MATHEMATICAL MODELS FOR FINANCIAL DATA

HAOSHU TIAN

A DISSERTATION
PRESENTED TO THE FACULTY
OF PRINCETON UNIVERSITY
IN CANDIDACY FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE
BY THE PROGRAM IN
APPLIED AND COMPUTATIONAL MATHEMATICS
ADVISERS: WEINAN E AND CHRISTOPHER A. SIMS

JUNE, 2015
Abstract

The first chapter studies the effects of the temporary short-selling ban on US financial stocks in 2008 by overcoming a key problem in earlier empirical works of having to use non-financial stocks as a control group since almost all financial stocks were banned. Alternatively, the control group is selected to be a synthetic portfolio of non-banned financial stocks constructed from the finance segments of large industrial companies that were not banned. With this control group, it is found that the ban leads to over-valuation of banned stocks that is highest at the beginning of the ban and steadily converges to zero at the end of the ban. To understand this dynamic, the model of Scheinkman and Xiong (2003) is solved in a finite trading horizon with zero trading cost.

The default of one bank can cause other banks to default through two channels: financial contagion in the inter-bank liability network and fire sale in the asset selling market. In the second chapter jointly written with Weinan E, a model that incorporates these two channels is developed and analyzed theoretically. An algorithm for finding the state in which both the inter-bank liability network and the market are in equilibrium is proposed and tested.

An oft-mentioned but under-studied feature of asset price bubbles is a surge of new entrants, retail investors who never invested, joining the bubble because of their friends or neighbors. The third chapter, jointly written with Harrison Hong, incorporates this viral element into an otherwise standard bubble model with forward-looking agents. Optimism spreads across households following an epidemic process and the participation rate rise as new entrants buy anticipating trending prices. Insiders or institutions gradually sell their shares, generating trading volume and moderating price growth. This model rationalizes several patterns in the data, which have been difficult to explain, including trending prices and volume peaking months before prices, participation rates, and short-selling in both stock market and housing bubbles.
Acknowledgment

This dissertation would not have been possible without support from many people. First of all, I am extremely indebted to my thesis advisers Professor Weinan E and Professor Christopher A. Sims. Without your guidance and encouragement, I cannot imagine how I would overcome the problems and difficulties in my research. You have taught me how to ask questions, how to solve problems, and how to think as an applied mathematician and as an economist.

I would like to thank my committee member Professor Harrison Hong for your generous help, inspiring comments, and encouraging advice. I can still remember how you inspired me when I showed you the rudimentary draft of the first chapter, and how you encouraged me when I told you my decision for graduation. It was you who helped build up my confidence and courage to work independently.

I also benefited tremendously from conversations with Professor Wei Xiong, Professor Jose A. Scheinkman, Dr. Robert Jiang, Dr. Hong Ni, Dr. Qianxiao Li, Dr. Yuk-Fung Chan, Dr. Chu Wang, Professor James M. Stone, Professor Amit Singer, Professor Derek Lidow, Professor Birgit Rudloff, Professor Ramon van Handel, Professor Yacine Ait-Sahalia, Professor Valentin Haddad, Professor Atif Mian and Professor Nobuhiro Kiyotaki. I am also grateful to Professor Pingwen Zhang and Professor Changping Wang for introducing me to this program, and Audrey Mainzer, Christina Lipsky, Valerie Marino and Karen Neukirchen for your patience and help. I also enjoyed exchanging ideas with friends and students from Program in Applied and Computational Math, Department of Economics, Bendheim Center for Finance and Department of Operations Research and Financial Engineering.

Last but not least, my special thanks go to my beloved Yu Feng, for your love, support, and faith in me. No one needs you more than I need you. This thesis is dedicated to you and my beloved parents, my mother Zhihong Zhu and my father Xuehe Tian.
To my parents and my wife.
Contents

Abstract ................................................................. iii
Acknowledgment ....................................................... iv

1 The Effects of a Temporary Short-selling Ban 1
1.1 Introduction ......................................................... 2
1.2 Literature Review .................................................. 9
1.3 Empirical Analysis ............................................... 12
  1.3.1 The Financial Factor for a Given Stock ..................... 12
  1.3.2 Stock Return Data and the Universe ......................... 14
  1.3.3 Ban Names ...................................................... 14
  1.3.4 The Effect of the Short-Selling Ban ......................... 15
1.4 The Model ......................................................... 19
  1.4.1 Information Structure and Beliefs .......................... 20
  1.4.2 Price during the Short-selling Ban .......................... 22
  1.4.3 Dynamics of the Speculative Bubble ....................... 25
  1.4.4 Overvaluation Caused by the Ban ......................... 29
1.5 Test the Implications of the Model ......................... 31
  1.5.1 Understand the Initial Bubble Size ....................... 31
  1.5.2 Shorting Activity around the Short-Selling Ban .......... 33
  1.5.3 2008 Short-Selling Ban and Speculative Betas ............ 35
1.6 Conclusion ......................................................... 38
# List of Figures

1.1 The Effect of the Short-selling Ban without Controlling to be Financial 4  
1.2 The Effect of the Short-selling Ban on Financial Stocks (Introduction) 7  
1.3 The Effect of the Short-selling Ban on Financial Stocks (Empirical Analysis) 17  
1.4 Test the Robustness of the Definition of Financial Factor 19  
1.5 Average Speculative Bubble during the Ban 27  
1.6 Average Speculative Bubble during the Ban 28  
1.7 Understand the Bubble Sizes by Interest Rates and Lengths of the bans 33  
1.8 Shorting Activity around the Ban 34  
1.9 Short Interest for Financial Stocks 35  
1.10 The Effect of Short-Selling Ban on Low and High Beta Stocks 37  
1.11 Beta vs. Overpricing 38  

2.1 Number of bank failures based on the data from Federal Deposit Insurance Corporation (FDIC) 48  
2.2 The equilibrium of the system requires both the market and the payment network are in equilibrium 50  
2.3 Each bank's balance sheet contains two parts with the same value: assets and liabilities 51  
2.4 RMBS price fell to a liquidity freezing state 57  
2.5 Banks get "frozen" when the market gets illiquid 64
2.6 Number of defaults for different network connectivity .......................... 65
2.7 Leverage can play different roles ...................................................... 66
2.8 The Brazilian interbank network (Source: Edson Bastos e Santosy and Rama Cont, The Brazilian Interbank Network Structure and Systemic Risk, Banco Central do Brasil, 2010) .................................................. 67
2.10 Number of defaults for different network structure .......................... 71

3.1 Price and Volume Dynamics of the U.S. Housing Bubble and the China’s Stock Bubble ......................................................... 83
3.2 Realized Price and the Average Price .................................................. 96
3.3 Volume Peaks before Price ............................................................... 97
3.4 Volume Peaks Earlier when Population Increases Faster .................. 98
3.5 Volume Peaks later when Population Increases Slower ..................... 98
3.6 Number of Accounts in China Shanghai A Market ........................... 99
3.7 The Model Predicts Price and Volume Well ....................................... 101
3.8 Fit the Account Number by an Epidemic Model ............................... 102
3.9 Predicting Price and Volume by the Epidemic Model ....................... 103
List of Tables

1.1 Control and Treated Groups ........................................ 3
1.2 An Example of Segment SIC Codes .......................... 6
1.3 Previous Research on the 2008 Short-selling Ban .......... 12
1.4 2008 Business Segment Data for Hawaiian Electric Inds . 13
1.5 Sample Selection Procedure ................................. 15
1.6 Descriptive Statistics of the Sample ....................... 15
1.7 The Effect of the Short Selling Ban on Financial Stocks .... 18
1.8 Lengths of the Bans, Interest Rate and Initial Bubble Size . 32

2.1 The parameters that characterize the Erdős-Rényi network .... 62
2.2 The values of the parameters in simulation ................ 63
2.3 The parameters that characterize the Erdős-Rényi network .... 69
2.4 The values of the parameters in simulation ................ 70

3.1 Parameters for the Model ...................................... 100
3.2 Parameters Estimated by the Regression (I) .............. 104
3.3 Parameters Estimated by the Regression (II) .............. 104
3.4 Parameters Estimated by the Regression (III) .............. 105
Chapter 1

The Effects of a Temporary Short-selling Ban

This chapter studies the effects of the temporary short-selling ban on US financial stocks in 2008 by overcoming a key problem in earlier empirical works of having to use non-financial stocks as a control group since almost all financial stocks were banned. Alternatively, I use as a control group a synthetic portfolio of non-banned financial stocks constructed from the finance segments of large industrial companies that were not banned. This control group shares all the properties of the financial stocks without the ban. With this control group, I find that the ban leads to overvaluation of banned stocks that is highest at the beginning of the ban and steadily converges to zero at the end of the ban. To understand this dynamic, I solve the model of Scheinkman and Xiong (2003) in a finite trading horizon with zero trading cost. The speculative bubble caused by the ban becomes dynamic and satisfies a partial differential equation (PDE) involving time. The closed form solution of this PDE has the same behavior as found in the empirical finding. Other implications of the model are also discussed and tested. This chapter is presented at China International Conference in Finance 2014, Princeton Civitas Finance Seminar 2014 and PRS 2013.
1.1 Introduction

On September 19, 2008, the U.S. Securities and Exchange Commission (SEC) surprised the market by issuing an emergency order to temporarily ban short-selling of almost all financial stocks. The government presumably did this because they believed that a ban could support the current price of financial stocks. Over the past decade, an active literature has argued that short-selling constraints lead to overpricing because pessimistic investors sit out of the market. Miller (1977) points out this idea; Harrison and Kreps (1978) prove this in a discrete time model; and Scheinkman and Xiong (2003) show this in a continuous-time model.

However, recent empirical research on this short-selling ban finds mixed results. Harris et al. (2013) and Autore et al. (2011) find this short-selling ban resulted in significant overpricing. Boehmer et al. (2013) and Battalio et al. (2012), in contrast, find little concrete evidence that the ban led to a temporary upward bump in prices. These studies find conflicting evidence because they use different benchmarks to evaluate the abnormal return of the banned stocks.

More importantly, none of their benchmarks exclude the difference of financial and non-financial stocks. To clarify, they do not compare the banned financial stocks with the non-banned financial stocks (Group 1 and Group 2 in Table 1.1); they compare the banned financial stocks with the non-banned non-financial stocks (Group 1 and Group 3). To be specific, Boehmer et al. (2013) use matched non-banned stocks as the control group, where all the matched stocks are non-financial. Harris et al. (2013) and Battalio et al. (2012) compare the return of banned stocks with an index extracted from all the non-banned stocks, the vast majority of which are non-financial.

---

1See Section 1.2 for a detailed summary of the methodologies and findings of these previous studies.

2For each stock subject to the ban, Boehmer et al. (2013) choose a non-banned stock that is listed on the same exchange, has the same options listing status and the smallest distance, as measured by the absolute value of the proportional market-cap difference plus the absolute value of the proportional dollar trading volume difference.
Beber and Pagano (2011) try both of these methods. Based on the Fama and French (1993) three-factor model, Autore et al. (2011) calculate the abnormal return for banned stocks relative to the entire market, where financial stocks only account for a small proportion. Thus, their benchmarks are almost all non-banned non-financial stocks.

Table 1.1: Control and Treated Groups
When evaluating the effect of the 2008 short-selling ban, previous studies compare the banned financial stocks with the non-banned non-financial stocks (Group 1 and Group 3 in this table). The conclusions drawn from these comparisons can be misleading because the fundamental of financial stocks could be quite different from non-financial stocks during the 2008 financial crisis. To obtain a concrete conclusion about the effect of this ban, I compare the banned with the non-banned stocks given they are all financial (Group 1 and Group 2 in this table).

<table>
<thead>
<tr>
<th></th>
<th>Banned</th>
<th>Non-banned</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial</td>
<td>Group 1</td>
<td>Group 2</td>
</tr>
<tr>
<td>Non-financial</td>
<td></td>
<td>Group 3</td>
</tr>
</tbody>
</table>

Using the improper benchmark of non-banned non-financial stocks, the ban seems have a significant and permanent effect. The cumulative return of banned and non-banned stocks diverges at the beginning of the ban and the difference continues after the ban in the following graph.
Figure 1.1: The Effect of the Short-selling Ban without Controlling to be Financial

The following graph plots the cumulative estimated return of the banned (solid) and the non-banned (dashed) stocks, $cum_t(R^B)$, $cum_t(R^N)$. The vertical lines identify the first and the last day of the ban. The cumulative returns of the banned and the non-banned stocks diverge at the beginning of the ban. Without controlling for financial stocks, the difference does not shrink and exists even after the ban.

Banned vs. Non-banned (Not control Financial)

However, the conclusion drawn from this comparison can be misleading because the fundamental of financial stocks could be quite different from the fundamental of non-financial stocks during the 2008 financial crisis. On one hand, both the short-selling ban and the concomitant TARP (Troubled Asset Relief Program that targeted on all financial companies) could cause overpricing for the banned financial stocks. Even if overpricing is observed, no concrete conclusion can be drawn without fixing the treated and control groups as financial stocks. On the other hand, the negative effect of the bad fundamental news of financial stocks and the positive effect of the short-selling ban may be canceled. Thus, even though some researchers report little overpricing effect for the banned financial stocks, it is not sufficient to show that the ban is ineffective. Beber and Pagano (2011) summarized this drawback of the
literature as leaving only financial stocks in the treated group and only non-financial stocks in the control group.

However, comparing banned and non-banned financial stocks is difficult because there were few financial stocks not subject to the ban. Battalio et al. (2012) mentioned, “Since almost all financial stocks were targeted by the ban, it is difficult (if not impossible) to find an appropriate benchmark against which to evaluate their returns.”

To find a proper benchmark for evaluating the return of banned financial stocks, I use companies’ business segment data to decompose stock returns and find the factor return of non-banned financial stocks. This benchmark shares all the properties of financial stocks without the ban. When establishing the short-selling ban, the SEC identified banks, insurance companies, and securities firms by a group of 31 target Standard Industrial Classification (SIC) codes. A SIC code is a four digit number identifying the business type of a given company. Stocks of companies with SIC codes in the target group were banned from short-selling. There were also non-financial companies with part of its business in the financial sector. Since their company-level SIC codes did not belong to the target group, they were not included on the list of banned stocks by SEC. However, using one company’s business segment data, we can know the business-level SIC code and the proportion for each segment of its business. Then, we can find a financial but non-banned segment inside this stock. By extracting and aggregating the returns of this kind of non-banned financial segments, an appropriate benchmark can be formed to evaluate the returns of banned financial stocks. As an example, two stocks’ information is presented in Table 1.2. Group 1 is banned because 100% of its business is financial. Group 2 is not banned because its

---

3 According to the SEC (2008), these SIC codes are 6000, 6011, 6020, 6021, 6022, 6025, 6030, 6035, 6036, 6111, 6140, 6144, 6200, 6210, 6211, 6231, 6282, 6305, 6310, 6311, 6320, 6321, 6324, 6330, 6331, 6350, 6351, 6360, 6361, 6712, 6719

4 Some stocks with SIC codes not in the target group were also banned by stock exchanges afterward. This will be discussed in section 1.3.3.
principal business is non-financial, but it has 34.82% of its business that is financial.

Table 1.2: An Example of Segment SIC Codes
I use companies’ business segment data to identify financial portions of non-banned stocks. In this example, the first stock is banned because 100% of its business is financial. The second one is not banned because its principal business is non-financial, but there is still a portion of 34.82% of its business that is financial.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AHL</td>
<td>6331</td>
<td>Yes</td>
<td>Yes</td>
<td>Insurance</td>
<td>100%</td>
<td>6331</td>
<td>Yes</td>
<td>Group 1 Banned financial</td>
</tr>
<tr>
<td>HE</td>
<td>4911</td>
<td>No</td>
<td>No</td>
<td>Electric Utility</td>
<td>65.18%</td>
<td>4911</td>
<td>No</td>
<td>Group 2 Non-banned financial</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bank</td>
<td>34.82%</td>
<td>6035</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

Using this proper benchmark of non-banned financial stocks, a significant, dynamic but temporary effect of the ban is found. There is significant overpricing for the banned stocks at the beginning of the ban, while this effect shrinks and disappears at the end of the ban.
Figure 1.2: The Effect of the Short-selling Ban on Financial Stocks (Introduction)

The following graph plots the cumulative estimated return of the banned (solid) and the non-banned (dashed) financial stocks, \( \text{cum}_t(\hat{R}_{BF}^t), \text{cum}_t(\hat{R}_{NF}^t) \). The vertical lines identify the first and the last day of the ban. The cumulative returns of the banned and the non-banned financial stocks diverge at the beginning of the ban. The difference shrinks and disappears at the end of the ban. This result is consistent with the prediction made by the model for the temporary short-selling ban.

To model the temporary effect of the ban, a continuous time stock pricing model is developed in which investors have heterogeneous beliefs and short-selling is banned during a finite period. I use the same information structure as in the model of Scheinkman and Xiong (2003). There is only one stock traded in continuous time. The stock’s fundamental moves according to a mean-reverting process. Investors cannot observe the fundamental directly but can observe three processes driven by it: the dividend process and two signal processes. Investors optimally filter the information of the three processes to get their best estimation of the fundamental. However, investors are divided into two groups. Each group overestimates the precision of one of the signal processes. Thus they get different estimates of the fundamental. Overconfidence generates heterogeneous beliefs.
It’s assumed that the investors are risk-neutral and there is a limited supply of the stock. If there is only one group of competing investors, the stock price will be the investors’ expected discounted future dividends until the end of the trading. Investors form their expectation of future dividends based on their current beliefs of the fundamental. When there are two groups of investors with different beliefs, the short-selling ban makes the optimistic group hold all the shares of the stock. If the current pessimistic investors become optimistic in the future, they will buy the stock from the current owners. Being aware of this resale option, current owners value the stock for a certain amount more than their expected future dividends. This is the speculative bubble caused by the short-selling ban.

Unlike Scheinkman and Xiong (2003), a finite duration of the ban is assumed. Then the size of the bubble becomes dynamic and the remaining time to the end of the ban becomes crucial. By assuming zero trading cost, a partial differential equation (PDE) for the dynamic bubble is derived. Using a Laplace transform and Kummer functions, a closed form solution for the bubble is obtained. Then the bubble is averaged based on the stationary distribution of the process of the difference of beliefs. It is also analyzed that how economic parameters influence the magnitude and shape of the bubble. It is shown that the temporary ban leads to over-valuation that is highest at the beginning of the ban and steadily converges to zero at the end of the ban.

The predictions of the model and the results of the empirical analysis match well: there is significant overpricing for the banned stocks at the beginning of the ban, while this effect shrinks and disappears at the end of the ban. The model can also predict the initial bubble size precisely for other countries where the interest rates and the lengths of the ban are different. Similar to Hong and Sraer (2012), the model predicts higher overpricing for high beta stocks. This is also verified by the empirical analysis.
Section 1.2 summarizes related literature. Section 1.3 shows the method to determine the business segment proportion (the financial factor) and the process to prepare data. A statistical model is built to compare the return of banned and non-banned financial stocks. Section 1.4 presents a continuous-time model of a temporary short-selling ban and proves the theorem of shrinking bubble. More implications of the model are discussed and tested in Section 1.5.

1.2 Literature Review

Assuming heterogeneous beliefs among investors, Miller (1977) points out that short-selling constraints will result in overvaluation. If there are no short-selling constraints, the equilibrium price will stay between the valuations of optimistic and pessimistic investors. While pessimistic investors cannot short, optimistic investors will dominate the market and drive the equilibrium price up to their valuation. Chen et al. (2002) directly follow this idea and build a model using breadth of ownership as a proxy for short-selling constraints. Following this intuition, Harrison and Kreps (1978) build a infinite-horizon model in discrete time to find the equilibrium price scheme under short-selling constraints. They find overvaluation in their model caused by upward bias and speculative behavior. Scheinkman and Xiong (2003) prove this in a continuous-time model. Berestycki et. al. (2013) study the viscosity solutions of this model when the horizon is finite and the trading cost is non-zero. Hong and Stein (2003) explain a variety of stylized facts about crashes by assuming heterogeneous investors facing short-selling constraints. Hong et al. (2006) examine the relationship between this kind of speculative bubbles and asset float in a dynamic setting. Hong and Sraer (2012) also explain why credit bubbles are quieter (low volatility and turnover) than equity bubbles based on investor disagreement and short-sales constraints.

