Essays on the Risks and Real Effects of Non-Bank Financial Institutions

Citation

Zeng, Yao. 2016. Essays on the Risks and Real Effects of Non-Bank Financial Institutions. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493529

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

Essays on the Risks and Real Effects of Non-Bank Financial Institutions

A dissertation presented

by

Yao Zeng

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

April 2016

© 2016 Yao Zeng

All rights reserved.

Dissertation Advisors:
Professor John Campbell
Professor Jeremy Stein

Author:
Yao Zeng

Essays on the Risks and Real Effects of Non-Bank Financial Institutions

Abstract

These essays explore the risks and real effects of various non-bank financial institutions, such as open-end mutual funds, hedge funds, and venture capital. In the first essay, “A Dynamic Theory of Mutual Fund Runs and Liquidity Management,” I show that mutual funds are subject to bank-run-like risks, but the underlying mechanism is different. In the second essay, "Investment Exuberance under Cross Learning,” written with Shiyang Wang, we argue that firm cross learning can lead to inefficient investment exuberance, in which institutional investors’ (like hedge funds) play a key role. In the third essay, “Financing Entrepreneurial Production: Security Design with Flexible Information Acquisition," written with Ming Yang, we explores the optimal security design when the investor (like venture capital) can acquire information to help the entrepreneur make better investment decisions.

Contents

Abstract iii
Acknowledgments viii

Introduction **1**

1 A Dynamic Theory of Mutual Fund Runs and Liquidity Management **3**

1.1 Introduction 3
1.2 The Model 13
1.2.1 Setup 13
1.2.2 Interpretation of the Setup 22
1.2.3 Roadmap and Solution Approach 23
1.3 Shareholder Runs 24
1.3.1 Stage-Game Equilibrium Definition and Preliminary Analysis 24
1.3.2 High Cash-to-Assets Ratio Region G_h 26
1.3.3 Low Cash-to-Assets Ratio Region G_l 32
1.3.4 Intermediate Cash-to-Assets Ratio Region G_m 37
1.3.5 The Differences from Bank Runs and Market Runs 38
1.4 Fund Liquidity Management in the Presence of Runs 41
1.4.1 Dynamic Equilibrium Definition and Preliminary Analysis 41
1.4.2 The No-Run and Extreme-Run Scenarios: $\theta = 0$ and $\theta = 1$ 44
1.4.3 The General Scenarios 53
1.4.4 The Time-Inconsistency Problem 55
1.4.5 Expected Total Fire Sale Losses 57
1.5 Policy Implications 58
1.5.1 MMF Reforms 58
1.5.2 Liquidity Requirements 59
1.5.3 Redemption Fees 61
1.5.4 In-Kind Redemptions 63
1.5.5 Redemption Restrictions 65
1.5.6 Credit Lines 66

1.5.7 Swing Pricing . . . 67
1.6 Extensions . . . 69
1.6.1 Flow-to-Performance Relationship . . . 69
1.6.2 Asset Price Correlations . . . 71
1.6.3 Persistent Price Impacts . . . 72
1.7 Conclusion . . . 73

Bibliography **75**

2 Investment Exuberance under Cross Learning **84**
2.1 Introduction . . . 84
2.2 The Model . . . 93
2.2.1 The Economy . . . 93
2.2.2 Capital Providers and Investment . . . 94
2.2.3 Speculators and Secondary Market Trading . . . 96
2.2.4 Discussion . . . 97
2.3 Cross-Learning Equilibrium . . . 98
2.3.1 Equilibrium Definition . . . 98
2.3.2 Equilibrium Characterization . . . 99
2.3.3 Self-Feedback Benchmark . . . 105
2.4 Systematic Risks . . . 107
2.4.1 Impacts of Speculators' Weight on Systematic Risks . . . 108
2.4.2 Endogenous Cross Learning and Systematic Risks . . . 109
2.5 Investment Inefficiency and Competition . . . 115
2.5.1 Overall Investment Efficiency . . . 116
2.5.2 Competition and Cross Learning . . . 118
2.6 Discussion and Extension . . . 126
2.6.1 Growth Opportunities and Systematic Risks under Cross Learning . . . 126
2.6.2 Over-investment under Cross Learning . . . 127
2.6.3 Industry Momentum under Cross Learning . . . 128
2.7 Conclusion . . . 130

Bibliography **131**

3 Financing Entrepreneurial Production: Security Design with Flexible Information Acquistion **137**
3.1 Introduction . . . 137
3.2 The Model . . . 143
3.2.1 Financing Entrepreneurial Production . . . 143

3.2.2 Flexible Information Acquisition 145
3.3 Security Design 149
3.3.1 Optimal Security without Inducing Information Acquisition 151
3.3.2 Optimal Security Inducing Information Acquisition 154
3.3.3 Project Transfer 161
3.4 Optimal Securities in Different Circumstances 163
3.4.1 NPV Dimension 163
3.4.2 The Efficiency Dimension 165
3.5 Comparative Statics of the Optimal Security 168
3.5.1 Project Profitability 168
3.5.2 Project Uncertainty 169
3.6 Conclusion 172

Bibliography **173**

Appendix A Appendix to Chapter 1 **179**
A.1 Institutional Background 179
A.2 The Analysis of the Intermediate Cash-to-Assets Ratio Region G_m 180
A.3 A Micro-foundation for the Pattern of Selling Prices 185
A.4 Proofs 189

Appendix B Appendix to Chapter 2 **213**
B.1 Proofs 213

Appendix C Appendix to Chapter 3 **229**
C.1 Derivation, Extensions, and Proofs 229
C.1.1 Derivation of Convertible Preferred Stock as the Optimal Security . . 229
C.1.2 General Allocation of Bargaining Power 233
C.1.3 Proofs 235

List of Figures

1.1 Timeline . . . 14
1.2 Sample Path of the Selling Prices . . . 19
1.3 Equilibrium Cash Rebuilding Policy When $\theta = 0$ . . . 47
1.4 Equilibrium Cash Rebuilding Policies When $\theta = 0$ (Left) and $\theta = 1$ (Right) . . . 50
1.5 Equilibrium Cash Rebuilding Policy . . . 55
1.6 The Time-Inconsistency Problem . . . 56
1.7 Price Path with Persistent Price Impacts . . . 72

2.1 Self-Feedback Benchmark . . . 86
2.2 Firm / Capital Provider Cross Learning . . . 87
2.3 Mechanical Effect on Systematic Risks . . . 111
2.4 Learning Effect on Systematic Risks . . . 112
2.5 Extended Cross-Learning Framework . . . 120
2.6 Competition on Cross Learning and Efficiency Loss . . . 125

3.1 The Unique Optimal Security without Information Acquisition . . . 153
3.2 The Unique Optimal Security with Information Acquisition . . . 159
3.3 Project Transfer as A Security . . . 162
3.4 Change of Investment with $\mathbb{E}[\theta] = 0.5, \mu = 0.2$ . . . 169
3.5 Change of Uncertainty: $k = 0.4 < \mathbb{E}[\theta] = 0.5$, $\mu = 0.2$ . . . 170
3.6 Change of Uncertainty: $k = 0.525 > \mathbb{E}[\theta] = 0.5$, $\mu = 0.2$ . . . 171

A.1 Trends in Corporate Bond Mutual Funds . . . 179
A.2 Hypothetical Equilibrium Cash Rebuilding Policy . . . 205

Acknowledgments

I am greatly indebted to my advisors for their constant advice, encouragement, and feedback. I am grateful to John Campbell for many conversations with tremendous patience since my Day 1 in graduate school, to Jeremy Stein for asking the initial questions that led my to write "A Dynamic Theory of Mutual Fund Runs and Liquidity Management" and the throughout guidance, to Andrei Shleifer for teaching me how to address research questions and how to link research to reality, to Emmanuel Farhi for helping noodle with modeling details and for teaching me how to present theoretical work, and to Josh Lerner (who continues to be my collaborator on a new project) for introducing me to entrepreneurial finance. The dissertation is never possible without the countless office hours with them.

I would like to thank many people who have gave me detailed advice, feedback and constant support in various ways before I went on the job market. They include Malcolm Baker, Juliane Begenau, Roland Benabou, Lauren Cohen, Josh Coval, Edu Davila, Darrell Duffie, Ben Friedman, Nicolae Garleanu, Simon Gervais, Itay Goldstein, Ben Golub, Robin Greenwood, Steve Grenadier, Sam Hanson, Oliver Hart, Zhiguo He, Ben Hebert, Victoria Ivashina, Wei Jiang, Steve Kaplan, Divya Kirti, David Laibson, Debbie Lucas, Yueran Ma, Matteo Maggiori, Steve Morris, Jonathan Parker, Gordon Philips, Tom Philipson, Uday Rajan, Adriano Rampini, David Schartstein, Hyun Song Shin, Alp Simsek, Jon Skinner, Erik Stafford, Adi Sunderam, Luis Veceira, Laura Veldkamp, Mike Woodford, Wei Xiong, Lilei Xu, and Haoxiang Zhu. I may have missed many other names here, but I tried my best to incorporate all their inputs into my dissertation, and I am very grateful to them.

I would like to thank my co-authors of Chapter 2 and Chapter 3 of this dissertation, Shiyang Huang and Ming Yang, respectively. I am also fortunate to work with Jianqing Fan, Kevin Pan, and Xin Tong on other projects before I went on the market. I am grateful for their many stimulating conversations and tremendous patience. I never felt alone.

Finally, I would like to thank my father, Shengyu Zeng, my mother, Min Gao, and my wife, Congcong Wan. I love doing economics. But without their love, I would never have learned what love means.

Introduction

The first chapter develops a model of an open-end mutual fund that invests in illiquid assets and show that shareholder runs can occur even with a fully flexible fund NAV. The key is the fund's dynamic management of its cash buffer. Holding more cash at time t helps the fund avoid fire sales of its illiquid assets if it experiences a significant net outflow. However, the need to rebuild the cash buffer at time $t+1$ after outflows at t implies predictable sales of illiquid assets and hence a predictable decline in NAV. This generates a first-mover advantage at t, leading to shareholder runs. This mechanism differs from that underlying bank runs, which relies on fixed-NAV claims. I then study the fund's optimal dynamic cash policy in the presence of run concerns, which gives rise to the following tension. Rebuilding the cash buffer more rapidly at $t+1$ can trigger runs at t. However, lack of cash re-building makes the fund more likely to suffer another round of fire sales in the future. This tension is aggravated by a time-inconsistency problem: the fund may want to pre-commit to a less rapid cash re-building policy to avoid runs but cannot credibly convince the shareholders absent a commitment device. Therefore, despite optimal liquidity management, mutual funds are not run-free and runs can lead to higher ex-ante fire sale losses. Appropriate design of policies aiming at reducing financial stability risks of mutual funds requires taking into account the dynamic interdependence of runs and liquidity management.

The second chapter, co-authored with Shiyang Huang, investigates how cross-learning by firms amplifies industry-wide investment exuberance. When investing, firms learn from each other's stock prices to gain more knowledge about a common shock, generating higher investment sensitivity to the shock. Speculators respond by putting more weight on the shock when trading, making prices even more informative about it. This spiral produces higher investment (and price) comovements in investment exuberance. Meanwhile, cross-learning generates a new externality by making other firms' prices less informative about their idiosyncratic shocks. This externality increases in the number of firms, suggesting that more competitive industries may exhibit more inefficient investment exuberance.

The third chapter, co-authored with Ming Yang, proposes a new theory of the use of debt and non-debt securities in financing entrepreneurial production, positing that the investor can acquire costly information on the entrepreneur's project before making the financing decision. We show that debt is optimal when information is not valuable for production, while the combination of debt and equity is optimal when information is valuable. These predictions are consistent with the empirical facts regarding the finance of entrepreneurial businesses. Flexible information acquisition allows us to characterize the payoff structures of optimal securities without imposing usual assumptions on feasible securities or belief distributions.

Chapter 1

A Dynamic Theory of Mutual Fund Runs and Liquidity Management

1.1 Introduction

There are rising concerns about the financial stability risks posed by open-end mutual funds, which promise daily liquidity to shareholders but have been increasingly holding illiquid assets such as corporate bonds.[1] Many regulators are worried about the potential for a bank-run-like scenario on mutual funds investing in illiquid assets,[2] and a number of funds, the Focused Credit Fund of Third Avenue as the most notable example, shut down redemptions in the middle of severe shareholder runs at the end of 2015.[3] However, despite the prominence of this issue, the theoretical mechanism of mutual fund runs is not well understood and the existence of runs is still in dispute. First, conventional wisdom suggests that mutual funds with a flexible end-of-day net asset value (NAV) should be immune to bank-run-like crises, which occur only with fixed-NAV claims. Second, observers also argue that careful fund liquidity management can mitigate first-mover advantages and hence prevent runs. With these two points in mind, can there really be runs on mutual funds?

---

[1] These patterns of liquidity mismatch are significant in corporate bond mutual funds and are also pervasive in funds investing in other illiquid assets. See Appendix A.1 for facts and institutional details.

[2] The U.S. SEC voted 5-0 on Sept 22, 2015 to approve a new proposal requiring mutual funds to manage liquidity risks. Also see "Potential Emerging Threats and Vulnerabilities," Ch. 7 in the *Annual Report*, the U.S. FSOC, May 2015, "Asset Management Industry and Financial Stability," Ch. 3 in the *Global Financial Stability Report*, the IMF, April 2015, and "Old and New Risks in the Financial Landscape", Ch. 6 in the *Annual Report*, the BIS, June 2015.

[3] The Third Avenue shutdown on Dec 10, 2015 was the first case since the 1940 Act that a U.S. mutual fund shut down redemptions without getting approval from the U.S. SEC. In particular, the Focus Credit Fund was

In this paper, I develop a model of an open-end mutual fund that invests in illiquid assets and show that shareholder runs can occur in equilibrium even with a fully flexible NAV. My main insight is that the combination of a flexible NAV and active fund liquidity management, both of which are viewed as means to mitigate financial stability risks, can make the fund prone to shareholder runs even without any fundamental shocks to the underlying assets.

The mechanism works as follows. Holding more cash at time t helps the fund avoid fire sales of its illiquid assets if it experiences a significant net outflow. However, the need to rebuild the cash buffer at time $t+1$ after outflows at t implies predictable sales of illiquid assets and hence a predictable decline in NAV at $t+1$. This generates a first-mover advantage at t, which leads to shareholder runs in equilibrium. This logic illustrates the key trade-off in the model: cash buffers mitigate fire sales today, but the need to rebuild cash buffers tomorrow triggers runs today.

The potential for shareholder runs further gives rise to a tension regarding the fund's optimal dynamic cash management. On the one hand, rebuilding the cash buffer more rapidly at $t+1$ by aggressively selling illiquid assets can trigger runs at t for the reasons outlined above. On the other hand, rebuilding the cash buffer less rapidly makes the fund more likely to suffer another round of fire sales in the future. This tension is aggravated by a time-inconsistency problem: the fund may want to pre-commit to a less rapid cash

---

the single largest holder of many high-yield corporate bonds, the fundamental of which were still good. This suggests that the liquidity mismatch and the resulting strategic considerations among shareholders must have played an important role in the run-up to its crisis. A number of the so-called "liquid-alternative" mutual funds, operated by hedge fund managers such as Whitebox Advisors, J.P. Morgan, and Guggenheim Partners also experienced shareholder runs and were forced to close in 2015.

rebuilding policy to avoid runs but cannot credibly convince the shareholders absent a commitment device. Thus, despite optimal liquidity management, mutual funds are not run-free and runs can lead to considerably higher ex-ante fire sale losses.

My theoretical predictions are consistent with new micro-level evidence. Chen, Goldstein and Jiang (2010), Feroli, Kashyap, Schoenholtz and Shin (2014), Goldstein, Jiang and Ng (2015), Shek, Shim and Shin (2015) and Wang (2015) document that current fund outflows predict a future decline in fund NAV, and the magnitude of the predictable decline in NAV is larger if the fund invests in more illiquid assets or has less cash.[4] My model provides a concrete mechanism to explain these documented patterns of run incentives and shows that they can indeed lead to runs in equilibrium. Moreover, I show that the potential for runs can in turn distort fund liquidity management, generating new testable predictions.

I formulate the ideas sketched above in a stochastic dynamic model of an open-end mutual fund with many shareholders, who may redeem their shares daily at the end-of-day flexible NAV. Section 3.2 lays out the model, which is built based on four realistic assumptions. First, the fund invests in both cash and many illiquid assets. Second, selling illiquid assets generates fire sale losses (Williamson, 1988, Shleifer and Vishny, 1992, 1997). Third, the fund minimizes total expected fire sale losses by managing its cash buffer over time. Finally, the fire sale prices are time-varying. Specifically, outflow-induced fire sales can create temporary price overshooting at t and subsequent reversal at $t+1$, as documented by Coval and Stafford (2007) and Duffie (2010).[5] This price pattern gives rise to a motive for fund liquidity management. After a significant outflow at t, the fund may voluntarily sell some assets to rebuild its cash buffer at $t+1$ when the selling price partially rebounds, in order to avoid potentially more severe fire sales in the future (i.e., at $t+2$) should another outflow shock comes. Other than this time-varying fire sale cost, my model does not feature other shocks or frictions at the asset market level.

---

[4] These papers differ in focus, but they all suggest a single point: the existence of run incentives among mutual fund shareholders. See the literature review for more detailed discussions of these papers.

[5] This assumption is consistent with many other empirical studies including Campbell, Grossman and Wang (1993), Pastor and Stambaugh (2003), Hendershott and Seasholes (2007), Mitchell, Pedersen and Pulvino (2007), Comerton-Forde, Hendershott, Jones, Moulton and Seasholes (2010), Greenwood and Thesmar (2011), Edmans, Goldstein, and Jiang (2012), Jotikasthira, Lundblad, and Ramadorai (2012), Lou (2012), Nagel (2012) and Hendershott and Menkveld (2014). Duffie (2010) ascribes this pervasive pattern to the idea of "slow-moving capital," which is in turn linked to the classic work of Grossman and Miller (1988). In Appendix A.3 I provide a micro-foundation where the price pattern emerges endogenously but in the main text I assume it to highlight the run mechanism in the paper.

I first show in Section 1.3 that the fund's desire to rebuild its cash buffer can induce shareholder runs, and more rapid cash rebuilding leads to more severe runs. This result is general: runs can occur in equilibrium regardless of whether the fund starts with a high cash position or a low one. However, the cost of runs, that is, the impact of runs on the risk of fire sales, is different in these two cases. The nature of strategic interactions among shareholders is also different in these two cases. Therefore, it is helpful to discuss them separately to clarify the run mechanism.

On the one hand, when the fund starts with a high cash-to-assets ratio, shareholder runs induced by fund cash rebuilding lead to higher risk of future fire sales. When a redemption shock occurs at t, the fund starting with a high cash position can satisfy the projected redemptions at both t and $t+1$ without incurring fire sales. This implies that, even if the shareholders who initially plan to redeem at $t+1$ ran at t, the fund would still have enough cash at t, and thus time-t NAV would not adjust. But since some cash is paid out at t, the fund may want to rebuild its cash buffer by voluntarily selling some illiquid assets at $t+1$. Thus, the shareholders who initially plan to redeem at $t+1$ would get a lower NAV if they waited until $t+1$, and hence may decide to run at t. Fundamentally, this occurs because cash rebuilding endogenously gives rise to strategic complementarity among shareholders. In this scenario, shareholder runs force the fund to pay more redeeming shareholders at the endogenously unchanged NAV at t, and thus the fund loses more cash.[6] Therefore, shareholder runs partially offset the fund's cash rebuilding efforts at $t+1$ and thus lead to higher risk of future fire sales at $t+2$.

On the other hand, if the fund starts with a low cash-to-assets ratio such that it cannot satisfy the projected redemptions at both t and $t+1$ without selling illiquid assets, shareholder runs further lead to higher current (i.e., time-t) fire sale losses in addition to higher risk of future (i.e., $t+2$) fire sales. Interestingly, in this scenario, a shareholder is less likely to run if more of other shareholders decide to run. This is because runs may force the fund to fire sell more of its illiquid assets at an extremely low price at t, and any shareholder who runs at t has to share that cost, that is, to get a lower NAV. This means that shareholders' run decisions can exhibit strategic substitutability. However, since the fund is already running out of cash and may voluntarily sell more assets at $t+1$ to rebuild its cash buffer, waiting may only give the shareholders an even lower NAV. Therefore, the fund's desire to rebuild its cash buffer reinforces a strong incentive for shareholders to redeem earlier despite the strategic substitutability. In this scenario, shareholder runs introduce a more severe cost by directly forcing the fund to fire sell more assets at t, a time when the fire sale price is extremely low. In addition, runs still offset the fund's cash rebuilding efforts and thus lead to higher risk of future fire sales.

---

[6] In other words, if the shareholders who initially plan to redeem at $t+1$ do not run, the fund can pay them a lower NAV when rebuilding its cash buffer and hence effectively carry more cash into $t+2$ under the same cash rebuilding policy.

Having analyzed the implications of cash rebuilding on shareholder runs for an arbitrary starting level of cash, I endogenize the dynamic cash rebuilding policy of the fund. I show in Section 1.4 that introducing the potential for runs gives rise to a tension absent in existing liquidity management theories. On the one hand, rebuilding the cash buffer more rapidly at $t+1$ can trigger shareholder runs at t. As described above, shareholder runs lead to higher risk of fire sales. This run concern makes a more rapid cash rebuilding policy less appealing. On the other hand, adopting a less rapid cash rebuilding policy at $t+1$ makes the fund more likely to suffer another round of future fire sales at $t+2$. Moreover, carrying less cash to $t+2$ also implies that the fund may ultimately have to rebuild its cash buffer more rapidly at time $t+3$, which can trigger future runs at $t+2$. With this tension, the fund's optimal dynamic cash rebuilding policy is significantly different from the benchmark case where there are no runs.

Moreover, I show that the potential for shareholder runs introduces a time-inconsistency problem for the fund, which aggravates the tension in choosing between a rapid or a slow cash rebuilding policy. When the cost of runs at t is relatively large, ex-ante, the fund may wish to commit itself to rebuilding its cash buffer less rapidly at $t+1$ to reduce run risks at t. However, ex-post, the fund may instead be tempted to adopt a more rapid cash rebuilding policy at $t+1$, because the time-t cost is sunk. Anticipating this, shareholders will always have strong incentives to run at t. In other words, in the absence of a commitment device, the fund cannot make credible announcement to convince shareholders not to run. I further show that, in certain circumstances, introducing a commitment device can help temper the run incentives at t and thus reduce total expected fire sale losses.

Overall, my paper provides theoretical underpinnings for understanding why open-end mutual funds may not be run-free, in contrast to what the conventional wisdom suggests. The potential for shareholder runs can considerably increase fire sale losses in expectation despite optimal cash management by the fund. The dynamic interdependence of shareholder runs and fund liquidity management uncovered in this paper plays a critical role in shaping these outcomes.

Fundamentally, shareholder runs in my model are driven by a key contractual property of mutual fund NAVs: they are flexible but not forward-looking. In other words, the NAV at time t does not take into account the predictable asset sales and price impact at $t+1$. With this contractual property, fund cash rebuilding gives rise to predictable declines in NAV and thus the potential for runs.

Although they are reminiscent of bank runs in some ways, fund shareholder runs differ from classic bank runs (Diamond and Dybvig, 1983) in terms of the underlying mechanism. In my model, the first-mover advantage does not come from an exogenous fixed-NAV claim at t (like the deposit at a bank). Because NAVs in my model flexibly and endogenously adjust, a shareholder redeeming at t realizes that more early withdrawals will potentially induce more fire sales at t and thus lower the proceeds she receives. Hence, if the fund did not rebuild its cash buffer at $t+1$, the net benefit of running over waiting could be decreasing as more shareholders run. Rather, it is the fund's desire to rebuild its cash buffer at $t+1$ and the resulting predictable decline in NAV that lead to a strong first-mover advantage. This mechanism highlights a dynamic interaction between the fund and its shareholders. Such an interaction is absent in bank run models, which focus on coordination failures among depositors themselves.

The mechanism in my model is also different from that underlying market runs. Bernardo and Welch (2004) and Morris and Shin (2004) argue that if an asset market features an downward-sloping demand, investors fearing future liquidity shocks will have an incentive to front-run, fire selling the asset earlier to get a higher price. One might imagine that introducing an intermediary that helps manage liquidity shocks can alleviate such problems. Indeed, in my model, fund cash management is beneficial to shareholders because it reduces fire sale losses. However, the key tension that I document is that the fund's cash rebuilding also endogenously gives rise to predictable declines in NAV and thus run incentives. In contrast, there is no role for liquidity management in market run models. In this sense, market run models focus on asset markets themselves while my theory focuses on the role of financial intermediaries. This allows me to distinguish between risks that come from active management of financial intermediaries and those that are only a reflection of market-level frictions and would occur in the absence of intermediaries.

My model generates new policy implications, which I explore in Section 1.5. First, the model suggests that to introduce a flexible NAV is not a fix to money market mutual fund (MMF) runs, as also argued in Hanson, Scharfstein and Sunderam (2015). I also consider many fund-level policies, including liquidity requirements, in-kind redemptions, redemption fees and restrictions, credit lines, and swing pricing, all of which aim at mitigating financial stability risks of mutual funds. Perhaps surprisingly, these policies do not necessarily improve shareholder welfare in equilibrium because they may distort fund liquidity management, and thus lead to more fire sales. Overall, my model suggests that policies should be designed taking into account the dynamic interdependence of runs and fund liquidity management.

Section 2.6 explores various extensions to the baseline model, including the flow-to-performance relationship, asset price correlations, and persistent price impacts. Section 3.6 concludes.

Related Literature. This paper first contributes to the burgeoning literature on financial stability risks posed by open-end mutual funds.[7] Empirically, Feroli, Kashyap, Schoenholtz and Shin (2014) find that fund outflows predict future declines in NAV, suggesting the existence of run incentives for shareholders. At a more micro level, Chen, Goldstein and Jiang (2010) find that the flow-to-performance relationship is stronger for funds investing in less liquid stocks. Goldstein, Jiang and Ng (2015) echo the message by showing that corporate bond funds even exhibit a concave flow-to-performance relationship. Shek, Shim and Shin (2015) explores the underlying channel by showing that outflows are associated with future discretionary bond sales and liquidity rebuilding in an emerging market bond fund context. Wang (2015) further finds that outflows predict a stronger decline in future NAVs when the fund has less cash or invests in more illiquid bonds. My model predictions are consistent with all of these facts. Chen, Goldstein and Jiang (2010) and Morris and Shin (2014) have addressed the potential for mutual fund runs from theoretical perspectives, but their focuses and approaches are different from this paper, and they do not consider fully flexible NAV adjustment or fund liquidity management.[8]

There is a broader literature on the costs of outflows to non-trading shareholders and to future fund performance.[9] Edelen (1999), Dickson, Shoven and Sialm (2000), Alexander, Cici and Gibson (2007) and Christoffersen, Keim, and Musto (2007) find that flow-induced trades hurt fund performance, and redeeming shareholders impose externalities on non-trading shareholders through trading-related costs (including commissions, bid-ask spreads, and taxes) that are not reflected in current NAVs. Coval and Stafford (2007), Ellul, Jotikasthira and Lundblad (2011) and Manconi, Massa and Yasuda (2012) further show this by highlighting the channel of flow-induced fire sales. In addition, Chernenko and Sunderam (2015) find that even careful liquidity management of mutual funds cannot alleviate those costs. These papers do not examine the potential for shareholder runs. Especially, since runs are induced by active fund liquidity management in my model, I am able to identify a new externality that even including all the current trading-related costs in NAVs (as suggested by swing pricing) cannot internalize.

---

[7] The notion of open-end mutual funds does not include MMFs, insurance companies, or pension funds. Recent literature has also documented run-like dynamics associated with those financial intermediaries. For example, Schmidt, Timmermann and Wermers (2014) provide a throughout investigation of shareholder runs on MMFs, and Strahan and Tanyeri (2014) find that MMFs experiencing more outflows also reallocate their portfolio more, consistent with my prediction in a mutual fund context. More recently, Foley-Fisher, Narajabad and Verani (2015) and Da, Larrain, Sialm and Tessada (2015) provide new evidence suggesting that runs can even happen to insurance companies and pension funds, respectively.

[8] In the appendix of Chen, Goldstein and Jiang (2010), the authors build a static global game model for the purpose of hypothesis development. In that model, fund NAV is not fully flexible: shareholders are assumed to get a fixed-value claim if they run (in the spirit of Diamond and Dybvig, 1983). Their model does not consider fund liquidity management. Morris and Shin (2014) build a model of runs by funds on the asset markets, focusing on fund managers' relative performance concerns. Their model does not distinguish between open- and closed-end funds and does not consider shareholder runs.

[9] See Christoffersen, Musto and Wermers (2014) for a comprehensive review of this literature.

