing in κ'. As a result,

$$\frac{\partial v_b}{\partial \kappa'} = \frac{\partial v_b}{\partial g}\frac{\partial g}{\partial \kappa'} \leq 0$$

where $\frac{\partial v_b}{\partial g} < 0$ and $\frac{\partial g}{\partial \kappa'} \geq 0$.

Then, by continuity, there exists a unique value of κ' satisfying the indifference condition: $v_b(g(\kappa^*)) = v_{alt}(\kappa^*)$. A unique threshold κ^* thus defines a unique flow of investors $g^* = g(\kappa^*)$ entering the market. But given a flow of investors entering the market, there exists unique equilibrium measures (η_b^*, η_0^*, η_s^*) of each type of investor, expected utilities (v_b^*, v_0^*, v_s^*) and price of the asset, p^*, as proved in Propositions 4.1 and 4.4. Consequently, market equilibrium, as presented in Definition 4.1, is unique. This proves Theorem 4.1. □

4.8.4 Proof of Proposition 4.6

Proof. The illiquidity discount is defined as:

$$k\frac{x}{r}$$

where $\frac{x}{r}$ is the present value of the holding cost x, $k = \frac{(r+\gamma+\lambda\eta_s)z+\gamma}{(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)}$ and η_s and η_b, as given by equations (4.5) and (4.6), are functions of the flow of investors g. Let us compute the partial derivative with respect to the flow of investors g entering the market:

$$\frac{\partial}{\partial g}\left(k\frac{x}{r}\right) = \frac{x}{r}\frac{\partial k}{\partial g} < 0$$

since $\frac{\partial k}{\partial g} < 0$ (as shown in subsection 4.8.2) and $x, r > 0$. Hence, illiquidity decreases in g or equivalently, market liquidity increases in the flow of investors entering our market. $\square$

References

Acharya, V. V. and L. H. Pedersen (2005). Asset Pricing with Liquidity. *Journal of Financial Economics 77*(2), 375–410.

Amihud, Y. and H. Mendelson (1986). Asset Pricing and Bid-ask Spread. *Journal of Financial Economics 17*(2), 223–249.

Amihud, Y., H. Mendelson, and L. H. Pedersen (2005). Liquidity and Asset Prices. *Foundations and Trends in Finance 1*(4), 269–364.

BIS (1999, October). A Review of Financial Market Events in Autumn 1998. *CGFS Publication* (12). http://www.bis.org/ publ/ cgfs12.htm.

Brunnermeier, M. K. (2008). Deciphering the 20078 Liquidity and Credit Crunch. *Mimeo, Princeton University*. http://www.princeton.edu/ ~markus/ index.htm.

Brunnermeier, M. K. and L. H. Pedersen (2007). Market Liquidity and Funding Liquidity. *Mimeo, Princeton University*. http://www.princeton.edu/ ~markus/ research/ papers/ liquidity.

Burden, R. L. and J. D. Faires (1993). *Numerical Analysis.* Boston (Massachusetts): PWS Publishing Company.

Cohen, B. H. and E. M. Remolona (2001, December). Overview: Financial Markets Prove Resilient. *BIS Quarterly Review*. http://www.bis.org/ publ/ r_qt0112a.pdf.

Cooper, R. and A. John (1988). Coordinating Coordination Failures in Keynesian Models. *The Quarterly Journal of Economics 103*(3), 441–463.

Cooper, R. W. (1999). *Coordination Games: Complementarities and Macroeconomics.* Cambridge (United Kingdom): Cambridge University Press.

Diamond, P. A. (1982a). Aggregate Demand Management in Search Equilibrium. *Journal of Political Economy 90*(5), 881–894.

Diamond, P. A. (1982b). Wage Determination and Efficiency in Search Equilibrium. *The Review of Economic Studies 49*(2), 217–227.