Diamond and Verrecchia (1987) build a model with rational learning market mak-
ers to show how short-selling constraints can affect the speed of price adjustment. They find that prohibiting traders from shorting reduces the adjustment speed of prices to private information. Since informed traders know the true value and market makers learn it rationally, the price will adjust to the true value sooner or later. Thus their model never generates overvaluation in equilibrium. However, in the model of this paper, a bubble is generated due to investors’ heterogeneous beliefs and speculative behaviors. This mechanism is different from previous bubble models (Allen et al. (2006), Tirole (1982), Blanchard and Watson (1982), Allen and Gorton (1993) and Allen et al. (1993))

During the financial crisis of 2008, a ban was placed on short-selling U.S. financial stocks from September 19 to October 8. Since only a portion stocks on the market were subject to the ban, this period offers a good opportunity for economists to evaluate the effect of the ban by comparing the banned and non-banned stocks in the same market.

Harris et al. (2013) compare the returns of the banned and the non-banned stocks by the Fama and French (1993) risk-factor model. The model coefficients for banned stocks are estimated with the one-year data before the ban. During the ban, factor returns are estimated by the returns of the non-banned stocks. Then the model coefficients and the factor returns are combined to generate the benchmark. They find price inflation of 10-12% in the banned stocks and attribute this to the ban. Autore et al. (2011) also use the Fama and French model, but estimate the factor returns by all the stocks. They find a positive abnormal return (2.70%) at the onset of the ban and a price reversal (-2.01%) at the removal of the ban. They also attribute the positive effect to the ban.

Battalio et al. (2012) compare the cumulative returns of the banned stocks with all the non-banned stocks. They find that the ban does little to slow the decline in the prices of financial stocks. For each stock subject to the ban, Boehmer et al.
choose a non-banned stock that is listed on the same exchange, has the same options listing status and has the smallest distance measured by the absolute value of the proportional market-cap difference plus the absolute value of the proportional dollar trading volume difference. They find that banned stocks rise by 6.68% on the first day of the ban, compared to an average 3.48% of matched non-banned stocks. Banned stocks fall by an average of 8.27%, compared to a price decline of 5.38% for the matched non-banned stocks. But the authors think that much of this is caused by good news about the fundamental, not the shorting ban. They come to this conclusion by looking at the subset of firms that are added to the ban list at a later date. Beber and Pagano (2011) test the effect of both the U.S. 2008 ban and other partial bans around the world. They compare the returns of the banned stocks with both the matched group and the whole non-banned group. In their Figure 6, the cumulative abnormal returns of the U.S. banned stocks increase from about 0.5% to 10% during the 14 trading days after the ban starts. However they find no significant abnormal return for banned stocks of other countries. Based on these results, they posit, “The effect of the ban on U.S. financial stock prices may be clouded by the concomitant TARP announcement, precisely aimed at supporting U.S. financial institutions”.

Three kinds of benchmarks are used as control groups in these studies: all the stocks, all the non-banned stocks and matched non-banned stocks. The vast majority (if not all) of the stocks in these control groups are non-financial. But the banned stocks in the treated group are almost all financial. However, the fundamental of financial stocks could be quite different from non-financial stocks during the 2008 financial crisis. Though they find more or less abnormal returns of the treated group, previous research disagree on whether it is the ban that causes this overpricing. Some of them attribute this overpricing instead to the concomitant Trouble Asset Relief Program (TARP). Finding a benchmark of non-banned financial stocks can resolve this debate.
Table 1.3: Previous Research on the 2008 Short-selling Ban

Three kinds of benchmarks are used as control groups in previous research: all the stocks, all the non-banned stocks and matched non-banned stocks. The vast majority (if not all) of the stocks in the control groups are non-financial. But the banned stocks in the treated group are almost financial. However, the fundamental of financial stocks could be quite different from non-financial stocks during the 2008 financial crisis. Though they find more or less abnormal returns of the treated group, previous research disagree on whether the ban causes this overpricing. Some of them attribute this overpricing to the concomitant Trouble Asset Relief Program (TARP). Finding a benchmark of non-banned financial stocks can resolve this debate.

<table>
<thead>
<tr>
<th>Control group</th>
<th>Methodology</th>
<th>Does the ban cause overpricing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harris et al. (2013)</td>
<td>All the non-banned stocks</td>
<td>Risk factor model</td>
</tr>
<tr>
<td>Autore et al. (2011)</td>
<td>All the stocks</td>
<td>Risk factor model</td>
</tr>
<tr>
<td>Battalio et al. (2012)</td>
<td>All the non-banned stocks</td>
<td>Cumulative return</td>
</tr>
<tr>
<td>Boehmer et al. (2013)</td>
<td>Matched non-banned stocks</td>
<td>Cumulative return</td>
</tr>
<tr>
<td>Beber and Pagano (2013)</td>
<td>All the non-banned stocks</td>
<td>Median excess return</td>
</tr>
</tbody>
</table>

1.3 Empirical Analysis

1.3.1 The Financial Factor for a Given Stock

As discussed in Section 1.1, the financial segments of non-financial firms are used as the benchmark to evaluate the effect of the ban. To calculate the proportion of financial segments in a firm, financial factor is defined as the following.

In the COMPUSTAT Database provided by S&P, the 2008 business segment data are used to calculate the financial factors. Each stock in this database has a SIC code to identify its principal business. Besides this, each stock has one or more business segments. For each segment, there is a specific SIC code,\(^5\) a record of segment asset,

\(^5\)There are segments with two SIC codes. When we use both the segment SIC and the additional SIC, we can identify 24 more targeted records than 565 records using only the first segment SIC (only 4.2% up). So we ignored the additional SIC code.
and a record of segment sales.\(^6\) (Table 1.4 is an example)

**Table 1.4: 2008 Business Segment Data for Hawaiian Electric Inds**

In the COMPUSTAT Database provided by S&P, each stock in this database has a SIC code to identify its principal business. Besides this, each stock has one or more business segments. For each segment, there is a specific SIC code, a record of segment asset, and a record of segment sales.

<table>
<thead>
<tr>
<th>Ticker</th>
<th>SIC</th>
<th>Segment</th>
<th>Segment SIC</th>
<th>Asset</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE</td>
<td>4911</td>
<td>Electric Utility</td>
<td>4911</td>
<td>3856.109</td>
<td>2860.177</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bank</td>
<td>6035</td>
<td>5437.12</td>
<td>358.553</td>
</tr>
</tbody>
</table>

In this paper, a stock or a business segment of a company is called financial if its SIC code belongs to the 31 financial SIC codes that the SEC aimed to ban.\(^7\) Since the SIC code 6035 in Table 1.4 is targeted, the bank segment of this company is called financial.

The financial factor for a given stock is defined by

\[
f = \frac{1}{2} \left( \frac{\sum_{\text{fin}} \text{Asset}}{\sum_{\text{all}} \text{Asset}} + \frac{\sum_{\text{fin}} \text{Sales}}{\sum_{\text{all}} \text{Sales}} \right)
\]

where \(\text{fin}\) is the index of all financial segments. The robustness of this definition is also tested in Figure 1.4. It turns out that very similar results are obtained for different weight that is assigned to asset and sales.

For the example above, we have

\[
f = \frac{1}{2} \left( \frac{5437.12}{3856.109 + 5437.12} + \frac{358.553}{2860.177 + 358.553} \right) = 34.82\%
\]

A stock with financial factor \(f = 1\) is a purely financial company. A stock with

\(^6\)There are segments created to record the eliminations due to accounting issue. These segments usually have names like “Elimination”, “Other”, “Adjustment” and so on. Sometimes, the Asset or Sales are negative for these segments. These segments are ignored during the calculation of financial factors. (Account for 2.1% of the total records)

\(^7\)According to the SEC (2008), these SIC codes are 6000, 6011, 6020, 6021, 6022, 6025, 6030, 6035, 6036, 6111, 6140, 6144, 6200, 6210, 6211, 6231, 6282, 6305, 6310, 6311, 6320, 6321, 6324, 6330, 6331, 6350, 6351, 6360, 6361, 6712, 6719
There are 5447 U.S. stocks in COMPUSTAT with segment data in 2008. Financial factors are calculated for 5446 stocks among them.\footnote{8}{One stock with ticker BAM has both financial SIC and non-financial SIC, but its Asset and Sales are all NA values. So its financial factor cannot be determined.}

1.3.2 Stock Return Data and the Universe

In the CRSP (the Center for Research in Security Prices) database, there are 3973 common U.S. stocks\footnote{9}{Common U.S. stocks are the stocks with share code SHRC\textup{D}=11 in CRSP database.} listed in the three major U.S. exchanges (NYSE, NASDAQ, and AMEX)\footnote{10}{Stocks listed in the 3 major exchanges are the stocks with exchange code EXCH\textup{C}D=1, 2, 3 in the CRSP database.} with full price data during the 58 trading days from August 19 (one month before the ban starts) to November 7, 2008 (one month after the ban ends\footnote{11}{November 8, 2008 is a Saturday. Thus the price data end on November 7, 2008.}) and with market capitalization above $20 million on September 18, 2008.

Based on a link between COMPUSTAT and CRSP, the intersection of the names with segment data and the names with return data is found. This intersection has 3139 names. The 3139 stocks in the 58 trading days are the universe of this research.\footnote{12}{See Table 1.5 for details of the sample selection procedure} The daily returns are recorded in a $3131 \times 58$ matrix $(r_{it})$.

1.3.3 Ban Names

The ban names come from two sources. The first source is the original 797 names listed by the SEC. All the 797 names were banned from the beginning of the ban. The second source is the stock exchange ban list.\footnote{13}{NYSE/NYSE Arca, NASDAQ, NYSE ALTERNEXT U.S./AMEX CONSOLIDATED HISTORICAL LIST OF ADDS/REMOVES for SS PROHIBITION http://www.nyse.com/attachment/CONSOLIDATED-SSPROHIBTION.xls} The SEC allowed NYSE, NASDAQ and AMEX to add and remove names during the ban.

All of the information about the short-selling ban is aggregated into a $3139 \times 58$ matrix. During the 14 days of the ban, if the $i^{th}$ stock among the 3139 stocks is banned on day $t$, then the ban indicator $b_{it} = 1$, otherwise 0. During the 23 trading
days before the ban, \( b_{it} = 1 \) if the stock \( i \) is banned from short-selling on the first day of the ban; during the 22 trading days after the ban, \( b_{it} = 1 \) if the stock \( i \) is banned from short-selling on the last day of the ban. The stocks with \( b_{it} = 1 \) before or after the ban serve as the placebo group: the banned and non-banned financial stocks should behave similarly before the ban, but differently during the ban.

**Table 1.5: Sample Selection Procedure**

In the CRSP (the Center for Research in Security Prices) database there are 3973 common stocks listed in the three major U.S. exchanges (NYSE, NASDAQ, and AMEX) with full price data during the 58 trading days from 19th Aug (one month before the ban starts) to 7th Nov 2008 (one month after the ban ends), and with market capitalization above $20 million on September 18, 2008. Based on a link between COMPUSTAT and CRSP, I find the intersection of the names with segment data and the names with return data. This intersection has 3139 names. The 3139 stocks in the 58 trading days are the universe of this research.

| Total Number of securities in CRSP with full return data during 8/19 to 11/7, 2008 | 6767 |
| Remove securities not in NYSE, NASDAQ, or AMEX | -266 |
| Remove securities other than common stocks | -2161 |
| Remove stocks with capitalization below $20 million | -367 |
| Remove stocks without business segment information | -842 |
| Stocks in final sample | 3139 |
| Sample stocks once banned from short-selling | 205 |

**Table 1.6: Descriptive Statistics of the Sample**

Stocks are grouped based on the financial factor \( f \). This table shows the number of banned and non-banned stocks in each \( f \)-group. One can find banned stocks in all the \( f \)-groups. The percentage of banned stocks is generally higher for larger \( f \).

<table>
<thead>
<tr>
<th>( f )</th>
<th>0&lt;( f )&lt;0.25</th>
<th>0.25&lt;( f )&lt;0.5</th>
<th>0.5&lt;( f )&lt;0.75</th>
<th>0.75&lt;( f )&lt;1</th>
<th>( f )=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Once Banned</td>
<td>44.00</td>
<td>5.00</td>
<td>10.00</td>
<td>7.00</td>
<td>36.00</td>
</tr>
<tr>
<td>Never Banned</td>
<td>2900.00</td>
<td>6.00</td>
<td>5.00</td>
<td>4.00</td>
<td>1.00</td>
</tr>
<tr>
<td>(Banned/All)</td>
<td>0.01</td>
<td>0.45</td>
<td>0.67</td>
<td>0.64</td>
<td>0.97</td>
</tr>
</tbody>
</table>

1.3.4 The Effect of the Short-Selling Ban

Daily return \( r_{it} \), ban indicator \( b_{it} \) and financial factors \( f_i \) are prepared for the
3139 stocks during the ban, \( i = 1, 2, \ldots, 3131 \), \( t = 1, 2, \ldots, 58 \). For each day \( t \), assume that the return of each stock can be decomposed into four parts: the factor return of Banned, Financial stocks \( R_{t}^{BF} \), the factor return of Non-banned Financial stocks \( R_{t}^{NF} \), the factor return of Banned, Non-financial stocks \( R_{t}^{BN} \), and the factor return of Non-banned Non-financial stocks \( R_{t}^{NN} \); then

\[
    r_{i,t} = f_i(b_{i,t}R_{t}^{BF} + (1 - b_{i,t})R_{t}^{NF}) + (1 - f_i)(b_{i,t}R_{t}^{BN} + (1 - b_{i,t})R_{t}^{NN}) + \epsilon_{i,t}
\]

Figure 1.2 shows the cumulative return of the banned financial stocks \( cum_t(\hat{R}^{BF}) \) and the cumulative return of the non-banned financial stocks \( cum_t(\hat{R}^{NF}) \) estimated by this linear regression. The vertical lines identify the first and the last day of the ban.
Figure 1.3: The Effect of the Short-selling Ban on Financial Stocks (Empirical Analysis)

The following graph plots the cumulative estimated return of the banned (solid) and the non-banned (dashed) financial stocks, $cum_t(\hat{R}^{BF})$, $cum_t(\hat{R}^{NF})$. The vertical lines identify the first and the last day of the ban. The cumulative returns of the banned and the non-banned financial stocks diverge at the beginning of the ban. The difference shrinks and disappears at the end of the ban. This result is consistent with the prediction made by the model for the temporary short-selling ban.

The cumulative returns of the banned and the non-banned financial stocks diverge at the beginning of the ban. The difference shrinks and disappears at the end of the ban. This result is consistent with the prediction made by the model for the temporary short-selling ban. Define $C_t = R_t^{BF} - R_t^{NF}$ to be the daily effect of the short-selling ban on financial stocks. A t-test shows that among the 58 days, only on the first day of the ban (September 19, 2008), the daily effect of the ban is statistically different from zero at the 1% level.\(^{14}\)

\(^{14}\)See Table 1.7 for more details
Table 1.7: The Effect of the Short Selling Ban on Financial Stocks
The factor return of banned and non-banned financial stocks are estimated by the statistical model in section 1.3. Define \( C_t := R^B_t - R^N_t \) to be the daily effect of the short-selling ban on financial stocks. A t-test shows that among the 58 days, only on the first day of the ban (20080919), the daily effect of the ban is statistically different from zero at 1% level. See Figure 1.2 for a plot of these two factor returns.

|        | Estimate | Std. Error | t value | Pr(>|t|) |
|--------|----------|------------|---------|---------|
| 20080904 | 0.004    | 0.008      | 0.449   | 0.653   |
| 20080905 | 0.003    | 0.007      | 0.375   | 0.707   |
| 20080908 | -0.026   | 0.011      | -2.478  | 0.013   |
| 20080909 | -0.005   | 0.010      | -0.557  | 0.578   |
| 20080910 | -0.002   | 0.009      | -0.177  | 0.859   |
| 20080911 | 0.009    | 0.008      | 1.174   | 0.241   |
| 20080912 | 0.018    | 0.009      | 2.037   | 0.042   |
| 20080915 | -0.005   | 0.011      | -0.435  | 0.663   |
| 20080916 | 0.008    | 0.012      | 0.672   | 0.501   |
| 20080917 | -0.003   | 0.011      | -0.288  | 0.773   |
| 20080918 | -0.004   | 0.018      | -0.213  | 0.832   |
| 20080919 | 0.048    | 0.016      | 3.028   | 0.002   |
| 20080922 | 0.021    | 0.015      | 1.361   | 0.174   |
| 20080923 | 0.002    | 0.013      | 0.164   | 0.870   |
| 20080924 | 0.003    | 0.013      | 0.254   | 0.799   |
| 20080925 | -0.005   | 0.012      | -0.389  | 0.698   |
| 20080926 | 0.009    | 0.014      | 0.646   | 0.518   |
| 20080929 | 0.013    | 0.021      | 0.621   | 0.535   |
| 20080930 | -0.001   | 0.020      | -0.066  | 0.947   |
| 20081001 | 0.018    | 0.015      | 1.189   | 0.234   |
| 20081002 | -0.016   | 0.016      | -1.002  | 0.317   |
| 20081003 | -0.012   | 0.014      | -0.888  | 0.375   |
| 20081006 | 0.020    | 0.019      | 1.042   | 0.298   |
| 20081007 | -0.034   | 0.018      | -1.856  | 0.063   |
| 20081008 | -0.005   | 0.019      | -0.285  | 0.776   |
| 20081009 | -0.047   | 0.022      | -2.159  | 0.031   |
| 20081010 | -0.024   | 0.031      | -0.782  | 0.434   |
| 20081013 | 0.036    | 0.030      | 1.189   | 0.235   |
| 20081014 | 0.041    | 0.023      | 1.814   | 0.070   |
| 20081015 | -0.016   | 0.021      | -0.762  | 0.446   |
| 20081016 | -0.025   | 0.025      | -0.987  | 0.324   |
| 20081017 | 0.007    | 0.018      | 0.372   | 0.710   |
| 20081020 | 0.011    | 0.018      | 0.612   | 0.540   |
| 20081021 | -0.011   | 0.017      | -0.638  | 0.524   |
| 20081022 | -0.015   | 0.019      | -0.769  | 0.442   |
The robustness of the definition of financial factor is also tested in the following figure.

**Figure 1.4: Test the Robustness of the Definition of Financial Factor**
The following graph tests the robustness of the definition of financial factor. Let 
\[ f_1 = \sum_{\text{Asset}} \frac{\text{fin}}{\text{all}} \text{Asset} \] and 
\[ f_2 = \sum_{\text{Sales}} \frac{\text{fin}}{\text{all}} \text{Sales} \], then the financial factor in this paper is defined as
\[ f = \frac{1}{2} f_1 + \frac{1}{2} f_2. \]
The cumulative return of banned financial stocks and non-banned financial stocks is plotted in the following graph when \( f_1, f_2 \) and \( f \) are used as the financial factor. It turns out that very similar results are obtained for different weight that is assigned to asset and sales.

**The Effect of the Ban on Financial Stocks**

1.4 The Model

In this model, the short-selling ban is effective during a finite period \([0, T]\). The information structure in section 1.4.1 is the same as Scheinkman and Xiong (2003). The price is dynamic because time matters in this finite period setting. A PDE is derived for the price and is solved in section 1.4.2 by Laplace transform and Kummer function. I prove that the speculative bubble caused by the ban shrinks during the ban and vanishes at the end of the ban.
Section 1.4.4 analyzes the trading in an infinite horizon $[0, \infty)$. The prices of two scenarios are compared. In the first scenario, short-selling is banned during $[0, T]$ with no ban afterwards. In the second scenario, there is no ban throughout the trading. Two components of the overvaluation are found: an upward bias caused directly by the ban and a bubble caused by speculative behaviors.