My paper also contributes to the literature of mutual fund liquidity management.[10] This literature suggests that holding cash is costly because funds have to give up other higher-yielding investment opportunities (Wermers, 2000), but cash can help them withstand redemption shocks and reduce fire sales (Edelen, 1999, Christoffersen, Keim, and Musto, 2007, Coval and Stafford, 2007). Recently, Simutin (2013), Ben-Rephael (2014) and Huang (2015) investigate the determinants of cash management and the implications on fund performance for equity funds, and Chernenko and Sunderam (2015) provide more comprehensive evidence on fund cash management covering both bond and equity funds. The most relevant theory is Chordia (1996) who shows in a static model that funds hold more cash when there is uncertainty about redemptions, but funds with load and redemption fees hold less cash. My paper document a novel aspect of fund liquidity management: rebuilding cash buffers by selling illiquid assets can induce shareholder runs, which can in turn distort fund liquidity management.

Dated back to Bryant (1980) and Diamond and Dybvig (1983), there is a vast bank run literature.[11] Beyond the discussion in the introduction, I review three branches of theories that are closely related.

---

[10] There is also a large dynamic corporate finance literature on general liquidity management for non-financial institutions. With the premise of costly external liquidity under agency problems, this literature focuses on the role of holding internal liquidity in making investments rather than meeting redemptions. I refer readers to Bates, Kahle and Stulz (2009) for empirical evidence and Bolton, Chen and Wang (2011) for a typical modern theoretical treatment. Hugonnier and Morellec (2014), Sundaresan and Wang (2014) and Della Seta, Morellec and Zucchi (2015), among others, further push this research agenda to the bank context with a focus on the relationship between liquidity management and bank default risks. In this literature, the bank can only accumulate internal liquidity by retaining its proceeds; partial liquidation (like in my model) is not allowed. The bank in question defaults abruptly when it chooses to liquidate assets.

First, pioneered by Cooper and Ross (1998), a literature looks at how banks manage their liquidity buffers to mitigate runs.[12] By combining liquidity management and fully flexible NAV adjustment, I uncover a cost of cash rebuilding in the mutual fund context: unintended shareholder runs. In contrast, when a comparable bank rebuilds its cash buffer by selling assets, the underlying deposit value will not be affected, and thus cash rebuilding by itself would not directly trigger depositor runs.

Second, there is a growing literature about runs on non-bank but leveraged financial institutions, for instance, Liu and Mello (2011) on leveraged hedge fund runs, Martin, Skeie and von Thadden (2014) on repo runs, and Parlatore (2015) on MMF runs. These theories resemble classic bank run models in that investors still get a fixed-value claim if they run.

Third, a burgeoning literature models bank runs in dynamic contexts. Some models imbed a Diamond-Dybvig type bank run model into a dynamic growth or economic fluctuation model (Ennis and Keister, 2003, Gertler and Kiyotaki, 2014). Others model investors' dynamic run decisions directly (He and Xiong, 2012, He and Manela, 2014).[13] These dynamic models make bank run theory appealing to more contexts like debt rollovers and rumor-based runs. For tractability reasons, those models do not feature liquidity management. By developing a new framework, I can model the interaction between mutual fund shareholder runs and fund liquidity management in a tractable manner.

Finally, it is also important to distinguish my mechanism from that in the market run literature (Bernardo and Welch, 2004, Morris and Shin, 2004), as aforementioned in the introduction. The idea of market run is also present in the "cash-in-the-market" theory in Allen and Gale (1994, 2005) and the predatory treading theory in Brunnermeier and Pedersen (2005) and Carlin, Lobo and Viswanathan (2007). Based on the search model of Lagos, Rocheteau and Weill (2011), Di Maggio (2015) investigates the implications of market runs in over-the-counter (OTC) markets. My run mechanism differs from theirs by focusing on the role of financial intermediaries rather than market-level frictions themselves.

---

[11] See Gorton and Winton (2003) for a comprehensive review.

[12] This literature has been growing recently (Vives, 2014, Diamond and Kashyap, 2015) given the emphasis of bank liquidity requirement in Basel III.

[13] This literature has been growing rapidly. See Cheng and Milbradt (2012) and Schroth, Suarez and Taylor (2014) for recent developments built upon He and Xiong (2012). Morris (2014) gives a theoretical treatment about how such dynamic bank run models can be reconciled with static, global-game-based bank run models (Goldstein and Pauzner, 2005) and synchronization games (Abreu and Brunnermeier, 2003).

1.2 The Model

I first introduce the baseline model of an open-end mutual fund investing in both cash and illiquid assets. The model features the key institutional setting of open-end mutual funds, that is, possible daily shareholder redemptions at a flexible end-of-day NAV. The mechanism of this baseline model is fairly general and can be readily extended to other settings.

1.2.1 Setup

Time is discrete and infinite. Discount rate is normalized to 1. There is a single open-end mutual fund with many existing shareholders, who may redeem their shares on ex-ante unknown dates. The fund invests in two types of assets: 1) cash, denoted by x, which is liquid and the only consumption good, and 2) a continuum of many illiquid assets, denoted by a. These illiquid assets are assumed to have a fundamental value $R > 1$, but they are all different, which means they can have different market prices in equilibrium. The illiquid assets pay off at the end of the game (specified later) and do not generate any interim cash flows.

Timeline. Each stage consists of two dates, an even date and an odd date.[14] I use $2t$ to denote an even date and $2t + 1$ to denote an odd date. I still use t to denote a stage or a general date when the difference between even and odd dates is not important. At the

---

[14] To make distinctions between even and odd dates is a common modeling tool in the theoretical literature to capture time-varying or alternating market conditions. See Woodford (1990) and Lagos and Wright (2005) for examples.

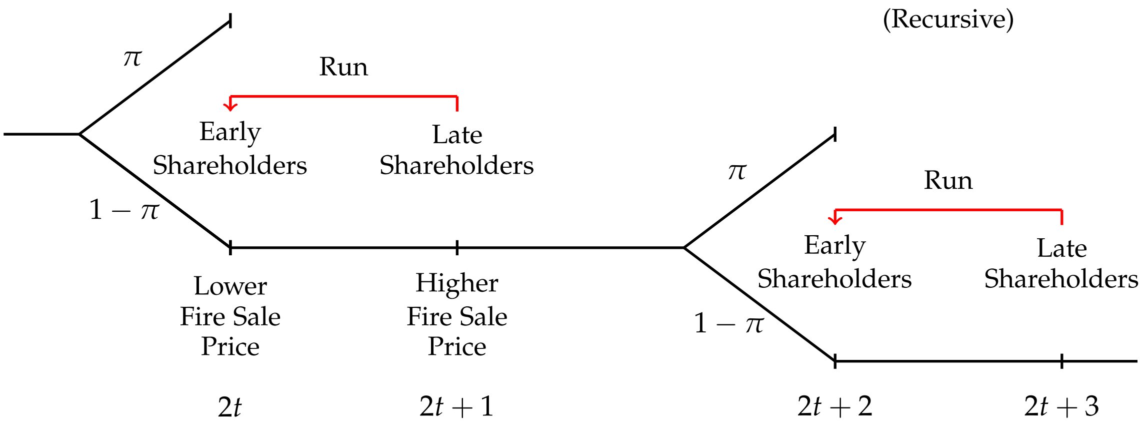

(Recursive)
π
Run
Early Shareholders
Late Shareholders
$1-\pi$
Lower Fire Sale Price
Higher Fire Sale Price
π
Run
Early Shareholders
Late Shareholders
$1-\pi$
$2t$
$2t+1$
$2t+2$
$2t+3$

Figure 1.1: *Timeline*

beginning of any date t, the fund has x_t cash and a_t unit of the basket of illiquid assets. The fund also has n_t existing shareholders, some of whom may exit the fund by redeeming their shares in future. Redemption needs must be met in cash, so the fund may be forced to sell its illiquid assets if running short of cash, but doing this will generate fire sale losses because of the underlying illiquidity problem. Specifically, the unit fire sale price for any illiquid assets on date t is p_t, which is lower than R. To focus on redemptions, I assume that the fund has no inflows or credit lines.[15]

At the beginning of each stage (i.e., right before an even date), a shock hits the economy. Specifically, with probability π the game ends; otherwise the game continues. Only if the game continues, will there be projected redemptions on the following two dates within the given stage. The random end-of-game event can be thought of as an upside event in which all the illiquid assets mature at their fundamental value, shareholders get paid off, and there will be no future redemption needs. Therefore, the shock structure parsimoniously captures the randomness of redemptions in reality, with a lower π implying that redemptions are more persistent. For these reasons, I call the shock redemption shock in what follows. Figure 1.1 shows the timeline with some elements to be explained shortly. In Section 1.2.2, I will give an intuitive interpretation of this setup and map it to real-world scenarios.

---

[15] I will relax this assumption in Section 1.5.6. As shown then, having credit lines cannot reduce potential financial stability risks of mutual funds but may instead aggravate them in some cases.

Flexible Fund NAV. The first important pillar of this model is a flexible fund NAV. The end-of-day flexible NAV will reflect all the asset sale losses during the given day. Specifically, if the fund does not sell any illiquid assets on date t, the end-of-day NAV will be

$$NAV_t = \frac{a_t R + x_t}{n_t}.$$

However, if the fund sells some assets on date t at the fire sale price p_t, the NAV will reflect that loss and hence become lower:

$$NAV_t = \frac{x_t + (a_t - a_{t+1})p_t + a_{t+1}R}{n_t}. \tag{1.2.1}$$

I explain the three terms in the numerator of (1.2.1) in order. The first term x_t is the amount of cash that the fund has initially at the beginning of date t. The second term represents the amount of cash the fund raises by selling $a_t - a_{t+1}$ unit of assets at the fire sale price p_t. The third term represents the value of the non-traded illiquid assets remained on the fund's balance sheet. In (1.2.1), the market prices of those different and non-traded assets will not change. This is true for illiquid assets that are different in nature, especially for those traded in OTC markets like corporate bonds. This is also consistent with the empirical evidence in Coval and Stafford (2007) that flow-induced fire sales only have temporary and local price impacts within the assets being sold. In practice, asset prices may be correlated, and mutual funds may also use matrix pricing for these non-traded assets based on the fire sale price p_t of the assets that are sold. But as I will show in Section 1.6.2 as an extension, asset price correlations or different accounting rules such as matrix pricing are not crucial for my model mechanism and will not change my results qualitatively. What is crucial in (1.2.1) is that the end-of-day NAV is flexible in the sense that it takes into account all the same-day price impact and asset sale losses, while that NAV is not forward-looking in the sense that it will not reflect any possible future price impacts and asset reallocation costs. These contractual features of fund NAV are robust regardless of the nature of different asset markets and accounting rules.

Fund Manager. The second important pillar of this model is active fund liquidity management. The fund has one fund manager. On any date t, the fund manager's objective is to minimize total expected fire sales from date t to the end of the game. The formal objective function will become clear shortly after I describe the asset market.

Since all the shareholders are ex-ante identical, having a fund manager who minimizes total expected fire sale losses implies that there is no agency friction between the fund manager and the shareholders as a whole. In this sense, the fund manager's objective parsimoniously captures the outcome of optimal contract design between investors and the asset manager (see Bhattacharya and Pfleiderer, 1985 for a classic treatment). Minimizing total fire sale losses also suggests that the fund manager's compensation is tied to the size or equivalently the assets under management (AUM) of the fund, which is common in practice.

Shareholders. Fund shareholders may redeem their shares when they have consumption needs. I define three groups of shareholders within each stage: early shareholders, late shareholders, and sleepy shareholders. Specifically, if the game continues on an even date $2t$ (with probability $1 - \pi$), $\mu_E n_{2t}$ early shareholders and $\mu_L n_{2t}$ late shareholders are hit by unanticipated consumption shocks and thus must consume, where $0 < \mu_E, \mu_L < 1$ and $0 < \mu_E + \mu_L < 1$. Since consumption shocks are unanticipated, the remaining $(1 - \mu_E - \mu_L)n_{2t}$ sleepy shareholders do nothing but wait until the next stage; they do not plan ahead for future stages although they may randomly become early or late shareholders in the future.[16] Both early and late shareholders do not have any cash in advance, so they have to redeem their shares and get the endogenous and flexible end-of-day NAV.

Early shareholders must consume on date $2t$, so they always redeem their shares at the endogenous end-of-day NAV on $2t$. Late shareholders prefer to consume on date $2t+1$, but can also choose to consume on date $2t$. Formally, late shareholders' utility function is:

---

[16] This is consistent with the observation that many mutual fund shareholders are mom-and-pop investors: they do not actively review their portfolios but only do so when subject to unanticipated liquidity shocks (for empirical evidence, see Agnew, Balduzzi and Sunden, 2003, Ameriks and Zeldes, 2004, Mitchell, Mottola, Utkus and Yamaguchi, 2006, Brunnermeier and Nagel, 2008, and Grinblatt and Keloharju, 2009). Institutional investors like insurance companies and pension funds also review and update their mutual fund asset portfolio infrequently. From a theoretical point of view, this helps construct a standard stochastic dynamic game with a long-run player (the fund manager) and many generations of short-run players (the shareholders). It also allows me to highlight the conflict of interests between different generations of shareholders from a dynamic perspective.

$$u_L(c_{2t}, c_{2t+1}) = \theta c_{2t} + c_{2t+1}\,, \tag{1.2.2}$$

where $0 \leqslant \theta \leqslant 1$. As late shareholders are risk neutral,[17] their consumption choice boils down to a binary problem: to redeem on date $2t$ or date $2t+1$. There is no outside storage technology, so if a late shareholder redeems on date $2t$, she gets the endogenous end-of-day NAV on $2t$ and must consume immediately; otherwise she gets the endogenous end-of-day NAV on $2t+1$ and consume then. If a late shareholder chooses to redeem and consume on date $2t$, I say that the late shareholder runs the fund. I allow late shareholders to choose mixed strategies: the run probability of late shareholder i is denoted by $\lambda^i_{2t} \in [0,1]$. As one can expect, a late shareholder's run decision will depend on the difference of NAV between the two dates, which will in turn depend on other late shareholders' run decisions and the fund manager's asset allocation decision in the given stage.

The preference parameter θ in (1.2.2) parsimoniously captures different types of shareholders with different propensities to run.[18] Intuitively, when θ is lower, late shareholders prefer late consumption more and they are less likely to run even if the NAV is lower on date $2t+1$. This setting implies that late shareholders' realized marginal utilities can be different on the two dates, a setting also commonly seen in the bank run literature (for example, Peck and Shell, 2003). My model mechanism works for any $\theta \in [0,1]$.

The presence of different types of shareholders, that is, the early, the late, and the sleepy ones, captures an important friction in my model: the conflict of interests among shareholders with different investment horizons. This conflict of interests is natural and well-documented in the empirical literature: short-term investors can impose negative impacts on a fund's long-term performance through the transaction and fire-sale costs they incur to the fund, and the fund has to manage its portfolio and liquidity carefully to mitigate such costs (for example, Edelen, 1999, Dickson, Shoven and Sialm, 2000, Alexander, Cici and Gibson, 2007, Christoffersen, Keim, and Musto, 2007, Coval and Stafford, 2007). This realistic friction plays an important role in driving the interaction between shareholder runs and fund liquidity management in my model.

---

[17] In Diamond-Dybvig type bank run models, depositors are usually assumed to be risk averse, and demandable deposit emerges as the optimal contract for risk-sharing between early and late depositors. Instead, I focus on the commonly observed contractual features of open-end mutual funds rather than optimal contract design, so I assume risk-neutrality to help better document the impact of flexible NAVs on shareholders' consumption choices.

[18] There are many plausible explanations for different types of shareholders to have different values of θ. For example, Chen, Goldstein and Jiang (2010) suggest that institutional investors may have a lower θ because they often have stricter investment targets and are more likely to internalize the market impact posed by own trading activities. Alternatively, Gennaioli, Shleifer and Vishny (2015) argue that mutual funds provide trust to their shareholders. For those shareholders who value such trust, if they choose to leave the fund early, they have to give up the trust premium so can also be viewed as having a lower θ.

Asset Market and Fire Sales. On any date, the fund manager can sell the illiquid assets to an outside investor at fire sale prices. Flow-induced fire sales are natural and pervasive (Williamson, 1988, Shleifer and Vishny, 1992, 1997), and they can create temporary price impacts (Coval and Stafford, 2007). Based on these evidence, I assume that the selling price of any unit of illiquid assets is $p_E = \delta_E R$ on date $2t$ and $p_L = \delta_L R$ on date $2t+1$, where $0 < \delta_E, \delta_L < 1$. Moreover, I assume that selling right after the shock (i.e., on date $2t$) incurs a higher price discount (and thus a lower selling price). Figure 1.2 illustrates a sample selling price path before the end of the game (assumed to be date 4 in this example).[19,20] Specifically, the following parameter assumption holds throughout the paper:

Assumption **1.** *The selling prices satisfy* $\delta_L > \delta_E + (1-\delta_E)(\mu_E + \mu_L)$.[21]

This selling price pattern can be micro-founded by the idea of slow-moving capital in illiquid asset markets (Duffie, 2010). When a redemption shock just hits the economy on date $2t$, because there are only a few liquidity providers available, it is very hard for the fund manager to find a good selling price. But if she can wait until the next date $2t+1$,

---

[19] Although my model features discrete time, the price path is intentionally depicted as a *cádlág* function that is everywhere right-continuous and has left limits everywhere to better reflect the perfectly elastic intra-day asset demand and daily price update.

[20] I also consider more persistent price impacts in Section 1.6.3 as an extension and show that they will only strengthen my results.

[21] Besides the fact that the fire sale cost on even dates is larger, the specific form in Assumption 1 is not crucial to my results.

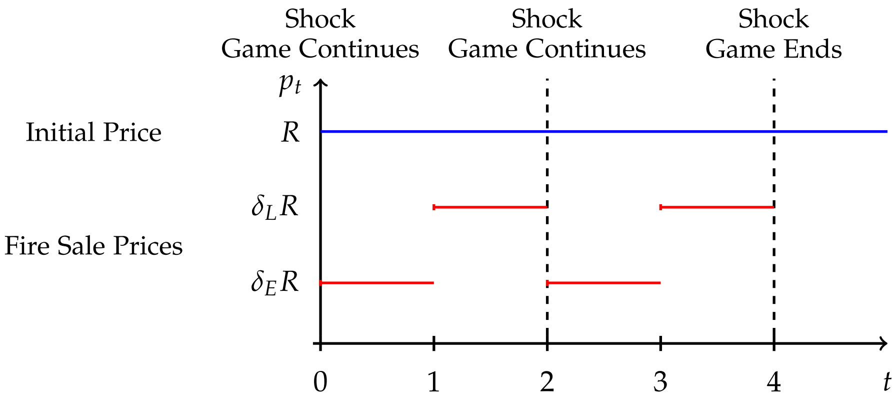


Figure 1.2: *Sample Path of the Selling Prices*

since more liquidity providers step in, it would be easier to find a higher selling price. If the game continues on date $2t+2$, that is, when another round of redemption shock hits the economy, the selling price drops again. These transitory price impacts induced by outflows and the associated price over-shooting and reversal have been empirically documented in various asset markets (for example, Campbell, Grossman and Wang, 1993, Pastor and Stambaugh, 2003, Coval and Stafford, 2007, Hendershott and Seasholes, 2007, Mitchell, Pedersen and Pulvino, 2007, Comerton-Forde, Hendershott, Jones, Moulton and Seasholes, 2010, Greenwood and Thesmar, 2011, Edmans, Goldstein, and Jiang, 2012, Jotikasthira, Lundblad, and Ramadorai, 2012, Lou, 2012, Nagel, 2012, Hendershott and Menkveld, 2014, among others). They are also consistent with the timeline and in particular with the nature of shocks in my economy. This price pattern will be formally micro-founded in Appendix A.3, based on the idea of slow-moving capital outlined above. To focus on redemption shocks and their price impacts, my model is intentionally abstracted away from any fundamental shocks of the underlying assets.

Fund Liquidity Management. The open-end mutual fund has to meet redemption needs in cash on a daily basis.[22] Hence, the fund manager needs to manage its liquidity carefully to keep an adequate cash position. Specifically, she manages the cash position of the fund both passively and actively.

---

[22] In Section 1.5 I discuss how this model can be extended to analyze emergency rules such as redemption restrictions and in-kind redemptions.

On the one hand, on any date t, if the fund does not have enough cash to meet date-t projected redemptions at the beginning-of-day NAV (i.e., NAV_{t-1}), the fund will be forced to raise cash until all redemption needs can be met. Since there are no inflows and the illiquid assets do not pay interim cash flows, the fund manager can only raise cash by selling illiquid assets passively[23] at the fire sale price p_t. However, redeeming shareholders will only get the end-of-day NAV, which will reflect asset sale losses within the given date, so that the fund manager only has to sell assets to a point at which the redemptions can be met at the end-of-day NAV. Denote the amount of illiquid assets that the fund has to sell passively by q_t, which will be endogenously determined in equilibrium.

On the other hand, in addition to selling assets passively for meeting redemptions, the fund can also manage its cash buffer actively. Specifically, the fund manager is able to voluntarily sell illiquid assets more than actual redemption needs to rebuild the cash buffer, also at the fire sale price p_t. Denote the amount of assets that the fund voluntarily sells on date t by s_t. As will become clear later, it is the time-varying fire sale price that gives rise to the motive for active cash rebuilding: the fund manager may optimally choose to rebuild the fund's cash buffer on odd dates when the fire sale price is high (and thus the fire sale loss is low), in order to avoid more severe fire sales in future. In what follows, I will call s_t the fund's cash rebuilding policy on date t. Intuitively, a larger s_t means that the fund is rebuilding its cash buffer more rapidly.

Now it is natural to specify the fund manager's objective function formally. Denote by T the random date on which the game ends.[24] The fund manager chooses a sequence $\{s_\tau\}$ date by date to maximize:

$$-\mathbb{E}_t \sum_{\tau=t}^{T-1} (q_\tau + s_\tau)(R - p_\tau)\,, \tag{1.2.3}$$

where the expectation is taken over the random variable T. In particular, selling illiquid assets (either passively or actively) at fire sale prices will always hurt the fund NAV that redeeming shareholders are able to get. This implies that late shareholders' redemption decisions will be directly affected by the fund's cash rebuilding policies. Specifically, the late shareholders in stage t rely on the difference of NAV between date $2t$ and date $2t+1$ to make redemption decisions, so they rationally form beliefs about the fund's cash rebuilding policies $\{s_{2t}, s_{2t+1}\}$ within that stage.[25] As one can expect, the fund's cash rebuilding policies will affect shareholders' run decisions, which will in turn affect the fund's optimal cash rebuilding policies in the dynamics. This interaction is only at play when fund NAV is flexible, the key institutional setting I highlight throughout this paper.

---

[23] In Section 1.5.6, an extension to the baseline model, I also allow the fund to have credit lines to raise cash.

[24] By construction, T must be an even number.

It is important to note that, although feasible, the fund manager will never rebuild the fund's cash buffer on even dates in any generic equilibrium, that is, $s_{2t} = 0$ for any t.[26] This is intuitive because all the purpose for the fund to manage its cash buffer is to avoid extremely costly fire sales on even dates, and hence it never makes sense for the fund manager to voluntarily sell assets on even dates. As a result, the fund's cash rebuilding policy in stage t is solely determined by s_{2t+1}, the amount of illiquid assets the fund manager voluntarily sells on the odd date $2t+1$. In what follows, I consider s_{2t+1} the only choice variable of the fund manager in any stage t.

Finally, it is worth noting that my model is admittedly not intended to be a general theory of mutual fund management. In general, a fund manager may engage in other management activities (see Wermers, 2000 for a comprehensive evaluation of fund management activities such as asset-picking, style-investing, and fee-setting). Also, in addition to meeting redemption needs, a fund keeps cash for other purposes, such as making timely investments in illiquid and high-yield assets without waiting for inflows (Chordia, 1996). To focus on the interaction between fund liquidity management and shareholder runs, I assume that the purchasing price of the illiquid assets (from outside dealers) is always the fundamental value

---

[25] In the U.S., mutual funds are required to disclose their asset positions quarterly, under the scrutiny of the SEC according to 17 CFR Parts 230, 232, 239, and 274. See "Enhanced Disclosure and New Prospectus Delivery Option for Registered Open-End Management Investment Companies, Investment Company Act Release No. 28584," The SEC, Jan 13, 2009. From a theoretical point of view these requirements allow shareholders to form consistent beliefs about a fund's cash rebuilding policies.

[26] This statement will be formally proved in Appendix C.1.3.

R. This is consistent with the existence of large bid-ask spreads in illiquid asset markets, and it implies that the fund will never repurchase the illiquid assets in my economy with net outflows only.

1.2.2 Interpretation of the Setup

The settings described above represent a mutual fund crisis management scenario, during which the fund experiences persistent redemption shocks before a random recovery time. In the setup, π measures how persistent the redemptions shocks are, or in other words how likely the economy is going to recover from a bad market condition. When π is lower, the game is more likely to continue, and thus the fund is more likely to experience redemptions and potential fire sales in the next stage. As the fund manager never knows when the game will end, liquidity management indeed helps the fund minimize its total expected fire sale losses, and it matters more when π is lower. Also, I use the end-of-game event to parsimoniously capture what can possibly happen in a normal-time scenario such as inflows to the fund and dividend payouts to shareholders. In a normal-time scenario, asset prices are also more likely to be high and correctly reflecting their fundamentals, as I postulate in the setting. In this sense, my model is not intended as a general boom-bust cycle model of mutual fund management but one focusing on crisis management and the economic forces involved. From a theoretical standpoint, many dynamic crisis management theories in other contexts employ similar structures of persistent shocks followed by a random recovery time, for example, Lagos, Rocheteau and Weill (2011) on crises in OTC asset markets, He and Xiong (2012) on corporate debt runs, and He and Milbradt (2015) on maturity choices in a debt rollover crisis, among others.[27]

More importantly, the crisis management scenarios of open-end mutual funds are pervasive in reality. They can occur at both long and short time horizons, and for both fundamental and non-fundamental reasons. Since the May of 2013, the flagship Total Return fund of the Pacific Investment Management Company (PIMCO), one of the largest fund in the U.S., has experienced net outflows for more than 28 consecutive months.[28] The Prudential M&G's flagship Optimal Income fund, one of the largest bond fund in the Europe, has experienced more than 50 consecutive trading days of net outflows in mid 2015.[29] Another example features Aberdeen Asset Management, the largest listed fund manager in Europe, has experienced net outflows for 15 consecutive months as of the end of 2015.[30] This is also true at the aggregate level: between Aug 20, 2015 and Aug 26, 2015, aggregate outflows from the entire equity fund sector happened for five consecutive trading days at the total amount of 29.5 billion U.S. dollars, the largest weekly outflow on record since fund flow data began being calculated in 2002. Not surprisingly, these scenarios happen to smaller funds as well. My model is designed to capture such scenarios and to explore the potential risks of shareholder runs and fire sales. Given the pervasiveness of the fund crisis management scenarios, my model is likely to have first-order implications on the potential financial stability risks of open-end mutual funds.

---

[27] These models all have quite distinct approaches and focuses from mine, but they share the similar structure of negative shocks followed by a random recovery time.

1.2.3 Roadmap and Solution Approach

Formally, the setup above defines an infinite-horizon stochastic game with a long-run player (the fund manager) and a sequence of short-run players (the late shareholders). I take a two-step approach to solve this game. First, in Section 1.3, I solve the stage game among the late shareholders who are hit by consumption shocks in stage t, given any generic cash policy of the fund within that stage. The late shareholders' equilibrium run decisions in that stage game will help determine how the asset and cash positions of the fund evolve over time. Then I turn to the fully stochastic dynamic game in Section 1.4, solving for the fund manager's optimal dynamic cash policies. I will offer formal equilibrium definitions accordingly in these two sections.

---

[28] "PIMCO Total Return Assets Drop Below 100 Billion," The *Reuters*, Sept 2, 2015.

[29] "Investors Pull 2.7 Billion from Giant M&G Bond Fund," The *Financial Times*, July 29, 2015.

[30] "Aberdeen Hit by Investor Outflows," The *Wall Street Journal*, December 1, 2015.

Both of the two steps provide new insights to the literature. The first step shows that when fund NAVs flexibly adjust, cash rebuilding by the fund can directly push shareholders to run, a point absent in classic bank run models. The second step further shows how the potential for shareholder runs distorts fund liquidity management, a point new to the mutual fund management literature.

1.3 Shareholder Runs

In this section, I focus on the two-date stage game, showing that the fund's desire to rebuild its cash buffer can trigger shareholder runs, and more rapid cash rebuilding leads to more severe runs. This result is very general: runs can occur in equilibrium regardless of the fund's initial cash position.