Dow, J. (2004). Is Liquidity Self-Fulfilling? *Journal of Business 77*(4), 895–908.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005). Over-the-Counter Markets. *Econometrica 73*(6), 1815–1847.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2007). Valuation in Over-the-Counter Markets. *Review of Financial Studies*, forthcoming.

Duffie, D. and Y. Sun (2007). Existence of Independent Random Matching. *Annals of Applied Probability 17*(1), 386–419.

Evans, M., N. Hastings, and B. Peacock (1993). *Statistical Distributions (Second edition).* New York: Wiley-Interscience.

Fender, I. and P. Hördahl (2007). Overview: Credit Retrenchment Triggers Liquidity Squeeze. *BIS Quarterly Review*. http://www.bis.org/ publ/ qtrpdf/ r_qt0709a.pdf.

Greenlaw, D., J. Hatzius, A. K. Kashyap, and H. S. Shin (2008). Leveraged Losses: Lessons from the Mortgage Market Meltdown. *Mimeo, Princeton University*. http://www.chicagogsb.edu/usmpf/docs/usmpf2008confdraft.pdf.

Gromb, D. and D. Vayanos (2002). Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs. *Journal of Financial Economics 66*(2-3), 361–407.

Huang, J. C. and J. Wang (2007). Market Liquidity, Asset Prices, and Welfare. *Mimeo, McCombs School of Business*. http://papers.ssrn.com/ sol3/ papers.cfm?abstract_id=1009568.

IMF (1998, December). World Economic Outlook and International Capital Markets: Interim Assessment. *Mimeo, McCombs School of Business*. http://www.imf.org/ external/ pubs/ ft/ weo/ weo1298/ index.htm.

Lagos, R. and G. Rocheteau (2007). Liquidity in Asset Markets with Search Frictions. *Mimeo, New York University*. http://rl561a.googlepages.com/ workingpapers.

McAndrews, J. J. and S. M. Potter (2002). Liquidity Effects of the Events of September 11, 2001. *Federal Reserve Bank of New York Economic Policy Review*, 59–79.

Milgrom, P. and J. Roberts (1990). Rationalizability, Learning and Equilibrium in Games with Strategic Complementarities. *Econometrica 58*(6), 1255–77.

Mortensen, D. T. (1982, J.J. Mccall, ed.). The Matching Process as a Non-Cooperative/Bargaining Game. *The Economics of Information and Uncertainty New York: NBER Conference Volume.*

Pagano, M. (1989). Trading Volume and Asset Liquidity. *The Quarterly Journal of Economics 104*(2), 255–274.

Phelps, E. (1972). *Microeconomic Foundations of Employment and Inflation Theory.* New York: W. W. Norton & Company, Inc.

Pissarides, C. A. (1985). Short-Run Equilibrium Dynamics of Unemployment, Vacancies, and Real Wages. *American Economic Review 75*(4), 676–690.

Pissarides, C. A. (2001). Economics of Search. *International Encyclopedia of the Social & Behavioral Sciences*, 13760–13768.

Plantin, G. (2004). Self-Fulfilling Liquidity and the Coordination Premium. *Mimeo, London Business School*. http://ideas.repec.org/p/cmu/ gsiawp/ 710428070.html.

Topkis, D. M. (1979, November). Equilibrium Points In Nonzero-Sum n-Person Submodular Games. *SIAM Journal on Control and Optimization 17*(6), 773–787.

Vayanos, D. (1998). Transaction Costs and Asset Prices: A Dynamic Equilibrium Model. *Review of Financial Studies 11*(1), 1–58.

Vayanos, D. and T. Wang (2007). Search and Endogenous Concentration of Liquidity in Asset Markets. *Journal of Economic Theory*, forthcoming.

Vayanos, D. and P.-O. Weill (2007). A Search-Based Theory of the On-the-Run Phenomenon. *Journal of Finance*, forthcoming.

Weill, P.-O. (2007). Liquidity Premia in Dynamic Bargaining Markets. *Mimeo, UCLA*. http://www.econ.ucla.edu/ poweill/ liquiditypremia.pdf.