1.4.1 Information Structure and Beliefs

The assumptions in this section is similar to Scheinkman and Xiong (2003) despite having the trading only occur during $[0, T]$.

Risk neutral investors in Group A and B trade a stock during a finite period $[0, T]$. The stock pays dividends $D_t$ during this period. The dividend process is driven by a factor $f_t$ with a noise.

\begin{equation}
    dD_t = f_t dt + \sigma_D dZ^D_t
\end{equation}

The factor $f_t$ fluctuates around a constant $\bar{f}$.

\begin{equation}
    df_t = -\lambda (f_t - \bar{f}) dt + \sigma_f dZ^f_t
\end{equation}

The investors know this structure (The SDEs and the parameters) but cannot observe $f_t$. They can observe the dividends and two more signals driven by the factor $f_t$.

\begin{equation}
    ds^A_t = f_t dt + \sigma_s dZ^A_t
\end{equation}

\begin{equation}
    ds^B_t = f_t dt + \sigma_s dZ^B_t
\end{equation}

Investors are overconfident about their own signal. Investors in Group A believe that the signal $s^A_t$ contains more information than it does in reality as follows.
\[
\begin{aligned}
\begin{cases}
 ds_t^A &= f_t dt + \sigma_s \phi dZ_t^f + \sigma \sqrt{1-\phi^2} dZ_t^A \\
 ds_t^B &= f_t dt + \sigma_s dZ_t^B
\end{cases}
\end{aligned}
\]

(1.5)

With this belief, they can optimally filter the signals \( D_t, s_t^A, \) and \( s_t^B \) to form their beliefs \( \hat{f}_t^A \) of the factor \( f_t \).

\[
\begin{aligned}
d\hat{f}^A &= -\lambda(\hat{f}^A - \bar{f}) dt + \frac{\phi \sigma_s \sigma_f + \gamma}{\sigma_s^2}(ds^A - \hat{f}^A dt) + \frac{\gamma}{\sigma_D^2}(ds^B - \hat{f}^A dt) + \frac{\gamma}{\sigma_D^2}(dD - \hat{f}^A dt)
\end{aligned}
\]

(1.6)

Group A investors respond more to surprises in signal \( s_t^A \) than in signal \( s_t^B \) due to overconfidence. Investors in Group B are overconfident about signal \( s_t^B \) and form their belief \( \hat{f}_t^B \) in a similar way.

The difference of beliefs \( g^A = \hat{f}^B - \hat{f}^A \) satisfies a simple stochastic differential equation (SDE) based on the proposition 1 in Schenkman and Xiong (2003).

\[
dg_i = -\rho g_i dt + \sigma dW_i
\]

(1.9)

where \( \rho \) and \( \sigma \) are constant.

---

15 Based on the optimal filtering theory, the conditional beliefs about \( f_t \) of investors in group A are Gaussian with mean \( \hat{f}_t^A \) and variance \( \gamma \), where \( \gamma \) is a constant \( \gamma \equiv -\lambda/(\lambda^2 + \lambda \sigma_f^2 + (1-\phi^2)^2) \). Following Schenkman and Xiong (2003), I call the mean \( \hat{f}_t^A \) to be the belief of Group A investors.

16 Investors in Group B believe that

\[
\begin{aligned}
\begin{cases}
 ds_t^A &= f_t dt + \sigma_s dZ_t^A \\
 ds_t^B &= f_t dt + \sigma_s \phi dZ_t^f + \sigma \sqrt{1-\phi^2} dZ_t^B
\end{cases}
\end{aligned}
\]

(1.7)

Based on this, they form their belief \( \hat{f}_t^B \) on the factor \( f_t \).

\[
\begin{aligned}
d\hat{f}^B &= -\lambda(\hat{f}^B - \bar{f}) dt + \frac{\gamma}{\sigma_s^2}(ds^A - \hat{f}^B dt) + \frac{\phi \sigma_s \sigma_f + \gamma}{\sigma_s^2}(ds^B - \hat{f}^B dt) + \frac{\gamma}{\sigma_D^2}(dD - \hat{f}^B dt)
\end{aligned}
\]

(1.8)

17 \( \rho = \sqrt{(\lambda^2 + \phi \sigma_f^2 + (1-\phi^2) \sigma_f^2 \phi^2)} \), \( \sigma = \sqrt{2 \phi \sigma_f} \)
1.4.2 Price during the Short-selling Ban

At time $t \in [0, T]$, Group A investors’ expectation of the discounted future dividends is\(^{18}\)

\[
E_t^A \left[ \int_t^T e^{-r(s-t)} dD_s \right] = \frac{\hat{f}}{r} (1 - e^{-r(T-t)}) + \frac{\hat{f}_t^A - \hat{f}}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)}) \tag{1.10}
\]

If there are only Group A investors in the market, then the price of the stock will equal to this expectation. However, there are also Group B investors in the market with a different expectation of future dividends

\[
E_t^B \left[ \int_t^T e^{-r(s-t)} dD_s \right] = \frac{\hat{f}}{r} (1 - e^{-r(T-t)}) + \frac{\hat{f}_t^B - \hat{f}}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)}) \tag{1.11}
\]

Since there is a short-selling ban, pessimistic investors with a lower expectation will sit out of the market. Optimistic investors with a higher expectation will be the owners of the stock and set the price. It is possible that the pessimistic investors will become optimistic in the future and want to buy the stock from the current owner. Thus when setting the price the current owner will also consider the value of the resale option. So the price is the current owner’s expectation of future dividends plus a resale option.

\[
p_t^p = p(\hat{f}_t^p, g_t^p, t) = \frac{\hat{f}}{r} (1 - e^{-r(T-t)}) + \frac{\hat{f}_t^p - \hat{f}}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)}) + q(g_t^p, t) \tag{1.12}
\]

\(^{18}\)The method to calculate the expected future dividends is as follows: since the expectations of all the stochastic parts are zero, I ignore the stochastic terms when calculating the expectation. Thus by equation (1.6), one has $d\hat{f}^A = -\lambda(\hat{f}^A - \hat{f})dt$; the solution to this ODE is $\hat{f}_s^A = \hat{f} + C e^{-\lambda(s-t)}$. The constant $C$ can be determined by letting $s = t$. Thus $C = \hat{f}_t^A - \hat{f}$. Thus $\hat{f}_s^A = \hat{f} + (\hat{f}_t^A - \hat{f})e^{-\lambda(s-t)}$. Ignoring the stochastic part, based on equation (1.1), one has $dD_s = f_s ds$. Thus $E_t^A[\int_t^T e^{-r(s-t)} dD_s] = \int_t^T e^{-r(s-t)} [\hat{f} + (\hat{f}_t^A - \hat{f})e^{-\lambda(s-t)}] ds = \frac{\hat{f}}{r} (1 - e^{-r(T-t)}) + \frac{\hat{f}_t^A - \hat{f}}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)})$
where the superscript "o" means owner and "\bar{o}" means the group who does not own the stock. \( o = A \) and \( \bar{o} = B \) if \( \hat{f}^A_t > \hat{f}^B_t \), and vice versa. The value of the resale option \( q(g^o_t, t) \) depends on the difference of beliefs \( g^o = \hat{f}^\bar{o} - \hat{f}^o \) between the two groups.

The owner determines an optimal time \( \tau \) to sell the stock to maximize the dividends collected and the capital gain. \(^{19}\)

\[
p^o_t = \sup_{t \leq \tau \leq T} \mathbb{E}^o_t \left[ \int_{t}^{\tau} e^{-r(s-t)} dD_s + e^{-r(\tau-t)} p^\bar{o}_{\tau} \right]
\]

(1.13)

At the end of the trading, the price is \( p^o_T = 0 \) because there are no more dividends to be collected and there are no more chances to profit from re-selling the stock. Further, the price is required to be smooth in its variables. This technical requirement guarantees the uniqueness of the solution.

Express \( p^o_t \) and \( p^\bar{o}_{\tau} \) in the optimal decision equation (1.13) by the candidate equilibrium price equation (1.12), one can derive the equation for the resale option. Together with the boundary condition and technical requirement, one has

\[
\begin{align*}
    q(g^o_t, t) &= \sup_{t \leq \tau \leq T} \mathbb{E}^o_t \{ e^{-r(\tau-t)} \left[ \frac{g^o_t}{1 + \lambda} (1 - e^{-(r+\lambda)(T-\tau)}) \right] + q(g^\bar{o}_t, \tau) \} \\
    q(g^o_t, T) &= 0 \\
    q &\text{ smooth}
\end{align*}
\]

(1.14)

This is an optimal execution problem. Let \( x = g^o_t - \hat{f}^0_t = \hat{f}^\bar{o}_t - \hat{f}^o_t \) be the current belief difference among investors. If \( x < 0 \), the owner is more optimistic and chooses to not to sell and wait\(^{20}\). Then the discount value of this resale option \( e^{-r\tau} q(g^o_t, t) \) is a

\(^{19}\)I do not consider trading cost in this model. This will simplify the model technically. However, the speculative bubble still exists like in Scheinkamm and Xiong (2003).

\(^{20}\)If there is a trading cost, investors will choose to wait if \( x < k \), where \( k \) is a positive number that depends on the cost and remaining time of trading. However, if there is no trading cost, as was assumed here, trading will occur whenever the beliefs cross. Thus the owner will wait if \( x < 0 \) and

23
martingale to the current holder. Based on Ito’s lemma and the evolution equation for \( g^\rho_t \), one has
\[
\frac{1}{2} \sigma^2 g_{xx} - \rho x g_x + q_t - rq = 0, \quad x < 0
\]

If the current belief difference is non-negative, \( x \geq 0 \), the owner is more pessimistic and chooses to sell right now. Then the optimal execution time is equal to current time \( \tau = t \) and the optimal value of the right hand side is realized in the first equation of (1.14) right now.

\[
q(x, t) = \frac{x}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)}) + q(-x, t), \quad x \geq 0
\]

The requirement of \( q(x, t) \) being smooth in equation (1.14) means \( q(x, t) \) and \( q_x(x, t) \) are continuous when \( x \) cross 0.

Since \( \lim_{x \to 0^+} q(x, t) = \lim_{x \to 0^+} \left[ \frac{x}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)}) + q(-x, t) \right] = q(0, t) \) and \( \lim_{x \to 0^-} q(x, t) = q(0, t) \), the continuation of \( q(x, t) \) is satisfied automatically.

Since \( \lim_{x \to 0^+} q_x(x, t) = \lim_{x \to 0^+} \left[ \frac{1}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)}) - q_x(-x, t) \right] = \frac{1}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)}) - \frac{1}{2(r + \lambda)} (1 - e^{-(r+\lambda)(T-t)}) \) and \( \lim_{x \to 0^-} q_x(x, t) = q_x(0, t) \), the continuation of \( q_x(x, t) \) means \( q_x(0, t) = \frac{1}{2(r + \lambda)} (1 - e^{-(r+\lambda)(T-t)}) \).

Thus the optimal execution problem is transformed to a PDE problem for \( x < 0 \) and \( 0 \leq t \leq T \).

\[
\begin{cases}
\frac{1}{2} \sigma^2 g_{xx} - \rho x g_x + q_t - rq = 0 \\
q(x, T) = 0 \\
q_x(0, t) = \frac{1}{2(r + \lambda)} (1 - e^{-(r+\lambda)(T-t)})
\end{cases}
\]

If the second order derivative term is \( \frac{1}{2} \sigma^2 g_{xx} \) instead of \( \frac{1}{2} \sigma^2 x^2 g_{xx} \), then it is the Black–Scholes equation and can be solved by transforming it to a diffusion equation. sell if \( x \geq 0 \). This is similar to the theorem 1 in Scheinkman and Xiong (2003).
Without the $x^2$ term, the problem becomes harder. Given the equation above, if the boundary condition is the Dirac delta function, physicists can solve it by performing a Fourier transform. With the more complicated boundary condition, the PDE becomes much harder to solve.

By carefully performing a Laplace transform on the equation and using the integral form of the Kummer function, I solved this equation. Based on the proof in the appendix, one has

**Theorem 1.** The solution of equation (1.15) is

$$u(x, t) = \frac{\sigma}{4(r + \lambda)\sqrt{\pi \rho}} \int_{N(t)}^{\infty} e^{-\frac{\rho}{\sigma^2} x^2 y} \frac{r\rho e^{-2y\rho}}{2\rho} (1 + y)^{-\frac{r\rho}{2\rho} \left[1 - e^{-(r+\lambda)(T-t-M(y))}\right]} dy$$

where $N(t) = \frac{e^{-2\rho(T-t)}}{1-e^{-2\rho(T-t)}}$, $M(y) = \frac{1}{2\rho} \ln(1 + \frac{1}{y})$, $x < 0$ and $0 \leq t \leq T$.

Based on Theorem 1, we know that the speculative bubble (the value of the resale option) is

$$q(x, t) = \begin{cases} \frac{x}{r+\lambda}(1 - e^{-(r+\lambda)(T-t)}) + u(-x, t) & x \geq 0 \\ u(x, t) & x < 0 \end{cases} \quad (1.16)$$

**Theorem 2.** The speculative bubble $q(x, t)$ is decreasing in $t$ and equals 0 when $t = T$.

From theorem 2, we know that this speculative bubble shrinks during the ban and vanishes at the end of the ban.

### 1.4.3 Dynamics of the Speculative Bubble

According to Theorem 1, the magnitude of the speculative bubble $q(x, t)$ is determined by time $t$ and $x$, the current difference of beliefs in the owners’ minds. Averaging the stochastic term $x$ in $q(x, t)$ can lead to the mean magnitude of the bubble $\bar{q}(t)$, which only depends on time.
The difference of beliefs in Group A investors’ minds $g^A = \hat{f}^B - \hat{f}^A$ satisfies a simple SDE based on equation (1.9)

$$dg^A = -\rho g^A dt + \sigma dW_g$$

The stationary distribution of this process is normal, $N(0, \eta^2)$, where $\eta^2 = \frac{\sigma^2}{2\rho}$. The same distribution applies to $g^B$ since $g^B = \hat{f}^A - \hat{f}^B = -g^A$. When Group A investors hold the stock, $\hat{f}^B < \hat{f}^A$, thus $x = g^A < 0$. When Group B investors hold the stock, $\hat{f}^A < \hat{f}^B$, thus $x = g^B = -g^A < 0$. So $x$ is always less than or equal to zero with a stationary distribution that doubles the left half of the normal distribution $N(0, \eta^2)$:

$$w(x)dx = 2 \cdot \frac{1}{\eta\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\eta^2}} dx \quad x < 0 \quad (1.17)$$

**Theorem 3.** Averaging the speculative bubble $q(x, t)$ in equation (1.16) with the stationary distribution $w(x)dx$ in equation (1.17) leads to the mean magnitude of the bubble

$$\bar{q}(t) = \frac{\sigma \sqrt{\rho}}{2\lambda \sqrt{\pi}} \left[ \frac{1}{r} (1 - e^{-r(T-t)}) - \frac{1}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)}) \right].$$

Based on theorem 3, the size of the bubble increases linearly in $\sigma$, the volatility of the disagreement process. It also increases linearly in $\sqrt{\rho}$, the square root of the mean reversion coefficient of the disagreement process. When the disagreement process is more volatile, it’s more likely to observe extreme disagreement and trading occurs more frequently. This leads to a higher speculative bubble. Increasing the mean reversion coefficient will also increase the chance for investors’ beliefs to cross. This increases the speculative bubble by increasing the chance of resale.

Generally, increasing $r$ will make the size of the bubble smaller because it decreases the present value of the future resale price, which makes investors less willing
to speculate. Increasing $\lambda$ will also decrease the size of the bubble because it makes the fundamental converge to the long-term mean faster, which decreases the excess dividends that optimistic investors were expecting to collect. This makes the optimistic investors not that optimistic and makes the bubble smaller. The following figures show the different dynamics.

Figure 1.5: Average Speculative Bubble during the Ban
The x-axis shows the time starting at 0 and ending at $T$. The y-axis shows the average overvaluation caused by the ban. Increasing $r$, the interest rate, will make the size of the bubble smaller because it decreases the present value of the resale price, which makes investors less willing to speculate.

![Dynamics of the Bubble for Different $r$](diagram.png)
The interest rate $r$ and the mean reversion coefficient of the true fundamental $\lambda$ not only influence the size of the bubble, but also influence the shape of the bubble as described in the following theorem.

**Theorem 4.** The average bubble $\bar{q}(t)$ is decreasing in $t$ and vanishes at the end of the ban. Let $\tau^* = \frac{1}{\lambda} \ln(1 + \frac{r}{\lambda})$; then the inflection point of $\bar{q}(t)$ is at $t^* = T - \tau^*$. When $t < t^*$, $\bar{q}(t)$ is concave; when $t^* < t < T$, $\bar{q}(t)$ is convex. At the inflection point, the bubble has the highest decreasing speed $\frac{d\bar{q}(t)}{dt} \big|_{t=t^*} = \frac{-\sigma \sqrt{\pi}}{2(r+\lambda) \sqrt{\pi}} (\frac{r}{r+\lambda})^\frac{3}{4}$. 

28
Based on Theorem 4, the remaining time when the bubble crosses an inflection point is $\tau^* = \frac{1}{\lambda} \ln(1 + \frac{\lambda}{r})$. When the remaining time $T - t$ is bigger than $\tau^*$, the curve of the bubble is concave. When it’s very close to the end of the ban (i.e., $T - t$ is smaller than $\tau^*$), the curve is convex. As $r \to +\infty$, $\tau^* \to 0$. This means the curve is always concave, the highest speed of decreasing occurs at the end of the ban. As $r \to 0$, $\tau^* \to +\infty$. The whole curve will be convex. As $\lambda \to +\infty$, $\tau^* \to 0$ and as $\lambda \to 0$, $\tau^* \to \frac{1}{r}$. $r$ and $\lambda$ together determine the position of the inflection point and thus the shape of the bubble curve.

Fixing the other parameters, increasing $T$ can be realized by moving the $t = 0$ point to the left (extend the curve to the left such that it takes time $T$ to vanish). Bigger $T$ leads to a higher initial bubble; however, the bubble won’t go to infinity. Actually as $T \to +\infty$, $\bar{p}(t) \equiv \bar{p}_{T=\infty} = \frac{\sigma \sqrt{\bar{p}}}{2(r + \lambda) \sqrt{\pi}}$. An infinite ban leads to a constant average bubble.