1.3.1 Stage-Game Equilibrium Definition and Preliminary Analysis

The two-date stage-game equilibrium is a mixed-strategy Nash equilibrium: in any stage t (consisting of dates $2t$ and $2t+1$), given the fund's initial portfolio position (a_{2t}, x_{2t}) and the late shareholders' common beliefs on the fund's cash rebuilding policy s_{2t+1}, a late shareholder's run strategy maximizes her utility given other late shareholders' strategies. Since all the late shareholders are identical, there is no loss of generality to consider symmetric equilibria when mixed strategies are allowed (Khan and Sun, 2002). Formally:

Definition 1. *Given μ_E, μ_L, δ_E, δ_L, R, a_{2t}, x_{2t}, and s_{2t+1}, a symmetric run equilibrium of the stage-t game is defined as a run probability $\lambda_{2t}(a_{2t}, x_{2t}, s_{2t+1}) \in [0, 1]$ such that*

i) given other late shareholders' run probability λ_{2t}, late investor i's optimal run probability $\lambda^i_{2t} = \lambda_{2t}$ maximizes her utility function (1.2.2), and

ii) all of the late shareholders have a common belief about the fund's cash rebuilding policy s_{2t+1}.[31]

By the Law of Large Numbers, in a symmetric run equilibrium, the total population of shareholders who redeem on date $2t$ is $(\mu_E + \lambda_{2t}\mu_L)n_{2t}$. Intuitively, this means when some late shareholders are going to run, there will be effectively more early shareholders and fewer late shareholders.

---

[31] For simplicity, when analyzing the stage game, I slightly abuse the notation s_{2t+1} to denote both the shareholders' common belief about the fund's cash rebuilding policy and the actual cash rebuilding policy itself.

Before solving for the stage-game equilibrium, I first describe three cases of the stage game according to the fund's starting cash position, x_{2t}. As will become clear shortly, the consequences of runs on the risk of fire sales are different when the initial cash position varies. Different x_{2t} also implies different nature of strategic interactions among late shareholders. Hence, it is useful to discuss these cases separately to clarify the mechanism. These cases can be characterized by the following three cash-to-assets ratio regions.

Cash-to-Assets Ratio Regions

Formally, assuming no cash rebuilding and no shareholder runs as the status quo, I characterize three different cash-to-assets ratio regions of the portfolio position space $\{(a_{2t}, x_{2t})\} \subseteq \mathbb{R}^2_+$. In these different regions, the amounts of illiquid assets that the fund is forced to sell on the two adjacent even and odd dates, that is, q_{2t} and q_{2t+1}, vary. I define the cash-to-assets ratio

$$\eta_t \equiv \frac{x_t}{a_x}$$

for any date t.

Lemma 1. *Suppose the fund does not rebuild its cash buffer and no late shareholder is going to run, that is, $s_{2t+1} = 0$ and $\lambda_{2t} = 0$. Then there are three regions of the cash-to-assets ratio η_{2t} in the stage-t game. In these three regions, the amounts of illiquid assets that the fund has to sell passively on dates $2t$ and $2t+1$ are characterized by:*

$$\begin{cases} \textit{High Region } G_h\textit{:} & q_{2t} = 0, q_{2t+1} = 0, \quad \textit{iff} \quad \eta_{2t} \geqslant \dfrac{(\mu_E + \mu_L)R}{1 - \mu_E - \mu_L}, \\ \textit{Intermediate Region } G_m\textit{:} & q_{2t} = 0, q_{2t+1} > 0, \quad \textit{iff} \quad \dfrac{\mu_E R}{1 - \mu_E} \leqslant \eta_{2t} < \dfrac{(\mu_E + \mu_L)R}{1 - \mu_E - \mu_L}, \\ \textit{Low Region } G_l\textit{:} & q_{2t} > 0, q_{2t+1} > 0, \quad \textit{iff} \quad \eta_{2t} < \dfrac{\mu_E R}{1 - \mu_E}. \end{cases}$$

The exact values of q_{2t} and q_{2t+1} in the intermediate and low regions will be solved in Section 1.3.3 and Section 1.3.4.

The three regions of η_{2t} are intuitive. When $\eta_{2t} \in G_h$, the fund has enough cash to meet all projected redemptions on both date $2t$ and $2t+1$, and thus no forced fire sales occurs. When $\eta_{2t} \in G_m$, the fund only has enough cash to meet redemptions on date $2t$ but not on date $2t+1$, so it has to passively fire sell its illiquid asset on $2t+1$. Finally, when $\eta_{2t} \in G_l$, the fund does not even have enough cash to meet redemption needs on date $2t$, and thus has to incur forced fire sales on both dates.

Lemma 1 also implies that the stage game is scale-invariant. Neither the absolute value of (a_{2t}, x_{2t}) nor the initial population of shareholders n_{2t} plays a role in determining the three regions. This allows me to use the single variable, the cash-to-assets ratio, to characterize shareholder runs in the stage game. This property of the stage game plays an important role in making the model transparent and tractable.[32] Without loss of generality, I assume $n_{2t} = 1$ throughout this section.

Although cash rebuilding by the fund unambiguously leads to shareholder runs, Lemma 1 suggests that the detailed reasons for runs can be different in these three regions because of the different initial amounts of fire sales. In what follows, I analyze equilibrium shareholder runs and their impact on the risk of fire sales in the three regions one by one.

1.3.2 High Cash-to-Assets Ratio Region G_h

When the stage game is in the high cash-to-assets ratio region G_h, Lemma 1 implies that the fund is not forced to sell any illiquid assets, that is, $q_{2t} = q_{2t+1} = 0$, if no shareholder decides to run. But what would happen if some shareholders decided to run, that is, when $\lambda_{2t} > 0$? The next lemma answers this question. It shows that even if all shareholders decide to run, there will be still no forced fire sales.

---

[32] Existing dynamic bank run models are usually assumed to be cashless, because adding variable cash positions would introduce a second state variable that makes a dynamic model intractable. See He and Xiong (2012) and Cheng and Milbradt (2012) for discussions about this point.

Lemma 2. *When* $\eta_{2t} \in G_h$*,* $q_{2t}(\lambda_{2t}) = q_{2t+1}(\lambda_{2t}) = 0$ *for any given* $\lambda_{2t} \in [0, 1]$ *.*

The intuition of Lemma 2 is clear. When some late shareholders decide to run, there will be effectively more early shareholders and fewer late shareholders, but the total population of redeeming shareholders in the given stage is not changed. Since the fund always has sufficient cash to meet all early and late redemption needs at the initial NAV, it indeed has enough cash on date $2t$ even if all of the late shareholders are going to run.

Lemma 2 has an important implication: NAV_{2t} will never change regardless of whether late shareholders run or not. In other words, in the high region, shareholders can effectively get a fixed-value claim on date $2t$ even though the NAV is flexible by nature. As long as the fund does not rebuild its cash buffer on date $2t + 1$, Lemma 2 further implies that $NAV_{2t+1} = NAV_{2t}$ regardless of λ_{2t}, suggesting that there is no strategic interaction among late shareholders absent fund cash rebuilding. As a result, late shareholders will never run if the fund does not rebuild its cash buffer.

However, given the endogenously fixed NAV_{2t}, late shareholders may decide to run if the fund rebuilds its cash buffer on date $2t + 1$ (i.e., $s_{2t+1} > 0$), which results in a predictable decline in NAV_{2t+1}. The following lemma characterizes late shareholders' strategic interaction when s_{2t+1} is positive.

Lemma 3. *When* $\eta_{2t} \in G_h$*, late shareholders' run decision* λ_{2t} *exhibits strategic complementarity if and only if* $s_{2t+1} > 0$*. Moreover, the strategic complementarity becomes stronger as* s_{2t+1} *increases. Mathematically, there are:*

$$\frac{\partial \Delta u_L(\lambda_{2t})}{\partial \lambda_{2t}} > 0 \ \textit{and} \ \frac{\partial^2 \Delta u_L(\lambda_{2t})}{\partial \lambda_{2t} \partial s_{2t+1}} > 0\,,$$

if and only if $s_{2t+1} > 0$*, where* $\Delta u_L(\lambda_{2t}) = u_L(\lambda_{2t,i} = 1; \lambda_{2t,-i} = \lambda_{2t}) - u_L(\lambda_{2t,i} = 0; \lambda_{2t,-i} = \lambda_{2t})$*, while*

$$\frac{\partial \Delta u_L(\lambda_{2t})}{\partial \lambda_{2t}} = 0\,,$$

when $s_{2t+1} = 0$.[33]

---

[33] For simplicity, in what follows when I state results about strategic complementarity and substitutability I omit the mathematical definitions because they are standard.

Lemma 3 suggests the existence of run incentives, which comes from the predictable decline in NAV when the fund rebuilds its cash buffer. Suppose the fund manager voluntarily sells $s_{2t+1} > 0$ unit of illiquid assets on date $2t+1$ and λ_{2t} of the late shareholders decide to run. Lemma 2 first suggests that no matter what λ_{2t} is, q_{2t} is always zero and thus we have

$$NAV_{2t} = \underbrace{Ra_{2t} + x_{2t}}_{NAV_{2t-1}} . \tag{1.3.1}$$

However, NAV_{2t+1} becomes lower due to the fire sales of illiquid assets on date $2t+1$:

$$\begin{aligned} NAV_{2t+1} &= \frac{R \overbrace{(a_{2t} - s_{2t+1})}^{\text{illiquid assets remained}} + \overbrace{x_{2t} - (\mu_E + \lambda_{2t}\mu_L)(Ra_{2t} + x_{2t})}^{\text{cash remained}} + \overbrace{\delta_L R s_{2t+1}}^{\text{cash rebuilt}}}{\underbrace{1 - (\mu_E + \lambda_{2t}\mu_L)}_{\text{shareholders remained on date } 2t+1}} && (1.3.2) \\ &= NAV_{2t} - \frac{(1-\delta_L)Rs_{2t+1}}{1 - \mu_E - \lambda_{2t}\mu_L} . && (1.3.3) \end{aligned}$$

The calculation in (1.3.2) reflects the fund manager's voluntary asset sales and the associated price impact on date $2t+1$ before the end-of-day NAV is determined. Consequently, a wedge, as shown in (1.3.3), emerges between the NAV of the two dates. It suggests a predictable decline in the NAV on date $2t+1$.

The predictable decline in NAV implies that the fund manager rebuilds its cash buffer at the expense of the late shareholders who initially plan to wait until date $2t+1$, giving rise to run incentives among these late shareholders. Specifically, the wedge in (1.3.3) is increasing in $1-\delta_L$ and s_{2t+1}, suggesting that the late shareholders who wait are hurt more if the price impact is larger or if the fund sells more. The wedge is also increasing in μ_E and λ_{2t}, suggesting that the late shareholders who wait are hurt more if more of others are running, and thus they have to bear a higher fire sale cost per share on date $2t+1$. Moreover, for given s_{2t+1} and λ_{2t}, the utility gain $\Delta u_L(\lambda_{2t})$ of running over waiting is $\theta NAV_{2t} - NAV_{2t+1}$, which is strictly increasing in λ_{2t} by (1.3.3). This illustrates the underlying strategic complementarity among late shareholders.

Lemma 3 suggests that both cash rebuilding and flexible NAV adjustment play crucial roles in generating the run incentives for late shareholders. If $s_{2t+1} = 0$, the stage game features no strategic interaction at all in the high region. If NAV_{2t+1} was fixed, which is the case for MMFs, rebuilding the cash buffer by selling illiquid assets would not generate a wedge of value between early and late shareholders.

I further show that the run incentives can indeed lead to shareholder runs in equilibrium. When the fund rebuilds its cash buffer rapidly enough, that is, when s_{2t+1} is large enough, late shareholders are going to run.

Proposition 1. *When $\eta_{2t} \in G_h$, late shareholders' equilibrium run behaviors are given by the following three cases:*

i) none of the late shareholders runs, that is, $\lambda_{2t} = 0$, if

$$s_{2t+1} < \underline{s}_h \equiv \frac{(1-\theta)(1-\mu_E-\mu_L)(Ra_{2t}+x_{2t})}{(1-\delta_L)R},$$

ii) all of the late shareholders run, that is, $\lambda_{2t} = 1$, if

$$s_{2t+1} > \bar{s}_h \equiv \frac{(1-\theta)(1-\mu_E)(Ra_{2t}+x_{2t})}{(1-\delta_L)R},$$

iii) $\lambda_{2t} \in \{0, \widetilde{\lambda}_{2t}, 1\}$, if

$$\underline{s}_h \leqslant s_{2t+1} \leqslant \bar{s}_h\,,$$

where $\widetilde{\lambda}_{2t}$ is the solution to

$$s_{2t+1} = \frac{(1-\theta)(1-\mu_E-\widetilde{\lambda}_{2t}\mu_L)(Ra_{2t}+x_{2t})}{(1-\delta_L)R}.$$

Moreover, there are $0 \leqslant \underline{s}_h \leqslant \bar{s}_h$.

Proposition 1 suggests that the fund's cash rebuilding indeeds leads to shareholder runs in equilibrium, and more rapid cash rebuilding can trigger more shareholders to run. The intuitions for the three cases are as follows. In Case i), when the fund sells only a few illiquid assets, NAV_{2t+1} is still high enough, the utility gain of running over waiting would be negative even if all of the late shareholders decided to run, so it turns out no one runs. In Case ii), when the fund manager voluntarily sells so many illiquid assets to a point where NAV_{2t+1} is so low and the utility gain of running would be positive even if others did not run, all of the late shareholders will run. Both Case i) and Case ii) feature a unique equilibrium. In Case iii), the utility gain of running is negative when no one runs but becomes positive when all of the late shareholders are going to run. Strategic complementarity implies that the utility gain of running is increasing when more late shareholders decide to run, so multiple equilibria emerge. Among the three possible equilibria, the partial run equilibrium $\widetilde{\lambda}_{2t}$ features a point where the utility gain of running is zero so that any late shareholder is indifferent between running or waiting.

One natural question is, would cash rebuilding (i.e., a positive s_{2t+1}) ever make sense from the fund manager's perspective? This question is sensible given that shareholder runs happen in equilibrium only if $s_{2t+1} > 0$. As I will show in Section 1.4, there can indeed be scenarios in which the fund manager optimally chooses to rebuild its cash buffer rapidly and bear some costs of shareholder runs, which supports the existence of run equilibria in the stage game.

The next question is: what kind of costs do shareholder runs impose on the fund? In other words, how do shareholder runs affect the fund's total risk of fire sales, given the fund manager's objective function as in (1.2.3)? I formally answer this question by looking at the law of motions of the fund's portfolio position (a_{2t}, x_{2t}). The law of motions is also important for the fully dynamic analysis in Section 1.4.

Corollary 1. *When $\eta_{2t} \in G_h$, the law of motions of (a_{2t}, x_{2t}) is given by*

$$\begin{aligned} a_{2t+2} &= a_{2t} - s_{2t+1}\,, \text{ and}\,, \\ x_{2t+2} &= \underbrace{x_{2t} - (\mu_E + \mu_L)(Ra_{2t} + x_{2t})}_{\text{cash remained if no cash rebuilding}} + \underbrace{\delta_L R s_{2t+1}}_{\text{cash rebuilt}} + \underbrace{\frac{(1-\lambda_{2t})\mu_L(1-\delta_L)Rs_{2t+1}}{1-\mu_E-\lambda_{2t}\mu_L}}_{\text{cash saved due to NAV adjustment}}\,, \end{aligned} \tag{1.3.4}$$

where λ_{2t} is the equilibrium run probability induced by (a_{2t}, x_{2t}) and s_{2t+1}, as characterized in Proposition 1.

I illustrate the implications of runs on the total risk of fire sales by interpreting the two laws of motions in Corollary 1. The law of motions of a_{2t} is straightforward by Lemma 2 because $q_{2t} = q_{2t+1} = 0$ regardless of λ_{2t}. This suggests that even though cash rebuilding can trigger shareholder runs, it will not induce any forced fire sales in the current stage when $\eta_{2t} \in G_h$.

However, shareholder runs can offset the fund's cash rebuilding efforts and lead to higher risk of future fire sales. This can be seen from the law of motions of x_{2t} in (1.3.4). To make this clear, I organize the terms in the right hand side of (1.3.4) in a way to better reflect the cost of shareholder runs. The first term denotes the amount of cash remained if the fund did not rebuild its cash buffer so that the fund paid the initial NAV to all of the early and late shareholders. The second term denotes the actual amount of cash the fund can get by selling s_{2t+1} illiquid assets. Neither of these two terms depends on λ_{2t}. The third term is more interesting. It reflects the fact that the fund can give the late shareholders less cash when it rebuilds its cash buffer on date $2t+1$. Specifically, when s_{2t+1} is positive, NAV_{2t+1} becomes lower as shown in (1.3.3). Thus, more cash remains on the fund's balance sheet than that indicated by the first term in (1.3.4). But the third term is strictly decreasing in λ_{2t}, suggesting that this benefit of cash saving to the fund becomes smaller when more late shareholders are running.[34] As a result, when more shareholders run in equilibrium, the fund loses more cash in the given stage, carries less cash to future stages under the same cash rebuilding policy s_{2t+1}, and thus faces higher risk of future fire sales.

With the intuition outlined above, it is convenient to combine the last two terms in (1.3.4) and define

$$\widehat{p}_L(\lambda_{2t}) \equiv \left[\delta_L + \frac{(1-\lambda_{2t})\mu_L(1-\delta_L)}{1-\mu_E-\lambda_{2t}\mu_L}\right] R \tag{1.3.5}$$

as the effective selling price on the odd dates $2t+1$. It is decreasing in λ_{2t}, meaning that the effective selling price on odd dates is lower when more shareholders are going to run. Intuitively, the fund prefers a higher effective selling price on odd dates because that helps reduce future risk of fire sales. I will frequently refer to this definition in the dynamic analysis in Section 1.4.

---

[34] There are two effects captured by the numerator and denominator of the third term in (1.3.4), respectively. On the one hand, as more late shareholders run on date $2t$, the fund has to pay the (endogenously) fixed NAV_{2t} to more shareholders. This is reflected in the numerator in the sense that only $1-\lambda_{2t}\mu_L$ late shareholders are left to bear the asset sale cost on date $2t+1$, so that the benefit of cash saving due to flexible NAV adjustment becomes lower to the fund. On the other hand, as reflected by the denominator as well as the wedge term in (1.3.3), NAV_{2t+1} becomes lower as more shareholders are going to run, suggesting that the fund can give less to each of those late shareholders who wait. In equilibrium, the former effect dominates, suggesting that shareholder runs impose an unambiguous negative effect on the net amount of cash the fund can get by selling s_{2t+1} illiquid assets.

1.3.3 Low Cash-to-Assets Ratio Region G_l

Now I turn to the low cash-to-assets ratio region G_l. In this region, the fund's starting cash position is so low that it cannot even meet the redemption needs of the early shareholders. Thus, it is forced to fire sell its illiquid assets on both dates $2t$ and $2t+1$. If late shareholders decide to run, the fund has to passively sell even more. The following lemma formally characterizes how late shareholder runs affect the fund's forced fire sales on the two dates.

Lemma 4. *When $\eta_{2t} \in G_l$, there are*

$$q_{2t}(\lambda_{2t}) = \frac{\overbrace{(\mu_E + \lambda_{2t}\mu_L)(Ra_{2t} + x_{2t}) - x_{2t}}^{\text{cash gap}}}{\underbrace{[\delta_E + (1-\delta_E)(\mu_E + \lambda_{2t}\mu_L)]\, R}_{\text{effective selling price with NAV adjustment}}}, \text{and}, \tag{1.3.6}$$

$$q_{2t+1}(\lambda_{2t}) = \frac{\overbrace{(1-\lambda_{2t})\mu_L \cdot \frac{R(a_{2t} - q_{2t})}{1 - \mu_E - \lambda_{2t}\mu_L}}^{\text{cash gap}}}{\underbrace{\left[\delta_L + \frac{(1-\lambda_{2t})\mu_L(1-\delta_L)}{1-\mu_E-\lambda_{2t}\mu_L}\right] R}_{\text{effective selling price with NAV adjustment}}}, \tag{1.3.7}$$

where q_{2t} is increasing in λ_{2t}, q_{2t+1} is decreasing in λ_{2t}, and $q_{2t} + q_{2t+1}$ is increasing in λ_{2t}.

I first interpret the intuition behind the expressions of $q_{2t}(\lambda_{2t})$ and $q_{2t+1}(\lambda_{2t})$. In determining the amounts of forced fire sales, one needs to know 1) the amount of cash that the fund is forced to raise (i.e., the "cash gap"), and 2) the price at which the fund can sell its assets. Specifically, at the beginning of each date, the cash gap is defined as the difference between the fund's initial cash position and the amount of cash needed to meet projected redemptions at the beginning-of-day NAV, as shown in the numerators of (1.3.6) and (1.3.7). However, the fund does not really have to raise that much cash in equilibrium, because the NAV goes down as the fund sells its assets, and redeeming shareholders are only entitled to the end-of-day NAV, which reflects those asset sale costs. Effectively, this is equivalent to a counterfactual in which the fund still sells assets to close the initial cash gap but at a higher effective selling price as the denominators of (1.3.6) and (1.3.7) indicate.

Thus, like (1.3.5), I also formally define the notion of effective selling price on the even dates $2t$:

$$\widehat{p}_E(\lambda_{2t}) \equiv [\delta_E + (1-\delta_E)(\mu_E + \lambda_{2t}\mu_L)]\, R\,. \tag{1.3.8}$$

Lemma 4 shows how shareholder runs affect the amounts of forced fire sales on the two dates, respectively. When more late shareholders decide to run, the fund has to meet more redemptions on date $2t$ while fewer redemptions on date $2t+1$. Hence, it is forced to fire sell more assets on date $2t$ while fewer assets on date $2t+1$.[35]

More importantly, Lemma 4 also illustrates that runs unambiguously lead to higher total amounts of forced fire sales in the given stage, as shown in the monotonicity of $q_{2t} + q_{2t+1}$ in λ_{2t}. This is because the effective selling price on date $2t$ is always lower than that on date $2t+1$,[36] which means that more early redemptions have to be met by selling assets at a lower effective selling price while fewer late redemptions will be met by selling assets at a higher effective selling price. Hence, the increase of q_{2t} will dominate the decrease of q_{2t+1} when more shareholders are going to run. As a result, shareholder runs, if occurring in equilibrium, lead to unambiguously more severe current-stage fire sales.

Lemma 4 implies that both NAV_{2t} and NAV_{2t+1} will be lower when shareholder runs

---

[35] Formally, for q_{2t}, there are two effects: the cash gap is larger when λ_{2t} is larger, but the effective selling price is also higher. The cash gap effect dominates so that q_{2t} is increasing in λ_{2t}. For q_{2t+1}, there are three effects. When λ_{2t} gets larger, fewer late shareholders choose to wait, and fewer illiquid assets remain as well (since q_{2t} becomes larger). Both lead to a smaller cash gap. However, the effective selling price is lower as well. Again the cash gap effect dominates so that q_{2t} is increasing in λ_{2t}.

[36] More formally, by the monotonicity of the effective selling prices $\widehat{p}_L(\lambda_{2t})$ and $\widehat{p}_E(\lambda_{2t})$, there is

$$\widehat{p}_E(\lambda_{2t}) \leqslant \widehat{p}_E(1) < \widehat{p}_L(1) \leqslant \widehat{p}_L(\lambda_{2t})$$

for any $\lambda_{2t} \in [0,1]$. In other words, the potential for runs may change the effective selling prices on the two adjacent dates in the given stage, but the effective selling price on $2t$ is still lower than that on $2t+1$ regardless of shareholder runs.

occur. There are

$$\begin{aligned} NAV_{2t}(\lambda_{2t}) &= R\overbrace{(a_{2t}-q_{2t}(\lambda_{2t}))}^{\text{illiquid assets remained}} + \overbrace{x_{2t}}^{\text{cash existed}} + \overbrace{\delta_E R q_{2t}(\lambda_{2t})}^{\text{cash raised}} \\ &= \underbrace{Ra_{2t}+x_{2t}}_{NAV_{2t-1}} - (1-\delta_E)Rq_{2t}(\lambda_{2t}), \end{aligned} \tag{1.3.9}$$

and

$$NAV_{2t+1}(\lambda_{2t}) = \frac{R\overbrace{(a_{2t}-q_{2t}(\lambda_{2t})-q_{2t+1}(\lambda_{2t})-s_{2t+1})}^{\text{illiquid assets remained}} + \overbrace{\delta_L R\,(q_{2t+1}(\lambda_{2t})+s_{2t+1})}^{\text{cash raised and rebuilt}}}{\underbrace{1-\mu_E-\lambda_{2t}\mu_L}_{\text{shareholders remained on date } 2t+1}}, \tag{1.3.10}$$

where $q_{2t}(\lambda_{2t})$ and $q_{2t+1}(\lambda_{2t})$ are given in (1.3.6) and (1.3.7).

Like in the high region, a predictable decline in NAV_{2t+1} emerges because of the fund's forced fire sales and active cash rebuilding on date $2t+1$, as shown in (1.3.10). However, different from the high region, $NAV_{2t}(\lambda_{2t})$ is no longer fixed but decreasing in λ_{2t}, as shown in (1.3.9). Intuitively, because the fund NAV flexibly adjusts, shareholders who choose to run have to bear all the fire sale costs incurred on date $2t$. This feature changes the nature of the stage game.

Lemma 5. *When $\eta_{2t} \in G_l$, late shareholders' run decision λ_{2t} exhibits strategic substitutability for any λ_{2t} satisfying $\theta NAV_{2t}(\lambda_{2t}) \geqslant NAV_{2t+1}(\lambda_{2t})$ and any feasible s_{2t+1}.*[37] *However, when s_{2t+1} increases, the strategic substitutability becomes weaker and $\theta NAV_{2t}(\lambda_{2t}) - NAV_{2t+1}(\lambda_{2t})$ becomes larger, reinforcing a stronger incentive to redeem earlier.*

The intuition behind Lemma 5 is as follows. On the one hand, the predictable decline in NAV_{2t+1} can give rise to an incentive to redeem earlier (i.e., to run on date $2t$). On the other hand, a late shareholder who decides to run has to accept a lower NAV_{2t} when more of other late shareholders decide to run. In particular, as more shareholders are going to run, the difference between $NAV_{2t}(\lambda_{2t})$ and $NAV_{2t+1}(\lambda_{2t})$ becomes smaller, implying strategic substitutability among late shareholders. Intuitively, a shareholder redeeming at t realizes that more early withdrawals will potentially induce more fire sales at t and thus lower the proceeds she receives, and thus the expected utility gain of running over waiting would be decreasing as more shareholders run.

---

[37] In equilibrium λ_{2t} will be an endogenous function of s_{2t+1}. But in showing the strategic interaction among late shareholders of the stage game, λ_{2t} should be treated as an independent variable. This also applies to the analysis of the intermediate region.

However, when the fund rebuilds its cash buffer, the strategic substitutability may not reduce the run incentive (when θ is large enough). Concretely, when the fund voluntarily sells a sufficiently large amount of assets to rebuild its cash buffer, the resulting large predictable decline in NAV_{2t+1} reinforces a sufficiently strong run incentive. As a result, a late shareholder may still decide to run even if all of the other late shareholders have already run, despite the strategic substitutability.

Consequently, the next proposition fully characterizes late shareholders' equilibrium run behaviors in the low region.

Proposition 2. *When $\eta_{2t} \in G_l$, late shareholders' equilibrium run behaviors are given by the following three cases:*

i) none of the late shareholders runs, that is, $\lambda_{2t} = 0$, if

$$s_{2t+1} < \underline{s}_l \equiv \frac{Ra_{2t} - \theta(1-\mu_E)(Ra_{2t}+x_{2t}) - (1-\theta(1-\delta_E)(1-\mu_E))Rq_{2t}(0)}{(1-\delta_L)R} - q_{2t+1}(0)\,,$$

ii) all of the late shareholders run, that is, $\lambda_{2t} = 1$, if

$$s_{2t+1} > \bar{s}_l \equiv \frac{Ra_{2t} - \theta(1-\mu_E-\mu_L)(Ra_{2t}+x_{2t}) - (1-\theta(1-\delta_E)(1-\mu_E-\mu_L))Rq_{2t}(1)}{(1-\delta_L)R}\,,$$

iii) some of the late shareholders run, that is, $\lambda_{2t} = \widetilde{\lambda}_{2t}$, if

$$\underline{s}_l \leqslant s_{2t+1} \leqslant \bar{s}_l\,,$$

where $\widetilde{\lambda}_{2t}$ is the solution to

$$s_{2t+1} = \frac{Ra_{2t} - \theta(1-\mu_E-\widetilde{\lambda}_{2t}\mu_l)(Ra_{2t}+x_{2t}) - (1-\theta(1-\delta_E)(1-\mu_E-\widetilde{\lambda}_{2t}\mu_l))Rq_{2t}(\widetilde{\lambda}_{2t})}{(1-\delta_L)R} - q_{2t+1}(\widetilde{\lambda}_{2t})\,.$$

All of the $q_{2t}(\lambda_{2t})$ and $q_{2t+1}(\lambda_{2t})$ are given in Lemma 4. Moreover, there are $\bar{s}_l \geqslant 0$ and $\bar{s}_l > \underline{s}_l$.