1.4.4 Overvaluation Caused by the Ban

When there is no short-selling ban, assume that the price is just the average of the two groups’ expected future dividends. One group longs the stock while the other group shorts. The price lies in the middle of their beliefs. Instead of ending the trading at time $T$, consider trading during $[0, \infty)$. At time $t \in [0, T]$, the expected sum of future dividends for Group A is

$$\mathbf{E}_{t}^{A} \left[ \int_{t}^{\infty} e^{-r(s-t)} dD_{s} \right] = \mathbf{E}_{t}^{A} \left[ \int_{t}^{T} e^{-r(s-t)} dD_{s} + \int_{T}^{\infty} e^{-r(s-t)} dD_{s} \right] = A_{t}^{T} + A_{T}^{\infty} \quad (1.18)$$

where $A_{t}^{T} = \frac{\bar{f}}{r} (1 - e^{-r(T-t)}) + \frac{\bar{f}^{A} - \bar{f}}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)})$ is Group A’s expectation of the dividends to be collected during the period $[t, T]$, $A_{T}^{\infty} = \frac{\bar{f}}{r} e^{-r(T-t)} + \frac{\bar{f}^{A} - \bar{f}}{r + \lambda} e^{-(r+\lambda)(T-t)}$ is Group A’s expectation of the dividends to be collected after $T$. Similarly, B’s expected future dividends come from two parts.
\[ B_t^\infty = B_t^T + B_t^\infty \]

When there is no short selling ban through \([0, \infty)\), the price is

\[ p_t^{NonBan} = \frac{1}{2} (A_t^T + B_t^T) + \frac{1}{2} (A_t^\infty + B_t^\infty). \]

When there is a short selling ban through \([0, T]\) and no ban afterwards, since investors know that the ban will end at \(T\), they agree with the average price of the dividends after the ban, \(\frac{1}{2} (A_T^\infty + B_T^\infty)\). Thus at \(t \in [0, T]\), the price comes from two parts: one is this average price of the dividends after the ban, the other is the ban-period price I’ve calculated in the previous sections. In the previous sections, I end the trading at \(T\) and find the price consists of the optimistic investors’ expectation of the dividends during the ban and the speculative bubble. Thus

\[ p_t^{Ban} = max(A_t^T, B_t^T) + q(g_t^o, t) + \frac{1}{2} (A_t^\infty + B_T^\infty). \]

Thus the overvaluation caused by the ban is

\[ V_t = p_t^{Ban} - p_t^{NonBan} = \frac{1}{2} |A_t^T - B_t^T| + q(g_t^o, t). \]

By the definition of \(A_t^T\) and \(B_t^T\), one has

\[ V_t = \frac{\bar{f}_t^A - \bar{f}_t^B}{2(r + \lambda)} (1 - e^{-(r+\lambda)(T-t)}) + q(g_t^o, t). \]

The first part of the overvaluation is the upward bias directly caused by the belief difference. It will shrink over the duration of the ban and vanish at the end of the ban because there is less time to collect dividends. The second part of the overvaluation is the bubble caused by speculative behaviors. It will also shrink over the duration of the ban and vanish at the end of the ban as was proved in the previous section. It shrinks
because there is less time to re-sell the stock to the counter-party speculatively.

1.5 Test the Implications of the Model

1.5.1 Understand the Initial Bubble Size

Based on theorem 3, the average size of the bubble is determined by the length of the ban, interest rate and other parameters of the underlying belief dynamics. The US, the UK, Canada and Germany had different interest rates and announced partial bans with different lengths. This section uses the length and interest rate data to understand the initial bubble size based on the model developed in this paper.

For Canada, the UK and Germany, the initial bubble size is measured by the abnormal return of banned financial stocks at the beginning of ban. The control group is constructed by the non-banned stocks with targeted financial SIC codes in the same exchange: the Toronto stocks exchange, the London stock exchange and the Frankfurt stock exchange. The returns are averaged by capitalization. The lengths of the bans are the lengths announced in the first announcement in each country. Some of the countries extended the ban, but this information should not be used to understand the initial bubble size. The interest rate is the central bank rate in each country during September 2008. The following table summarizes the information of these bans.
Table 1.8: Lengths of the Bans, Interest Rate and Initial Bubble Size

For Canada, the UK and Germany, the initial bubble size is measured by the abnormal return of banned financial stocks at the beginning of ban. The control group is constructed by the non-banned stocks with targeted financial SIC codes in the same exchange: the Toronto stocks exchange, the London stock exchange and the Frankfurt stock exchange. The returns are averaged by capitalization. The lengths of the bans are the lengths announced in the first announcement in each country. Some of the countries extended the ban afterwards, but this information should not be used to understand the initial bubble size. The interest rate is the central bank rate in each country during September 2008.

<table>
<thead>
<tr>
<th></th>
<th>Length Announced (days)</th>
<th>Interest Rate (%)</th>
<th>Initial Bubble Size(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>9</td>
<td>3.00</td>
<td>2.30</td>
</tr>
<tr>
<td>US</td>
<td>10</td>
<td>2.00</td>
<td>4.80</td>
</tr>
<tr>
<td>Germany</td>
<td>73</td>
<td>4.25</td>
<td>10.20</td>
</tr>
<tr>
<td>UK</td>
<td>84</td>
<td>5.00</td>
<td>7.60</td>
</tr>
</tbody>
</table>

I assume the parameters of the underlying belief dynamics are the same across all the countries. I recover them by applying the least square method on the data of interest rates, lengths of the bans and observed initial bubble sizes. Then the predicted and observed initial bubble size is plotted. The shape of the curve predicted by the model match the data very well.
1.5.2 Shorting Activity around the Short-Selling Ban

The overpricing effect of the two treated groups is attributed to the short-selling ban. It is good to test whether the treated groups really suffer from less shorting activity than usual. During the 28 days from September 10 to October 18, 2008 there are 2282 NYSE stocks with full daily short transaction data.\footnote{The data are taken from the Princeton University database of NYSE short sales.} These 2282 stocks
are put into two groups: one group of 261 stocks that were banned throughout or at some point during the ban and another group of 2021 stocks that were never banned. For each group the daily median shorting activity is plotted in the Figure 1.8. The shorting activity is defined to be the number of round lots transacted in all the shorting trades of stock \( i \) on day \( t \). The vertical lines are the first and last day of the short-selling ban.

**Figure 1.8: Shorting Activity around the Ban**

For both the once-banned stocks and the never-banned stocks, the daily median shorting activity is plotted here. The shorting activity \( s_{it} \) is defined to be the number of round lots transacted in all the shorting trades of stock \( i \) on day \( t \). The vertical lines are the first and last day of the short-selling ban. For the group of stocks that were once banned from short-selling (solid line), shorting activity was high before and after the ban, but was low during the ban. For the Never-Banned group (dashed line), shorting activity was constantly low. The shorting activity for banned stocks was lower than the non-banned stocks but was not zero because market makers were still allowed to provide liquidity by shorting.

For the group of stocks that were once banned from short-selling (solid line), shorting activity was high before and after the ban but was low during the ban. For the Never-Banned group (dashed line), shorting activity was constantly low. The following link provides more details on the data:

http://dss2.princeton.edu/data/3151/2008/nyseshortsales200809/
shorting activity for banned stocks was lower than the non-banned stocks but was not zero because market makers were still permitted to provide liquidity by shorting.\textsuperscript{22}

Similar results are found if the two groups are controlled to be financial.

\textbf{Figure 1.9: Short Interest for Financial Stocks}

Short interest can also be decomposed: \( s_{it} = f_i b_i S_{iF}^t + f_i (1-b_i) S_{iF}^{NF} + (1-f_i) b_i S_{iF}^{BN} + (1-f_i)(1-b_i) S_{iF}^{NN} + \epsilon_{it} \). This is the graph for the short interest of banned and non-banned financial stocks \( S_{iF}^t \) and \( S_{iF}^{NF} \) ("financial" is fixed; compare the once banned and never banned groups). Shorting activity of the banned financial group was high before and after the ban, but was low during the ban. Shorting activity of the non-banned financial group was constantly low.

1.5.3 2008 Short-Selling Ban and Speculative Betas

Are high beta assets more prone to speculative overpricing? Hong and Sraer (2012) gave a positive answer in their model: if investors have different beliefs about the common factor of the stocks, high beta will amplify this divergence of beliefs while low beta will narrow this divergence. With the short-selling ban, there should be more overpricing for high beta stocks based on the overvaluation theory of Miller (1977).

\textsuperscript{22}According to the SEC (2008), they were providing a limited exception for certain bona fide market makers.
This section tests their predictions based on the factor model developed in this paper. The betas for the 3139 stocks in my universe are calculated based on the winsorized 1-year daily return\textsuperscript{23}. The factor model can distinguish the effect of the ban on both financial and non-financial stocks based on the segment data and detailed ban names. For example, when only inputting the information of the low beta stocks, the model decomposes the returns of the low beta stocks into four parts: Banned Financial, Banned Non-financial, Non-banned Financial, and Non-banned Non-financial. By comparing these four return series, one can tell the magnitude of the overpricing on low beta stocks during the ban.

Figure 1.10 shows the cumulative returns of non-financial stocks during the ban: the left one with the lowest beta (bottom 20%) and the right one with the highest beta (top 20%). $C_t = R_t^{BF} - R_t^{NF}$ is defined to be the effect of the short-selling ban on financial stocks. The dashed blue line shows one standard deviation of the cumulative effect of the short-selling ban $\text{cum}_t(C)$\textsuperscript{24}. On the left graph, the cumulative return of banned stocks (blue solid line) and the non-banned stocks (red solid line) cross each other several times and remain close. On the right graph, the banned stocks always have a higher price than the non-banned stocks. A significant overpricing effect can be observed according to the one standard deviation line. This is consistent with the prediction by Hong and Sraer (2012).

\textsuperscript{23}CRSP daily return data from September 18, 2007 to September 18, 2008 are used. Both 1\% and 0.1\% winsorizing thresholds are tested in this analysis. They yield the same results.

\textsuperscript{24}Assume that $C_t \sim N(\hat{C}_t, \hat{\sigma}_C^2)$ are independent for $t = 1, 2, ..., 14$, then the cumulative effect of short selling ban is $\text{cum}_t(C) = \sum C_S \sim N(\sum \hat{C}_t, \sum \hat{\sigma}_C^2)$. Thus the standard deviation of $\text{cum}_t(C)$ is $\hat{\sigma}_{\sum C_s} = \sqrt{\sum \hat{\sigma}_C^2}$. 

Figure 1.10: The Effect of Short-Selling Ban on Low and High Beta Stocks

The following two graphs show the cumulative returns of non-financial stocks during the ban: the left graph shows returns with the lowest beta (bottom 20%), while the right graph shows returns with the highest beta (top 20%). Define $C_t := R_{BF}^t - R_{NF}^t$ be the effect of the short-selling ban on financial stocks. The dashed blue line shows the one standard deviation spread of the cumulative effect of the short-selling ban $cum_t(C)$. In the left graph, the cumulative return of banned stocks (blue solid line) and the non-banned stocks (red solid line) cross each other several times and stay within one standard deviation of each other. In the right graph, the banned stocks always have a higher price level than the non-banned stocks. A significant overpricing effect can be observed according to the one standard deviation line (blue dashed line). This is consistent with the prediction by Hong and Sraer (2012).

In order to be concrete, the stocks are separated into 5 groups based on their beta values, from bottom 20% beta group to top 20% beta group. For each group, the stock information is inputted to my factor model and the cumulative returns are decomposed into four parts. The average daily difference of the cumulative returns between the banned and non-banned stocks is used to measure the size of the bubble. Figure 1.11 shows the bubble size of different beta groups: left for financial stocks and right for non-financial stocks (with different y-axes). Generally, high beta stocks are more prone to speculative overpricing during the 2008 ban.
Figure 1.11: Beta vs. Overpricing
The stocks are separated into 5 groups based on their beta values, from bottom 20% beta group to top 20% beta group. For each group, the stock information is given to the factor model and the returns are decomposed into four portions. The average daily difference of the cumulative returns between the banned and non-banned stocks is used to measure the size of the bubble. The following two graphs show the bubble size of different beta groups, left for financial stocks and right for non-financial stocks (note the different scales of the y-axes). Generally, high beta stocks are more prone to speculative overpricing during the 2008 ban.

1.6 Conclusion

By solving Scheinkman and Xiong (2003)'s model in a finite trading horizon with zero trading cost, I obtain a closed form solution for the bubble caused by a temporary short-selling ban. I average the bubble based on the stationary distribution of the difference of beliefs and analyze how economic parameters influence the magnitude and shape of the bubble. I show that the temporary ban leads to over-valuation that is highest at the beginning of the ban and steadily converges to zero at the end of the ban.

I verify this prediction using the 2008 temporary ban on US financial stocks. I overcome a key problem in earlier empirical works which use non-financial stocks as
a control group (since all financial stocks were banned). Instead, I use as a control group a synthetic portfolio of non-banned financial stocks constructed from the finance segments of large industrial companies which were not banned. The predictions of the model and the results from my empirical analysis match qualitatively.
1.7 Appendix

1.7.1 Proof of Theorem 1

The PDE to be solved is

\[
\begin{aligned}
\frac{1}{2}\sigma^2 q_{xx} - \rho x q_x + q_t - rq &= 0 \\
q(x, T) &= 0 \\
q_x(0, t) &= \frac{1}{2(r+\lambda)}(1 - e^{-(r+\lambda)(T-t)})
\end{aligned}
\tag{1.19}
\]

where \( x < 0 \) and \( 0 \leq t \leq T \). The boundary condition is at the end of the ban when \( t = T \). By the transform \( \tau = T - t \), one can make the boundary condition lie at the beginning of the new time \( \tau \). Let \( l(x, \tau) = q(x, T - \tau) \), then \( l(x, \tau) \) satisfies

\[
\begin{aligned}
\frac{1}{2}\sigma^2 l_{xx} - \rho x l_x - l_\tau - rl &= 0 \quad \text{PDE} \\
l(x, 0) &= 0 \quad \text{Initial condition} \\
l_x(0, \tau) &= \frac{1}{2(r+\lambda)}(1 - e^{-(r+\lambda)\tau}) \quad \text{Boundary condition}
\end{aligned}
\tag{1.20}
\]

where \( x < 0 \) and \( 0 \leq \tau \). The PDE and the boundary condition is well-defined for all \( \tau > T \), thus we can solve it for all \( \tau \geq 0 \) and then pick the \( 0 \leq \tau \leq T \) part.

Consider performing a Laplace transform on this PDE. The Laplace transform of a function \( f(\tau) \) is \( F(s) = \mathcal{L}[f(\tau)] = \int_0^\infty e^{-st} f(\tau) d\tau \). Let \( L(x, s) = \mathcal{L}[l(x, \tau)] \). Performing a Laplace transform on the PDE of \( l(x, \tau) \) and using the following properties of the Laplace transform

\[
\mathcal{L}[af(\tau) + bg(\tau)] = aF(s) + bG(s)
\]

\[
\mathcal{L}[f'(\tau)] = sF(s) - f(0)
\]
one gets

\[ \frac{1}{2} \sigma^2 g_{xx} - \rho x L_x - [sL(x, s) - l(x, 0)] - rL = 0 \]

Using the initial condition \( l(x, 0) = 0 \), one has

\[ \frac{1}{2} \sigma^2 g_{xx} - \rho x L_x - (s + r)L = 0 \]

For a fixed \( s \), this is an ODE of \( x \). This ODE has a class of solutions of the form

\[ L(x, s) = \beta(s) h(x, s) \tag{1.21} \]

where \( \beta(s) \) is an arbitrary function of \( s \) and

\[ h(x, s) = U\left( \frac{r + s}{2\rho}, \frac{1}{2}, \frac{\rho}{\sigma^2} x^2 \right) \tag{1.22} \]

where \( U(\cdot) \) is the Kummer U function\(^{25}\)

\[ U(a, b, z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zy} y^{a-1} (1 + y)^{b-a-1} dy. \tag{1.23} \]

We can use the boundary condition to pin down the function \( \beta(s) \). Using linearity and the following property of Laplace transform

\[ \mathcal{L}[e^{at}] = \frac{1}{s - a} \]

one has the Laplace transform of the boundary condition

\[ L_x(0, s) = \mathcal{L}[l_x(0, \tau)] = \mathcal{L}\left[ \frac{1}{2(r + \lambda)} \left( 1 - e^{-(r+\lambda)\tau} \right) \right] = \frac{1}{2(r + \lambda)} \left( \frac{1}{s} - \frac{1}{s + r + \lambda} \right). \]

\(^{25}\)Scheinkman and Xiong (2003) use this Kummer U function to express the solution of a similar ODE. They use the sum of series for the U function. I use the integration form of the Kummer U function because it makes the inverse Laplace transform possible.

41
On the other hand, we know \( L_x(0, s) = \beta(s)h_x(0, s) \) and \( h_x(0, s) = \frac{\pi \sqrt{\rho}}{\sigma \Gamma(\frac{r+\lambda}{2\rho}) \Gamma(\frac{3}{2})} \).

Thus we can solve \( \beta(s) \)

\[
\beta(s) = \frac{\sigma \Gamma(\frac{r+\lambda}{2\rho}) \Gamma(\frac{3}{2})}{2(r + \lambda) \pi \sqrt{\rho}} \left( \frac{1}{s} - \frac{1}{s + r + \lambda} \right). \tag{1.24}
\]

Based on equations (1.21), (1.22), (1.23) and (1.24) we have

\[
L(x, s) = \int_0^\infty A(x, y) \left( \frac{1}{s} - \frac{1}{s + r + \lambda} \right) e^{-M(y)s} dy
\]

where \( A(x, y) = \frac{\sigma}{4(r + \lambda) \sqrt{\rho}} e^{-\frac{\rho}{2\sigma} x^2 y} y^\frac{r-2\rho}{2\rho} (1 + y)^{-\frac{r+\rho}{2\rho}} \) and \( M(y) = \frac{1}{2\rho} \ln(1 + \frac{1}{y}) \)

With the following property of the inverse Laplace transform, we can get \( l(x, \tau) \) from \( L(x, s) \):

\[
\mathcal{L}^{-1}\left[ \frac{1}{s + a} e^{-bs} \right] = u(\tau - b)e^{-a(\tau-b)}
\]

where \( u(x) \) is the unit step function: if \( x > 0 \), \( u(x) = 1 \); if \( x < 0 \), \( u(x) = 0 \).

\[
l(x, \tau) = \mathcal{L}^{-1}[L(x, s)] = \int_0^\infty A(x, y)[1 - e^{-(r+\lambda)(\tau-M(y))}] dy
\]

where \( \tilde{N}(\tau) = \frac{e^{-2\rho\tau}}{1 - e^{-2\rho \tau}} \).

Thus, the solution to our original PDE (1.19) is \( u(x, t) = l(x, T-t) \). So

\[
u(x, t) = \frac{\sigma}{4(r + \lambda) \sqrt{\rho}} \int_0^\infty e^{-\frac{\rho}{2\sigma} x^2 y} y^\frac{r-2\rho}{2\rho} (1 + y)^{-\frac{r+\rho}{2\rho}} \left[ 1 - e^{-(r+\lambda)(T-t-M(y))} \right] dy
\]

where \( x < 0, 0 \leq t \leq T, N(t) = \frac{e^{-2\rho(T-t)}}{1 - e^{-2\rho(T-t)}} \) and \( M(y) = \frac{1}{2\rho} \ln(1 + \frac{1}{y}) \).

### 1.7.2 Proof of Theorem 2

First, we prove \( u_t(x, t) < 0 \).

Since
\[ u(x,t) = \int_{N(t)}^{\infty} A(x,y)[1-e^{-(r+\lambda)(T-t-M(y))}]dy \]

where \( A(x,y) = \frac{\sigma}{4r+\lambda}\sqrt{\frac{\sigma}{2\pi}} e^{-\frac{\sigma^2}{2}(x^2+y^2)}(1+y)^{-\frac{r+\lambda}{2\rho}} \), \( N(t) = \frac{e^{-2\rho(T-t)}}{1-e^{-2\rho(T-t)}} \) and \( M(y) = \frac{1}{2\rho} \ln(1+\frac{1}{y}) \), we have

\[ u_t(x,t) = -A(x,N(t))[1-e^{-(r+\lambda)(T-t-M[N(t)])}] \frac{dN(t)}{dt} \]

\[ - \int_{N(t)}^{\infty} A(x,y)(r+\lambda)e^{-(r+\lambda)(T-t-M(y))}dy \]

where \( A(x,N(t))[1-e^{-(r+\lambda)(T-t-M[N(t)])}] \frac{dN(t)}{dt} = 0 \) because \( M[N(t)] = T-t \). The integral \( \int_{N(t)}^{\infty} A(x,y)(r+\lambda)e^{-(r+\lambda)(T-t-M(y))}dy \) is positive, because

\[(r+\lambda)e^{-(r+\lambda)(T-t-M(y))} > 0\]

by definition and when \( y > N(t) \), we have \( A(x,y) > 0 \). Thus \( u_t(x,t) < 0 \).

Since

\[ q(x,t) = \begin{cases} \frac{x}{r+\lambda}(1-e^{-(r+\lambda)(T-t)}) + u(-x,t) & x \geq 0 \\ u(x,t) & x < 0 \end{cases} \]

we know

\[ q_t(x,t) = \begin{cases} -xe^{-(r+\lambda)(T-t)} + u_t(-x,t) & x \geq 0 \\ u_t(x,t) & x < 0 \end{cases} \]

Thus \( q_t(x,t) < 0 \).

On the other hand,

\[ q(x,T) = \begin{cases} u(-x,T) & x \geq 0 \\ u(x,T) & x < 0 \end{cases} \]
Thus \( q(x, T) = 0 \) because \( u(x, T) = 0 \).

### 1.7.3 Proof of Theorem 3

Since \( x \leq 0 \),

\[
\bar{q}(t) = \int_{-\infty}^{0} u(x, t) \cdot w(x) dx
\]

where, according to equation (1.17),

\[
w(x)dx = 2 \cdot \frac{1}{\eta \sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\eta^2}} dx
\]

with \( \eta^2 = \frac{\sigma^2}{2\rho} \) and

\[
u(x, t)_{\text{homonolith}}^1 \frac{1}{2} e^{2\sigma y} \cdot w(x) dx = (1 + y)^{-\frac{1}{2}} \cdot \frac{2}{\phi \sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{x^2}{2\eta^2}} dx = (1 + y)^{-\frac{1}{2}}.
\]

With the term containing \( x \) integrated,

\[
\bar{p}(t) = \frac{\sigma}{4(r + \lambda)\sqrt{\pi \rho}} \int_{N(t)}^{\infty} (1 + y)^{-\frac{1}{2}} \frac{e^{2\sigma y}}{2\rho} (1 + y)^{-\frac{r+\rho}{2\rho}} [1 - e^{-(r+\lambda)(T-t-M(y))}] dy.
\]

Plugging \( M(y) = \frac{1}{2\rho} \ln(1 + \frac{1}{y}) \) into this equation and reorganizing the terms,
\( \bar{p}(t) = \frac{\sigma}{4(r + \lambda)\sqrt{\pi} p} \int_{N(t)}^{\infty} y^{-\frac{1}{2p}}(1+y)^{-\frac{1}{2p}-1} dy - e^{-(r+\lambda)(T-t)} \int_{N(t)}^{\infty} y^{-\frac{1}{2p}}(1+y)^{-\frac{1}{2p}-1} dy \]

With the integration formula \( \int y^a(1+y)^{-a-1} dy = \frac{1}{a} y^a(1 - y)^{-a} \), \( \bar{p}(t) \) can be simplified as

\[ \bar{p}(t) = \frac{\sigma}{4(r + \lambda)\sqrt{\pi} p} \left[ \frac{2p}{r} y^{\frac{1}{2p}}(1+y)^{-\frac{1}{2p}} + e^{-(r+\lambda)(T-t)} \frac{2p}{\lambda} y^{-\frac{1}{2p}}(1+y)^{\frac{1}{2p}} \right]_{N(t)}. \]

Plugging \( N(t) = \frac{e^{-2p(T-t)}}{1-e^{-2p(T-t)}} \) into this equation and reorganizing the terms,

\[ \bar{q}(t) = \frac{\sigma}{2\lambda\sqrt{\pi}} \left[ \frac{1}{r} (1 - e^{-r(T-t)}) - \frac{1}{r + \lambda} (1 - e^{-(r+\lambda)(T-t)}) \right]. \]

### 1.7.4 Proof of Theorem 4

Based on the equation of \( \bar{q}(t) \)

\[ \frac{dp(t)}{dt} = -\frac{\sigma\sqrt{p}}{2\lambda\sqrt{\pi}} e^{-r(T-t)} [1 - e^{-\lambda(T-t)}] < 0 \]

since all the parameters are positive and \( 0 \leq t \leq T \).

It is also easy to verify that \( \bar{q}(T) = 0 \). Thus the bubble shrinks and vanishes at the end of the ban.

Since

\[ \frac{d^2 p(t)}{dt^2} = -\frac{\sigma\sqrt{p}}{2\lambda\sqrt{\pi}} [r - (r + \lambda)e^{-\lambda(T-t)}] e^{-r(T-t)} \]

Letting \( \frac{d^2 p(t)}{dt^2} = 0 \), one can solve for the inflection point \( t^* = T - \tau^* \), where \( \tau^* = \frac{1}{\lambda} ln(1 + \frac{\lambda}{r}) \).

The size of the bubble at the inflection point is \( \bar{p}(t^*) = \frac{\sigma\sqrt{p}}{2\lambda\sqrt{\pi} (r + \lambda)} [1 - \left( \frac{r}{r+\lambda}\right)^{\frac{r}{r+\lambda}} 2\lambda]. \)
The decreasing speed at the inflection point is 
\[ \frac{d\rho(t)}{dt} \bigg|_{t=t^*} = \frac{-\sigma \sqrt{\rho}}{2(r+\lambda)\sqrt{\pi}} \left( \frac{r}{r+\lambda} \right) \xi. \]
Chapter 2

Fire Sale in Financial Networks

The default of one bank can cause other banks to default through two channels: financial contagion in the inter-bank liability network and fire sale in the asset selling market. When the defaulted bank cannot fully pay its debt, the loss is transmitted to other banks. When banks rush to sell the same asset simultaneously, they may fall into a Nash equilibrium in which banks compete for liquidity and sell their assets at an artificially low price. In this paper, a model that incorporates these two channels is developed and analyzed theoretically. An algorithm for finding the state in which both the inter-bank liability network and the market are in equilibrium is proposed and tested. This chapter is published in Tian and E (2014), and presented at the INFORMS Annual Meeting 2014 (INFORMS: Institute for Operations Research and the Management Sciences), PACM Colloquium 2014 (PACM: the Program in Applied and Computational Mathematics at Princeton University) and Center for Science of Information 2013 (CIOS: NSF Science and Technology Center).
2.1 Introduction

At the time of the bankruptcy of Lehman Brothers in September 2008, several money funds with significant exposure to Lehman fell below $1 per share following losses on their holdings of Lehman assets (Gullapalli et al., 2008). Meanwhile, the prospect for Lehman’s $4.3 billion in mortgage securities getting liquidated sparked a sell-off in the CMBS market (Wei and Corkery, 2008). During the 3 years that ensued, more than 300 U.S. banks defaulted, while only 3 banks defaulted in 2007 (Federal Deposit Insurance Corporation: Failed Bank List, http://www.fdic.gov/).

Figure 2.1: Number of bank failures based on the data from Federal Deposit Insurance Corporation (FDIC)

![Monthly Number of Bank Failures (2000-2014)](image)

what is the normal figure?

The default of one bank can cause other banks to default through two channels

1. **Financial contagion**: When the defaulted bank cannot fully pay its debt to other banks, the loss is transmitted to other banks and this increases the risk for them to default.

2. **Fire sale**: The defaulted bank has to sell its assets to raise cash for payment.
When banks rush to sell the same asset, they may fall into a Nash equilibrium in which banks compete for liquidity and sell their assets at an artificially low price.

To understand the first channel, financial contagion, Eisenberg and Noe (2001) analyzed a model of the inter-bank clearing payment network and proved the existence and uniqueness of the equilibrium state under certain conditions. Researchers from the Bank of England carried out simulations and investigated how financial conditions affect the equilibrium state (Nier et al., 2007 and Anand et al., 2012). Shin (2010) related this equilibrium to credit boom-and-bust and stressed the role of the system leverage. Nier et al. (2007) and Chen et al. (2013) use a decreasing price function to test the effects of illiquidity on the financial system. However, they do not model banks’ optimal selling strategy. In particular, the Nash equilibrium of fire sale has not considered. For the second channel, the fire sale, Coval and Stafford (2007) investigated how selling creates price pressure based on mutual fund transactions data. Merrill et al. (2012) discovered empirical evidence of artificially low asset price during the 2008 financial crisis. Shleifer and Vishny (2011) related fire sale to more general finance and macroeconomic concepts.

In this paper, we incorporate these two channels into one model. During the clearing payment process in a standard inter-bank liability network, banks are allowed to sell their assets in an illiquid market. Banks compete for liquidity and may sell their assets at a low price given that other banks are selling at a low price. This state is not optimal for the system but nevertheless is the Nash equilibrium state. The intuition for this equilibrium is similar to Cournot competition (Cournot, 1838), but it is more complicated to solve since banks have heterogeneous cash demand and shares of assets available. The illiquid market and the payment system interact with each other. The amount of cash that banks are willing to raise in the market depends
on the realized payment, while the payment a bank can make is partially influenced by how much cash it can raise. We will develop an algorithm for finding the state in which both the market and the payment system are in equilibrium.

Figure 2.2: The equilibrium of the system requires both the market and the payment network are in equilibrium

2.2 Model

2.2.1 Bank’s Balance Sheet

Each bank’s balance sheet contains two parts with the same value: assets and liabilities. Banks raise money through liabilities and spend it to buy assets. Liabilities include deposit $d$, capital borrowed from other banks $b$ and share holders’ equity $e$. Assets include cash $h$, loans lent to other banks $l$ and also $s$ shares of illiquid assets with face value $1$ per share.
Figure 2.3: Each bank’s balance sheet contains two parts with the same value: assets and liabilities.

Balance sheet identity requires assets equal to liabilities

\[ d + b + e = h + s \cdot 1 + l \]

Suppose there are \( N \) banks in the financial system. Let \( L \) be the liability matrix. Then \( L_{ij} > 0 \) means that bank \( i \) borrow the amount of \( L_{ij} \) from bank \( j \). Bank \( i \) does not borrow from itself, thus \( L_{ii} \equiv 0 \). By definition, the capital that bank \( i \) borrowed from other banks is \( b_i = \sum_j L_{ij} \); the loans that bank \( i \) lent to other banks is \( l_i = \sum_k L_{ki} \).

On the payment date, banks use their assets to pay for their debts. When some of the inter-bank debts are not paid in face value \( \hat{L}_{ki} < L_{ki} \), bank \( i \) collects less money from other banks than it expects \( \hat{l}_i < l_i \). When the total value of cash and payment from other banks is less than the total value that bank \( i \) has to pay its depositors and other banks, bank \( i \) has to sell its illiquid assets to raise cash. The amount of cash needed is

\[ c_i = (d_i + b_i) - (h_i + \hat{l}_i) \]

where \((d_i + b_i)\) is the total value that bank \( i \) has to pay its depositors and all other
banks; \((h_i + \hat{l}_i)\) is the total value of cash and the realized payment bank \(i\) collected from other banks.

If bank \(i\) cannot raise enough cash, it will default. As a result, it will pay less to its creditors. Each creditor \(j\) gets

\[
\beta_i \cdot \frac{L_{ij}}{b_i}
\]

where \(\beta_i\) is the total amount that bank \(i\) pays to other banks and \(\frac{L_{ij}}{b_i}\) is the proportion that bank \(j\) gets.

Paying less may lead to default of other banks. This in turn may lead to even less money collected by bank \(i\), \(\hat{l}_i < \hat{l}_i < l_i\).

The mathematical problem is to find the payment vector for the liability network and the Nash equilibrium in the illiquid market. In the following sections, we will model the illiquid market and find the state when the system are in both of these equilibriums.

2.2.2 Price of the Illiquid Asset

Bank \(i\) owns \(s_i\) shares of the illiquid asset. If bank \(i\) sells \(x_i \in [0, s_i]\) shares, the total supply of shares by the bank system is

\[
X = \sum_i x_i
\]

Investors outside the bank system will buy these shares with a discounted price since they are bearing the risk of holding these assets. The mechanism is modeled in the following framework.

One share of the asset will pay a dividend

\[
D = 1 + \epsilon
\]

where 1 is the expected value of the asset. \(\epsilon\) is normally distributed with mean zero.
and variance $\sigma^2$. The variance $\sigma^2$ models the risk of the return of this asset.

Investors outside the bank system maximize CARA (Constant Absolute Risk Aversion) utility

$$U(W) = 1 - e^{-\alpha W}$$

where $\alpha$ is the risk aversion coefficient; $W$ is the wealth of the investor.

If the price of the asset is $p$ per share, then for each share of the asset owned by investors, the profit is $D - p$. If the investors buy $Q$ shares of the asset, then their profit is

$$Q \cdot (D - p)$$

Viewed in today, this is random because the dividend is a random variable. Investors would like to maximize their expected utility on this. Thus their optimal demand solves the following maximization problem

$$\max_Q \{E[U(Q \cdot (D - p))}\}$$

By the first order condition of this optimization problem, one can solve investors’ optimal demand of the asset

$$Q(p) = \frac{1 - p}{\alpha \sigma^2}$$

When the price is lower than the expected value of the dividend ($p > 1$), investors buy the asset ($Q$ is positive). When the dividend is less risky ($\sigma^2$ is small) or the investors are less risk averse ($\alpha$ is small), investors demand more of the asset given a fixed price. When the price is higher than the expected value of the dividend ($p < 1$), investors short-sell the asset.

As defined at the beginning of this section, the total supply of shares by the bank system is $X$. Then the market clearing condition $Q(p) = X$ yields an equilibrium
price

\[ p(X) = 1 - \alpha \sigma^2 X \]

The price is lower than its expected value because investors bear the risk of holding it. When the total supply \( X \) increases, the price of the asset decreases because investors are holding more asset and bearing higher risk.

### 2.2.3 Optimal Selling Strategy

Each bank \( i \) has its own cash demand \( c_i = (d_i + b_i) - (h_i + \hat{l}_i) \) to avoid default in the payment process. Given other banks’ current selling amount \( x_j \), bank \( i \) optimally sells \( x_i \in [0, s_i] \) such that

- Bank \( i \) can raise enough cash\(^1\):

\[ x_i \cdot p(X) = c_i \]

- If bank \( i \) cannot raise enough cash (for all \( x_i \in [0, s_i], x_i \cdot p(X) < c_i \)), bank \( i \) will choose \( x_i \) to maximize the amount of cash it raised

\[ x_i = \arg\max_{x_i \in [0, s_i]} [x_i \cdot p(X)] \]

Since \( x_i \cdot p(X) \) is quadratic in \( x_i \), the solution to the optimization problem is either on the boundary \( x_i = s_i \) or at the climax \( \frac{\partial x_i \cdot p(X)}{\partial x_i} = 0 \). The situation of optimally selling at the climax is more interesting. It characterizes the situation when banks compete to raise cash but are frozen at a point where a lot of shares are sold at a very low price. The following example explains this intuition.

---

\(^1\)If bank \( i \) could raise the same amount cash \( c_i \) by selling \( x'_i \) or \( x''_i \), bank \( i \) will optimally sell less (choose the smaller \( x_i \))
2.2.4 An Example of Two Banks

To simplify, let $\alpha \sigma^2 = 1$ and write the price function as

\[ p = 1 - z \]

where $z$ is the total shares sold by the bank system to the market.

Then the money that the bank system raised by selling $z$ shares is

\[ L = pz = (1 - z)z \]

When $z = \frac{1}{2}$, $L$ peaks at the value $L_{\text{max}} = \frac{1}{4}$.

Suppose there are two banks. Each bank has $s = 1$ shares of illiquid assets (In this example, $x_i < s$ automatically holds). Suppose the two banks have the same cash demand $c$. Then the total demand is $C = 2c$. When $C > L_{\text{max}}$, the market cannot provide enough liquidity to the bank system. One possibility is that each bank sells $\frac{1}{4}$ such that the total shares sold is $x = \frac{1}{2}$ and they get all the liquidity $L_{\text{max}}$ from the market. However, this situation is not a Nash equilibrium.

Consider the situation when $c$ is big enough so that banks want to raise as much cash as possible. Given bank A sells $x = \frac{1}{4}$, the money that bank B can raise by selling $y$ shares is

\[ L_B = (1 - (\frac{1}{4} + y)) \cdot y \]

With the first order condition, we know that when $y = \frac{3}{8}$, B can raise the most liquidity

\[ L_B(x = \frac{1}{4}, y = \frac{3}{8}) = \frac{9}{64} \]

Since this is better than

\[ L_B(x = \frac{1}{4}, y = \frac{1}{4}) = \frac{1}{8} \]
B will choose to sell \( \frac{3}{8} \) instead of \( \frac{1}{4} \).

Given that B chooses \( \frac{3}{8} \), A will optimally choose \( \frac{5}{16} \). This process will continue because none of the state is stable (or Nash equilibrium).

To find the equilibrium state, note that A wants to maximize

\[
L_A(x, y) = (1 - (x + y))x
\]

thus we know that the optimal supply of A is

\[
x = \frac{1 - y}{2}
\]

B wants to maximize

\[
L_B(x, y) = (1 - (x + y))y
\]

thus we know that the optimal supply of B is

\[
y = \frac{1 - x}{2}
\]

Solving this system, we have

\[
\begin{aligned}
x^* &= \frac{1}{3} \\
y^* &= \frac{1}{3}
\end{aligned}
\]

Given A sells \( \frac{1}{3} \), the optimal strategy for B is to sell \( \frac{1}{3} \), and vise versa. This is the Nash equilibrium. The total amount of cash that the bank system can raise is

\[
L^* = (1 - (\frac{1}{3} + \frac{1}{3}))(\frac{1}{3} + \frac{1}{3}) = \frac{2}{9} < \frac{1}{4} = L_{max}
\]

We call this state liquidity freezing: Assets are sold at artificially low price due to competition, causing the total liquidity that the bank system raised to decrease.
When more banks are involved in this state, the effect is even stronger (see Theorem ??).

In particular, we see that \( x = \frac{1}{4} \) and \( y = \frac{1}{4} \) is not a Nash equilibrium, even though it is good for the whole bank system: Individual banks will move away from this state if other banks are in this state. Here is the evidence during the 2008 financial crisis. Merrill et. al. (2012) plotted the transaction prices of non-agency residential mortgage-backed security through time from 1998 to 2009. Before 2008, the price is very close to 100. During 2008, the price began to drop. During late 2008 and early 2009, there is a dense cluster of prices close to 0. Price does not go to this low level gradually. Instead, it suddenly drop to this low level. This matches our intuition that financial institutions compete to raise liquidity to avoid bankrupts.

![Figure 2.4: RMBS price fell to a liquidity freezing state](image)

Transaction Prices on Non-Agency RMBS through Time.

Source: Craig B. Merrill, Taylor D. Nadauld, Rene M. Stulz, and Shane M. Shefend, Why did financial institutions sell RMBS at fire sale prices during the financial crisis? 2012
2.3 Find the Equilibrium for the System

2.3.1 Nash Equilibrium in the Illiquid Market

We search the state space to find the Nash equilibrium of the illiquid market. There are 3 possible states ($\theta_i = 0, 1, 2$) for each bank $i$:

- $\theta_i = 0$: Bank $i$ raises enough cash
  
  $x_i p(X) = c_i, \quad x_i \in [0, s_i]$

- $\theta_i = 1$: Bank $i$ sells all its shares but fail to raise enough cash
  
  $x_i = s_i, \quad x_i p(X) < c_i$

- $\theta_i = 2$: Bank $i$ is competing to sell
  
  $\frac{\partial x_i \cdot p(X)}{\partial x_i} = 0, \quad x_i p(X) < c_i$

Given one realization of the state vector of all banks, we can calculate the shares sold by each bank by the following theorem.

**Theorem 5.** Let there be $N$ banks. Let their cash demand be $c = (c_1, c_2, ..., c_N)$ and their shares available be $s = (s_1, s_2, ..., s_N)$. If the state vector $\theta^* = (\theta_1^*, \theta_2^*, ..., \theta_N^*)$ is given, then the shares that bank $i$ sells are
\[ x_i = \begin{cases} 
    c_i \cdot w & \theta_i^* = 0 \\
    s_i & \theta_i^* = 1 \\
    \frac{1}{n+1} (\frac{1}{\alpha^2} - \bar{c} \cdot w - \bar{s}) & \theta_i^* = 2
\end{cases} \]

where

\[ \bar{c} = \sum_{\theta_i^* = 0} c_i \]

is the sum of cash demand of banks with \( \theta_i^* = 0 \); 

\[ \bar{s} = \sum_{\theta_i^* = 1} s_i \]

is the sum of shares available of banks with \( \theta_i^* = 1 \); \( n \) is the number of banks with \( \theta_i^* = 2 \); and \( w \) is the asset needed to raise one unit of cash by type-0 banks (\( \theta_i^* = 0 \)).

\[
    w = \frac{(1 - \alpha \sigma^2 \bar{s}) - \sqrt{(1 - \alpha \sigma^2 \bar{s})^2 - 4\alpha \sigma^2 \bar{c}(n + 1)}}{2\alpha \sigma^2 \bar{c}}
\]

However, the shares sold determined by Theorem 5. may not be a Nash equilibrium. We just calculated the shares each bank would sell given its state, but the bank may want to change its state based on other banks’ current states. The following theorem characterizes the selling vector when all banks are in a Nash equilibrium.