Like Proposition 1, Proposition 2 also suggests that the fund's cash rebuilding leads to shareholder runs in equilibrium, and more rapid cash rebuilding can trigger more shareholders to run, despite the strategic substitutability. In Case i), when the fund does not rebuild its cash buffer or only sells a few illiquid assets, NAV_{2t+1} can be still higher regardless of shareholders' redemption decisions, so that late shareholders will not run. In Case ii), when the fund voluntarily sells so many illiquid assets, NAV_{2t+1} is so low that the utility gain of running over waiting is positive even if all of the late shareholders have already run. In other words, fund cash rebuilding can reinforce a strong run incentive despite the strategic substitutability. In addition, all of the late shareholders do not run unless the fund rebuilds its cash buffer (since $\bar{s}_l \geqslant 0$), suggesting that only cash rebuilding by the fund can push all the shareholders to run in this mutual fund context.[38] In Case iii), the utility gain of running over waiting is positive when no shareholder runs but becomes negative when all of the late shareholders are going to run. In this case, there exists some run equilibrium in which the utility gain of running over waiting is zero, so that late shareholders are indifferent between running or waiting.

Again, I show how shareholder runs affect the risk of forced fire sales, by looking at the law of motions of the fund portfolio position (a_{2t}, x_{2t}).

Corollary **2.** *When $\eta_{2t} \in G_l$, the law of motions of (a_{2t}, x_{2t}) is given by*

$$a_{2t+2} = a_{2t} - \overbrace{(q_{2t}(\lambda_{2t}) + q_{2t+1}(\lambda_{2t}))}^{\textit{forced fire sales}} - s_{2t+1}\,, \textit{ and}\,, \tag{1.3.11}$$

$$x_{2t+2} = \underbrace{\delta_L R s_{2t+1}}_{\textit{cash rebuilt}} + \underbrace{\frac{(1-\lambda_{2t})\mu_L(1-\delta_L)R s_{2t+1}}{1-\mu_E-\lambda_{2t}\mu_L}}_{\textit{cash saved due to NAV adjustment}} = \widehat{p}_L(\lambda_{2t}) s_{2t+1}\,, \tag{1.3.12}$$

where λ_{2t} is the equilibrium run probability induced by (a_{2t}, x_{2t}) and s_{2t+1}, as characterized in Proposition 2.

The two laws of motions here are different from those in Corollary 1, but can be unified under the same intuition. The law of motions of a_{2t} in (1.3.11) is different because of the newly introduced forced fire sales terms: shareholder runs result in more forced fire sales in the given stage. The law of motions of x_{2t} in (1.3.12) loses the first term in (1.3.4). This is natural because by construction no cash would remain in the low region if the fund did not rebuild its cash buffer. Perhaps surprisingly, the two terms in (1.3.12) are exactly the same as the last two terms in (1.3.4) despite the more complicated forced asset sales and NAV updating in the low region. But this is still intuitive: all the proceeds from forced fire sales accrue to the redeeming shareholders, and thus they will not affect the amount of cash that the fund can carry into future stages. This suggests that, in rebuilding its cash buffer the fund is still selling at the effective selling price $\widehat{p}_L(\lambda_{2t})$.

---

[38] This is not true for a comparable bank with fixed-value deposits, in which all shareholders can run in equilibrium even if the bank does not do anything by itself.

Importantly, Corollary 2 implies two different costs of shareholder runs. First, shareholder runs force the fund to fire sell more illiquid assets in the current stage (remember Lemma 4 shows that $q_{2t} + q_{2t+1}$ is increasing in λ_{2t}). Second, like that in the high region, shareholder runs lead to a lower effective selling price on date $2t + 1$ when the fund rebuilds its cash buffer. This means that runs partially offset the fund's efforts of cash rebuilding and thus lead to higher risk of future fire sales. The second cost is also present in the high region as shown in (1.3.4).

Compared to the analysis for the high region in Section 1.3.2, Proposition 2 and Corollary 2 suggest that starting with a low cash position makes a fund financially more fragile. Being in the low cash-to-assets ratio region makes the fund more prone to forced fire sales initially. Even worse, because the fund is running out of cash, it is likely to rebuild its cash more rapidly (as formally shown in Section 1.4), leading to more severe runs. In particular, runs in the low region are more detrimental to the fund: they not only increase current-stage forced fire sales but also lead to higher risk of future fire sales.

1.3.4 Intermediate Cash-to-Assets Ratio Region G_m

I then analyze the intermediate cash-to-assets ratio region G_m. In this region, the fund's starting cash position is moderate in the sense that it can meet all of the early shareholders' redemption needs but then falls short for late shareholders' redemption requests.

In the intermediate region, we still have the universal results of shareholder runs as those in the high and low regions. Specifically, the fund's cash rebuilding can lead to runs in equilibrium, and more rapid cash rebuilding implies more severe runs. Also, there are two costs of runs as those in the low region: more current-stage forced fire sales and higher risk of future fire sales.

However, the underlying strategic interaction among shareholders becomes more involved in the intermediate region. When only a few late shareholders decide to run, the fund will not be forced to fire sell its illiquid assets on date $2t$, and thus NAV_{2t} will be endogenously fixed. However, when many late shareholders decide to run, the fund will be forced to sell its assets on date $2t$, and thus both NAV_{2t} and NAV_{2t+1} vary. In this sense, the stage game in the intermediate region can be viewed as a hybrid of one game in the high region and another one in the low region, which can switch from strategic complementarity to substitutability as more shareholders decide to run. However, it is still the fund's cash rebuilding and the resulting predictable decline in NAV_{2t+1} that reinforce a strong run incentive. Given the results in Sections 1.3.2 and 1.3.3, I defer the full investigation of the intermediate region to Appendix A.2. The formal results are stated there as Lemma 21, Proposition 27, and Corollary 6.

1.3.5 The Differences from Bank Runs and Market Runs

Before analyzing the fully dynamic model of fund cash management, it is useful to contrast the shareholder run mechanism to those underlying classic bank runs (Diamond and Dybvig, 1983) and market runs (Bernardo and Welch, 2004, Morris and Shin, 2004). This helps highlight the contribution of this paper to the existing run literature.

First, fund shareholder runs differ from classic bank runs in terms of the underlying mechanism. In my model, the first-mover advantage does not come from an exogenous fixed-NAV claim on date $2t$ like the deposit at a bank. Instead, it is the fund's desire to rebuild its cash buffer on date $2t + 1$ and the resulting predictable decline in NAV_{2t+1} that lead to a strong first-mover advantage and thus the potential for shareholder runs. In contrast, when a comparable bank rebuilds its cash buffer by selling assets, the underlying deposit value will not change,[39] and thus bank cash rebuilding by itself cannot directly generate depositor runs.

To illustrate the differences between fund shareholder runs and classic bank runs more in depth, it is again helpful to separate the cases with a high and a low starting cash position.

When the fund starts with a high cash position, late shareholders are going to run only if the fund voluntarily sells a sufficient amount of assets to rebuild its cash buffer, which generates a large enough predictable decline in NAV. On the one hand, although the late shareholders who run are expecting to get a fixed NAV on date $2t$, NAV_{2t} is fully endogenous and flexible by nature. This is in contrast to bank run models in which a fixed-value claim is either exogenously assumed or derived as the optimal contract in an outer risk-sharing problem. On the other hand, in a typical two-date bank run model, strategic complementarity arises because run-induced asset liquidation on the early date hurts what a waiting shareholder can get on the late date; no asset liquidation happens on the late date. But in my model, no fire sales ever occur on date $2t$ when the fund is in the high cash-to-assets region. Instead, runs hurt shareholders who wait because fewer of them are left on date $2t + 1$ and thus each has to bear a higher asset sale cost per share when the fund rebuilds its cash buffer. As will be clear in the dynamic analysis, when more shareholders run, the fund loses more cash on date $2t$ and thus may want to rebuild it cash buffer more rapidly, creating an negative externality on the late shareholders who wait.

In contrast, when the fund starts with a low cash position, the stage game can even feature strategic substitutability, but the fund's cash rebuilding reinforces a strong run incentive. In classic two-date bank run models, the game exhibits strategic substitutability only if the bank goes bankruptcy on the early date and an equal sharing rule is applied among depositors. In other words, strategic substitutability emerges only because more depositors are going to share the fixed liquidation value on the early date, and thus everyone gets less. Such strategic substitutability is absent as long as the bank is still solvent. However, in my mutual fund model with a fully flexible NAV, the fund never goes bankruptcy. Strategic substitutability instead emerges from the fact that the fund may be forced to fire sell more illiquid assets on date $2t$ and thus NAV_{2t} becomes lower. Indeed, some fund managers refer to this strategic substitutability to deny the existence of fund shareholder runs. But as I have already showed, the fund's cash rebuilding and the resulting predictable decline in NAV can lead to shareholder runs in equilibrium despite the strategic substitutability.

---

[39] To be precise, banks are always financed by some equity, so that the asset sale costs go to the equity holders. This is consistent with the reality in the sense that deposit banks are always subject to some capital requirements.

Regardless of the starting cash position, the shareholder run mechanism further highlights a dynamic interaction between the fund and its shareholders, which is absent in bank run models that focus on coordination failures among depositors themselves. Specifically, a shareholder's run decision not only depends on her belief about other shareholders' run decisions, but also depends on her belief about the fund's cash rebuilding policy,[40] which in turn depends on all the future generations of shareholders' run decisions in the fully dynamic model.

Observationally, fund shareholder runs also tend to be "slow-moving," in the sense that their costs are reflected in higher fire sale losses over time rather than an abrupt bankruptcy as that in bank run models.[41] This nature suggests that the negative impacts of shareholder runs on fund performance can be gradual but long-lasting.

The run mechanism in my model is also different from that underlying market runs. The notion of market runs is formally proposed by Bernardo and Welch (2004) and Morris and Shin (2004), which independently argue that if an asset market features an exogenous downward-sloping demand curve, investors fearing future liquidity shocks will have an incentive to front-run, selling the asset earlier to get a higher selling price. This run mechanism leads to massive fire sales in equilibrium. In my model, shareholders get access to the underlying assets only through the fund, and fund cash management is indeed beneficial to shareholders because it helps reduce total expected fire sale losses. However, the key tension that I document is that the fund's dynamic cash rebuilding also endogenously gives rise to a predictable decline in NAV and thus a new kind of run incentives. No cash management implies no runs in my model. In contrast, there is no role for cash management in market run models. In this sense, market run models focus on the asset market itself while my theory focuses on the role of financial intermediaries. This allows me to distinguish between risks that come from the active management of financial intermediaries and those that are only a reflection of the underlying asset market frictions.

---

[40] In this sense, the shareholder run mechanism within the stage game more closely resemble the idea of fundamental runs as argued by Chari and Jagannathan (1988), Jacklin and Bhattacharya (1988) and Allen and Gale (1998).

[41] In my model, as long as the entire population of shareholders are not going to run (i.e., $\mu_E + \mu_L < 1$), the fund will stay solvent. This is not true for a bank, which can be forced to go bankruptcy even if only a fraction of depositors run.

1.4 Fund Liquidity Management in the Presence of Runs

In this section, I turn to the fully dynamic game and endogenize the fund's optimal dynamic cash rebuilding policy. I show that the potential for runs gives rise to a new tension: rebuilding the cash buffer more rapidly can trigger runs, while rebuilding it less rapidly puts the fund at higher risk of future fire sales as well as future runs. I then show that a time-inconsistency problem further complicates this tension, leading to severe fire sales in expectation despite optimal cash management by the fund.

1.4.1 Dynamic Equilibrium Definition and Preliminary Analysis

The dynamic equilibrium is a Markov perfect equilibrium[42]: in any stage t (consisting of dates $2t$ and $2t+1$), as long as the game continues, both the fund manager and the late shareholders' strategies are functions of the state variables a_{2t} and x_{2t}, the fund's starting portfolio position, and the strategy profile is subgame perfect. Formally, I have:

---

[42] As shown in Bhaskar, Morris and Mailath (2013), an equilibrium is purifiable if some close-by behavior is consistent with equilibrium when agents' payoffs in each stage are perturbed additively and independently, and for infinite stochastic games with at most one long-run player all purifiable equilibria are Markov. My model can be viewed as a special case of the general class of games described in Bhaskar, Morris and Mailath (2013). As the fund manager is the only long-run player in my model, to restrict attention to Markov equilibria involves no loss of generality in the sense of finding all purifiable equilibria.

Definition 2. *Given μ_E, μ_L, δ_E, δ_L, and R, a Markov perfect equilibrium is defined as a combination of the fund manager's optimal cash rebuilding policy function $s^*_{2t+1}(a_{2t}, x_{2t})$ and the late shareholders' run decision $\lambda_{2t}(a_{2t}, x_{2t}, s_{2t+1})$ such that*

i) given any state (a_{2t}, x_{2t}) and any generic common belief of the cash rebuilding policy $s_{2t+1}(a_{2t}, x_{2t})$, the late investors' run decision $\lambda_{2t}(a_{2t}, x_{2t}, s_{2t+1}) \in [0, 1]$ constructs a symmetric run equilibrium as defined in Definition 1, which also determines q_{2t} and q_{2t+1} in any stage,

ii) the fund manager's optimal cash rebuilding policy function $s^*_{2t+1}(a_{2t}, x_{2t}, \lambda_{2t})$ solves the following Bellman equation:

$$V(a_{2t}, x_{2t}) = -(1-\delta_E)Rq_{2t} - (1-\delta_L)Rq_{2t+1} + \max_{s^*_{2t+1}}[-(1-\delta_L)Rs_{2t+1} + (1-\pi)V(a_{2t+2}, x_{2t+2})], \tag{1.4.1}$$

iii) the state variables (a_{2t}, x_{2t}) are govern by the endogenous laws of motions as described in Corollaries 1, 2 and 6, according to the respective cash-to-assets ratio regions.

I use a guess-and-verify approach to solve for the equilibrium.[43] Specifically, I will first characterize some important properties of the value function $V(a_{2t}, x_{2t})$. With the help of these properties, I solve for the equilibrium for different parameter values of θ, which governs the late shareholders' propensity to run.

The stage game may admit multiple equilibria in some circumstances, and thus an equilibrium selection mechanism is needed. Since equilibrium selection is not crucial to my main point about the dynamic interdependence between shareholder runs and fund liquidity management, I assume that late shareholders will coordinate to the worst equilibrium whenever multiple equilibria occur.[44] This equilibrium selection mechanism can be motivated by that the fund manager may be ambiguity averse to the potential for shareholder runs, so that she wants to find the most conservative cash rebuilding policies.

---

[43] Formally, the usual first-order approach does not apply here, because the value function is only piecewise-differentiable.

[44] See Postlewaite and Vives (1987), Allen and Gale (1998), Cooper and Ross (1998), among other more recent papers, for a similar treatment; some of those papers assume shareholders to coordinate to the best equilibrium to justify the existence of banks.

Alternative equilibrium selection mechanisms such as selecting the best equilibrium or the static global game approach (Goldstein and Pauzner, 2005) will not qualitatively change my results.

The following proposition establishes the existence and some important analytical properties of the value function.

Proposition 3. *A value function $V(a_{2t}, x_{2t})$ exists under the Markov strategies proposed in Definition 8. In particular, $V(a_{2t}, x_{2t})$ is homogeneous of degree one (HD1) in (a_{2t}, x_{2t}).*

The fact that $V(a_{2t}, x_{2t})$ is HD1 in (a_{2t}, x_{2t}) is important. It implies that the dynamic game is also scale-invariant, and thus the cash-to-assets ratio η_{2t} becomes the single effective state variable of the fully dynamic model. This property will make the analysis of the fund's dynamic cash rebuilding policies more transparent.

Finally, before analyzing the fully dynamic equilibrium in general, I analyze how different values of π, the probability at which the game ends, shape the fund's optimal cash rebuilding policy. Intuitively, when the shock is less persistent (i.e., π is large), the future risk of fire sales is small, and thus cash buffers become less valuable. Therefore, when π is sufficiently large, it makes little sense for the fund to rebuild its cash buffer ex-ante, because doing do only induces current sales of assets for sure but generates little future benefit. As a result, the model admits a type of equilibria in which the fund finds it optimal not to rebuild its cash buffer at all. In these equilibria, ex-post, the fund may be forced to fire sell many illiquid assets in future, because it may not have enough cash buffer when the game continues. This is consistent with the view that if agents underestimate the probability of bad shocks they are likely to suffer huge losses ex-post (Gennaioli, Shleifer and Vishny, 2012, 2013).

Lemma 6. *When π is sufficiently large, the equilibrium features $s^*_{2t+1}(a_{2t}, x_{2t}) = 0$ for any starting portfolio position (a_{2t}, x_{2t}).*

To better illustrate the key trade-off involved in the dynamic model, in the following analysis, I will consider an arbitrarily small (but still positive) π. Intuitively, this means that the redemption shocks are sufficiently persistent, consistent with the crisis management scenarios discussed in Section 1.2.2. This will introduce significant future risk of fire sales and thus give rise to a significant trade-off between current runs and future fire sales.[45]

1.4.2 The No-Run and Extreme-Run Scenarios: $\theta = 0$ and $\theta = 1$

First, I consider two extreme scenarios, which are sufficient to illustrate the key trade-off underlying the fund's optimal cash management. One is the scenario of $\theta = 0$, in which there are completely no runs. The other is the scenario of $\theta = 1$, in which late shareholders are indifferent between early and late consumptions so that they have the strongest propensity to run.

I show that, in both scenarios, the fund optimally rebuilds its cash buffer when its cash position falls below some threshold. But the threshold and the optimal amount of cash rebuilding are different in these cases, reflecting different trade-offs between current runs and future fire sales.

I start by defining some new notations to streamline the presentation. First, the dynamic game is scale-invariant according to Proposition 3, so it is convenient to define

$$\sigma_{2t+1} \equiv \frac{s_{2t+1}}{a_{2t}},$$

the fraction of illiquid assets that the fund voluntarily sell on odd dates $2t+1$ (relative to the beginning-of-stage asset position a_{2t}), to denote the cash rebuilding policy. Moreover, Corollaries 1, 2, and 6 suggest that η_{2t+2} is uniquely determined by (a_{2t}, x_{2t}) and σ_{2t+1}.[46] Given (a_{2t}, x_{2t}) as the state variables and η_{2t} as the only effective state variable, it is also convenient to use η_{2t+2} to denote the fund's cash rebuilding policy.

I also recap the definitions of different cash-to-assets ratio regions. In particular, I further divide the high region G_h into three different sub-regions: the high-low region G_{hl}, the high-intermediate region G_{hm}, and the high-high region G_{hh}:

---

[45] It should be noted that those equilibria characterized by Lemma 6 are still intuitive and consistent with the model settings. They are just less relevant to the main point of this paper: the dynamic interdependence of shareholder runs and fund liquidity management in a crisis management scenario.

[46] Keep in mind that the fixed stage-game equilibrium selection mechanism is used when needed.

Region	Cash-to-Assets Ratio
G_l	$\eta_{2t} < \dfrac{\mu_E R}{1-\mu_E}$
G_m	$\dfrac{\mu_E R}{1-\mu_E} \leqslant \eta_{2t} < \dfrac{(\mu_E+\mu_L)R}{1-\mu_E-\mu_L}$
G_{hl}	$\eta_{2t} \geqslant \dfrac{(\mu_E+\mu_L)R}{1-\mu_E-\mu_L}$ and $\eta_{2t+2} < \dfrac{\mu_E R}{1-\mu_E}$ if $\sigma_{2t+1} = 0$
G_{hm}	$\eta_{2t} \geqslant \dfrac{(\mu_E+\mu_L)R}{1-\mu_E-\mu_L}$ and $\dfrac{\mu_E R}{1-\mu_E} \leqslant \eta_{2t+2} < \dfrac{(\mu_E+\mu_L)R}{1-\mu_E-\mu_L}$ if $\sigma_{2t+1} = 0$
G_{hh}	$\eta_{2t} \geqslant \dfrac{(\mu_E+\mu_L)R}{1-\mu_E-\mu_L}$ and $\eta_{2t+2} \geqslant \dfrac{(\mu_E+\mu_L)R}{1-\mu_E-\mu_L}$ if $\sigma_{2t+1} = 0$

The three sub-regions of the high region G_h are defined from a dynamic perspective, and they will be useful in describing the optimal dynamic cash rebuilding policy. When the fund starts from G_h and does not rebuild its cash buffer, by definition, after meeting redemptions in the given stage the fund still has a non-negative cash position in the next stage. If the fund ends up into the low region G_l in the next stage, I say that the fund starts from the high-low region G_{hl}. If instead the fund ends up into the intermediate region G_m in the next stage, I say that the fund starts from the high-intermediate region G_{hm}. The high-high region G_{hh} is defined in the same manner. Clearly, there is

$$G_h = G_{hl} \cup G_{hm} \cup G_{hh}\,.$$

Now I characterize the fully dynamic equilibria when $\theta = 0$ and $\theta = 1$. First, I analyze the behavior of shareholder runs in these two scenarios, given any generic and feasible cash rebuilding policy of the fund. These results are useful not only because they illustrate whether and when shareholders will run in equilibrium, but also because they can help pin down all the possible off-equilibrium paths.

Lemma 7. *When $\theta = 0$, none of the late shareholders run in stage t, that is, $\lambda_{2t}(a_{2t}, x_{2t}) = 0$ for any (a_{2t}, x_{2t}) and any cash rebuilding policy σ_{2t+1}.*

Lemma 7 is straightforward. If a shareholder gets nothing when running, they will never run. In other words, there will be completely no runs in this scenario.

Lemma 8. *When $\theta = 1$, all of the late shareholders run in stage t, that is, $\lambda_{2t}(a_{2t}, x_{2t}) = 1$ for any (a_{2t}, x_{2t}) and any positive and feasible cash rebuilding policy $\sigma_{2t+1} > 0$.*

Lemma 8 shows that when the shareholders' propensity to run is the highest (i.e., $\theta = 1$), all of the late shareholders decide to run even if the fund only sells a small amount of assets to rebuild its cash buffer, regardless of the fund's initial cash position. In intuitive terms, in this case shareholders are extremely sensitive to the fund's cash rebuilding. This is because when $\theta = 1$ the late shareholders simply compare between NAV_{2t} and NAV_{2t+1} to decide whether to run. As long as the fund rebuilds its cash buffer, there will be a predictable decline in NAV_{2t+1} regardless of either the fund's initial cash position or other shareholders' run behavior, and thus all of the late shareholders are going to run.

Now I turn to the fund's equilibrium cash rebuilding policy. With the help of Lemma 7, the following proposition first characterizes the optimal cash rebuilding policy when $\theta = 0$.

Proposition 4. *When $\theta = 0$, the equilibrium cash rebuilding policy of the fund is characterized by:*[47]

i) if $\eta_{2t} \in G_l \cup G_m \cup G_{hl}$, the fund chooses $\sigma^*_{2t+1} > 0$ such that

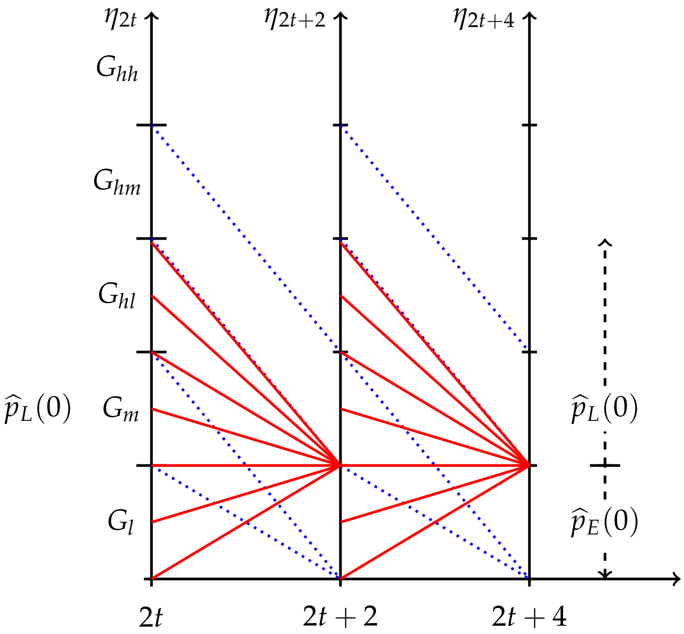
$$\eta^*_{2t+2} = \frac{\mu_E R}{1 - \mu_E}\,, \textit{ and}\,,$$

ii) if $\eta_{2t} \in G_{hm} \cup G_{hh}$, the fund does not rebuild its cash buffer, that is, $\sigma^*_{2t+1} = 0$.

The fund's optimal dynamic cash rebuilding policy when $\theta = 0$, as characterized in Proposition 4, is illustrated in Figure 1.3. In this figure, the horizontal axis denotes date, while the vertical axis denotes the cash-to-assets ratio. The blue dotted line depicts the evolution of the cash-to-assets ratio if the fund does not rebuild its cash buffer at all.

---

[47] Keep in mind that the dynamic equilibrium requires sequential optimality. In other words, the fund's cash rebuilding policy is optimal in a stage only when in the next stage the fund also follows its optimal cash rebuilding policy, which is again conditional on the fund's optimal cash rebuilding policy in the following stage, and so on.



The horizontal axis denotes date. The vertical axis denotes the cash-to-assets ratio (the effective state variable) by different regions. The blue dotted line depicts the evolution of the cash-to-assets ratio if the fund does not rebuild its cash buffer at all. The red line depicts the equilibrium evolution of the cash-to-assets ratio when the fund follows the optimal dynamic cash rebuilding policy. From the perspective of any stage t (including dates $2t$ and $2t+1$), the effective selling price $\widehat{p}_L(\cdot)$ on the left side is that at which the fund can rebuilds its cash buffer in the current stage, while the effective selling prices $\widehat{p}_E(\cdot)$ and $\widehat{p}_L(\cdot)$ on the right side are those at which the fund can raise cash in the next stage in different cash-to-assets regions. See the main text for more explanations.

Figure 1.3: *Equilibrium Cash Rebuilding Policy When* $\theta = 0$

The red line depicts the evolution of the cash-to-assets ratio when the fund follows the optimal dynamic cash rebuilding policy in equilibrium. Because of the scale-invariance and the resulting stationarity of the dynamic game, the equilibrium cash rebuilding policy (conditional on the effective state variable, the cash-to-assets ratio η_{2t}) always follows the same pattern in different stages as long as the game continues.

Proposition 4 says that as long as its initial cash position falls below the high-intermediate region G_{hm}, the fund optimally rebuilds its cash buffer until the next-stage cash-to-assets ratio η_{2t+2} reaches the cutoff between the low region G_l and the intermediate region G_m. But if the fund starts with a higher initial cash position, it will not rebuild its cash buffer. In either case, the fund has a strictly positive cash target for the next stage.

Since there are no runs in equilibrium (by Lemma 7), the main insight behind Proposition 4 is a trade-off between current-stage active asset sales (under a policy of more rapid cash rebuilding) and future-stage forced fire sales (under a policy of no or less rapid cash rebuilding). Intuitively, because the fund manager cares about total expected fire sale losses, it is worthwhile for her to voluntarily sell more assets at the current stage (on date $2t+1$), if the cash buffer rebuilt can help avoid more severe fire sales in the next stage (on dates $2t+2$).

Importantly, due to flexible NAV adjustment, what matter for the dynamic trade-off of current and future amounts of fire sales are not the physical but the effective selling prices as defined in (1.3.5) and (1.3.8). To see this, on the one hand, suppose there is a cash gap $\Delta x_{2t+2} > 0$ on date $2t+2$. In other words, the difference between the fund's initial cash position x_{2t+2} and the amount of cash needed to meet projected redemptions on date $2t+2$ at the beginning-of-day NAV (i.e., NAV_{2t+1}) is Δx_{2t+2}. As a result, there will be forced fire sales on date $2t+2$. But since redeeming shareholders on date $2t+2$ will only get a lower end-of-day NAV, the fund manager can effectively sell at the effective selling price on date $2t+2$ to meet the initial cash gap Δx_{2t+2}. On the other hand, the fund manager can choose to actively sell more assets on date $2t+1$, also at the corresponding effective selling price, to rebuild Δx_{2t+2} unit of cash buffer in advance on date $2t+1$, carry it to date $2t+2$, and thus avoid forced fire sales on date $2t+2$. Moreover, the physical selling price on date $2t+1$ is always higher than that on date $2t+2$. Hence, when the effective selling price on date $2t+1$ is higher than that on date $2t+2$, the fund can sell a smaller amount of assets at a higher physical price on date $2t+1$ in order to avoid higher fire sale losses on date $2t+2$. Hence, a more rapid cash rebuilding policy is better for the purpose of minimizing total fire sale losses, and thus it is optimal for the fund to do so.

I discuss different cash-to-asset ratio regions to illustrate this trade-off in more depth.

First, the fund always rebuilds its cash buffer when starting from the low, the intermediate, or the high-low cash-to-assets ratio region (i.e., $\eta_{2t} \in G_l \cup G_m \cup G_{hl}$). The reason is the following. If the fund did not rebuild its cash buffer, it would end up in the low region in the next stage (i.e., $\eta_{2t+2} \in G_l$). Since the fund will be forced to fire sell its assets then (as the game continues with a high probability $1-\pi$), the fund manager may want to rebuild its cash buffer on date $2t+1$ to avoid fire sales on $2t+2$. Specifically, because late shareholders never run (by Lemma 7), the effective selling price to rebuild cash buffers actively on date $2t+1$ is $\widehat{p}_L(0)$, while the effective selling price to raise cash passively on date $2t+2$ is $\widehat{p}_E(0)$. As $\widehat{p}_L(0) > \widehat{p}_E(0)$, the fund manager always finds it optimal to rebuild its cash buffer on date $2t+1$.

Given that the fund rebuilds its cash buffer, what is the optimal amount of active asset sales? In equilibrium, the fund manager will rebuild the cash buffer up to a point where η_{2t+2} just hits the cut-off between the low and the intermediate region. This is because, on the one hand, a lower cash target still implies forced fire sales on date $2t+2$ at a lower effective selling price $\widehat{p}_E(0)$ and thus is not optimal. On the other hand, any more cash rebuilding on date $2t+1$ means the fund will still have a strictly positive cash buffer on date $2t+3$ after outflows on date $2t+2$. This is also not optimal because that cash buffer is excessive from the perspective of date $2t+1$. In other words, even if asset sales occur on date $2t+3$, the fund manager will be able to sell at the higher effective selling price $\widehat{p}_L(0)$ then. Since the game only has a less than one probability to continue, selling at the same effective price $\widehat{p}_L(0)$ on date $2t+1$ to build that excessive cash buffer is not profitable.

Then, it is straightforward to understand the equilibrium cash rebuilding policy in the high-intermediate and the high-high regions (i.e., $\eta_{2t} \in G_{hm} \cup G_{hh}$) with the intuition outlined above. Here, even if the fund does not rebuild its cash buffer, it will end up at least in the intermediate region, where the fund can raise cash at the effective selling price $\widehat{p}_L(0)$ when needed. As a result, any cash buffer rebuilt on date $2t+1$ (by selling assets also at the effective selling price $\widehat{p}_L(0)$) is excessive, and thus the fund finds $\sigma^*_{2t+1} = 0$ to be optimal.

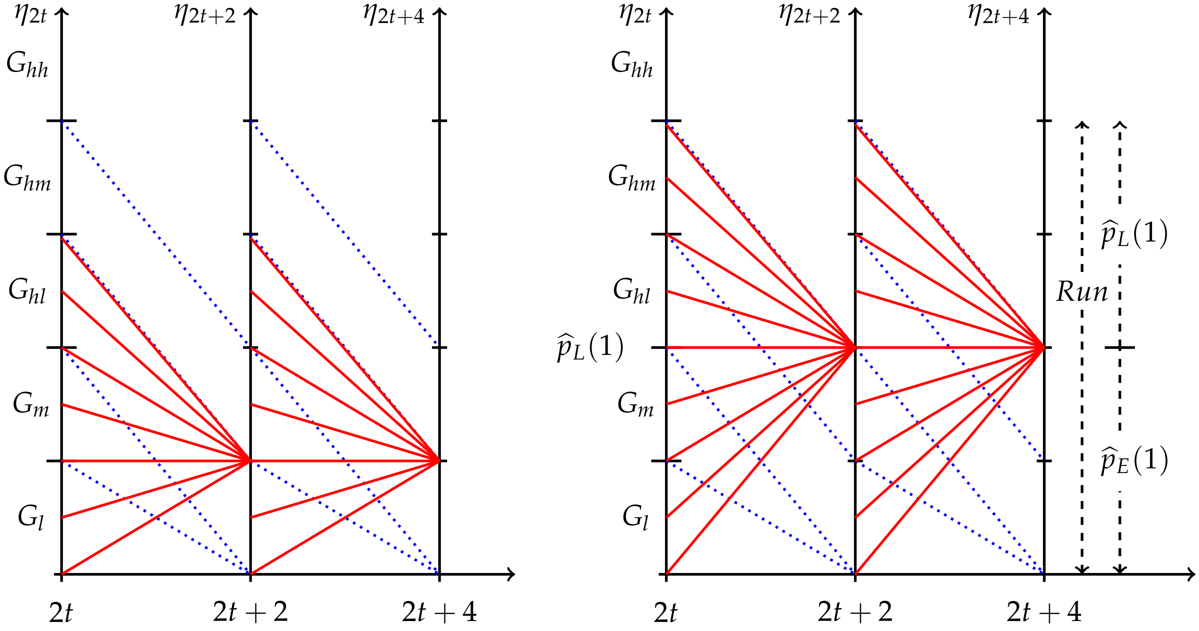
Next, I characterize the optimal cash rebuilding policy when $\theta = 1$ and contrast it to that in the scenario of $\theta = 0$. This illustrates how the potential for runs distorts a fund's dynamic liquidity management.

PROPOSITION 5. *When $\theta = 1$, the equilibrium cash rebuilding policy of the fund is characterized by:*

i) if $\eta_{2t} \in G_l \cup G_m \cup G_{hl} \cup G_{hm}$*, the fund chooses* $\sigma^*_{2t+1} > 0$ *such that*

$$\eta^*_{2t+2} = \frac{(\mu_E + \mu_L)R}{1 - \mu_E - \mu_L}, \text{ and},$$

ii) if $\eta_{2t} \in G_{hh}$*, the fund does not rebuild its cash buffer, that is,* $\sigma^*_{2t+1} = 0$*.*



η_{2t}
η_{2t+2}
η_{2t+4}
G_{hh}
G_{hm}
G_{hl}
G_m
G_l
$2t$
$2t+2$
$2t+4$
η_{2t}
η_{2t+2}
η_{2t+4}
G_{hh}
G_{hm}
G_{hl}
$\widehat{p}_L(1)$
G_m
G_l
$\widehat{p}_L(1)$
Run
$\widehat{p}_E(1)$
$2t$
$2t+2$
$2t+4$

Figure 1.4: *Equilibrium Cash Rebuilding Policies When* $\theta = 0$ *(Left) and* $\theta = 1$ *(Right)*

The right panel of Figure 1.4 illustrates the equilibrium cash rebuilding policy when $\theta = 1$. To recap and better show the difference, I illustrate the equilibrium cash rebuilding policy when $\theta = 0$ on the left. As one can see, the equilibrium when $\theta = 1$ differs significantly from that when $\theta = 0$ because of the interdependence between shareholder runs and fund liquidity management.

Proposition 5 says that the fund starts to rebuild its cash buffer at a higher starting cash position, and it also rebuilds the cash buffer more rapidly, compared to the scenario without runs. Specifically, as long as the fund's initial cash position falls below the high-high region G_{hh}, it rebuilds its cash buffer until the next-stage cash-to-assets ratio η_{2t+2} reaches the cutoff between the intermediate region G_m and the high region G_h.

Although Proposition 5 still features the trade-off between current- and future-stage fire sales, this trade-off becomes more subtle in the presence of runs. By Corollaries 1, 2 and 6, runs in equilibrium result in less effective cash rebuilding (i.e., a lower effective selling price) on odd dates and more forced fire sales on even dates. Thus, when current-stage run risks are relatively high, the fund wants to choose a less rapid cash rebuilding policy. In contrast, when future-stage risk of fire sales is relatively high, in particular, when future-stage runs lead to more severe future-stage fire sales, the fund prefers a more rapid cash rebuilding policy.

Again, I discuss different cash-to-asset ratio regions to illustrate how shareholder runs interact with fund cash rebuilding in equilibrium.

First, suppose the fund starts from the low or the intermediate region (i.e., $\eta_{2t} \in G_l \cup G_m$). By Lemma 8, all of the late shareholders are going to run on date $2t$ (because $NAV_{2t} > NAV_{2t+1}$) in this case, which implies a lower effective selling price (for cash rebuilding) $\widehat{p}_L(1)$ on date $2t+1$. However, if the fund did not rebuild its cash buffer, it would end up in the low region in the next stage, where the fund would have to fire sell at an effective price $\widehat{p}_E(1)$. Because $\widehat{p}_L(1) > \widehat{p}_E(1)$, the risk of future fire sales is relatively larger. Hence, the fund still finds it optimal to rebuild its cash buffer on date $2t+1$ to avoid more costly fire sales on date $2t+2$, despite the runs on date $2t$.

However, different from the scenario when $\theta = 0$, when $\theta = 1$ the fund does not stop rebuilding its cash buffer even if the next-stage cash-to-assets ratio hits the cutoff between the low and the intermediate region. The reason is as follows. If the fund ended up in the intermediate region in the next stage (i.e., $\eta_{2t+2} \in G_m$), again by Lemma 8, all of the late shareholders in the next stage will run on date $2t+2$ too. Thus, the fund would be forced to fire sell its assets on date $2t+2$ at the effective selling price $\widehat{p}_E(1)$ even if starting in the intermediate region then. Fundamentally, future-stage runs lead to higher risk of future fire sales. As a result, the fund will keep rebuilding its cash buffer even when $\eta_{2t+2} \in G_m$.

In equilibrium, the fund manager will rebuild the cash buffer up to a point where η_{2t+2} hits the cutoff between the intermediate and the high region, which is higher than the counterpart when $\theta = 0$ (as shown in Proposition 5). As analyzed above, a lower cash target implies forced fire sales on date $2t + 2$ at a lower effective selling price $\widehat{p}_E(1)$ and thus is not optimal. Also, a higher cash target becomes excessive despite runs in the next stage. Specifically, a higher cash target implies that the fund would end up in the high region in the next stage (i.e., $\eta_{2t+2} \in G_h$), where runs only lead to a lower effective selling price $\widehat{p}_L(1)$ on date $2t + 3$. Since the game only has a less than one probability to continue, selling at the same effective price $\widehat{p}_L(1)$ on date $2t + 1$ to build that excessive cash buffer is not profitable.

Second, suppose the fund starts from the high-low or the high-intermediate region (i.e., $\eta_{2t} \in G_{hl} \cup G_{hm}$). If the fund did not rebuild its cash buffer, it would end up in the low or the intermediate region (i.e., $\eta_{2t+2} \in G_l \cup G_m$), where there would be expected forced fire sales in the next stage. Since the risk of next-stage fire sales is still relatively higher, the fund still finds it optimal to suffer current-stage runs (i.e., to incur a lower effective selling price) and choose a more rapid cash rebuilding policy. For the same reason as discussed above, the fund still optimally rebuilds its cash buffer to the cutoff between the intermediate and the high region, and any more cash buffer would be excessive.

Lastly, I consider the high-high region (i.e., $\eta_{2t} \in G_{hh}$). Without cash rebuilding, the fund would end up in the high-low or high-intermediate region in the next stage (i.e., $\eta_{2t} \in G_{hl} \cup G_{hm}$) as long as the starting cash position is not sufficiently high.[48] As discussed above, all the next-stage late shareholders would decide to run on date $2t + 2$ in those two regions, because the fund would rebuild its cash buffer on date $2t + 3$. Would it make sense for the fund to rebuild its cash buffer on date $2t + 1$ to prevent these runs on date $2t + 2$? The answer is no in this case. This is because in the future stage the fund will be at least in the high region, where runs will not result in forced fire sales on date $2t + 2$. Although future-stage runs on date $2t + 2$ still lead to a lower effective selling price $\widehat{p}_L(1)$ for cash rebuilding on date $2t + 3$, the effective selling price on date $2t + 1$ would also be $\widehat{p}_L(1)$ if the fund chose to rebuild the cash buffer and triggered runs on date $2t$. Since the game only has a less than one probability to continue, the risk of future-stage fire sales becomes relatively smaller while current-stage costs of runs become relatively larger. As a result, the fund optimally chooses not to rebuild its cash buffer in the high-high region.

---

[48] If η_{2t} is sufficiently high, there can be $\eta_{2t+2} \in G_{hh}$ with $\sigma_{2t+1} = 0$. I can further divide region G_{hh} into three separate regions to make the argument more precise. But doing this adds little to provide new insights.

Overall, compared to the scenario without runs (i.e., $\theta = 0$), Proposition 5 and the intuition above suggest that the trade-off in fund cash rebuilding becomes more complicated in the presence of runs. When the starting cash position is lower, future risk of fire sales (in particular, future-stage forced fire sales induced by future runs) is relatively higher, and thus the fund optimally chooses a more rapid cash rebuilding policy. On the contrary, when the starting cash position is higher, current-stage costs of runs are relatively higher, and thus the fund optimally chooses a less rapid cash rebuilding policy or does not rebuild the cash buffer at all.

Finally, it is helpful to contrast the equilibrium cash rebuilding policies here to those in the classic (s, S)-type inventory problem.[49] In the (s, S)-type inventory problems, a firm holds inventory because it helps the firm to meet future demand more easily. However, to accumulate inventory is also costly. It may incur additional adjustment costs, which are often assumed to be exogenous or following a quadratic form. When there are no runs, my fund liquidity management framework resembles a (s, S)-type inventory problem with a zero fixed cost and different variable costs on different dates. However, when the potential for runs is introduced, it features a novel cash inventory problem in which the cost structure is endogenously determined by shareholders' run decisions, which are in turn endogenously determined by the fund's cash rebuilding policy itself.

1.4.3 The General Scenarios

I proceed to characterize the fully dynamic equilibria in the general scenarios when $\theta \in (0, 1)$. In these general scenarios, the equilibrium cash policies and the resulting run behaviors become more involved. This is because when the shareholders have a moderate propensity to run, they become less sensitive to the fund's cash rebuilding than they would in the $\theta = 1$ scenario, while a sufficiently rapid cash rebuilding policy can still push them to run. This in turn shapes the fund's optimal cash rebuilding policy in equilibrium.

---

[49] See Stokey and Lucas (1989) for a textbook treatment of the classic (s, S)-type inventory problems and Strebulaev and Whited (2012) for a review on the modern applications of (s, S)-type problems to dynamic corporate liquidity management. None of the existing (s, S)-type problems considers a mutual fund context as I do.

However, despite the complexity of the general scenarios, all the equilibrium results can be still unified under the same trade-off between current-stage runs and future-stage fire sales as discussed in Section 1.4.2. The formal result is stated in Proposition 6.

Proposition 6. *When $\theta \in (0,1)$, there exist two endogenously determined thresholds $0 < \underline{\theta} < \overline{\theta} < 1$, such that*

i) if $\theta \in (0, \underline{\theta}]$, the equilibrium cash rebuilding policy is characterized by Proposition 4, that is, the cash rebuilding policy follows that in the scenario of $\theta = 0$,

ii) if $\theta \in (\underline{\theta}, \overline{\theta})$, the equilibrium cash policies are characterized by

a) if $\eta_{2t} \in G_l \cup G_m \cup G_{hl} \cup \underline{G_{hm}}$, the fund chooses $\sigma^*_{2t+1} > 0$ such that

$$\eta^*_{2t+2} = \eta(\check{\lambda}) \equiv \frac{(\mu_E + \check{\lambda}\mu_L)R}{1 - \mu_E - \check{\lambda}\mu_L},$$

where $\check{\lambda}$ is given by

$$\begin{cases} \lambda^*_{2t} > 0 & \text{iff} \quad \eta_{2t} < \eta(\check{\lambda}), \\ \lambda^*_{2t} = 0 & \text{iff} \quad \eta_{2t} \geqslant \eta(\check{\lambda}), \end{cases}$$

in which λ^*_{2t} denotes the equilibrium run behaviors under the optimal cash rebuilding policy σ^*_{2t+1}, and

$$\underline{G_{hm}} \equiv \left\{ \eta_{2t} | \eta_{2t} \geqslant \frac{(\mu_E + \mu_L)R}{1 - \mu_E - \mu_L} \text{ and } \frac{\mu_E R}{1 - \mu_E} \leqslant \eta_{2t+2} < \frac{(\mu_E + \check{\lambda}\mu_L)R}{1 - \mu_E - \check{\lambda}\mu_L} \text{ for } \sigma_{2t+1} = 0 \right\},$$

b) if $\eta_{2t} \in \overline{G_{hm}} \cup G_{hh}$, then $\sigma^*_{2t+1} = 0$, where $\overline{G_{hm}} = G_{hm} / \underline{G_{hm}}$,

iii) if $\theta \in [\overline{\theta}, 1)$, the equilibrium cash rebuilding policy is characterized by Proposition 5, that is, the cash rebuilding policy follows that in the scenario of $\theta = 1$.

Figure 1.5 illustrates the optimal cash rebuilding policies when θ varies. As suggested by Proposition 6, when θ is close to 0, the equilibrium is the same as that when $\theta = 0$, while when θ approaches 1 the equilibrium is the same as that when $\theta = 1$. As θ increases, shareholder runs spread to more cash-to-assets regions, and the fund also chooses a more rapid cash rebuilding policy in equilibrium to better avoid future-stage fire sales induced by future-stage runs. Figure 1.5 illustrates the scenarios with a moderate value of θ, in which the equilibrium is different from the two extreme scenarios when $\theta = 0$ and $\theta = 1$.

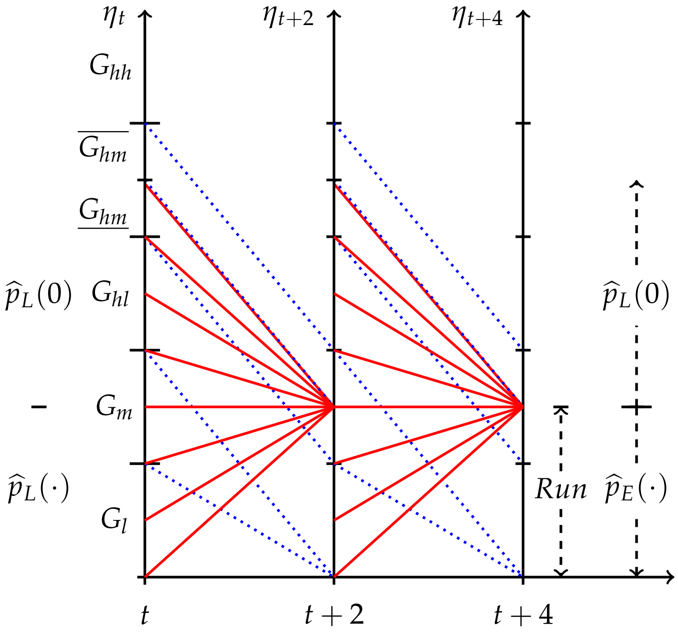

η_t
η_{t+2}
η_{t+4}
G_{hh}
$\overline{G_{hm}}$
$\underline{G_{hm}}$
$\widehat{p}_L(0)$
G_{hl}
G_m
$\widehat{p}_L(\cdot)$
G_l
$\widehat{p}_L(0)$
Run
$\widehat{p}_E(\cdot)$
t
$t+2$
$t+4$

Figure 1.5: *Equilibrium Cash Rebuilding Policy*

1.4.4 The Time-Inconsistency Problem

I illustrate the time-inconsistency problem associated with fund cash rebuilding by asking the following question. From Propositions 4, 5, and 6, it can be seen that for any θ and in any equilibrium path, the fund never allows its target of next-stage cash-to-assets ratio below the intermediate region G_m regardless of the starting cash position. Why? In other words, can there be any circumstances in which the fund finds it optimal to adopt a less rapid cash rebuilding policy such that the next stage game falls into the low region (i.e., $\eta^*_{2t+2} \in G_l$)?

This question is very valid in views of the trade-off between current runs and future fire sales. Especially, as suggested by Corollaries 2 and 6, more shareholder runs result in more severe current-stage forced fire sales when the fund starts from the low or the intermediate region. Why does not the fund choose a less rapid cash rebuilding policy to prevent current-stage runs and thus reduce those forced fire sale losses?

Proposition 7 gives an answer. It suggests that, in the absence of a commitment device, such a less rapid cash rebuilding policy as mentioned above will never appear in any equilibrium path. But it may indeed be optimal if the fund can credibly announce and commit to such a policy on date $2t$. Figure 1.6 illustrates this problem.

Proposition 7. *A cash rebuilding policy involving*

$$\eta^*_{2t+2} < \frac{\mu_E R}{1-\mu_E}$$

cannot happen in any equilibrium path unless the fund is able to credibly commit to such a policy.

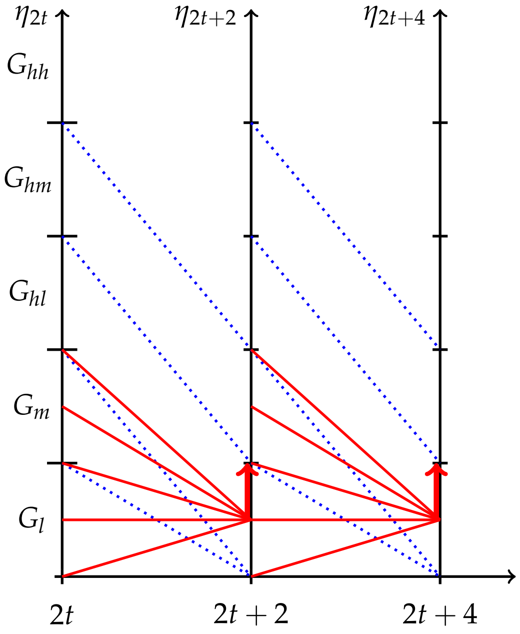


Figure 1.6: *The Time-Inconsistency Problem*

The intuition behind Proposition 7 is a time-inconsistency problem, which aggravates the tension in choosing between a rapid or a slow cash rebuilding policy. Starting from the low region or the intermediate region, the fund indeed has a relatively large current-stage cost of shareholder runs because they lead to severe current-stage forced fire sales. Thus, on date $2t$, the fund may wish to commit itself to rebuilding its cash buffer less rapidly on date $2t+1$ to reduce such run risks on date $2t$. However, on date $2t+1$, because all the date-$2t$ costs of runs are sunk, the fund may instead be tempted to adopt a more rapid cash rebuilding policy on date $2t+1$. Importantly, what matters for shareholders' run decisions on date $2t$ are their beliefs about the fund's cash rebuilding policy on date $2t+1$. In equilibrium, they can always anticipate the fund manager's date-$2t+1$ temptation to rebuild the cash buffer more rapidly, and thus will always have strong incentives to run on date $2t$. Mathematically, the intuition outlined above can be also seen from the dynamic equilibrium definition (Definition 8) and in particular from the Bellman equation (1.4.1) in the non-commitment benchmark.

Proposition 7 suggests a fundamental difficulty in reducing fund shareholder runs in practice, in which a commitment device can be hard to implement. In the mutual fund context, shareholders decide to run not only because they expect other shareholders to run at the same time, but more importantly because they expect the fund to rebuild its cash buffer in the future, which gives rise to the predictable NAV and thus the run incentives. As will be shown in Section 1.5, policies that are effective in preventing bank runs may fail in preventing fund shareholder runs, because they are not designed taking into account the dynamic interdependence of shareholder runs and fund liquidity management.

1.4.5 Expected Total Fire Sale Losses

Finally, I show in Proposition 8 that the potential for shareholder runs can lead to unambiguously higher total fire sale losses ex-ante, regardless of the fund's initial portfolio position. This occurs in a world where both the fund manager and the shareholders are rational, and the fund's cash rebuilding policy is optimal. It suggests that the potential financial stability risks induced by mutual fund shareholder runs can be significant and thus should not be overlooked.

Proposition 8. *When θ increases, the ex-ante total fire sale losses become higher for any positive starting portfolio position (a_{2t}, x_{2t}).*

As suggested by Proposition 7, the lack of a commitment device contributes to the occurrence of run problems despite optimal liquidity management by the fund. I show in Proposition 9 that introducing a commitment device can indeed help reduce total fire sale losses in expectation by tempering shareholders' run incentives.

Proposition 9. *When the fund can pre-commit to a cash policy s_{2t+1} on date $2t$, the ex-ante total fire sale losses become lower for any positive (a_{2t}, x_{2t}) and any $\theta > 0$.*

Intuitively, introducing a commitment device helps reduce total fire sales through two ways. On the one hand, as suggested by Proposition 7, since the fund is able to pre-commit to a less rapid cash rebuilding policy, it can directly reduce current-stage forced fire sales by reducing shareholder runs. On the other hand, from a dynamic perspective, the risk of future-stage fire sales also becomes lower thanks to less severe future runs, and thus the fund is also more comfortable in choosing a less rapid cash rebuilding policy by selling assets less aggressively in the current stage.

1.5 Policy Implications

Many regulators and practitioners have proposed fund-level policies, aiming at mitigating potential financial stability risks of open-end mutual funds. By documenting the dynamic interdependence between shareholder runs and fund liquidity management, my model delivers new policy implications. Rather than making definitive policy prescriptions, I emphasize how the mechanism in this model adds new dimensions to the current policy debates.

1.5.1 MMF Reforms

First of all, my model contributes to the recent debate about MMF reforms. A major proposal of reforming the MMFs is to adopt floating NAV accounting, which will effectively make

MMFs like regular open-end mutual funds.[50] However, as discussed by Hanson, Scharfstein and Sunderam (2015), flexible NAV adjustment may not be a fix to run problems on MMFs. My formal model also suggests the same view. In particular, MMFs adopting a flexible NAV are no longer prone to an abrupt "breaking-the-buck," but will be prone to a new type of shareholder runs as I have shown. This is in particular relevant in bad times when funds' cash positions are low while redemption shocks are large and persistent.

In the following, I turn to several other fund-level policies that are specific to regular open-end mutual funds. I show that, perhaps surprisingly, some of them are less effective than commonly thought in mitigating potential financial stability risks of mutual funds. The key insight is again the dynamic dependence between runs and liquidity management.

1.5.2 Liquidity Requirements

I first discuss the proposal of imposing liquidity requirements on open-end mutual funds. Although no formal liquidity requirements have been proposed to mutual funds, many have been imposed on banks and MMFs by the U.S. SEC[51] and the International Regulatory Framework for Banks (Basel III)[52] since the past financial crisis. Recently, the U.S. SEC voted 5-0 on Sept 22, 2015 to approve a new proposal requiring mutual funds to better manage their liquidity risks. Also, the IMF and the BIS have both proposed potential stress tests for mutual funds in 2015,[53] which resemble fund liquidity requirements. These developments suggest that formal liquidity requirements for mutual funds are likely to be introduced in the future, given the increasing concerns about potential financial stability risks.

---

[50] This proposal has been introduced by the SEC in 2013 in amendments to its 2a-7 rules and then formally adopted in 2014. It applies to institutional prime money market funds. See Hanson, Scharfstein and Sunderam (2015) for a comprehensive discussion.

[51] In 2010, the U.S. SEC changed the 2a-7 rules to institute overnight and weekly liquidity thresholds for MMFs. Specifically, the U.S. SEC mandated that any MMF holds at least 10% of their assets in securities maturing within one business day and at least 30% of assets in securities maturing within one week.

[52] The Basel III accord of 2010 and 2011 require banks to hold a certain amount of high-quality liquid assets to meet expected outflows. Given typical projections of outflows, the threshold for retail banks is set to be 3% of stable retail deposits.

[53] "IMF Seeks Stress Tests for Asset Managers," The *Financial Times*, April 8, 2015, and "Asset Managers' Push into Bonds Prompts Regulatory Scrutiny," The *Financial Times*, June 2, 2015.

My model offers a new building block to help assess the effectiveness of liquidity requirements for mutual funds. From a bank run perspective, Vives (2014) and Diamond and Kashyap (2015) are among the first attempts to evaluate the effectiveness of these new liquidity requirements in mitigating bank runs. Interestingly, as I have already shown, in a bank context, selling illiquid assets to meet liquidity requirements would not induce runs per se as long as the bank is solvent, but doing so can indeed lead to shareholder runs in a mutual fund context. This suggests that designing any appropriate mutual fund liquidity requirements needs to take account of the dynamic interdependence of shareholder runs and fund liquidity management.

Specifically, the results in Proposition 6 suggest state-contingent and fund-specific optimal cash targets for mutual funds. On the one hand, when persistent redemption shocks are likely to occur in the future, mutual fund managers should be warned about the possibility of future shareholder runs. Hence, a more stringent liquidity requirement is preferred. This ensures that when future redemption shocks realize, mutual funds will have higher cash-to-assets ratios and thus can better avoid severe fire sales and runs.

On the other hand, when persistent redemption shocks have already hit the economy as my model describes, the optimal liquidity requirement design should be more fund-specific. For funds whose shareholders have a lower propensity to run (i.e., θ is lower), a more stringent liquidity requirement is likely to help the fund better avoid future fire sales. However, for funds whose shareholders have a higher propensity to run (i.e., θ is higher), a less stringent liquidity requirement as suggested by Proposition 7 is likely to be appropriate, which can better mitigate the fund's time-inconsistency problem and thus reduce shareholder runs.

Of course, the design of any implementable fund liquidity requirements calls for a more comprehensive assessment of other relevant factors. At the bottom line, the results of my model suggest that the one-size-fits-all liquidity requirements as those currently designed for banks and MMFs are unlikely to be appropriate for mutual funds.[54] Instead, the dynamic interdependence of shareholder runs and fund liquidity management should be better taken into account.