**Theorem 6.** A selling vector \( \mathbf{x} = (x_1, x_2, ..., x_N) \) is a Nash equilibrium if and only if for each \( i \), the following three inequalities hold and the equality sign holds for at least one of them.

\[
\begin{cases}
    x_i \cdot p(X) \leq c_i \\
    x_i \leq s_i \\
    \frac{\partial x_i \cdot p(X)}{\partial x_i} \geq 0
\end{cases}
\]
With these two theorems, one can find all equilibriums of the illiquid market by traverse the state space. Since the state space has $3^N$ points, this may seem like an impossible task. However, the following method finds an equilibrium faster by only testing a sub-space with $O(N^3)$ points:

1. Rank the banks by their cash demand such that
   
   $c_1 < c_2 < \ldots < c_N$

2. Pick $i \in \{0, 1, \ldots, N\}$. Re-order the first $i$ banks by $\frac{c_i}{s_i}$ and re-order the last $N-i$ banks by $s_i$

3. Pick $j \in \{0, 1, \ldots, i\}$ and $k \in \{i, i+1, \ldots, N\}$

4. Let banks $\{1, 2, \ldots, j\}$ be in state 0.
   Let banks $\{j+1, j+2, \ldots, k\}$ be in state 1.
   Let banks $\{k+1, k+2, \ldots, N\}$ be in state 2.
   (the sets can be empty if the first element is greater than the last element)

5. Calculate the selling vector based on this state vector (Theorem 5.)

6. Test whether the selling vector is a Nash equilibrium (Theorem 6.)

2.3.2 Equilibrium of the System

Let

$$\Pi_{ij} = \frac{L_{ij}}{b_i}$$

be the payment proportion matrix. $L_{ij}$ is the amount that bank $i$ should pay bank $j$. $b_i = \sum_j L_{ij}$ is the total amount that bank $i$ should pay all other banks. If bank $i$
pays $\beta_i$ to all other banks, bank $j$ will get $\Pi_{ij} \cdot \beta_i$ from bank $i$. The vector $\Pi^T \cdot \beta$ is the amount of money that each bank can get if they pay according to the payment vector $\beta$. For example, the $i^{th}$ element of the vector is $(\Pi^T \cdot \beta)_i = \sum_j \Pi_{ji} \cdot \beta_j$. This is the sum of the amount that all other banks pay bank $i$. We can find the equilibrium payment vector by the following iteration:

1. Let $\beta[1] = b$, i.e. start with the situation that banks pay all their inter-bank debt initially.

2. Given the $m^{th}$ payment vector $\beta[m]$, let the cash demand vector be $c = (d + b - \Pi^T \cdot \beta[m] - h)^+$, the amount of cash needed to avoid default.

3. Based on cash demand $c$ and the inventory of illiquid asset $s$, solve the Nash equilibrium of the illiquid market to get the optimal selling vector $x$ and asset price $p$. Then the vector $p \cdot x$ records the cash raised by each bank.

4. Let the new payment vector be

   $$\beta[m + 1] = \min(\Pi^T \cdot \beta[m] + h + p \cdot x, \beta[m])$$

   where $\min(\cdot)$ picks the smaller element of the two vectors to generate a new vector.

Similar to Eisenberg and Noe (2001), we can prove

There exists some $m^*$ such that $\beta[m^*] = \beta[m^* + 1]$. The payment vector $\beta[m^*]$ is the clearing vector for this system.
2.4 Equilibrium in the Erdős-Rényi Network

2.4.1 Set up the Financial Network

One way to model the financial network is to use the Erdős-Rényi network. In this network, the probability that bank \(i\) borrow from bank \(j\) is fixed to be a constant. The Erdős-Rényi network is characterized by the following parameters:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N)</td>
<td>number of banks</td>
</tr>
<tr>
<td>(p_{ER})</td>
<td>Erdős–Rényi probability</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>equity-to-asset ratio of each bank</td>
</tr>
<tr>
<td>(\phi)</td>
<td>proportion of banks with illiquid assets</td>
</tr>
<tr>
<td>(\alpha\sigma^2)</td>
<td>illiquidity of the assets</td>
</tr>
</tbody>
</table>

First, we generate a 0-1 matrix \(L_{ij}\) such that for \(i \neq j\), \(P(L_{ij} = 1) = p_{ER}\) and \(L_{ii} = 0\). When \(p_{ER}\) gets bigger, the banks are more connected. For each bank \(i\), we have the total amount borrowed from other banks \(b_i = \sum_j L_{ij}\) and the total amount lent to other banks \(l_i = \sum_k L_{ki}\). If bank \(i\) lends more than it borrows, we use the deposit to fund the lending and let \(d_i = l_i - b_i\), otherwise let \(d_i = 0\) for simplicity. Pick \(e_i\) to match the equity-to-asset rate \(\frac{e_i}{e_i + b_i + d_i} = \gamma\). To balance assets and liabilities, we will allocate cash and illiquid assets to each bank. The total amount that bank \(i\) needs is \(a_i = e_i + b_i + d_i - l_i\). Randomly pick \(\phi\) percent of banks and offer them half cash and half illiquid assets \(h_i = s_i = \frac{1}{2}a_i\). For other banks, only offer them cash \(h_i = a_i\). After assigning these values, we get the balance sheets for all banks. The inter-bank system is characterized by these inter-bank balance sheets and the inter-bank liability matrix.
2.4.2 Simulation Results

To start the simulation, we randomly pick one bank and take away half of its cash \( \frac{1}{2} h_i \). The following table shows the values of the parameters in our simulation (The values of the first three parameters are picked according to Nier et al. (2007))

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Benchmark Value</th>
<th>Range of Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>20</td>
<td>Fixed</td>
</tr>
<tr>
<td>( p_{ER} )</td>
<td>0.2</td>
<td>0 to 1</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>5%</td>
<td>0 to 50%</td>
</tr>
<tr>
<td>( \phi )</td>
<td>50%</td>
<td>10% to 100%</td>
</tr>
<tr>
<td>( \alpha \sigma^2 )</td>
<td>0.15</td>
<td>0 to 0.2</td>
</tr>
</tbody>
</table>

2.4.2.1 Liquidity freezing

Since the asset price is determined by the total supply linearly \( p(X) = 1 - \alpha \sigma^2 X \), we can use \( \alpha \sigma^2 \) to measure the illiquidity of the market. When this number is big, a slight increase in supply will press the price down significantly. When \( \alpha \sigma^2 = 0 \), price is always 1 and does not drop due to selling pressure. The following graph is generated based on one fixed financial network. The x-axis is the illiquidity parameter \( \alpha \sigma^2 \). The y-axis is the number of banks fall in different categories. When the market is liquid, all banks that need cash can raise enough cash from the market by selling their assets. When the market gets illiquid, more and more banks sell all their assets but cannot raise enough cash. When the market gets very illiquid, a subset of banks fall into the liquidity freezing equilibrium. They sell their assets with a very low price but it is optimal for themselves given that others are also selling.
2.4.2.2 Network connectivity

When the Erdős–Rényi probability is very low, banks are rarely connected. Then the default of one bank has little effect on other banks. When the Erdős–Rényi probability is very high, banks are very much connected. Any shock on a single bank will be absorbed by the whole system. Thus the number of defaults is also low. This is confirmed by the numerical results (see Figure 3). In the middle of the Figure 3, shocks are transmitted but are not very well absorbed. Thus there are more defaults. This mechanism is more significant for the simulations in which 100% banks have illiquid assets. When banks sell their assets in the market to raise cash, liquidity freezing leads to a very low asset price and few of them can raise enough cash. For each Erdős–Rényi probability $p_{ER}$ we randomly generate 50 liability networks and pick $\phi$ percent of banks randomly to allocation illiquid assets. Then we randomly attack one bank and find the equilibrium of the system. When there are more than
one equilibrium, we pick the equilibrium with the highest asset price. The solid line is the average number of defaults. The colored area shows the one standard deviation region.

Figure 2.6: Number of defaults for different network connectivity

2.4.2.3 Leverage

When Equity-to-Asset ratio is low, banks are operating with a high leverage. Then they are more likely to default when one of them is attacked. The interesting phenomenon here is that the curve of default numbers shows different shapes when the illiquid assets have different coverage rate. When only 10% of the banks have illiquid assets, the system is quite stable for most of the leverage levels. Only when the leverage is extremely high, the system fails. However, when all banks have illiquid assets, the number of defaults increases linearly when the Equity-to-Asset ratio gets smaller. This offers a new perspective to understand the effect of leverage on systemic risk. Ignoring the illiquidity of the market, especially during the crisis, one may be overly optimistic about the current leverage rate and the systemic stability. Whenever some banks start fire sale, the asset price will drop severely. Other banks with the same asset will suffer from low valuation in their assets temporarily. Their leverage
can appear to be artificially high and the system can be dragged down into a very risky state.

Figure 2.7: Leverage can play different roles

![Leverage and Bank Defaults](image)

### 2.5 Equilibrium in the Core-periphery Network

#### 2.5.1 Set up the Financial Network

The Erdős-Rényi network offers a simple and direct way to characterize the connectivity of the network and analyze its effect on the stability of the financial system. However, banks in the financial system are not homogeneous. Big banks highly connected with each other, while small banks are merely inter-connected. Each big bank connects to several small banks in the same region or the same system (agriculture system, mortgage system and etc.).

Santos and Cont (2010) plotted the the inter-bank network in Brazil in December 2007 (Santosy and Cont, 2010). Most of the banks are in the core and interconnected. But there are also banks in the periphery connected to some of the banks in the core.
Minoiu and Reyes (2011) plotted the inter-bank network of the global banking system in 2007. Top GDP countries like United States, Japan, Germany are in the core and inter-connected with each other. Other countries connect with the top GDP countries in their continent. Some of the countries are even not connected with other countries.
To model this network structure observed in real world, we define the core-periphery network. In this network, the probability that bank $i$ borrows from bank $j$ is determined by two parameters $p_i$ and $p_j$. For the $m$ banks in the core, each bank $i$ has $p_i = p_c$. For the $n$ banks in the periphery, each bank $i$ has $p_i = p_p$. The probability that bank $i$ borrows from bank $j$ is then defined as

$$P(i \rightarrow j) = p_ip_j$$

The core-periphery network is characterized by the following parameters:
Table 2.3: The parameters that characterize the Erdős-Rényi network

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>number of banks in the core</td>
</tr>
<tr>
<td>$n$</td>
<td>number of banks in the periphery</td>
</tr>
<tr>
<td>$p_c$</td>
<td>the connecting probability for banks in the core</td>
</tr>
<tr>
<td>$p_p$</td>
<td>the connecting probability for banks in the periphery</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>equity-to-asset ratio of each bank</td>
</tr>
<tr>
<td>$\phi$</td>
<td>proportion of banks with illiquid assets</td>
</tr>
<tr>
<td>$\alpha\sigma^2$</td>
<td>illiquidity of the assets</td>
</tr>
</tbody>
</table>

First, we generate a 0-1 matrix $L_{ij}$ such that for $i \neq j$, $P(L_{ij} = 1) = p_ip_j$ and $L_{ii} \equiv 0$. For each bank $i$, we have the total amount borrowed from other banks $b_i = \sum_j L_{ij}$ and the total amount lent to other banks $l_i = \sum_k L_{ki}$. If bank $i$ lends more than it borrows, we use the deposit to fund the lending and let $d_i = l_i - b_i$, otherwise let $d_i = 0$ for simplicity. Pick $e_i$ to match the equity-to-asset rate $\frac{e_i}{e_i + b_i + d_i} = \gamma$. To balance assets and liabilities, we will allocate cash and illiquid assets to each bank. The total amount that bank $i$ needs is $a_i = e_i + b_i + d_i - l_i$. Randomly pick $\phi$ percent of banks and offer them half cash and half illiquid assets $h_i = s_i = \frac{1}{2}a_i$. For other banks, only offer them cash $h_i = a_i$. After assigning these values, we get the balance sheets for all banks. The inter-bank system is characterized by these inter-bank balance sheets and the inter-bank liability matrix.

2.5.2 Simulation Results

To start the simulation, we randomly pick one bank and take away half of its cash $\frac{1}{2}h_i$. The following table shows the values of the parameters in our simulation.
Table 2.4: The values of the parameters in simulation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Benchmark Value</th>
<th>Range of Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m$</td>
<td>10</td>
<td>5 to 15</td>
</tr>
<tr>
<td>$n$</td>
<td>10</td>
<td>5 to 15</td>
</tr>
<tr>
<td>$p_c$</td>
<td>0.9</td>
<td>0 to 1</td>
</tr>
<tr>
<td>$p_p$</td>
<td>0.3</td>
<td>0 to 1</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>5%</td>
<td>0 to 50%</td>
</tr>
<tr>
<td>$\phi$</td>
<td>50%</td>
<td>10% to 100%</td>
</tr>
<tr>
<td>$\alpha\sigma^2$</td>
<td>0.15</td>
<td>0 to 0.2</td>
</tr>
</tbody>
</table>

When Equity-to-Asset ratio is low, banks are operating with a high leverage. Then they are more likely to default when one of them is attacked. The interesting phenomenon here is that the curve of default numbers shows different shapes when the networks have different structures. When only 5 of the banks are in the periphery, the system is quite stable for most of the leverage levels. Only when the leverage is extremely high, the system fails. However, when 15 of the banks are in the periphery, the number of defaults increases linearly when the Equity-to-Asset ratio gets smaller. This offers a new perspective to understand the effect of the structure of the financial network. When most of the banks are in the core, they absorb and bear the shock together. This suggests that a more centralized network could be more stable.
2.6 Conclusion

This chapter analyzed the two channels through which the default of one bank can be transmitted in the bank system.

1. **Financial contagion**: When the defaulted bank cannot fully pay its debt to other banks, the loss is transmitted to other banks and this increases the risk for them to default.

2. **Fire sale**: The defaulted bank has to sell its assets to raise cash for payment. When banks rush to sell the same asset, they may fall into a Nash equilibrium in which banks compete for liquidity and sell their assets at an artificially low price.

We have incorporated these two channels into one model. During the clearing payment process in a standard inter-bank liability network, banks are allowed to sell their assets
in an illiquid market. Banks compete for liquidity and sell their assets at a low price
given that other banks are selling at a low price. This state is not optimal for the
system but nevertheless is the Nash equilibrium state. The illiquid market and the
payment system interact with each other. The amount of cash that banks are willing
to raise in the market depends on the realized payment, while the payment a bank
can make is partially influenced by how much cash it can raise.

An algorithm for finding the state in which both the market and the payment
system are in equilibrium is developed. We found that when the market gets very
illiquid, a subset of banks fall into the liquidity freezing equilibrium. They sell their
assets with a very low price but it is optimal for themselves given that others are also
selling. As for the coverage of the illiquid assets, when only 10% of the banks have
illiquid assets, the system is quite stable for most of the leverage levels. Only when
the leverage is extremely high, the system fails. However, when all banks have illiquid
assets, the number of defaults increases linearly when the Equity-to-Asset ratio gets
smaller. As for the structure of the network, when most of the banks are in the core,
they absorb and bear the shock together. A more centralized network could be more
stable.

2.7 Appendix

2.7.1 Proof of Theorem 5

We have derived that

\[ p(X) = 1 - \alpha \sigma^2 X \]

Generally, price has the form
\[ p(X) = A - B \cdot X \quad (2.1) \]

where \( A = 1 \) and \( B = \alpha \sigma^2 \) are all positive.

Let \( w \) be the asset needed to raise one unit of cash by type-0 banks \((\theta_i^* = 0)\), then for all these banks that raised enough cash, we have

\[ x_i = w \cdot c_i \]

Let the total cash raised by these banks to be

\[ \bar{c} = \sum_{\theta_i^* = 0} c_i \]

Then we know the total supply of these banks

\[ \sum_{\theta_i^* = 0} x_i = \bar{c}w \]

For all banks that sold all assets but not raised enough cash \((\theta_i^* = 1)\), their supply is just

\[ \sum_{\theta_i^* = 1} x_i = \sum_{\theta_i^* = 1} s_i =: \hat{s} \]

For banks that are competing \((\theta_i^* = 2)\), their supply is the same due to symmetry. Let the supply of a competing bank to be \( x \) and let the number of the competing banks to be \( n \), then we have

\[ \sum_{\theta_i^* = 2} x_i = nx \]

Thus, total supply is
\[ X = \sum_{i} x_i = \sum_{\theta_i^* = 0} x_i + \sum_{\theta_i^* = 1} x_i + \sum_{\theta_i^* = 2} x_i = \bar{c}w + \hat{s} + nx \quad (2.2) \]

For a competing bank \( i \), we have

\[ \frac{\partial p(X)}{\partial x_i} \cdot x_i = 0 \]

that is

\[ \frac{\partial (A - B(\bar{c}w + \hat{s} + \sum_{\theta_i^* = 2, j \neq i} x_j + x_i)) \cdot x_i}{\partial x_i} = 0 \]

that is

\[ -Bx_i + A - B(\bar{c}w + \hat{s} + \sum_{\theta_i^* = 2, j \neq i} x_j + x_i) = 0 \]

thus we have

\[ x_i = \frac{A}{B} - (\bar{c}w + \hat{s} + \sum_{\theta_i^* = 2} x_j) \]

Since all competing banks have the same supply \( x_i = x \), we have

\[ x = \frac{A}{B} - (\bar{c}w + \hat{s} + nx) \]

thus we have

\[ x = \frac{1}{n+1} (\frac{A}{B} - \bar{c}w - \hat{s}) \quad (2.3) \]

Let the equilibrium price to be \( p^* \). Then we know for type-0 banks (\( \theta_i^* = 0 \))

\[ p^* x_i = c_i \]
This means

\[ p^* w = 1 \quad (2.4) \]

Put (2.1)(2.2)(2.3) into (2.4), we have

\[ \{ A - B [\bar{c}w + \hat{s} + \frac{n}{n + 1} (\frac{A}{B} - \bar{c}w - \hat{s})] \} \cdot w = 1 \]

When \( \Delta := (A - B\hat{s})^2 - 4B\bar{c}(n + 1) > 0 \), this equation has two roots. Since non-competing banks would choose to sell less when raising enough cash, the smaller root is used.

\[ w = \frac{(A - B\hat{s}) - \sqrt{(A - B\hat{s})^2 - 4B\bar{c}(n + 1)}}{2B\bar{c}} \]

Thus we have

\[ x_i = \begin{cases} 
  c_iw & \theta^*_i = 0 \\
  s_i & \theta^*_i = 1 \\
  \frac{1}{n+1} (\frac{A}{B} - \bar{c}w - \hat{s}) & \theta^*_i = 2 
\end{cases} \]

When \( \Delta := (A - B\hat{s})^2 - 4B\bar{c}(n + 1) < 0 \), no equilibrium price can be founded since too much banks were set to be in state-0 (\( \theta^*_i = 0 \)). This kind of situation will be filtered out when searching for the real equilibrium.

### 2.7.2 Proof of Theorem 6

We first prove that if a selling vector \( \mathbf{x} = (x_1, x_2, ..., x_N) \) is a Nash equilibrium, then for each \( i \), the following three inequalities hold and the equality sign holds for at least one of them.
A Nash Equilibrium is a state that given other banks’ selling amount $x_j$, bank $i$ do not want to change its current selling amount $x_i$ based on its optimum. If $\mathbf{x} = (x_1, x_2, ..., x_N)$ is a Nash equilibrium, we know that each bank $i$ reaches its optimal selling amount. The optimum is defined as: given other banks’ current selling amount $x_j$, bank $i$ optimally sells $x_i \in [0, s_i]$ such that

- Bank $i$ can raise enough cash\(^2\):

$$x_i \cdot p(X) = c_i$$

- If bank $i$ cannot raise enough cash (for all $x_i \in [0, s_i]$, $x_i \cdot p(X) < c_i$), bank $i$ will choose $x_i$ to maximize the amount of cash it raised

$$x_i = \arg\max_{x_i \in [0, s_i]} [x_i \cdot p(X)]$$

From that we know that bank $i$ either raises enough cash or raises less cash than its demand. That is

$$x_i \cdot p(X) \leq c_i$$

On the other hand, $x_i$ is bounded by $s_i$, thus

\(^2\)If bank $i$ could raise the same amount cash $c_i$ by selling $x'_i$ or $x''_i$, bank $i$ will optimally sell less (choose the smaller $x_i$)
\[ x_i \leq s_i \]

If the bank can raise enough cash, then we know that

\[ x_i \cdot p(X) = c_i \]

Since the function \( y(x_i) = x_i \cdot P(X) \) is a parabola pointing downwards, we know that the horizon line \( y(x_i) = c_i \) cut the parabola at one or two points. When they intersect at one point, we know that

\[ \frac{\partial x_i \cdot p(X)}{\partial x_i} = 0 \]

When they intersect at two points \( x'_i < x''_i \), we know that

\[ \frac{\partial x_i \cdot p(X)}{\partial x_i} \bigg|_{x_i=x'_i} > 0, \quad \frac{\partial x_i \cdot p(X)}{\partial x_i} \bigg|_{x_i=x''_i} < 0 \]

When bank \( i \) can raise the same amount of cash by selling \( x'_i \) or \( x''_i \), bank \( i \) will optimally choose to sell less. Thus when there are two intersection, we also have

\[ \frac{\partial x_i \cdot p(X)}{\partial x_i} > 0 \]

The bound \( x_i \leq s_i \) can only make the final selling amount smaller. Then the first order derivative can only get bigger and keep positive. Thus

\[ \frac{\partial x_i \cdot p(X)}{\partial x_i} \geq 0 \]

When the bank cannot raise enough cash, it optimally sells \( x_i \) such that \( \mathbf{v} = (x_1, x_2, \ldots, x_N) \)

77
\[
\frac{\partial x_i \cdot p(X)}{\partial x_i} = 0
\]

For the same reasoning, the bound \( x_i \leq s_i \) can only make the derivative bigger and keep it positive. Thus we have proved that if \( x = (x_1, x_2, ..., x_N) \) is a Nash equilibrium, then

\[
\left\{
\begin{aligned}
    x_i \cdot p(X) &\leq c_i \\
    x_i &\leq s_i \\
    \frac{\partial x_i \cdot p(X)}{\partial x_i} &\geq 0
\end{aligned}
\right.
\]

Now we want to prove that the equality sign holds for at least one of them. If none of the equality sign holds, then

\[
\left\{
\begin{aligned}
    x_i \cdot p(X) &< c_i \\
    x_i &< s_i \\
    \frac{\partial x_i \cdot p(X)}{\partial x_i} &> 0
\end{aligned}
\right.
\]

From \( x_i \cdot p(X) < c_i \), we know that the bank doesn’t raise enough cash. Since \( x_i < s_i \), the bank can sell more. Since \( \frac{\partial x_i \cdot p(X)}{\partial x_i} > 0 \), the bank can raise more cash by sell more. This means the current selling amount \( x_i \) is not optimal for bank \( i \). Given other banks current selling amount, bank \( i \) do want and can change its selling amount to raise more cash. Thus this is not a Nash equilibrium. Thus we have proved that the equality sign holds for at least one of them.

Now we prove that if for each \( i \), the following three inequalities hold and the equality sign holds for at least one of them,
then the selling vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ is a Nash equilibrium. When the three inequalities hold, we’ll argue that if either of the equality sign holds for bank $i$, then it reaches its optimal.

If the first equality sign holds ($x_i \cdot p(X) = c_i$), then bank $i$ raise enough cash and reaches optimal.

If the first equality sign does not hold ($x_i \cdot p(X) < c_i$), then bank $i$ will raise as much cash as it can. If $x_i < s_i$ and $\frac{\partial x_i \cdot p(X)}{\partial x_i} > 0$, bank $i$ has more inventory and can raise more by selling more. Thus it cannot be optimal. Thus at least one of the last two equality sign will hold.

In summary, a selling vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ is a Nash equilibrium if and only if for each $i$, the following three inequalities hold and the equality sign holds for at least one of them.

\[
\begin{cases}
  x_i \cdot p(X) \leq c_i \\
  x_i \leq s_i \\
  \frac{\partial x_i \cdot p(X)}{\partial x_i} \geq 0
\end{cases}
\]
Chapter 3

Modeling Bubbles Caused by New Entrants

An oft-mentioned but under-studied feature of asset price bubbles is a surge of new entrants, retail investors who never invested, joining the bubble because of their friends or neighbors. We incorporate this viral element into an otherwise standard bubble model with forward-looking agents. Optimism spreads across households following an epidemic process and the participation rate rise as new entrants buy anticipating trending prices. Insiders or institutions gradually sell their shares, generating trading volume and moderating price growth. Our model rationalizes several patterns in the data, which have been difficult to explain, including trending prices and volume peaking months before prices, participation rates, and short-selling in both stock market and housing bubbles. This chapter is presented at the Bank of International Settlements and Hong Kong University of Science and Technology.
3.1 Introduction

Accounts of speculative asset price bubbles invariably point to a surge of new entrants, retail investors or households who never owned a stock before, joining the bubble. For instance, the South Sea Bubble of 1720 saw a significant increase in new investors including Isaac Newton as enthusiasm for the South Sea Companies reached a significant fraction of the population. Other historical bubble episodes such as the 1920’s boom in the US stock market before the Great Depression (Mackay and Galbraith) and most recently the Internet Bubble of 1996-2000 and the Housing Bubble of 2002-2007 also witnessed a gradual rise in investor participation rates. Household participation rates in the US stock market rose from 40% in the early nineties to an all-time peak of nearly 50% during the peak of the Internet Bubble. During the Housing Bubble, investment home ownership rates also climbed substantially as investment homes, which account for 17% of residential transactions in a typical year to 28% during the peak of the bubble.

This gradual entrance of new investors is tied to stories of friends and relatives who got rich off of the gradual run-up in prices of the bubble asset. Indeed, for many economists, a bubble is often synonymous with an epidemic or viral spreading of optimism about the newest investment across the population. In his Nobel lecture and book, Irrational Exuberance, Robert Shiller puts an epidemic nature at the center of a definition of a bubble: “A situation in which news of price increases spurs investor enthusiasm which spreads by psychological contagion from person to person, in the process amplifying stories that might justify the price increase and bringing in a larger and larger class of investors, who, despite doubts about the real value of the investment, are drawn to it partly through envy of others’ successes and partly through a gambler’s excitement.”

There is both anecdotal and panel data evidence on the importance of word-of-mouth communication as a transmission mechanism in asset price bubbles. The
story of Isaac Newton’s foray into South Sea Companies is supposedly related to hearing of stories of his friends and relatives getting rich. The recent speculative excesses of Twitter also point toward the importance of inexperienced new investors gradually piling into shares as stories of the gains are spread through the population. The Wall Street Journal article, entitled “Why one woman is eager to buy Twitter shares”, profiles Mrs. Watkins, a 56-year old administrative assistant and avid Mother Jones reader who opened a TD Ameritrade account so she could buy about 50 shares of Twitter. She also doesn’t really trust the market and says it’s like gambling, but with the addition of lies and subterfuge. “I’m just buying because everybody’s talking about Twitter,” she said. “I’m just gonna take a chance.” This word-of-mouth element in stock markets has been documented in panel data studies of household participation decisions during the the Internet Bubble (Hong, Kubik and Stein).

Yet the leading bubble theories do not explicitly incorporate time-varying market participation rates and the epidemic element. Even in the first generation rational bubble models of Blanchard and Watson and overlapping generations models of Tirole and Woodford, which allowed for new generations of investors, focused on stationary equilibrium as opposed to the non-stationary rise in participation rates. Noise trader, sentiment models or disagreement and short-sales constraint models of Harrison and Kreps, Morris, Scheinkman and Xiong, Hong and Sraer assume the same investors are always in the market as they speculate on their heterogeneous beliefs.

But as we will argue below, incorporating this epidemic element is crucial as existing models cannot explain a number of key facts about asset price bubbles. Two of these facts can be gleaned from Figure 1 where we plot both the trading volume and prices for the Housing Bubble of 2002-2007 (Panel A) and the Chinese Stock Market Bubble of 2005-2008 (Panel B). The second bubble episode is well known and the Chinese Bubble, which we will discuss in more detail below, arose as a result of the share reforms of 2005 which led to a windfall for early retail investors and
stories of these windfalls and how the Chinese government would not let the stock market crash before the 2008 Olympics spread across the retail population.

Figure 3.1: Price and Volume Dynamics of the U.S. Housing Bubble and the China’s Stock Bubble

In these cases, first notice that there are trending prices for a sustained period of time followed by a crash. It is well known that there is significant price momentum in both stock and housing markets not only in the time series and also in the cross-section that are more prominent during these bubble episodes. Second, and most importantly from our perspective, trading volume peaks noticeably and significantly ahead of prices. For the Housing Bubble, the peak of residential transactions occurs three months ahead of peak of the prices.

And in the case of China (Panel B), we also see that trading volume peaks before prices. Most interestingly, we can also plot the household participation rate by month, which jumps from 30 million accounts (or around 3% of the population) in 2005 to nearly 55 million accounts in 2007 (or 5.5% of the population). Notice that the participation rate rises according to an S-shape, which we argue is consistent with an epidemic and drives both price and trading volume patterns in markets.

Indeed, existing theories have a hard time naturally generating these patterns.
For instance, in the disagreement and short-sales constraints models, there are no trending prices and price and volume are highly correlated in time since an increase in overconfidence leads to higher prices and volume. The momentum trading models can of course deliver trending prices but volume in these models increase the more sustained the price increases, which implies that volume should peak at the same time or even after prices.

To capture these important and extant stylized facts, we extend a bubble model which consists of two types of forward-looking and mean-variance investors who disagree about the mean of a liquidating dividend paid at a known future date $T$. Retail investors who are optimistic and insiders (or short-sellers) who are less-optimistic or well-calibrated. We start at $t = 0$ with a number of $m$ insiders and $n_0$ retail investors. We assume that the number of optimistic retail investors enters the market following an epidemic process in which the change of the number of retail investors is governed by the standard epidemic equation

$$\Delta n_t = s n_{t-1}(1 - n_t)$$

Agents otherwise are optimizing and understand the structure of the model.

In the equilibrium, retail investors as they enter the market optimally only buy and never sell since they are optimistic and expect prices to rise as more retail investors enter the market. Without insiders, there would be no trade and prices will adjust instantly to the most optimistic value. But insiders are pessimistic and would ideally sell at these high prices. However, they recognize that prices will trend though they cannot be sure that the trend will continue for sure. So they will do some selling which then moderates the price growth.

At the beginning, there are not many retail investors to spread the optimism and so volume is low. Volume hits a peak when almost half the population has entered the
market since epidemic is maximized when there are a lot of retail investors to spread the news an enough to receive it. But this is well before price peaks since price peaks when most of the population of retail investors has entered. When price reaches a peak, volume is also low and falling since there are few new entrants remaining.

When news about the terminal dividend gets announced, prices fall as investors’ beliefs converge. This can then capture why there is also little volume during the end of bubbles since the beliefs of the investors have converged.

Our result is reminiscent of Abreu and Brunemeier where arbitrageurs delay their shorts because they do not know when others will short. Here the uncertainty is over the population dynamics.

Our paper is most related to Hong, Hong and Ungureanu who use such an exogenous epidemic process to understand momentum and Bursnide et.al who examine how epidemics influence housing booms and busts.

In our empirical work, we provide evidence for our model in a few ways. The first is to use the Chinese data, where we have detailed monthly data on account openings between 2005 and 2012, to do a calibration. We fit an epidemic process to the data and use both the actual and predicted estimates and input these into our model to then generate price and volume predictions. In particular, we estimate the uncertainty in the population growth as use this to calibrate the degree of insider selling. We can compare our predictions for these aggregates to the actual data. We show that the fit is very good.

This chapter proceeds as follows. In Section 2, we develop our model and derive key testable predictions. In Section 3, we test the implications of our model in a variety of ways. We conclude in Section 4.
3.2 Model

3.2.1 Settings

Two groups of investors trade a stock on day 0, 1, ..., $T - 1$. There are $S$ shares of the stock. Each share pays dividends $d_1, d_2, ..., d_T$ on day 1, 2, ...$T$. Dividends are independent realizations of a normal distribution

$$d_t \sim N(\mu, \frac{1}{\tau})$$

where precision $\tau$ is known to all investors, while $\mu$ is unknown. Investors rationally learn from the paid dividends to estimate $\mu$ and to predict future dividends.

Label the two groups of investors by $i = 0$ and $i = 1$. On day 0, before any dividend is paid, investors in group $i$ believe

$$\mu \sim N(\mu^i_0, \frac{1}{\tau_0})$$

where the precision $\tau_0$ of the prior distribution is the same for the two groups, while the two groups have heterogeneous prior beliefs

$$\mu^0_0 \neq \mu^1_0$$

On day $t$, dividends $d_1, d_2, ..., d_t$ are realized. Investors in group $i$ update their beliefs by the Bayes’ rule

$$\mu \sim N(\mu^i_t, \frac{1}{\tau_t})$$

where

$$\tau_t = \tau_0 + t \cdot \tau$$
\[ \mu_t^i = \frac{\tau_0 \mu_0^i + \tau (d_1 + d_2 + ... + d_t)}{\tau_t} \]

Based on the new belief about \( \mu_i \), investors can predict the next dividend

\[ d_{t+1} \sim N(\mu_t^i, \sigma_t^2) \]

where\(^1\)

\[ \sigma_t^2 = \frac{1}{\tau_t} + \frac{1}{\tau} \]

The population of group 0 is always \( m \). While the population of group 1 will increase from \( n_0 \) to \( N \) based on the epidemic model

\[ n_{t+1} - n_t = s \cdot n_t \cdot (N - n_t) \]

which means that the growth of population is proportional to the number of people who are already in the market and the number of people who are not in the market yet. \( s \) is the spreading coefficient.

On day \( t \), an investor in group \( i \) wants to maximize

\[ E_i^T[W_T^i] - \frac{1}{2} \gamma Var_i[W_T^i] \]

where

\[ W_T^i = W_t^i + \theta_t^i(d_{t+1} + P_{t+1} - P_t) + \theta_{t+1}^i(d_{t+2} + P_{t+2} - P_{t+1}) + ... + \theta_{T-1}^i(d_T + 0 - P_{T-1}) \]

All investors know all the assumptions mentioned above.

3.2.2 Equilibrium price and holding on day $T - 1$

<table>
<thead>
<tr>
<th>Day $T - 1$</th>
<th>Realized Variable</th>
<th>Random Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>dividend</td>
<td>$d_1, d_2, ..., d_{T-1}$</td>
<td>$d_T$</td>
</tr>
</tbody>
</table>

An investor in group $i$ believes

$$d_T \sim N(\mu^i_{T-1}, \sigma^2_{T-1})$$

where $\mu^i_{T-1}$ is a result of learning

$$\mu^i_{T-1} = \frac{\tau_0 \mu_0 + \tau d_1 + \tau d_2 + ... + \tau d_{T-1}}{\tau_{T-1}}$$ \hspace{1cm} (3.1)

The final wealth of an investor in group $i$ is

$$W^i_T = W^i_{T-1} + \theta^i_{T-1}(d_T + 0 - P_{T-1})$$

Since the only random term is $d_T$ and we only care about terms with $\theta^i_{T-1}$, we have

$$E^i_{T-1}[W^i_T]_{\text{with } \theta^i_{T-1}} = \theta^i_{T-1}[\mu^i_{T-1} - P_{T-1}]$$

$$Var^i_{T-1}[W^i_T]_{\text{with } \theta^i_{T-1}} = (\theta^i_{T-1})^2 \sigma^2_{T-1}$$

Let

$$\frac{d}{d\theta^i_{T-1}}(E^i_{T-1}[W^i_T] - \frac{1}{2} \gamma Var^i_{T-1}[W^i_T]) = 0$$

We have the optimal holding given current price $P_{T-1}$

$$\theta^i_{T-1} = \frac{\mu^i_{T-1} - P_{T-1}}{\gamma \sigma^2_{T-1}}$$
Market clearing condition

\[ \theta_{T-1}^1 \cdot n_{T-1} + \theta_{T-1}^0 \cdot m = S \]

yields

\[ P_{T-1}^* = \frac{n_{T-1}}{n_{T-1} + m} \mu_{T-1}^1 + \frac{m}{n_{T-1} + m} \mu_{T-1}^0 - \frac{\gamma \sigma_{T-1}^2 S}{n_{T-1} + m} \]

The equilibrium holding for an investor in group \( i \) is

\[
\theta_{T-1}^{i*} = \theta_{T-1}^i(P_{T-1}^*) \\
= \frac{1}{\gamma \sigma_{T-1}^2} \left[ \frac{n_{T-1}}{n_{T-1} + m} (\mu_1 - \mu_0) \frac{\tau_0}{\tau_{T-1}} + \frac{m}{n_{T-1} + m} (\mu_1 - \mu_0) \frac{\tau_0}{\tau_{T-1}} \right] + \frac{S}{n_{T-1} + m}
\]

The nice property here is that \( \theta_{T-1}^{i*} \) does not depend on the realized dividends \( d_1, d_2, \ldots, d_{T-1} \). The terms with dividends are eliminated when we substitute the equilibrium price \( P_{T-1}^* \) into the demand function \( \theta_{T-1}^i(P_{T-1}) \).

### 3.2.3 Equilibrium price and holding on day \( T - 2 \)

<table>
<thead>
<tr>
<th>Day ( T - 2 )</th>
<th>Realized Variable</th>
<th>Random Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>dividend</td>
<td>( d_1, d_2, \ldots, d_{T-2} )</td>
<td>( d_{T-1}, d_T )</td>
</tr>
<tr>
<td>Other terms</td>
<td></td>
<td>( \mu_{T-1}, P_{T-1}^* )</td>
</tr>
</tbody>
</table>

An investor in group \( i \) believes

\[ d_{T-1}, d_T \sim N(\mu_{T-2}^i, \sigma_{T-2}^2) \]
where $\mu^i_{T-2}$ is a result of learning

$$
\mu^i_{T-2} = \frac{\tau_0 \mu^i_0 + \tau d_1 + \tau d_2 + \ldots + \tau d_{T-2}}{\tau_{T-2}} \quad (3.2)
$$

Since $d_{T-1}$ is random viewed by today, the beliefs on day $T-1$ are also random viewed by today:

$$
\mu^i_{T-1} = \frac{\tau_0 \mu^i_0 + \tau d_1 + \tau d_2 + \ldots + \tau d_{T-2} + \tau d_{T-1}}{\tau_{T-1}} \quad (3.3)
$$

Since tomorrow’s price is a function of tomorrow’s beliefs, the price is also random.

$$
P^*_T = \frac{n_{T-1}}{n_{T-1} + m} \mu^i_{T-1} + \frac{m}{n_{T-1} + m} \mu^0_{T-1} - \frac{\gamma \sigma^2_{T-1} S}{n_{T-1} + m}
$$

Re-write the price as

$$
P^*_T = C_{T-1} d_{T-1} + D^i_{T-1}
$$

where $C_{T-1} = \frac{\tau}{\tau_{T-1}}$ is the coefficient of $d_{T-1}$ and $D^i_{T-1}$ is the sum of all other terms that do not contain $d_{T-1}$.