1.5.3 Redemption Fees

The second policy proposal is to increase or eliminate the cap on redemption fees. Open-end mutual funds can charge their shareholders redemption fees when they redeem their shares. Currently, the SEC requires mutual fund redemption fees to be lower than 2%.[55] Therefore, some observers argue that to increase or eliminate the cap, at least in crisis times, is likely to mitigate potential financial stability risks of mutual funds.

My model suggests that higher redemption fees may help reduce shareholder runs. Suppose $1-\kappa$ of the redemption proceeds are collected as redemption fees, where $\kappa \in (0,1)$. Thus, any shareholder who redeems on date t only gets κNAV_t.[56] Also, redemption fees are paid back directly to the fund, implying that the fund can save $(1-\kappa)NAV_t$ cash per share redeemed. To better contrast to the baseline model without redemption fees, I consider $\theta = 1$, that is, when the shareholders' propensity to run is the highest. The following proposition shows that the introduction of redemption fees can lead to less shareholder runs in equilibrium.

PROPOSITION **10.** *For any given starting portfolio position* (a_{2t}, x_{2t}), *any feasible cash rebuilding policy* s_{2t+1}, *and any proportional redemption fee* $1-\kappa > 0$, *there is* $\lambda_{2t}^{\kappa} \leqslant \lambda_{2t}$, *where* λ_{2t}^{κ} *is the equilibrium run probability in the game with the redemption fee while* λ_{2t} *is that in the game without redemption fees, all other things being equal.*

When the stage game starts from the high cash-to-assets region, redemption fees have a stronger effect. In contrast to the baseline model where any cash rebuilding (i.e., any $s_{2t+1} > 0$) leads to shareholder runs when $\theta = 1$, with the redemption fee there can be completely no runs in equilibrium when s_{2t+1} is small, that is, when the fund only sells a few assets to rebuild its cash buffer.

---

[54] The recent work of Diamond and Kashyap (2015) suggests that those one-size-fits-all liquidity requirement rules may not be optimal for banks as well, but for different reasons.

[55] This is according to Rule 22c-2 of the Investment Company Act of 1940.

[56] According to Rule 22c-2, the U.S. SEC prohibits discriminative redemption fees solely conditional on shareholder identities as those would effectively create classes of shareholder seniority. This implies that, in my model, the fund cannot intentionally impose different redemption fees on early and late shareholders.

Corollary **3.** *For any given starting portfolio position* (a_{2t}, x_{2t}) *satisfying* $\eta_{2t} \in G_h$ *and any proportional redemption fee* $1 - \kappa > 0$*, there exists a strictly positive* $\bar{s} > 0$ *such that* $\lambda^{\kappa}_{2t} = 0$ *constructs the unique equilibrium when* $s_{2t+1} \leqslant \bar{s}$*.*

Proposition 10 and Corollary 3 suggest that redemption fees can directly reduce shareholders' run incentives. Intuitively, with redemption fees, redeeming shareholders effectively get a value lower than the prevailing NAV, implying a wealth transfer from redeeming shareholders to staying ones. Moreover, in any stage, redemption fees allow the fund to save more cash proportionally, making it easier to meet redemption needs without incurring fire sales.

However, as suggested by Proposition 10, redemption fees do not directly alter the dynamic interdependence between runs and fund liquidity management. They cannot solve the time-inconsistency problem associated with the fund's dynamic cash rebuilding policy either. Only when the fund imposes a 100% redemption fee, or equivalently, chooses to be closed-ending, can shareholder runs be completely prevented. But closed-end mutual funds cannot provide liquidity service to their shareholders (Stein, 2005), and thus it will be obviously not optimal to push all the mutual funds to be closed-ending just for preventing runs.

Redemption fees may also be less effective in practice for the following reasons. First, in my model, redemption fees are introduced ex-ante. However, if redemption fees are first introduced on an odd date $2t + 1$ but are expected on the previous date $2t$, the late shareholders will have higher incentives to run to avoid the fees. This represents a real-world concern that imposing higher redemption fees by itself can lead to one-time market turmoil. Other unmodeled but plausible reasons include negative effects on future fund share sales and on the reputation of fund managers. In practice, many funds still stick to zero or lower than 2% redemption fees even when market volatility or outflows are high (Nanda, Narayanan and Warther, 2000), suggesting that to increase or eliminate the 2% cap on redemption fees may only have limited effects.

1.5.4 In-Kind Redemptions

By nature, open-end mutual funds can satisfy redemption requests by delivering a portion of the underlying basket of assets invested, including cash, which is known as "in-kind redemptions." Many practitioners argue that the option to elect to in-kind redemptions can largely mitigate any financial stability risks of mutual funds, at least during crisis times. Are in-kind redemptions really a relief?

My model suggests that in-kind redemptions can be very effective in preventing fund shareholder runs within a fund, but perhaps surprisingly, they do not necessarily help reduce total fire sale losses or improve total shareholder welfare. The following proposition offers a sufficient condition for the episodes in which the negative effects of in-kind redemptions dominate.

PROPOSITION **11.** *Electing to in-kind redemptions completely prevents shareholder runs, that is,* $\lambda(a_{2t}, x_{2t}) = 0$ *for any* (a_{2t}, x_{2t})*. However, when* θ*,* μ_L *are sufficiently small and* δ_L *is sufficiently larger than* δ_E*, in-kind redemptions lead to higher total fire sale loss ex-ante than a counterfactual in which the fund sticks to cash redemptions, all other things being equal.*

The intuition behind Proposition 11 relies on three progressive reasons. First, adopting in-kind redemptions completely eliminates any run incentives. This is because late shareholders always get the same basket of assets regardless of the time they redeem, and they would have to sell the illiquid assets at a lower price p_E for consumptions if they ran, so they prefer not to run. Second, since the fund manager only cares about total fire sale losses at the fund level, liquidity management becomes irrelevant. In other words, the fund will never rebuild its cash buffer, and the initial cash-to-assets ratio η_0 will never change. Third, early shareholders have to fire sell the illiquid assets they get at the extremely low price p_E for consumptions. These fire sale losses could have been avoided if the fund manager actively managed its cash buffer. If these fire sale losses are significant, shareholders will become worse-off than the counterfactual with cash redemptions.

At a fundamental level, Proposition 11 suggests that in-kind redemptions are not a free lunch even during crisis times, because shareholders who ask their fund to elect to in-kind redemptions effectively give up any benefit they could get from active liquidity management by the fund. This point again highlights the dynamic interdependence between shareholder runs and fund liquidity management. In addition, given that in-kind redemptions are obviously costly during normal times since they discourage the sales of shares, the overall benefit of adopting in-kind redemptions can be even more ambiguous. Moreover, in reality, in-kind redemptions can be hard to implement; they are seldom used for both legal and practical reasons.[57]

Furthermore, the analysis about in-kind redemptions here sheds new light on the potential financial stability risks of exchange-traded funds (ETFs). Observers may argue that ETFs should be immune to shareholder runs because they are not directly subject to outflows from shareholders. Specifically, unlike open-end mutual funds, ETF sponsors can issue and redeem shares only with market-making firms known as authorized participants (APs). In other words, APs are the only market player who can directly redeem ETF shares by trading with fund sponsors. In particular, transactions between an ETF sponsor and an AP are typically settled in-kind, where the AP delivers or receives a basket of assets almost identical to the ETF's holdings,[58] known as the creation/redemption basket. Therefore, by the logic underlying Proposition 11, there will be no direct shareholder or AP runs on ETFs, and ETFs hold less cash buffer than their mutual-fund counterparts do.[59] However, when

---

[57] Rule 18f-1 of the Investment Company Act of 1940 only enables mutual funds to limit in-kind redemptions, which implies that in-kind redemptions will not be effected unless specific approval is first obtained from the SEC. This rule is intended to facilitate mutual fund share sales in jurisdictions where in-cash redemptions are required.

[58] There are a few exceptions where cash transactions may be required. This is because some ETF holdings, such as some specific emerging market assets, are subject to legal restrictions that prevent in-kind transactions.

[59] The fact that ETFs on average hold less liquidity buffer than open-end mutual funds do is documented by new empirical evidence (Ben-David, Franzoni and Moussawi, 2014). This allows ETFs to better track the underlying benchmark while keeping a minimal tracking error.

APs redeem the shares they have bought (presumably at a discount to fund NAV) and get the underlying assets, they are less likely to hold them on their balance sheets for a long time. Rather, they typically sell the underlying assets immediately, trying to lock in the arbitrage profits. This mechanism, known as the AP arbitrage mechanism of ETFs, plays an important role to keep ETF prices as close as possible to fund NAVs. But as suggested by Proposition 11, this may potentially lead to more fire sales (by APs) of the underlying illiquid assets and potentially larger price impacts in the underlying markets. Overall, this analysis suggests that ETFs may also generate previously overlooked financial stability risks, the mechanism of which can be unified under the dynamic interaction between shareholder runs and fund liquidity management.[60]

1.5.5 Redemption Restrictions

A similar emergency rule is redemption restrictions, which give a fund the right to suspend redemptions in given periods as permitted by regulators, for example, the U.S. SEC.[61] Can redemption restrictions prevent shareholder runs?

In my framework, I model redemption restrictions by assuming that the fund is able to deny any individual shareholder's redemption request on any date with probability $1-\zeta$, $\zeta \in (0,1)$. To better contrast to the baseline model without redemption fees, I also consider $\theta = 1$ when the shareholders' propensity to run is the highest. The following proposition characterizes the nature of this game with redemption restrictions.

Proposition **12.** *For any given starting portfolio position* (a_{2t}, x_{2t}) *and any redemption restriction* $1-\zeta > 0$, *there is* $\lambda_{2t}^{\zeta} \leqslant \lambda_{2t}$, *where* λ_{2t}^{ζ} *is the equilibrium run probability in the game with the redemption restriction while* λ_{2t} *is that in the game without redemption restrictions, all other things being equal.*

---

[60] In a complementary paper, Pan and Zeng (2015) explores this AP arbitrage mechanism of ETFs in greater detail and documents the financial stability risks that arise from this mechanism.

[61] According to Rule 22e of the Investment Company Act, an open-end mutual fund is generally prohibited from suspending the right or redemption or postponing the payment of redemption proceeds for more than seven days. However, the SEC has the right to deem emergency periods during which a fund is able to suspend redemptions.

Proposition 12 suggests that redemption restrictions can help reduce shareholder runs. Interestingly, introducing redemption restrictions closely resembles the introduction of redemptions fees as analyzed in Proposition 10 and Corollary 3. The intuition for Proposition 12 is clear, because by the Law of Large Numbers, only ζ of the redeeming shareholders can get cash out of the fund. Therefore, there will be effectively fewer redemptions. But like the introduction of redemption fees, the introduction of redemption restrictions cannot fully prevent shareholder runs or solve the time-inconsistency problem associated with fund liquidity management.

In addition, my model suggests that redemption restrictions can be indeed hard to implement in reality, given that they have to be introduced at the discretion of regulators. Due to the "slow-moving" nature of shareholder runs, no abrupt bankruptcy or events like "breaking-the-buck" by MMFs can be observed. Even in scenarios where run-induced fire sales are extremely severe, as suggested by Proposition 6, the fund can still keep a positive cash-to-assets ratio in each stage, making it hard for the regulatory authority to deem such periods emergent. In practice, unsurprisingly, the U.S. SEC has seldom deemed a period to be an emergency to allow open-end mutual funds to use redemption restrictions. One and the only recent example for the SEC to permit redemption restrictions happened in 2008 for the Reserve Primary Fund,[62] which was an MMF.

1.5.6 Credit Lines

Although my baseline model focuses on a crisis management scenario during which there are no net inflows, mutual funds may turn to pre-established credit lines to raise cash. For example, in 2015 BlackRock increased the amount that its mutual funds can collectively borrow to meet redemptions to $2.1 billion, a historically high level.[63] Can fund credit lines prevent shareholder runs?

My model suggests that using credit lines may temporarily mitigate the negative effects of current-stage shareholder runs, but can induce more severe fire sales and runs in the future. Specifically, in stage t, suppose the fund uses pre-established credit lines (rather than selling assets) when it is in the low or the intermediate cash-to-assets region. Thus, the fund does not have to fire sell any illiquid assets in meeting redemptions on dates $2t$ and $2t+1$. As a result, the NAV will not change within stage t, that is, $NAV_{2t+1} = NAV_{2t} = NAV_{2t-1}$, and thus $(\lambda_E + \lambda_L)n_{2t}$ shareholders leave the fund with such an intact NAV. However, in the next stage (if the game continues) the fund will have no cash to start (i.e., $\eta_{2t+2} = 0$). If the credit lines have a sufficiently long maturity, the fund does not have to pay back the debts immediately, but have to face more severe fire sales unless it can borrow more. What is worse, if the fund is required to pay back its debts first on date $2t+2$, it will have to fire sell even more illiquid assets or simply default. Intuitively, using credit lines makes a fund' life easier temporarily, but the fund also forgoes the option value of cash rebuilding, which is higher when the redemption shocks are more persistent (i.e., π is smaller). This idea resembles that outlined for in-kind redemptions in Section 1.5.4, suggesting that it will be naive to shut down a fund's active liquidity management when attempting to prevent runs.

---

[62] See Investment Company Act Release No. 28,487.

[63] "BlackRock Leads Funds Raising Credit Lines Amid Review," The *Bloomberg Business*, January 21, 2015.

Moreover, credit lines may expose a fund back to the risk of debt runs as suggested by He and Xiong (2012) and by other bank run models. This is in particular relevant when the redemption shocks are persistent so that the fund has to repeatedly turn to credit lines established with multiple creditors or to rollover existing debts. In a crisis management scenario, creditor banks may also be subject to aggregate risks, making credit lines riskier and less reliable than cash buffers (Acharya, Almeida and Campello, 2013). Thus, having credit lines is hardly a relief for open-end mutual funds.

1.5.7 Swing Pricing

Some observers argue that swing pricing, which allows current NAVs to reflect commissions to asset brokers and dealers, bid-ask spreads, taxes, and other trading-related charges, can reduce the negative externalities imposed by redeeming shareholders on non-trading ones. A growing number of open-end mutual funds has been adopting swing pricing, while as of September 2015 the use of swing pricing is still voluntary and not required by the U.S. SEC.[64] Will full swing pricing prevent shareholder runs?

My model suggests that the answer is no. In fact, swing pricing, in its currently observed form, has already been incorporated into my baseline model, This is because flow-induced fire sales are the only type of trading-related costs in the model, and current NAVs have already taken them into account. However, my model suggests that, since they still do not incorporate future asset sale costs, they are not able to mitigate the risk of runs induced by active fund liquidity management. In this sense, my theory identifies a form of negative externality that even introducing swing pricing cannot internalize.

Rather than swing pricing in its current form, my model suggests that forward-looking NAVs may help reduce shareholder runs. This is equivalent to requiring shareholders to contract on future NAVs directly. However, from the perspective of market incompleteness, shareholders cannot fully contract on future NAVs because mutual funds promise to provide daily liquidity service to their shareholders. In other words, if shareholders instead contracted on future NAVs and they had common and rational beliefs on future NAVs, they would effectively go back to "separate accounts," or equivalently direct holdings of the underlying assets by the shareholders, and there will be no liquidity service provided by the funds. In this sense, there is no point to have a mutual fund in the first place. As a result, forward-looking NAV rules may be hard to implement in reality, and runs can be viewed as a cost that shareholders have to bear to have mutual funds engage in liquidity transformation.[65] To investigate optimal mechanism design for liquidity provision in a mutual fund context (like Green and Lin, 2003 and Peck and Shell, 2003 in a bank context) is beyond the scope of this paper, and thus I leave it for future research.

---

[64] "Fund Investors May Pay Fees for Withdrawals Amid Turmoil," The *Bloomberg Business*, September 11, 2015.

[65] One can go further along this line to ask why mutual funds provide liquidity. Gorton and Pennacchi (1990) argue that banks and bank-like financial intermediaries are the best candidate to provide liquidity because debt value is the least sensitive to asset-side value fluctuations. But as suggested by Stein (2005), in competing for funding, mutual funds are also eager to provide liquidity (i.e., to adopt open-ending) to shareholders through the means of equity. To make such liquidity appealing to shareholders, mutual funds invest in higher-yielding but illiquid assets with the help of deliberate cash management, effectively engaging themselves in liquidity transformation. On the one hand, this logic suggests that mutual funds are not a simple pass-through; they are not as "plain-vanilla" as practitioners usually argue. On the other hand, more importantly, as fund shareholders get the promised liquidity at the flexible NAV, their liquidity is more sensitive to asset-side fluctuations. This nature makes mutual funds immune to the usual notion of debt runs that stem from fixed claims, but prone to

1.6 Extensions

My baseline model parsimoniously captures the novel interdependence between shareholder runs and fund liquidity management. The underlying mechanism is fairly general and robust to other aspects which may either aggravate or mitigate mutual fund financial stability risks. Here I explore several extensions, one at a time, with a focus on their interactions with the main mechanism of the baseline model.

1.6.1 Flow-to-Performance Relationship

The baseline model assumes random redemption shocks, but the realized population of redeeming shareholders in each stage is taken as exogenous. One may argue that future fund flows are likely to be positively correlated with past returns, known as the flow-to-performance relationship. Earlier research finds that future flows mostly respond to past good performance (Ippolito, 1992, Sirri and Tufano, 1998), but recent evidence suggests that they also respond to bad performance in particular when the underlying assets are illiquid (for example, Spiegel and Zhang, 2013, Goldstein, Jiang and Ng, 2015). Does such flow-to-performance relationship interact with fund shareholder runs?

My model suggests that, in the presence of shareholder runs, introducing the flow-to-performance relationship implies higher total expected fire sale losses despite optimal cash management by the fund. To see this, I can incorporate the flow-to-performance relationship into my baseline model. For any stage, I define the fund return as

$$r_{2t+1} = \frac{NAV_{2t+1}}{NAV_{2t}},$$

---

a new notion of equity runs that result from asset-side value adjustment.

which is positive but no greater than one in my baseline model.[66] I then assume that in any stage t the populations of early and late shareholders are $\gamma_{2t}\lambda_E n_{2t}$ and $\gamma_{2t}\lambda_L n_{2t}$ for the even date $2t$ and the odd date $2t+1$, respectively, where $\gamma_{2t}(r_{2t-1}) \geqslant 1$ for $t > 1$ is a decreasing function of r_{2t-1} satisfying $\gamma_{2t}(1) = 1$, and $\gamma_0 = 1$. This implies that, if current fund return is lower, there will be more shareholders redeeming in the next stage if the game continues, capturing the flow-to-performance relationship.

In this extended setting, the flow-to-performance relationship does not directly alter shareholders' run incentives in any stage game, but will complicate the tension in choosing between a rapid or slow cash rebuilding policy by the fund manager. This can be seen from Proposition 5. Suppose the fund starts from the joint region $G_l \cup G_m \cup G_{hl} \cup G_{hm}$ where it is optimal to sell some illiquid assets to rebuild the cash buffer (i.e., $\sigma^*_{2t+1} > 0$). When the flow-to-performance relationship is introduced, σ^*_{2t+1} suggested by Proposition 5 is no longer optimal. To see this, notice that $\sigma^*_{2t+1} > 0$ implies $r_{2t+1} < 1$ and then $\gamma_{2t+2} > 1$. As a result, the fund either has to increase σ_{2t+1} to prevent more severe future fire sales due to a larger population of redeeming shareholders in the next stage, or to decrease σ_{2t+1} to sustain a higher current fund return but suffer higher risk of future fire sales. Either way, the fund incurs higher risk of shareholder runs and higher total expected fire sale losses as well.

At a fundamental level, the extended setting suggests a new amplification mechanism to explain fund performance persistence in bad times. The flow-to-performance relationship first implies that it is harder for the fund to manage its cash buffer. Due to the interdependence of shareholder runs and fund liquidity management, this further suggests more severe runs and fire sales, leading to worse performance. Only those funds with a sufficiently high cash-to-assets ratio are likely to withstand these hard times without incurring shareholder runs and fire sales.

---

[66] But it can be larger than one in the model with redemption fees or redemption restrictions.

1.6.2 Asset Price Correlations

In the baseline model, flow-induced fire sales will not impact the market prices of the non-traded assets that still remain on the fund's balance sheet. This is realistic given that mutual funds invest in many different illiquid assets, and flow-induced fire sales only have local and temporary price impacts (Coval and Stafford, 2007). One may argue that asset prices can be correlated with each other, and the fund manager may want to use alternative accounting rules such as matrix pricing for these non-traded assets. Will those differences have any qualitative effects on shareholder runs?

My model suggests no. To see why, I assume that asset prices are perfectly correlated (while still keeping the realistic assumption that the price impacts induced by fire sales are temporary; they only last for one stage). This can be effectively viewed as a world with only a single illiquid asset. In this alternative setting, if the fund sells some assets on date t at the fire sale price p_t, the end-of-day flexible NAV will be:

$$NAV_t = \frac{x_t + (a_t - a_{t+1})p_t + a_{t+1}p_t}{n_t}.[67] \tag{1.6.1}$$

Clearly, the only difference between (1.6.1) and the baseline model NAV (1.2.1) is the last term in the numerator, which reflects the fact that the market prices of non-traded assets are also updated to p_t, the temporary fire-sale price, in this extended setting. In both (1.2.1) and (1.6.1), the NAV is flexible in the sense that it takes into account all the same-day price impact and asset sale losses, while it is not forward-looking in the sense that it will not reflect any possible future price impacts and asset reallocation costs. By similar analysis as that in Section 1.3, as long as these contractual features of fund NAV are present, fund cash rebuilding still gives rise to a predictable decline in NAV and thus the run incentives. As a result, introducing asset price correlations or alternative accounting rules at the fund level would not change my results qualitatively.[68]

---

[67] To be more precise, in this case the fire sale price p_t will be assumed to be a decreasing function of the amount of fire sales within the given date, and the slope of this intra-day downward-sloping demand curve is steeper on an even date than that on an odd date.

[68] Depending on how correlated asset prices are, the run incentives can be quantitatively different in

1.6.3 Persistent Price Impacts

Although the price impacts induced by fire sales tend to be temporary, in some asset classes there can be more persistent or even long-term price impacts. For example, in illiquid asset markets where dealers actively manage their inventory, a higher inventory level implies a higher price concession to compensate for their inventory risks. Would persistent price impacts change shareholders' run behaviors?

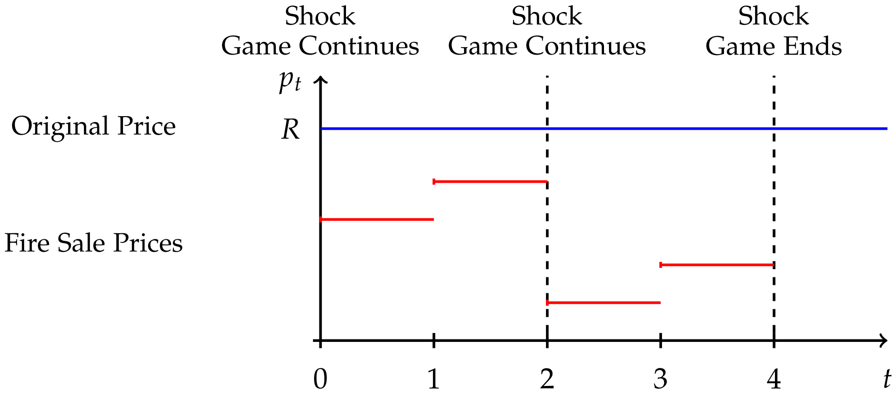


Figure 1.7: *Price Path with Persistent Price Impacts*

To answer this question, I can incorporate persistent price impacts into the baseline model. I illustrate that there can be more severe shareholder runs and fire sales in equilibrium in this extended setting, despite optimal cash management by the fund.

Specifically, I assume that in stage t, the fire sale price on the even date $2t$ is $\beta^t \delta_E R$, while that on the odd date $2t+1$ is $\beta^t \delta_L R$, where $\beta \in (0,1]$. When $\beta = 1$, this goes back to the baseline model. Figure 1.7 illustrates a sample path of selling prices with persistent price impacts. There are two important ingredients of this price pattern. First, in the short run (i.e., within each stage), it still features temporary price overshooting and subsequent reversal. This is consistent with the baseline model.[69] Second, in the long run (i.e., across

---

equilibrium. This quantitative difference is not crucial for the key mechanism of this model.

[69] Note that, this extended setting still keeps the structure of one shock per two dates, and thus the microfoundation in Appendix A.3 is still valid.

different stages) but before the game ends, the selling prices for the illiquid assets become lower over time.

Like the analysis in Section 1.6.1, persistent price impacts do not directly change shareholders' run incentives, but will significantly change the fund's dynamic optimal cash rebuilding policy, which in turn alters shareholders' equilibrium run behaviors. Again, this can be seen from Proposition 6. In the baseline model, the fund rebuilds its cash buffer on the odd date $2t+1$ to prevent forced fire sales on the next even date $2t+2$. The fund is not worried about potential asset sales on the next odd date $2t+3$ because of the same asset selling price there. However, in the extended setting with persistent price impacts, the asset selling price on the next odd date $2t+3$ will become lower, which gives rise to an incentive for the fund to sell more assets on date $2t+1$. Therefore, by the results in Propositions 1, 2 and 27, more rapid cash rebuilding unambiguously leads to more severe run problems and potentially higher total fire sale losses in expectation.

This extension has two implications. On the one hand, it suggests that shareholder runs and the dynamic interdependence between runs and fund liquidity management do not rely on the existence of persistent price impacts described here. In other words, there can be shareholder runs in equilibrium without persistent price impacts. On the other hand, the existence of persistent price impacts can make the run problems more severe. Given that persistence price impacts can indeed exist in some illiquid asset markets, this extension suggests that the concerns about shareholder runs can be indeed very relevant.

1.7 Conclusion

In this paper, I build a model of an open-end mutual fund with a flexible NAV, and show that shareholder runs can occur in equilibrium despite optimal liquidity management by the fund. With a flexible NAV, fund cash rebuilding by selling illiquid assets implies a predictable decline in NAV and thus a first-mover advantage, leading to runs. The presence of shareholder runs further complicates the fund's efforts in liquidity management, leading to higher total fire sale losses in expectation. Hence, appropriate design of policies aiming for mitigating financial stability risks of mutual funds should take into account the dynamic interdependence of shareholder runs and fund liquidity management.

At a fundamental level, shareholder runs are driven by a key contractual property of mutual fund NAVs: they are flexible but not forward-looking. Specifically, the NAV at t does not take into account the predictable asset sales and price impact at $t+1$. This contractual property comes from a form of market incompleteness that shareholders cannot fully contract on future NAVs. This property implies that cash rebuilding can give rise to predictable declines in NAV and thus the potential for runs.

Finally, my model sheds new light on potential systemic risks posed by mutual funds. As mutual fund runs can lead to more fire sales, the underlying asset markets may become even more illiquid. As suggested by Stein (2014) and formally shown in He and Milbradt (2014), this effect can cause more corporate bond defaults and impose considerate risks on real economic activities. This channel becomes increasingly relevant given the bank-bond substitution after the crisis (Becker and Ivashina, 2014, Crouzet, 2015) as well as the increasing "reaching-for-yield" behavior in corporate bond markets (Becker and Ivashina, 2015). To be clear, I do not claim that mutual fund runs cause more systemic risks. The systemic implications of mutual fund runs depend not only on the contagion from secondary-market fire sales to primary-market investment losses, but also on how non-bank financial intermediaries interact with other bank-like financial institutions. A thorough investigation covering all these issues is beyond the scope of this paper, but the results here can naturally serve as a building block for future research on these issues.

Bibliography

Abreu, Dilip and Markus Brunnermeier (2003). "Bubbles and Crashes." *Econometrica*, 71: 173-204.

Acharya, Viral, Heitor Almeida and Murillo Campello (2013). "Aggregate Risk and the Choice Between Cash and Lines of Credit." *Journal of Finance*, 68: 2059-2116.

Agnew, Julie, Pierluigi Balduzzi and Annika Sunden (2003). "Portfolio Choice and Trading in A Large 401(k) Plan." *American Economic Review*, 93: 193-215.

Alexander, Gordon, Gjergji Cici and Scott Gibson (2007). "Does Motivation Matter When Assessing Trade Performance? An Analysis of Mutual Funds." *Review of Financial Studies*, 20: 125-150.

Allen, Franklin and Douglas Gale (1994). "Limited Participation and Volatility of Asset Prices." *American Economics Review*, 84: 933-955.

Allen, Franklin and Douglas Gale (1998). "Optimal Financial Crises." *Journal of Finance*, 53: 1245-1284.

Allen, Franklin and Douglas Gale (2005). "From Cash-in-the-Market Pricing to Financial Fragility." *Journal of the European Economic Association*, 3: 535-546.

Ameriks, John and Stephen Zeldes (2004). "How Do Household Portfolio Shares Vary with Age?" Working Paper.

Bates, Thomas, Kathleen Kahle and Rene Stulz (2009). "Why Do U.S. Firms Hold so Much More Cash than They Used to?" *Journal of Finance*, 64: 1985-2021.

Ben-David, Itzhak, Francesco Franzoni and Rabih Moussawi (2014). "Do ETFs Increase Volatility?" Working Paper.

Ben-Rephael, Azi (2014). “Flight-to-Liquidity, Market Uncertainty, and the Actions of Mutual Fund Investors.” Working Paper.

Bernardo, Antonio and Ivo Welch (2004). “Liquidity and Financial Market Runs.” *Quarterly Journal of Economics*, 119: 135-158.

Becker, Bo and Victoria Ivashina (2014). “Cyclicality of Credit Supply: Firm Level Evidence.” *Journal of Monetary Economics*, 62: 76-93.

Becker, Bo and Victoria Ivashina (2015). “Reaching for Yield in the Bond Market.” *Journal of Finance*, 70: 1863-1901.

Bhaskar, V., Stephen Morris and George Mailath (2013). “A Foundation for Markov Equilibria in Sequential Games with Finite Social Memory.” *Review of Economic Studies*, 80: 925-948.

Bhattacharya, Sudipto and Paul Pfleiderer (1985). “Delegated Portfolio Management.” *Journal of Economic Theory*, 36: 1-25.

Bolton, Patrick, Hui Chen and Neng Wang (2011). “A Unified Theory of Tobin’s q, Corporate Investment, Financing, and Risk Management.” *Journal of Finance*, 66: 1545-1578.

Brunnermeier, Markus and Stefan Nagel (2008). “Do Wealth Fluctuations Generate Time-Varying Risk Aversion? Micro-Evidence on Individuals’ Asset Allocation.” *American Economic Review*, 98: 713-736.

Brunnermeier, Markus and Lasse Pedersen (2005). “Predatory Trading.” *Journal of Finance*, 60: 1825-1863.

Bryant, John (1980). “A Model of Reserves, Bank Runs, and Deposit Insurance.” *Journal of Banking and Finance*, 4: 335-344.

Campbell, John, Sanford Grossman and Jiang Wang (1993). “Trading Volume and Serial Correlation in Stock Returns.” *Quarterly Journal of Economics*, 108: 905-939.

Carlin, Bruce, Miguel Lobo and S. Viswanathan (2007). “Episodic Liquidity Crises: Cooperative and Predatory Trading.” *Journal of Finance*, 62: 2235-2274.

Chari, V. V. and Ravi Jagannathan (1988). “Banking Panics, Information, and Rational Expectations Equilibrium.” *Journal of Finance*, 43: 749-761.

Chen, Qi, Itay Goldstein and Wei Jiang (2010). “Payoff Complementarities and Financial Fragility: Evidence from Mutual Fund Outflows.” *Journal of Financial Economics*, 97: 239-262.

Cheng, Ing-Haw and Konstantin Milbradt (2012). “The Hazards of Debt: Rollover Freezes, Incentives, and Bailouts.” *Review of Financial Studies*, 25: 1070-1110.

Chernenko, Sergey and Adi Sunderam (2015). “Liquidity Transformation in Asset Management: Evidence from the Cash Holdings of Mutual Funds.” Working Paper.

Chordia, Tarun (1996). “The Structure of Mutual Fund Charges.” *Journal of Financial Economics*, 41: 3-39.

Christoffersen, Susan, Donald Keim and David Musto (2007). “Valuable Information and Costly Liquidity: Evidence from Individual Mutual Fund Trades.” Working Paper.

Christoffersen, Susan, David Musto and Russ Werners (2014). “Investor Flows to Asset Managers: Causes and Consequences.” *Annual Review of Financial Economics*, 6: 289-310.

Comerton-Forde, Carole, Terrence Hendershott, Charles Jones, Pamela Moulton, and Mark Seasholes (2010). “Time Variation in Liquidity: The Role of Market-Maker Inventories and Revenues.” *Journal of Finance*, 65: 295-332.

Cooper, Russell and Thomas Ross (1998). “Bank Runs: Liquidity Costs and Investment Distortions.” *Journal of Financial Economics*, 41: 27-38.

Coval, Joshua and Erik Stafford (2007). “Asset Fire Sales (and Purchases) in Equity Markets.” *Journal of Financial Economics*, 86: 479-512.

Crouzet, Nicolas (2015). “Aggregate Implications of Corporate Debt Choices.” Working Paper.

Da, Zhi, Borja Larrain, Clemens Sialm and Jose Tessada (2015). “Price Pressure from Coordinated Noise Trading: Evidence from Pension Fund Reallocations.” Working Paper.

Della Seta, Macro, Erwan Morellec and Francesca Zucchi (2015). “Debt Structure, Rollover Traps, and Default Risk.” Working Paper.

Diamond, Douglas and Philip Dybvig (1983). “Bank Runs, Deposit Insurance, and Liquidity.” *Journal of Political Economy*, 91: 401-419.

Diamond, Douglas and Anil Kashyap (2015). "Liquidity Requirements, Liquidity Choice and Financial Stability." Working Paper.

Di Maggio, Marco (2015). "Market Turmoil and Destabilizing Speculation." Working Paper.

Dickson, Joel, John Shoven and Clemens Sialm (2000). "Tax Externalities of Equity Mutual Funds." *National Tax Journal*, 53: 607-628.

Duffie, Darrell (2010). "Presidential Address: Asset Price Dynamics with Slow-Moving Capital." *Journal of Finance*, 65: 1237-1267.

Duffie, Darrell, Nicolae Garleanu and Lasse Pedersen (2005). "Over-the-Counter Markets." *Econometrica*, 73: 1815-1847.

Duffie, Darrell, Nicolae Garleanu and Lasse Pedersen (2007). "Valuation in Over-the-Counter Markets." *Review of Financial Studies*, 20: 1865-1900.

Edelen, Roger (1999). "Investor Flows and the Assessed Performance of Open-End Mutual Funds." *Journal of Financial Economics*, 53: 439-466.

Edmans, Alex, Itay Goldstein and Wei Jiang (2012). "The Impact of Prices on Takeovers." *Journal of Finance*, 67: 933-971.

Ellul, Andrew, Chotibhak Jotikasthira and Christian Lundblad (2011). "Regulatory Pressure and Fire Sales in the Corporate Bond Market." *Journal of Financial Economics*, 101: 596-620.

Ennis, Huberto and Todd Keister (2003). "Economic Growth, Liquidity, and Bank Runs." *Journal of Economic Theory*, 109: 220-245.

Feroli, Michael, Anil Kashyap, Kermit Schoenholtz and Hyun Song Shin (2014). "Market Tantrums and Monetary Policy." Working Paper.

Foley-Fisher, Nathan, Borghan Narajabad and Stephane Verani (2015). "Self-Fulfilling Runs: Evidence from the US Life Insurance Industry." Working Paper.

Gennaioli, Nicola, Shleifer, Andrei and Vishny, Robert (2012). "Neglected Risks, Financial Innovation and Financial Fragility." *Journal of Financial Economics*, 104: 452-468.

Gennaioli, Nicola, Shleifer, Andrei and Vishny, Robert (2013). "A Model of Shadow Banking." *Journal of Finance*, 68: 1331-1363.

Gennaioli, Nicola, Shleifer, Andrei and Vishny, Robert (2015). “Money Doctors.” *Journal of Finance*, 70: 91-114.

Gertler, Mark and Nobuhiro Kiyotaki (2014). “Banking, Liquidity and Bank Runs in an Infinite Horizon Economy.” Working Paper.

Goldstein, Itay, Hao Jiang and David Ng (2015). “Investor Flows and Fragility in Corporate Bond Funds.” Working Paper.

Goldstein, Itay and Ady Pauzner (2005). “Demand Deposit Contracts and the Probability of Bank Runs.” *Journal of Finance*, 60: 1293-1328.

Gorton, Gary and George Pennacchi (1990). “Financial Intermediaries and Liquidity Creation.” *Journal of Finance*, 45: 49-71.

Gorton, Gary and Andrew Winton (2003). “Financial Intermediation.” in Milton Harris, George Constantinides, and Rene Stulz, ed., *Handbook of the Economics of Finance: Corporate Finance*, 1: 431-552.

Green, Edward and Ping Lin (2003). “Implementing Efficient Allocations in A model of Financial Intermediation.” *Journal of Economic Theory*, 109: 1-23.

Greenwood, Robin and David Thesmar (2011). “Stock Price Fragility.” *Journal of Financial Economics*, 102: 471-490.

Grinblatt, Mark and Matti Keloharju (2009). “Sensation Seeking, Overconfidence, and Trading Activity.” *Journal of Finance*, 64: 549-578.

Grossman, Sanford and Merton Miller (1988). “Liquidity and Market Structure.” *Journal of Finance*, 43: 617-633.

Hanson, Samuel, David Scharfstein and Adi Sunderam (2015). “An Evaluation of Money Market Fund Reform Proposals.” *IMF Economic Review*, forthcoming.

He, Zhiguo and Asaf Manela (2014). “Information Acquisition in Rumor-based Bank Runs.” *Journal of Finance*, forthcoming.

He, Zhiguo and Konstantin Milbradt (2014). “Endogenous Liquidity and Defaultable Bonds.” *Econometrica*, 82: 1443-1508.

He, Zhiguo and Konstantin Milbradt (2015). “Dynamic Debt Maturity.” Working Paper.

He, Zhiguo and Wei Xiong (2012). “Dynamic Debt Runs.” *Review of Financial Studies*, 25: 1799-1843.

Hendershott, Terrence and Mark Seasholes (2007). “Market Maker Inventories and Stock Prices.” *American Economic Review*, 97: 210-214.

Hendershott, Terrence and Albert Menkveld (2014). “Price Pressures.” *Journal of Financial Economics*, 114: 405-423.

Huang, Jiekun (2015). “Dynamic Liquidity Preferences of Mutual Funds.” Working Paper.

Hugonnier, Julien and Erwan Morellec (2014). “Bank Capital, Liquid Reserves, and Insolvency Risk.” Working Paper.

Ippolito, Richard (1992). “Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry.” *Journal of Law and Economics*, 35: 45-70.

Jacklin, Charles and Sudipto Bhattacharya (1988). “Distinguishing Panics and Information-Based Bank Runs: Welfare and Policy Implications.” *Journal of Political Economy*, 96: 568-592.

Jotikasthira, Chotibhak, Christian Lundblad and Tarun Ramadorai (2012). “Asset Fire Sales and Purchases and the International Transmission of Financial Shocks.” *Journal of Finance*, 67: 2015-2050.

Kechris, Alexander (1995). *Classical Descriptive Set Theory*. Springer, New York.

Khan, Ali and Yeneng Sun (2002). “Non-Cooperative Games with Many Players.” in Robert Aumann and Sergiu Hart, ed., *Handbook of Game Theory with Economic Applications*, 3: 1761-1808.

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

Lagos, Ricardo, Guillaume Rocheteau and Pierre-Olivier Weill (2011). “Crashes and Recoveries in Illiquid Markets.” *Journal of Economic Theory*, 146: 2169-2205.

Liu, Xuewen and Antonio Mello (2011). “The Fragile Capital Structure of Hedge Funds and the Limits to Arbitrage.” *Journal of Financial Economics*, 102: 491-506.

Lou, Dong (2012). "A Flow-Based Explanation for Return Predictability." *Review of Financial Studies*, 25: 3457-3489.

McCabe, Patrick (2009). "The Economics of the Mutual Fund Trading Scandal." Working Paper.

Manconi, Alberto, Massimo Massa and Ayako Yasuda (2012). "The Role of Institutional Investors in Propagating the Crisis of 2007-2008." *Journal of Financial Economics*, 104: 491-518.

Martin, Antoine, David Skeie and Ernst-Ludwig von Thadden (2014). "Repo Runs." *Review of Financial Studies*, 27: 957-989.

Mitchell, Mark, Lasse Pedersen and Todd Pulvino (2007). "Slow Moving Capital." *American Economic Review*, 97: 215-220.

Mitchell, Olivia, Gary Mottola, Stephen Utkus and Takeshi Yamaguchi (2006). "The Inattentive Participant: Portfolio Trading Behavior in 401(k) Plans." Working Paper.

Morris, Stephen and Hyun Song Shin (2004). "Liquidation Black Holes." *Review of Finance*, 8: 1-18.

Morris, Stephen (2014). "Coordination, Timing and Common Knowledge." *Research in Economics*, 68: 306-314.

Morris, Stephen and Hyun Song Shin (2014). "Risk-Taking Channel of Monetary Policy: A Global Game Approach." Working Paper.

Nagel, Stefan (2012). "Evaporating Liquidity." *Review of Financial Studies*, 25: 2005-2039.

Nanda, Vikram, M. P. Narayanan and Vincent Warther (2000). "Liquidity, Investment Ability, and Mutual Fund Structure." *Journal of Financial Economics*, 57: 417-443.

Parlatore, Cecilia (2015). "Fragility in Money Market Funds: Sponsor Support and Regulation." Working Paper.

Pan, Kevin and Yao Zeng (2015). "Arbitrage under Liquidity Mismatch: Theory and Evidence from the Authorized Participants of Bond ETFs." Working Paper.

Pastor, Lubos and Robert Stambaugh (2003). "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy*, 111: 642-685.

Peck, James and Karl Shell (2003). "Equilibrium Bank Runs." *Journal of Political Economy*, 111: 103-123.

Postlewaite, Andrew and Xavier Vives (1987). "Bank Runs as An Equilibrium Phenomenon." *Journal of Political Economy*, 95: 485-491.

Schmidt, Lawrence, Allan Timmermann and Russ Wermers (2014). "Runs on Money Market Mutual Funds." Working Paper.

Schroth, Enrique, Gustavo Suarez and Lucian Taylor (2014). "Dynamic Debt Runs and Financial Fragility: Evidence from the 2007 ABCP Crisis." *Journal of Financial Economics*, 112: 164-189.

Shek, Jimmy, Ilhyock Shim and Hyun Song Shin (2015). "Investor Redemptions and Fund Manager Sales of Emerging Market Bonds: How Are They Related?" Working Paper.

Shleifer, Andrei and Robert Vishny (1992). "Liquidation Values and Debt Capacity: A Market Equilibrium Approach." *Journal of Finance*, 47: 1343-1366.

Shleifer, Andrei and Robert Vishny (1997). "The Limits of Arbitrage." *Journal of Finance*, 52: 35-55.

Simutin, Mikhail (2013). "Cash Holdings and Mutual Fund Performance." *Review of Finance*, 18: 1425-1464.

Sirri, Erik and Peter Tufano (1998). "Costly Search and Mutual Fund Flows." *Journal of Finance*, 53: 1589-1622.

Spiegel, Matthew and Hong Zhang (2013). "Mutual Fund Risk and Mutual Share-Adjusted Fund Flows." *Journal of Financial Economics*, 108: 506-528.

Stein, Jeremy (2005). "Why Are Most Funds Open-End? Competition and the Limits of Arbitrage." *Quarterly Journal of Economics*, 120: 247-272.

Stein, Jeremy (2014). "Comments on 'Market Tantrums and Monetary Policy,' a paper by Michael Feroli, Anil Kashyap, Kermit Schoenholtz, and Hyun Song Shin." Remarks at the US Monetary Policy Forum.

Stokey, Nancy and Robert Lucas (1989). *Recursive Methods in Economic Dynamics*. Harvard University Press.

STRAHAN, PHILIP and BASAK TANYERI (2014). "Once Burned, Twice Shy: Money Market Fund Responses to A Systemic Liquidity Shock." *Journal of Financial and Quantitative Analysis*, forthcoming.

STREBULAEV, ILYA and TONI WHITED (2012). "Dynamic Models and Structural Estimation in Corporate Finance." *Foundations and Trends in Finance*, 6: 1-163.

SUN, YENENG (2006). "The Exact Law of Large Numbers via Fubini Extension and Characterization of Insurable Risks." *Journal of Economic Theory*, 126: 31-69.

SUNDARESAN, SURESH and ZHENYU WANG (2014). "Bank Liability Structure." Working Paper.

VIVES, XAVIER (2014). "Strategic Complementarity, Fragility, and Regulation." *Review of Financial Studies*, 27: 3547-3592.

WANG, JEFFREY (2015). "Asset Managers and Financial Instability: Evidence of Run Behavior and Run Incentives in Corporate Bond Funds." Working Paper.

WEILL, PIERRE-OLIVIER (2007). "Leaning Against the Wind." *Review of Economic Studies*, 74: 1329-1354.

WERMERS, RUSS (2000). "Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transaction Costs, and Expenses." *Journal of Finance*, 55: 1655-1695.

WILLIAMSON, OLIVER (1988). "Corporate Finance and Corporate Governance." *Journal of Finance*, 43: 567-591.

WOODFORD, MICHAEL (1990). "Public Debt as Private Liquidity." *American Economic Review*, 80: 382-388.

Chapter 2

Investment Exuberance under Cross Learning[1]

2.1 Introduction

Industry-wide investment exuberance has been commonly observed in history, especially after the arrival of major innovations that involve large uncertainty.[2] Existing theories usually focus on boom-bust cycles, over-investments and price bubbles whereas ignore the second-moment implications of investment exuberance: high systematic risks. Specifically, in exuberant industries, a firm's real investment and asset price comove greatly with those of other firms, as well as with the market (see Rhodes-Kropf, Robinson and Viswanathan, 2005, Pastor and Veronesi, 2009, Hoberg and Phillips, 2010, Bhattacharyya and Purnanandam, 2011, Greenwood and Hanson, 2014, for recent empirical evidence in different contexts, and the herding literature, for example, Scharfstein and Stein, 1990, for early rational theories). These joint comovements are hard to be explained by fundamental shocks. More surprisingly, more competitive industries exhibit more inefficient investment exuberance with higher systematic risks, in the sense that their investment exuberance is more predictable and their firms' real and financial performances also comove more (Hoberg and Phillips, 2010, Greenwood and Hanson, 2014). Our paper provides a new rational amplification theory that helps unify these observations of industry-wide investment exuberance.

---

[1] Co-authored with Shiyang Huang.

[2] The most typical examples include the "railway mania" of the UK in the 1840s, the rapid development of automobiles and radio in the 1920s, and the surge of the Internet in the 1990s, among many others. Other notable examples include major financial innovations like asset-backed securities (ABS) and credit default swaps (CDS), as well as the Mississippi Scheme and the South Sea Bubble, in which market structures experienced dramatic changes.

Our theory features a natural fact, firm cross learning, meaning that firms learn from other firms' asset prices (in addition to their own asset prices) in making investment decisions.[3] Although firm cross learning has been explicitly documented by recent empirical work, Foucault and Fresard (2014), Ozoguz and Rebello (2013), and more broadly Graham and Harvey (2001),[4] it is overlooked in the theoretical literature. Our model attempts to fill this gap by providing a micro-foundation for firm cross learning and more importantly exploring its aggregate implications.

Our theory builds on the burgeoning literature that highlights the feedback from asset prices to investment decisions (see Bond, Edmans and Goldstein, 2012, for a survey). Specifically, since secondary market participants may have incremental information that is unavailable to firms or primary market participants, firms or their capital providers may learn from the asset prices in the secondary markets for making investment decisions, and this in turn affects the asset prices (see Luo, 2005, Chen, Goldstein, and Jiang, 2007, Bakke and Whited, 2010, Edmans, Goldstein, and Jiang, 2012, Foucault and Fresard, 2012, among many others, for the empirical evidence). The feedback literature, however, has not explored the multi-firm context and the cross-learning mechanism we emphasize which help generate and amplify industry-wide investment exuberance.

---

[3] This definition of cross learning differs from the observational learning literature (for example, Banerjee, 1992, Bikhchandani, Hirshleifer, and Welch, 1992), which suggests that agents may directly observe other agents' actions and then take similar actions. We do not deny the importance of direct observational learning and spill-over in the firm investment setting, but we put it aside in order to highlight our cross learning channel, which is less understood.

[4] In the survey by Graham and Harvey (2001), CFOs of firms report that they tend to rely on other firms' prices in making capital budgeting decisions, and this in turn affects CEOs' investment decisions.

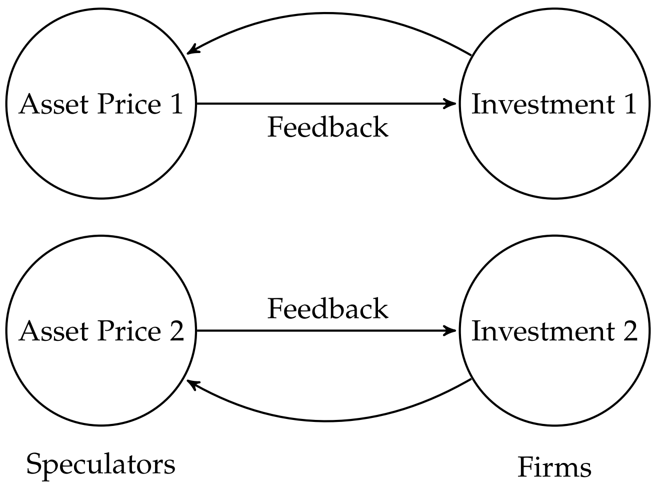


Figure 2.1: *Self-Feedback Benchmark*

To address industry-wide investment exuberance and cross learning, we construct a feedback model to admit multiple firms. We postulate that firms in an industry (or an economy) have correlated investment opportunities, so that they have incentives to learn from each other's asset prices. To clarify the idea, Figure 1 depicts the existing feedback models without this feature, even if they can literally accommodate many firms. Although these firms can learn from their own asset prices for making better investment decisions, they are essentially separated in segmented economies, in turn making other firms' asset prices irrelevant. This is the reason that the existing feedback models usually feature one representative firm or one single asset.[5]

Instead, the novelty of our work is developing a tractable multi-firm model featuring two-way cross learning of firms from other firms' asset prices (in addition to their own), and identifying a new pecuniary externality involved. Inspired by the two-factor feedback

---

[5] One exception is Subrahmanyam and Titman (2013), in which a private firm learns from the stock price of another public firm to make investment decision. The private firm's investment affects the profitability of the public firm through competition, which further generates interesting macroeconomic implications. But the public firm does not invest by itself and the private firm also does not have its own asset price. Hence, their model still features the standard feedback channel as shown in Figure 1. Their formal model also admits two private firms, which introduces an additional externality in terms of investment complementarity that amplifies their feedback effect. But as the authors have claimed, the introduction of two private firms is not essential for most of their results.

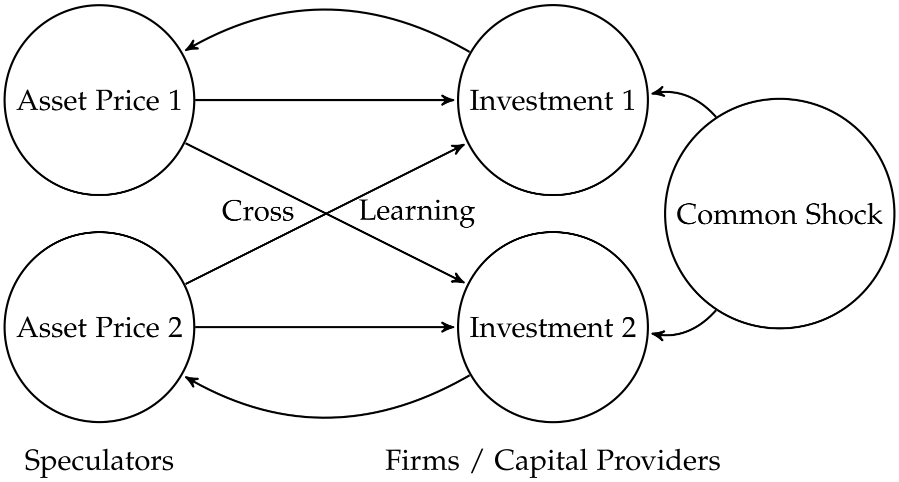


Figure 2.2: *Firm / Capital Provider Cross Learning*

framework introduced by Goldstein and Yang (2014a,b), we explicitly model correlated investment opportunities in our multi-firm setting. Specifically, the fundamental of each firm's asset is subject both to a common productivity shock (an industry shock)[6] and an idiosyncratic shock (a firm-specific shock). As speculators trade firms' assets, the prices aggregate their private information about the two types of shocks. Therefore, for each firm, other firms' asset prices become noisy but informative signals about the common shock. Consequently, the firm in question uses other firms' asset prices (and its own) to gain more knowledge about the common shock and make investment decisions accordingly, while other firms do the same. Such cross learning makes firms' investments more sensitive to the common shock, encouraging secondary market speculators to weight information about the common shock more. This in turn makes firms' asset prices more informative about the common shock, resulting in an even higher investment sensitivity to the common shock. Eventually, a tiny shock can be amplified significantly.[7] Moreover, when one firm learns from other firms' asset prices, it does not internalize a negative pecuniary externality where the prices become less informative about other firms' idiosyncratic shocks due to the speculators' endogenous over-weight on the information about the common shock (and thus under-weight on the information about the idiosyncratic shocks). This externality leads to high investment inefficiency. Interestingly, the new pecuniary externality takes effect through the informativeness rather than the level of prices. Figure 2 illustrates the idea of cross learning and contrasts it to the existing feedback models. Empirically, firm cross learning has been documented by recent studies like Foucault and Fresard (2014)[8] and Ozoguz and Rebello (2013), and the magnitude is shown to be considerable, serving as a strong support to our theory.

---

[6] In a broader sense, our common shock can also be interpreted as a shock to the entire economy. Hence, our model speaks to not only industry-wide investment exuberance but also more broadly economy-wide investment exuberance.

[7] This mechanism resembles rational herding in financial markets (see Scharfstein and Stein, 1990 for an example) and more closely the market-based two-task corporate investment setting in Aghion and Stein (2008), but we explicitly model learning from prices and our cross-learning mechanism does not rely on any form of

The predictions of our model are consistent with many empirical regularities. Compared with a benchmark in which firms cannot learn from others' asset prices, cross learning results in a higher weight of speculators on the information of the common shock in trading and firms' higher investment sensitivity to the common shock.[9] We further show that under cross learning (plus learning from own prices), a firm's investment and price comove more greatly with 1) other firms' investments and prices, and 2) with the common productivity shock. We interpret these patterns as amplification of systematic risks. Since the cross-learning mechanism relies on observable prices in public financial markets, it is further supported by Maksimovic, Phillips and Yang (2013)'s empirical findings that systematic risks are stronger among public firms than private ones. To the best of our knowledge, our work is the first to show that firms' two-way cross learning has significant aggregate

---

short-termism.

[8] For the purpose of developing empirical hypotheses, Foucault and Fresard (2014) build a model, featuring one-way learning: one focal firm may learn from its peer firm's price while not the other way around, and the peer firm does not invest. That model lays out a nice foundation for their empirical analysis. However, it does not generate inefficient multi-firm investment exuberance or comovements with the common productivity shock as emphasized in our paper. The setup and mechanisms of their model are also different from ours.

[9] See Peng, Xiong, and Bollerslev (2007), and more broadly Rhodes-Kropf, Robinson and Viswanathan (2005), Hoberg and Phillips (2010), Bhattacharyya and Purnanandam (2011) for empirical evidence. Complementary to another related theoretical literature that highlights investors' higher attention to the common shock (see Veldkamp, 2006, Peng and Xiong, 2006), our work further speaks to its endogenous origin from firm cross learning as well as its real consequences on firms' investment decisions.

amplification effects on investment, prices, and systematic risks.

We investigate many circumstances in which varying economic conditions amplify systematic risks through the cross-learning mechanism. First, an increasing uncertainty on the common shock, most typically associated with the introduction of major innovations, leads to more weight on the information of the common shock and amplification of systematic risks. This is consistent with the messages in Brunnermeier and Nagel (2004) and Pastor and Veronesi (2006, 2009) that systematic risks are too high to be explained by fundamental shocks during investment exuberance like the so-called Nasdaq bubble. This amplification also fits in line with the broader evidence of investment exuberance in Rhodes-Kropf, Robinson and Viswanathan (2005), Hoberg and Phillips (2010) and Bhattacharyya and Purnanandam (2011) in various contexts. Second, an improvement of capital providers' knowledge about the common shock leads to higher systematic risks. Last, lower market liquidity or higher variances of idiosyncratic non-fundamental trading activities also lead to higher systematic risks. These empirical regularities have frequently been ascribed to separate behavioral explanations in the previous literature, whereas our work provides a unified rational explanation.

Furthermore, our framework offers a novel perspective to look at the relationship between inefficient investment exuberance and industrial competition. This relationship is particularly helpful to separate aggregate effects specific to cross learning from other potential mechanisms. Due to the unaligned interests of firms and speculators in feedback and the new pecuniary externality associated with cross learning, the investment exuberance is inefficient. In particular, we show that, when the number of firms in an industry increases, that is, when the industry potentially becomes more competitive, the extent of cross learning becomes stronger, leading to more severe pecuniary externalities. This suggests a rationale for the puzzling facts identified in Hoberg and Phillips (2010) and Greenwood and Hanson (2014) that investment exuberance is more predictable and firms' investment (and price) comovements are greater in more competitive industries. According to Hoberg and Phillips (2010), no single existing theory can accommodate their findings. Our cross-learning mechanism with the new pecuniary externality implies that firms and investors in more competitive industries are more likely to overweight common shock whereas to underweight idiosyncratic shocks, leading to more inefficient investment exuberance with higher systematic risks. Such industry-level heterogeneity helps differentiate the cross-learning mechanism from other potential ones that may also generate investment exuberance. In this regard, Ozoguz and Rebello (2013) have identified that firms in more competitive industries have higher investment sensitivity to stock prices of their peers, which directly supports our predictions.