The final wealth of an investor in group $i$ is

$$
W^i_T = W^i_{T-2} + \theta^i_{T-2} (d_{T-1} + P^*_T - P_{T-2}) + \theta^i_{T-1} (d_T + 0 - P^*_T)
$$

Re-write final wealth as a function of random terms

$$
W^i_T = W^i_{T-2} + \theta^i_{T-2} (d_{T-1} + C_{T-1} d_{T-1} + D^i_{T-1} - P_{T-2}) + \theta^i_{T-1} (d_T + (0 - C_{T-1}) d_{T-1} - D^i_{T-1})
$$

Then

90
\[ E_{T-2}^i[W_T^i]_{\text{with } \theta_{T-2}} = \theta_{T-2}^i[\mu_{T-2} + C_{T-1}\mu_{T-2} + D_{T-1}^i - P_{T-2}] \]
\[ = \theta_{T-2}^i[E_{T-2}^i[d_{T-1} + P_{T-1}^*] - P_{T-2}] \]
\[Var_{T-2}^i[W_T^i]_{\text{with } \theta_{T-1}} = Var_{T-1}^i(\theta_{T-2}^i(1 + C_{T-1})d_{T-1}) + 2Cov_{T-1}^i(\theta_{T-2}^i(1 + C_{T-1})d_{T-1}, \theta_{T-1}^i(0 - C_{T-1})d_{T-1}) \]

To understand the variance here, first notice that only the random term \((d_{T-1} \text{ and } d_T)\) in the final wealth \(W_T^i\) can affect the variance. Then notice that we only care about the terms with \(\theta_{T-2}^i\) in the final wealth \(W_T^i\). Since \(Cov(d_{T-1}, d_T) = 0\), the only two terms remaining is \(Var_{T-1}^i(\theta_{T-2}^i(1 + C_{T-1})d_{T-1})\) and \(2Cov_{T-1}^i(\theta_{T-2}^i(1 + C_{T-1})d_{T-1}, \theta_{T-1}^i(0 - C_{T-1})d_{T-1})\). The second term captures the exposure of today and tomorrow’s holding to the random dividend \(d_{T-1}\). On day \(T - 1\), when the dividend is realized with a high value, investors will make more money from the daily return \(d_{T-1} + P_{T-1}^* - P_{T-2}\) by the holding \(\theta_{T-2}^i\) they hold on day \(T - 2\). However, the high price \(P_{T-1}^*\) will decrease the daily return \(d_T - P_{T-1}^*\). Thus, the holding \(\theta_{T-2}^i\) and the holding \(\theta_{T-1}^i\) has an opposite exposure to the dividend \(d_{T-1}\). With this term in investors’ utility, they will try to hedge the risk by adjusting today’s position \(\theta_{T-2}^i\) optimally. Write the variance in the form with \(\sigma_{T-2}^2\)

\[Var_{T-2}^i[W_T^i]_{\text{with } \theta_{T-1}} = [(\theta_{T-2}^i)^2(1 + C_{T-1})^2 + 2\theta_{T-2}^i(1 + C_{T-1})\theta_{T-1}^i(0 - C_{T-1})]\sigma_{T-2}^2 \]

Let
\[
\frac{d}{d\theta_{T-2}^i}(E_{T-2}^i[W_T^i]_{\text{with } \theta_{T-1}}) - \frac{1}{2}\gamma Var_{T-2}^i[W_T^i]_{\text{with } \theta_{T-1}} = 0
\]

We have the optimal holding given current price \(P_{T-2}\)

\[\theta_{T-2}^i = \frac{E_{T-2}^i[d_{T-1} + P_{T-1}^*] - P_{T-2}}{(1 + C_{T-1})^2\gamma\sigma_{T-2}^2} + \frac{H_{T-1}}{(1 + C_{T-1})}\]

where \(H_{T-1}\) is the term caused by hedging the exposure of future holdings on the
next dividend $d_{T-1}$:
\[ H^i_{T-1} = -[\theta^i_{T-1}(0 - C_{T-1})] \]

Market clearing condition

\[ \theta^1_{T-2} \cdot n_{T-2} + \theta^0_{T-2} \cdot m = S \]

yields

\[ P^*_{T-2} = \frac{n_{T-2}}{n_{T-2} + m} (E^i_{T-2}[d_{T-1} + P^*_{T-1}] + (1 + C_{T-1})\gamma \sigma^2_{T-2} H^1_{T-1}) \]
\[ + \frac{m}{n_{T-2} + m} (E^0_{T-2}[d_{T-1} + P^*_{T-1}] + (1 + C_{T-1})\gamma \sigma^2_{T-2} H^0_{T-1}) \]
\[ - \frac{1}{n_{T-2} + m} (1 + C_{T-1})^2 \gamma \sigma^2_{T-2} S \]

The equilibrium holding for an investor in group $i$ is

\[ \theta^i_{T-2} = \theta^i_{T-2}(P^*_{T-2}) \]
\[ = \frac{1}{(1+C_{T-1})^2 \gamma \sigma^2_{T-2}} \left[ \frac{n_{T-2}}{n_{T-2} + m} (\mu^i - \mu^0) \tau_{T-2} + \frac{m}{n_{T-2} + m} (\mu^i - \mu^0) \tau_{T-2} \right] \]
\[ + \frac{1}{(1+C_{T-1})} \left[ \frac{n_{T-2}}{n_{T-2} + m} (H^i_{T-1} - H^1_{T-1}) + \frac{m}{n_{T-2} + m} (H^i_{T-1} - H^0_{T-1}) \right] \]
\[ + \frac{S}{n_{T-2} + m} \]

Again, $\theta^i_{T-2}$ does not depend on the realized dividends $d_1, d_2, ..., d_{T-2}$.

3.2.4 Equilibrium price and holding on day $t$

By solving the price and holding recursively, we have

\[ P^*_t = \frac{n_t}{n_t + m} (E^i_t[d_{t+1} + P^*_{t+1}] + (C_{t+1} + 1)\gamma \sigma^2_{t} H^1_{t+1}) \]
\[ + \frac{m}{n_t + m} (E^0_t[d_{t+1} + P^*_{t+1}] + (C_{t+1} + 1)\gamma \sigma^2_{t} H^0_{t+1}) \]
\[ - \frac{1}{n_t + m} (C_{t+1} + 1)^2 \gamma \sigma^2 S \]

where
\[ H_{t+1}^i = -[\theta_{t+1}^{i*}(C_{t+2} - C_{t+1}) + \theta_{t+2}^{i*}(C_{t+3} - C_{t+2}) + \ldots + \theta_{T-1}^{i*}(C_T - C_{T-1})] \]

\[ P_T^* = 0 \quad \text{and} \quad H_T^i = 0 \]

\[ C_t = \frac{(T - t)r}{\tau_t} \quad \text{for} \ t = 0, 1, ..., T \]

\[ H_{t+1}^i > 0 \] appears because investors optimally hedge the risk of their future holdings on \( d_{t+1} \). For all \( s \geq t + 1 \), future prices \( P_s^* \) depend on the realization of \( d_{t+1} \) linearly. The coefficient of \( d_{t+1} \) in \( P_s^* \) is \( C_s \).

\[ \theta_t^{i*} = \theta_t^i(P_t^*) \] is the equilibrium holding of an investor in group \( i \) on day \( t \)

\[ \theta_t^{i*} = \theta_t^i(P_t^*) = \frac{1}{\gamma(C_{t+1})^2\tau_t^m} \left[ \frac{1}{n_t+m} \left( \mu_0 - \mu_0^1 \right) n_t^m + \frac{1}{n_t+m} \left( \mu_0^i - \mu_0^1 \right) n_t^m \right] + \frac{1}{\gamma(C_{t+1})^2\tau_t^m} \left[ \frac{1}{n_t+m} (H_{t+1}^i - H_{t+1}^1) + \frac{m}{n_t+m} (H_{t+1}^i - H_{t+1}^0) \right] + \frac{s}{n_t+m} \]

### 3.3 Calibrations and Empirical Tests

#### 3.3.1 Simulation method

As we have solved, \( \theta_t^{i*} \) depends on \( H_{t+1}^i \), while \( H_{t+1}^i \) depends on \( \theta_{t+1}^{i*}, \theta_{t+2}^{i*}, ..., \theta_{T-1}^{i*} \).

Since we already solved \( \theta_{T-1}^{i*} \), we can solve the \( \theta_t^{i*} \) and \( H_{t+1}^i \) sequences together recursively. It turns out that these two sequences do not depend on the realization of dividends. Thus they are determined.

Solving the price is more complicated because it is not only recursive, but also depends on realized dividends and investors’ heterogeneous expectations. Here is the pseudo-code to solve this problem:

```plaintext
price<-function(dv, t){
    if (t==T) {
        ...
    }
    ...
}
```
return(0)
}

else {
E1d<-get.next.Ed(dv,mu1)
E0d<-get.next.Ed(dv,mu0)
dv1<-c(dv,E1d)
#add the new element E1d at the end of the vector dv
dv0<-c(dv,E0d)

Calculate pricet
#based on E1d, price(dv1,t+1), E2d and price(dv0,t+1)
return(pricet)
}
}

price() is a recursive function (it calls itself). It takes the current day $t$ and the realized dividend vector $dv = (d_1, d_2, ..., d_t)$ as input. If $t < T$, it will calculate group $i$'s expectation of the next dividend $d_{t+1}$ based on the realized dividends $d_1, d_2, ..., d_t$ and group $i$'s prior belief $\mu_{0i}$

$$E^{i}_{t}[d_{t+1}] = \mu_{t}^{i} = \frac{\tau_{0}\mu_{0i} + \tau(d_1 + d_2 + ... + d_t)}{\tau_t}$$

Then, we have group $i$'s expectation of tomorrow's dividend vector

$$dvi = (d_1, d_2, ..., d_t, E^{i}_{t}[d_{t+1}])$$

Then price(dvi,t+1) is just group $i$'s expectation of the next price $E^{i}_{t}[P_{t+1}^*]$. When executing the function price(), it calls itself twice. Then each of the called function will call the price function twice. Until $t = T$, it will return 0 and then functions with input $t = T - 1$ can execute the calculation and send the results to functions with input $t = T - 2$. After all results are aggregated to the current price function with input $t = t$, it can generate the current price.

Since the dividends are paid during trading, price $P_{t}^*$ is just the valuation of the
future dividends $d_{t+1}, d_{t+2}, ..., d_T$. To compare price across different days, we normalize the price by the number of unpaid dividends

$$p_t^* = \frac{1}{T-t}p_t^*$$

All the prices plotted in the following sections are the adjusted price $p_t^*$. Volume is calculated as

$$V_t = \frac{1}{2}[n_{t-1}|\theta_t^1 - \theta_{t-1}^1| + (n_t - n_{t-1})|\theta_t^1| + m|\theta_t^0 - \theta_{t-1}^0|]$$

where $n_{t-1}$ is the population of group 1 on day $t-1$. For each of them, the change of holding is $|\theta_t^1 - \theta_{t-1}^1|$. The population of new coming investors is $(n_t - n_{t-1})$. Each of them get $|\theta_t^1|$ shares when coming into the market. $m$ is the population of group 0. For each of them, the change of holding is $|\theta_t^0 - \theta_{t-1}^0|$.

3.3.2 Simulation results

Let group 0’s prior belief be $\mu_0^0 = 0$ and group 1’s prior belief be $\mu_0^1 = 1$. Let the true mean of the dividends be $\mu = \frac{1}{2}$. The precision of investors’ prior is $\tau_0 = 1$, and the precision of the dividend is $\tau = 2$.

The population of group 0 is $m = 10$. During the $T = 10$ days, the population of group 1 increases from $n_0 = 1$ to $n_{max} = 100$ with an epidemic coefficient $s = 2$.

The total supply is normalized to $S = 1$ and the risk averse coefficient is set to be $\gamma = 1$.

To simulate the realized price curve, we generate random realizations of dividends according to its true distribution

$$d_t \sim N(\mu, \frac{1}{\tau})$$

Here is a plot based on 100 runs. The volume curve (blue) is not random because
the equilibrium holdings do not depend on the realization of dividends. Each green curve is a realization of the price. The average price is plotted in black.

**Figure 3.2: Realized Price and the Average Price**

\[
T=10; \ m=10, n_0=1, n_T=100, s=2; \ \text{tau}_0=1, \ \text{tau}=2; \ S=1, \ \text{gamma}=1
\]

In the following graph, we realize each dividend by its true mean \( d_t = \frac{1}{2} \). The dashed line is the posterior beliefs of the two groups. Beliefs start at their original level 0 or 1 and converge to \( \frac{1}{2} \) gradually. The red line between the dashed line reflect the population balance \( \frac{n_t}{m+n_t} \). When it touches the lower dashed line, almost all investors in the market are from group 0; when it touches the upper dashed line, all investors in the market are from group 1. If total shares available is zero \( S = 0 \), then the price will start at some point between investors' prior beliefs and converge to the posterior belief of group 1 because more and more group 1 investors are coming and all investors know it. When there is positive supply \( S > 0 \), price gets lower due to the risk sharing effect. This effect is more significant at the beginning of trading because investors feel more risky when they have less information. When population grows fast, there are more trading volume because new coming investors are buying.
Volume peaks before price.

Figure 3.3: Volume Peaks before Price

\[ T=10; \ m=10, \ n_0=1, \ n_T=100, \ s=2; \ \tau_0=1, \ \tau=2; \ S=1, \ \text{gamma}=1 \]

In the following two graphs, we pick different \( s \) such that they have different speed of population growth. When population increases faster, volume and price peaks earlier.
Figure 3.4: Volume Peaks Earlier when Population Increases Faster

\[ T=20; \ m=10, \ n_0=1, \ n_T=1000, \ s=1; \ \tau_0=1, \ \tau=0.03; \ S=1, \ \gamma=1 \]

![Graph showing volume peaks earlier with population increases faster.](image)

Figure 3.5: Volume Peaks later when Population Increases Slower

\[ T=20; \ m=10, \ n_0=1, \ n_T=1000, \ s=0.5; \ \tau_0=1, \ \tau=0.03; \ S=1, \ \gamma=1 \]

![Graph showing volume peaks later with population increases slower.](image)
3.3.3 Calibrate China’s stock bubble

Chinese news report in February 2007:

“There was a wealth effect of 120% return of Chinese A-share market in the last year 2006. This wealth effect was sufficiently spread during the New Year holiday. After the New Year holiday, the number of new accounts opening reached the historical high, 55,000 accounts per day, 1.7 times of the daily account opening before the New Year. After this, the number of account opening kept increase...In this January, brokers and related bank systems got over-loaded. There were long queues...”

The following graph plots the number of accounts in Chinese A share market. We pick the interval between the two vertical lines in the following graph. During this time window, account number increases slowly, fast and then slowly. This is one cycle of the previous cited epidemic. (From Sept 2006 to Apr 2008)

Figure 3.6: Number of Accounts in China Shanghai A Market

Number of Accounts in China Shanghai A Market

\[\text{http://www.hsg168.com/article.aspx?id=177}\]
We use the real number of accounts to understand the price and volume shape during this epidemic. Parameters are selected based on the following table.

Table 3.1: Parameters for the Model

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Unit</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_t$</td>
<td>36,610,000</td>
<td>accounts</td>
<td>Number of accounts of individual investors</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>57,570,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m$</td>
<td>200,000</td>
<td>accounts</td>
<td>Number of accounts of institutional investors</td>
</tr>
<tr>
<td>$S$</td>
<td>203,500,000,000</td>
<td>shares</td>
<td>Total tradable shares in Sept 2006</td>
</tr>
<tr>
<td>$\tau_0$</td>
<td>1</td>
<td></td>
<td>Precision of investors’ prior belief on $\mu$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>0.1</td>
<td></td>
<td>Precision of true dividend distribution</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.01</td>
<td></td>
<td>Risk aversion coefficient</td>
</tr>
</tbody>
</table>

In the following graph, volume not only matches the hump shape, but also matches the detail fluctuations. Since the model predicts the average price, it’s also nice to see that it captures the general trend of the real price.
We can fit the account number data by epidemic model as shown in the following graph. (optimally choose $n_0$, $n_{\text{max}}$ and $s$ to minimize the residual)
Based on the population generated by the epidemic model, we can generate the price and volume dynamics.
Since the population predicted by the epidemic model grows smoothly, the volume here is also smooth. The hump shape matches the real data, but the detail fluctuations are not captured by the epidemic model.

### 3.3.4 Understand account number by epidemic model

Normalize account number from Jan 2005 to Dec 2012 by Chinese population ($P_C = 126,583 \times 10^4$, the unit of account is also $10^4$)

$$k_t = \frac{\text{account}_t}{P_C}$$

Let

$$\Delta k_t = k_{t+1} - k_t$$

the regression $\Delta k_t = a + b \times [k_t (1 - k_t)]$ shows
Table 3.2: Parameters Estimated by the Regression (I)

|       | Estimate | Std. Error | t value | Pr(>|t|) |
|-------|----------|------------|---------|----------|
| a     | -0.00    | 0.00       | -0.38   | 0.71     |
| b     | 0.02     | 0.01       | 2.09    | 0.04     |

Then, we can control for the lag returns:

$$\Delta k_t = a + b \times [k_t(1 - k_t)] + c_0 \times r_t + c_1 \times r_{t-1} + c_2 \times r_{t-2} + c_3 \times r_{t-3} + c_4 \times r_{t-4} + c_5 \times r_{t-5}$$

Table 3.3: Parameters Estimated by the Regression (II)

|       | Estimate | Std. Error | t value | Pr(>|t|) |
|-------|----------|------------|---------|----------|
| a     | -0.00    | 0.00       | -1.45   | 0.15     |
| b     | 0.02     | 0.01       | 2.81    | 0.01     |
| c_0   | 0.00     | 0.00       | 2.25    | 0.03     |
| c_1   | 0.00     | 0.00       | 1.69    | 0.10     |
| c_2   | 0.00     | 0.00       | 1.87    | 0.07     |
| c_3   | 0.00     | 0.00       | 2.39    | 0.02     |
| c_4   | 0.00     | 0.00       | 1.24    | 0.22     |
| c_5   | 0.00     | 0.00       | 0.50    | 0.62     |

Without the $k(1 - k)$ term, we have

$$\Delta k_t = a + c_0 \times r_t + c_1 \times r_{t-1} + c_2 \times r_{t-2} + c_3 \times r_{t-3} + c_4 \times r_{t-4} + c_5 \times r_{t-5}$$

Generally, we can see from the following table that the epidemic model explains the account number data very well, even when controlling the past returns.
Table 3.4: Parameters Estimated by the Regression (III)

|   | Estimate | Std. Error | t value | Pr(|t|) |
|---|----------|------------|---------|--------|
| a | 0.00     | 0.00       | 4.80    | 0.00   |
| c0| 0.00     | 0.00       | 1.81    | 0.08   |
| c1| 0.00     | 0.00       | 1.29    | 0.20   |
| c2| 0.00     | 0.00       | 1.53    | 0.13   |
| c3| 0.00     | 0.00       | 2.14    | 0.04   |
| c4| 0.00     | 0.00       | 1.29    | 0.20   |
| c5| 0.00     | 0.00       | 0.57    | 0.57   |

3.4 Conclusion

We have proved that incorporating the epidemic element is crucial as existing models cannot explain two key facts about asset price bubbles: 1. the trending prices for a sustained period of time followed by a crash and the significant price momentum in both stock and housing markets. 2. trading volume peaks noticeably and significantly ahead of prices. We incorporate this viral element into an otherwise standard bubble model with forward-looking agents. There are not many retail investors to spread the optimism and so volume is low. Volume hits a peak when almost half the population has entered the market since epidemic is maximized when there are a lot of retail investors to spread the news an enough to receive it. But this is well before price peaks since price peaks when most of the population of retail investors has entered. When price reaches a peak, volume is also low and falling since there are few new entrants remaining. It turns out that our model can explain both the price and volume patterns of the recent bubbles very well.
Bibliography


[WEI and CORKERY, ] WEI, L. and CORKERY, M. After Lehman, Banks Jettison Commercial Property Debt.