Fundamentally, the amplification effect of firm cross learning stems from two-way endogenous strategic complementarities across the primary and secondary markets, which are absent in the previous literature. Unlike the existing literature involving complementarities in financial markets (see Veldkamp, 2011, for an extensive review), our mechanism does not posit any exogenous complementarity in actions or externality but a well-documented natural fact that firms learn from their own and other firms' asset prices.

Related Literature. Our framework contributes to the burgeoning feedback theories (see Bond, Edmans and Goldstein, 2012, for a survey).[10] Most related to the present approach are Goldstein, Ozdenoren and Yuan (2013), Goldstein and Yang (2014a,b), and Sockin and Xiong (2014a,b), all of which highlight the feedback from secondary market speculators' information aggregation to primary market capital providers' scale-varying investment in various contexts. In particular, pioneering the two-factor feedback framework, Goldstein and Yang (2014a) uncover new relationship between market efficiency and real efficiency, and Goldstein and Yang (2014b) deliver new implications regarding the effectiveness of disclosure policies. Our contribution is to develop a two-way cross-learning framework with multiple firms, which generates the new pecuniary externality and various aggregate implications regarding inefficient investment exuberance and competition. Methodologically, our approach is highly tractable despite the complexity of the equilibrium, offering a framework for future research involving market feedback with multiple firms.

---

[10] For instance, Dow and Gorton (1997) on market and real efficiencies, Fulghieri and Lukin (2001) on security design, Subrahmanyam and Titman (2001, 2013) on information cascades and shock propagation, Hirshleifer, Subrahmanyam and Titman (2006) on irrational trading, Foucault and Gehrig (2008) on cross listing, Goldstein and Guembel (2008) on manipulative short selling, Ozdenoren and Yuan (2008) on excess volatility, Edmans (2009) on blockholders trading, Bond, Goldstein and Prescott (2010) on corrective managerial actions, among many others.

Our work also contributes to the literature of rational models on investment exuberance by exploring new implications.[11] In particular, this literature has been paying increasing attention to the role of learning in financial markets.[12] In the rational learning model of Rhodes-Kropf and Viswanathan (2004), which shares a similar signal extraction problem with ours, managers cannot differentiate between common misevaluation and possible idiosyncratic synergies, leading to herding and merger and acquisition exuberance. Pastor and Veronesi (2009) propose an explicit learning model, in which the uncertain productivity of a new technology is subject to learning. Learning and the ensuing technology adoption make the uncertainty from idiosyncratic to systematic, generating rational herding and investment exuberance. Our contribution to this literature is three-fold. First, our model features multiple firms and their cross learning explicitly. This allows us to study the second-moment implications of industry-wide investment exuberance, which is less understood. Second, we cast the microstructure of public asset markets explicitly, ensuring us to reflect the indispensable role of public financial markets as suggested by Maksimovic, Phillips and Yang (2013). Lastly, our model identifies a new externality regarding the use of information about common shocks and idiosyncratic shocks in making investment decisions.

Identifying the externality associated with cross learning contributes to the large pecuniary externality literature.[13] The classical pecuniary externality takes effect through the level of prices: under various frictions agents do not internalize the impacts of their actions on equilibrium price levels, leading to a welfare loss. Instead in our multi-firm cross-learning framework, firms that make real investment decisions do not fully internalize the impact of cross learning on equilibrium price informativeness. This leads to a "tragedy of the commons" regarding the use of the price system as an information source. In this sense, our pecuniary externality is reminiscent of the learning externality in the early dynamic learning and herding literature (for example, Vives, 1997) that an agent, when responding to her private information, does not take into account the benefit of increased informativeness of public information in the future.[14]

---

[11] The bubble literature is the first to provide an explanation for investment exuberance in the time series (see Brunnermeier and Oehmke, 2013, Xiong, 2013, for surveys of various models and evidence). These theories have focused on the over-investment or over-valuation of one single firm, and have often referred to behavioral aspects. The modern literature of macro-finance (see Brunnermeier, Eisenbach and Sannikov, 2013, for an extensive survey) also generates various forms of over-investment, over-borrowing, and over-lending in the time series, by highlighting agency or financial frictions. The focuses of those literatures are mostly on the time series and thus different from ours. Our model is not intended as a general dynamic theory of booms and busts.

[12] Other notable rational models on investment exuberance include DeMarzo, Kaniel, and Kremer (2007), featuring endogenous relative wealth concerns, Carlson, Fisher and Giammarino (2004), featuring investments as real options, and Ozdenoren and Yuan (2013), featuring contractual externalities.

[13] See Stiglitz (1982), Greenwald and Stiglitz (1986), and Geanakoplos and Polemarchakis (1985).

Our work is also related to the literature on the interaction across different asset markets, in particular the models that highlight learning. This literature has focused on speculator learning rather than firm cross learning as we model. Cespa and Foucault (2014) consider the contagion of illiquidity across segmented markets by introducing a concept of cross-asset learning among speculators. By cross-asset learning, speculators trading in one market can potentially learn from the asset price in another market, which in turn generates propagation. Goldstein and Yang (2014c) model an environment in which different speculators are informed of different fundamentals affecting one single asset. Trading on information about the two fundamentals exhibits complementarity, suggesting that greater diversity of information improves price informativeness. Our model complements those papers by focusing on the implications of firm cross learning on both real investments and asset prices, in contrast to their exchange economy setting that focuses on trading.

Finally, our framework is broadly related to a macroeconomic literature focusing on dispersed information and boom-bust cycles. Closely related are Angeletos, Lorenzoni and Pavan (2012) on the role of beauty contest and Amador and Weill (2010) on the crowding-out effect of exogenous public information provision to the use of private information.

---

[14] The recent study of Vives (2014) further combines the classical pecuniary externality (through the level of prices) and the learning externality associated with exogenous public information in an industrial competition context. This is different from our new pecuniary externality through endogenous price informativeness on the two shocks. Its focus is also on the strategic interaction in product markets instead of our endogenous cross learning in financial markets.

Compared with Angeletos, Lorenzoni and Pavan (2012), our cross-learning framework with more detailed financial market structures can generate the new second-moment implications of investment exuberance. Our externality is also different from theirs, which instead features beauty contest in signaling and higher-order uncertainties. Complementary to Amador and Weill (2010), the externality in our model derives from a different mechanism and suggests a new crowding-out effect: endogenous over-weight on the common shock crowds out the use of information about the idiosyncratic shocks. Relatedly, Liu (2014) has embedded a single-firm feedback and disclosure mechanism (in the spirit of Goldstein and Yang, 2014b,c) into a macroeconomic setting. None of the papers aforementioned have distinguished between common and idiosyncratic shocks or considered firm cross learning.

Section 3.2 lays out the model, featuring correlated investment opportunities and cross learning. Section 2.3 characterizes the cross-learning equilibrium and benchmarks it to the self-feedback case. Section 2.4.2 investigates important implications of cross learning with a focus on systematic risks in investment exuberance. Section 3.4.2 explores the externality and the relationship between investment inefficiency and competition. Section 2.6 discusses some extensions of the model. All the proofs are delegated to Appendix unless otherwise noted.

2.2 The Model

2.2.1 The Economy

Inspired by Goldstein, Ozdenoren and Yuan (2013) and Goldstein and Yang (2014a,b), we build a new multi-firm feedback model to study industry-wide investment exuberance. To focus on the second moments and inefficiency of investment exuberance, which are less understood in the literature, we set aside a full dynamic investigation of the boom-bust cycles.

We consider a continuum of 1 of firms, $i \in [0,1)$, each having an asset traded in a secondary market. Each firm i's corresponding asset market is occupied by a mass 1 of exclusive informed risk-neutral speculators and uninformed noisy traders. We index speculators for firm i by (i, j), with $j \in [0, 1)$.[15] Each firm also has an exclusive capital provider i who decides how much capital to provide to the firm for investment purposes.

There are three dates, $t = 0, 1, 2$. At date 0, the speculators trade in their corresponding asset market based on their private information, and the asset price aggregates their information. At date 1, the capital providers observe the asset prices of both their own firm and all the other firms. Having observed all the prices and received their private information, the capital providers decide the amount of capital to provide to their corresponding firms and the firms undertake investment accordingly. All the cash flows are realized at date 2.

2.2.2 Capital Providers and Investment

All the firms in the economy have an identical linear production technology: $Q(I_i) = AF_iI_i$, where I_i is the amount of capital provided by capital provider i to firm i, and A and F_i are two stochastic productivity shocks. Specifically, shock A captures an industry-wide common productivity shock, and shock F_i captures an idiosyncratic productivity shock for firm i only. Denote by a and f_i the natural logs of these shocks, and assume that they are normal and mutually independent:

$$a \sim N(0, 1/\tau_a)\,, \text{ and } f_i \sim N(0, 1/\tau_f)\,,$$

where τ_a and τ_f are positive and $i \in [0, 1)$.

The introduction of multiple firms and of the two different productivity shocks plays an important role in necessitating firm cross learning. Specifically, if the investment opportunities are uncorrelated, cross learning does not make sense. On the other hand, if the investment opportunities are perfectly correlated, all assets become identical and thus it is less interesting to consider firm cross learning. To flesh the cross-learning mechanism out, we intentionally abstract away from any possible industrial structure of the firms' product market.[16]

---

[15] Since the speculators do not have a diversification motive, our results are unaffected if we assume that they can trade all assets. In other words, market segmentation regarding speculator trading plays no roles in our model.

At date 1, all the capital providers choose the amount of capital I_i simultaneously in their respective primary markets. Capital provider i captures a proportion $\kappa \in (0,1)$ of the output $Q(I_i)$ by providing I_i, which incurs a private quadratic adjustment cost, $C(I_i) = \frac{1}{2}cI_i^2$.[17] Thus, capital provider i's problem at $t = 1$ is

$$\max_{I_i} \mathbb{E}\left[\kappa A F_i I_i - \frac{1}{2}cI_i^2 | \Gamma_i\right], \tag{2.2.1}$$

where Γ_i is the information set of capital provider i at $t = 1$. It consists of the price of firm i's own asset, P_i, and those of all the other firms' assets, denoted by the set $\{P_{-i}\}$ for brevity, formed at date 0, as well as their private signals about the (log) productivity shocks a and f_i. Specifically, we assume that each capital provider i gets a private noisy and independent signal $s_{a,i}$ about the (log) common productivity shock a with precision τ_s, and another private noisy and independent signal $s_{f,i}$ about its own (log) idiosyncratic productivity shock f_i with precision τ_f:

$$s_{a,i} = a + \varepsilon_{a,i}\,, \text{ where } \varepsilon_{a,i} \sim N(0, 1/\tau_{sa})\,, \text{ and}$$

$$s_{f,i} = f_i + \varepsilon_{f,i}\,, \text{ where } \varepsilon_{f,i} \sim N(0, 1/\tau_{sf})\,.$$

Hence, for capital provider i, the information set is $\Gamma_i = \{P_i, \{P_{-i}\}, s_{a,i}, s_{f,i}\}$.

Different from existing literature, one major novelty of our setup is to allow capital providers to learn from other firms' asset prices as well as own firms' prices, which we formally call cross learning. As highlighted later, although the capital providers only care about their own firms, they use the prices of other firms' assets for making better investment decisions.

---

[16] However, in Session 3.4.2, we extend the baseline model to admit a finite number of n firms and use the number of firms to proxy the extent of competition. We discuss the plausibility of doing this in greater detail in Session 3.4.2.

[17] The objective function of the capital providers is inherited from Goldstein, Ozdenoren and Yuan (2013), where the capital providers are interpreted as banks, private managers, or large shareholders of the firms. The specification helps ensure the tractability of the model.

2.2.3 Speculators and Secondary Market Trading

At date 0, the remaining cash flow $(1-\kappa)Q(I_i)$, as an asset, is traded in a separate competitive secondary market for each firm i. For firm i, denote the price of this asset by P_i. To focus on capital providers' cross learning, we do not consider any possible monetary transfers from the secondary market to the firm, but instead highlight the information revealed in the secondary market trading.[18] In the asset market of firm i, each speculator (i,j) has two private and independent signals about the common shock and the respective idiosyncratic shock. Specifically, the first signal is about the common shock:

$$x_{ij} = a + \varepsilon_{x,ij}, \text{ where } \varepsilon_{x,ij} \sim N(0, 1/\tau_x)\,,$$

and the second signal is about the firm-specific idiosyncratic shock:

$$y_{ij} = f_i + \varepsilon_{y,ij}, \text{where } \varepsilon_{y,ij} \sim N(0, 1/\tau_y)\,.$$

In other words, the information set of speculator (i,j) is $\Gamma_{ij} = \{x_{ij}, y_{ij}\}$. During trading, their private information will be aggregated into asset prices.

Based on their private information, the speculators submit limited orders (so that their information set does not consist of the asset prices),[19] with an additional constraint that each speculator can buy or sell up to a unit of the asset.[20] Formally, the speculators maximize their expected trading profit given the market price. Their problems at $t=0$ are

$$\max_{d_{ij}\in[-1,1]} d_{ij}\mathbb{E}\left[(1-\kappa)AF_iI_i - P_i|\,\Gamma_{ij}\right], \tag{2.2.2}$$

where d_{ij} is speculator (i,j)'s demand. The aggregate demand in market i is given by

---

[18] Hence, this asset can be interpreted as either equity of the firm or a derivative on the return from the firm's investment.

[19] The fact that the speculators submit limited orders is not essential. For any firm i, even if its speculators can submit market orders, all of our results are unaffected. See Goldstein and Yang (2014b) for another exposition of this point in a similar framework.

[20] As discussed in the survey by Brunnermeier and Oehmke (2013), the specific size of this position limit is not essential for the results, as long as speculators cannot take unlimited positions, otherwise the prices would be fully revealing. This constraint can be easily justified by their capital or borrowing constraints.

$D_i = \int_0^1 d_{ij} \mathrm{d}j$.

We assume that the noisy supply in asset market i takes the following form:[21]

$$\Delta(\zeta, \xi_i, P_i) = 1 - 2\Phi(\zeta + \xi_i - \lambda \log P_i)\,,$$

where

$$\zeta \sim N(0, \tau_\zeta^{-1})\,, \text{ and } \xi_i \sim N(0, \tau_\xi^{-1})\,.$$

We elaborate the noisy supply. $\Phi(\cdot)$ denotes the cumulative standard normal distribution function. The first term ζ captures a common noisy supply shock that can be viewed as industry-wide sentiment or industry-wide fund flow. The second term ξ_i captures the idiosyncratic noisy supply shock in market i that can be viewed as market-specific liquidity shocks. The presence of a common noisy supply not only makes our framework more general, but more importantly prevents the aggregate price from fully revealing the common productivity shock. Both noisy supply shocks ζ and ξ_i are independent and also independent of other shocks in the economy. Meanwhile, λ in the noisy supply function captures price elasticity and can be viewed as market liquidity. When λ is high, the demand from speculators can be easily absorbed and thus their aggregate demand has little impact on the asset prices.

Finally, in equilibrium, the prices clear each asset market by equalizing the aggregate speculator demand to the noisy supply in each asset market i:

$$D_i = \Delta(\zeta, \xi_i, P_i)\,. \tag{2.2.3}$$

2.2.4 Discussion

Before proceeding, we discuss several crucial differences of our settings from the existing feedback literature. First, to lay out a foundation for cross learning, our model features a continuum of many firms. Modeling multiple firms and two-way cross learning imposes new technical challenges in terms of finding closed form solutions. Despite this difficulty, our model provides a tractable approach not only suited for our purpose but potentially useful for future work in other directions.

---

[21] This functional form resembles those used in Hellwig, Mukherji and Tsyvinski (2006) and Dasgupta (2007).

Second, built upon the multiple-firm setup, our economy features two different productivity shocks. In most existing papers, there is only one productivity shock. Goldstein and Yang (2014a,b) first introduce two factors to the cash flow in feedback models. Their two factors differ in an informational sense that the capital providers perfectly observe one but not the other. Instead, in the multi-firm setting, our two shocks differ in a productivity-related sense that one is common to all firms while the other is firm-specific. The presence of such two different shocks helps deliver new aggregate implications regarding firm cross learning and investment.

Finally, in contrast to the theory for hypotheses development in Foucault and Fresard (2014), our framework fully features two-way cross learning instead of one-way learning by a focal firm from its peer firm. The one-way learning channel in Foucault and Fresard (2014) gives clear predictions on how the peer firm's stock price may affect the focal firm's investment, but the peer firm itself does not invest or learn. Our framework with two-way cross learning as well as more detailed real and financial market structures captures the new strategic complementarities and spiral toward the common shock. This new mechanism is crucial in generating industry-wide inefficient investment exuberance, which are consistent with the empirical regularities.

2.3 Cross-Learning Equilibrium

2.3.1 Equilibrium Definition

We formally introduce the equilibrium concept. We focus on symmetric linear equilibria that are commonly used in the literature. Specifically, the speculators in market i long one share of the corresponding asset when $\phi_i x_{ij} + y_{ij} > \mu_i$, and short one share otherwise, where ϕ_i and μ_i are two constants to be determined in equilibrium. Since agents are risk neutral and firms are symmetric in our framework, symmetry further implies that $\phi_i = \phi$ and $\mu_i = \mu$, which mean that all the speculators use symmetric trading strategies in all asset markets, and the information contents of all asset prices are also symmetric.

Definition 3. *A (symmetric) cross-learning equilibrium is defined as a combination of a price function for each firm i,* $P_i(a, f_i, \zeta, \xi_i)$*:* $\mathbb{R}^4 \to \mathbb{R}$*, an investment policy for each capital provider i,* $I_i(s_{a,i}, s_{f,i}, P_i, \{P_{-i}\})$*:* $\mathbb{R}^2 \times \mathbb{R}^\infty \to \mathbb{R}$*, and a linear monotone trading strategy for each speculator* (i, j)*,* $d_{ij}(x_{ij}, y_{ij}) = \mathbf{1}(\phi_i x_{ij} + y_{ij} > \mu_i) - \mathbf{1}(\phi_i x_{ij} + y_{ij} \leqslant \mu_i)$*, such that*

i) each capital provider i's investment policy $I_i(s_{a,i}, s_{f,i}, P_i, \{P_{-i}\})$ *solves problem (2.2.1),*

ii) each speculator (i, j)*'s trading strategy* $d_{ij}(x_{ij}, y_{ij})$ *is identical and solves problem (2.2.2), and*

iii) market clearing condition (2.2.3) is satisfied for each market i.

2.3.2 Equilibrium Characterization

We characterize the equilibrium, featuring the capital providers' cross learning. We follow a step-by-step approach to streamline the presentation.

Step 1. First we solve for the price functions, which help characterize the information contents of prices from the capital providers' perspective. We have the following lemma:

Lemma 9. *The speculators' trading leads to the following equilibrium price of each asset i:*

$$P_i = \exp\left(\frac{\phi_i}{\lambda\sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}}a + \frac{1}{\lambda\sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}}f_i + \frac{\zeta + \xi_i}{\lambda} - \frac{\mu_i}{\lambda\sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}}\right). \tag{2.3.1}$$

Hence, from any capital provider i's perspective, the price for its own firm i's asset is equivalent to the following signal in predicting the common shock a:

$$z_a(P_i) = \frac{\lambda\sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}\log P_i + \mu_i}{\phi_i} = a + \frac{1}{\phi_i}f_i + \frac{\sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}}{\phi_i}(\zeta + \xi_i), \tag{2.3.2}$$

and is equivalent to the following signal in predicting the corresponding idiosyncratic shock f_i:

$$z_f(P_i) = \lambda\sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}\log P_i + \mu_i = f_i + \phi_i a + \sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}(\zeta + \xi_i). \tag{2.3.3}$$

Lemma 9 not only helps specify the information contents of a firm's asset price to its own capital provider, but it also hints those to other firms' capital providers. It suggests that capital providers would like to cross learn from each other when feasible. The next step formulates the idea.

Step 2. We then characterize the informational implications of cross learning. Specifically, we show that, when cross learning is feasible, that is, capital provider i's information set includes both P_i and $\{P_{-i}\}$, the capital provider will endogenously rely on an aggregate price as well as the target firm's own asset price (in addition to the private signals) in inferring the two productivity shocks. To formally show the results, we impose the symmetry conditions $\phi_i = \phi$ and $\mu_i = \mu$ to conditions (2.3.1), (2.3.2) and (2.3.3) as we focus on symmetric equilibria, and we also define the aggregate price as

$$\overline{P} = \int_0^1 P_i \mathrm{d}i\,.^{22}$$

Lemma **10.** *For capital provider i, when her information set includes both P_i and $\{P_{-i}\}$, these asset prices are informationally equivalent to the following two signals:*

i) a signal based on the aggregate price $\overline{P}$:

$$z_a(\overline{P}) = a + \frac{\sqrt{\tau_x^{-1}\phi^2 + \tau_y^{-1}}}{\phi}\zeta \tag{2.3.4}$$

for predicting the common shock a, with the precision

$$\tau_{\overline{pa}} = \frac{\tau_x \tau_y \tau_\zeta \phi^2}{\tau_x + \tau_y \phi^2}\,, \tag{2.3.5}$$

which is increasing in ϕ, and

ii) a signal based on the own asset price P_i as well as the aggregate price $\overline{P}$:

$$z_{f,i}(\overline{P}) = z_f(P_i) - \phi z_a(\overline{P}) = f_i + \sqrt{\tau_x^{-1}\phi^2 + \tau_y^{-1}}\xi_i \tag{2.3.6}$$

---

[22] The fact that the asset prices are equally weighted in calculating the aggregate price is inessential to their information contents. Our results carry through even if we choose arbitrarily positive weights.

for predicting the corresponding idiosyncratic shock f_i*, with the precision*

$$\tau_{\overline{pf}} = \frac{\tau_x \tau_y \tau_\xi}{\tau_x + \tau_y \phi^2}, \tag{2.3.7}$$

which is decreasing in ϕ*.*

Along with Lemma 9, Lemma 10 implies that cross learning changes the feedback channel in which a capital provider uses asset prices to infer the two productivity shocks: she now uses the aggregate price $\overline{P}$ to infer the common shock a and still uses her target firm's own price P_i to infer the idiosyncratic shock f_i. Intuitively, for capital provider i, other firms' asset prices $\{P_{-i}\}$ are uninformative about the idiosyncratic shock f_i but informative about the common shock a. Hence, when other firms' asset prices are observable, which is natural in reality, the capital provider of the firm in question uses them to make better inference about the common shock. In particular, in a symmetric equilibrium, all prices are symmetric regarding their information contexts, so that the aggregate price is equally weighted. By the law of large numbers, $\overline{P}$ only aggregates information about the common shock a: all the information about idiosyncratic shocks and the idiosyncratic noisy supply shocks gets wiped out, while the presence of the common noisy supply shock still prevents the aggregate price from fully revealing the fundamental. This makes $z_a(\overline{P})$, as characterized in (2.3.4), the most informative signal about the common shock a. Moreover, knowing $z_a(\overline{P})$, the capital provider also eliminates the information about the common shock and about the common noisy supply shock when she uses her own price P_i to infer the idiosyncratic shock f_i, as characterized in (2.3.6).

Step 3. We then solve for the capital providers' optimal investment policy under cross learning. This indicates the real consequences of cross learning. Lemma 10 implies that, under cross learning, capital provider i uses the new signal $z_a(\overline{P})$ and her private signal $s_{a,i}$ to infer the common shock a, and the new signal $z_{f,i}(\overline{P})$ and the private signal $s_{f,i}$ to infer the idiosyncratic shock f_i. Thus, we have the following lemma.

Lemma 11. *Observing $s_{a,i}, s_{f,i}, P_i$ and $\{P_{-i}\}$, capital provider i's optimal investment policy is*

$$I_i = \frac{\kappa}{c} \exp\left[\frac{\tau_{sa}s_{a,i} + \tau_{\overline{pa}} z_a(\overline{P})}{\tau_a + \tau_{sa} + \tau_{\overline{pa}}} + \frac{1}{2(\tau_a + \tau_{sa} + \tau_{\overline{pa}})} + \frac{\tau_{sf}s_{f,i} + \tau_{\overline{pf}} z_{f,i}(\overline{P})}{\tau_f + \tau_{sf} + \tau_{\overline{pf}}} + \frac{1}{2(\tau_f + \tau_{sf} + \tau_{\overline{pf}})}\right]. \tag{2.3.8}$$

The investment policy is intuitive. First, the optimal amount of investment is higher when the share κ of capital provider is higher while lower when the investment cost c is higher. Moreover, the capital providers infer the two productivity shocks a and f_i independently but simultaneously in making investment decisions, reflected in the first and third terms in the parenthesis. In particular, the capital providers find it optimal to learn from both their target firms' own asset prices as well as other firms' prices, which are summarized in the two new signals $z_a(\overline{P})$ and $z_{f,i}(\overline{P})$. This fits quite in line with the recent empirical facts about firm and capital provider cross learning behavior (Foucault and Fresard, 2014, Ozoguz and Rebello, 2013).

According to Lemma 11, we propose the following intuitive concept of investment sensitivity to capture how the capital providers' investment decision responds to the two productivity shocks under cross learning.

Definition 4. *For capital providers, the investment sensitivity to the common productivity shock and the idiosyncratic productivity shock are defined as:*

$$S_a(\tau_{\overline{pa}}) = \frac{\tau_{sa} + \tau_{\overline{pa}}}{\tau_a + \tau_{sa} + \tau_{\overline{pa}}}, \text{ and } S_f(\tau_{\overline{pf}}) = \frac{\tau_{sf} + \tau_{\overline{pf}}}{\tau_f + \tau_{sf} + \tau_{\overline{pf}}},$$

respectively. We call S_a the common investment sensitivity and S_f the idiosyncratic investment sensitivity.

We highlight that the investment sensitivity depends not only on the capital providers' private signals about the corresponding shock, but also on the new endogenous price signals coming from cross learning as characterized in Lemma 10. In particular, these two notions of investment sensitivity are increasing functions of $\tau_{\overline{pa}}$ and $\tau_{\overline{pf}}$, respectively, which are in turn affected by the speculators' trading strategy. Hence, by Lemma 10, we have the following lemma that bridges the capital providers' investment sensitivity and the speculators' weight

ϕ on the signal of the common productivity shock.

Lemma 12. *The common investment sensitivity $S_a(\tau_{\overline{pa}})$ is increasing in ϕ while the idiosyncratic investment sensitivity $S_f(\tau_{\overline{pf}})$ is decreasing in ϕ.*

Lemma 12 is helpful because it offers an intuitive look at the real consequences of cross learning from asset prices in the economy with two productivity shocks. When the weight ϕ is higher, the speculators put more weight on the information about the common shock, and thus asset prices become more informative about the common shock while less informative about the idiosyncratic shocks. This in turn leads to a more sensitive investment policy in response to the common shock while less sensitive in response to the idiosyncratic shock.

Step 4. We finally close the model by solving for the speculators' equilibrium trading strategy, characterized by the weight ϕ and the constant μ. This also pins down other equilibrium outcomes since they are all functions of ϕ.

For speculator (i, j), her expected profit of trading given her available information is

$$\mathbb{E}\left[(1-\kappa)AF_iI_i - P_i | x_{ij}, y_{ij}\right] , \tag{2.3.9}$$

in which I_i and P_i have been characterized by conditions (2.3.8) and (2.3.1) respectively.

It is easy to show that speculators' expected profit (2.3.9) of trading asset i can be expressed as

$$\mathbb{E}\left[(1-\kappa)AF_iI_i - P_i | x_{ij}, y_{ij}\right] = \frac{\kappa(1-\kappa)}{c} \exp\left(\alpha_0 + \alpha_1 x_{ij} + \alpha_2 y_{ij}\right) - \exp\left(\gamma_0 + \gamma_1 x_{ij} + \gamma_2 y_{ij}\right) ,$$

where $\alpha_0, \alpha_1, \alpha_2, \gamma_0, \gamma_1$, and γ_2 are all functions of ϕ:

$$\begin{aligned}
\alpha_1 &= (S_a + 1)\frac{\tau_x}{\tau_a + \tau_x} , \\
\alpha_2 &= \left(S_f + 1\right)\frac{\tau_y}{\tau_f + \tau_y} , \\
\gamma_1 &= \frac{\phi}{\lambda\sqrt{\tau_x + \tau_y\phi^2}}\frac{\tau_x}{\tau_a + \tau_x} , \\
\gamma_2 &= \frac{1}{\lambda\sqrt{\tau_x + \tau_y\phi^2}}\frac{\tau_y}{\tau_f + \tau_y} .
\end{aligned}$$

By definition, in a symmetric cross-learning equilibrium with cross learning, we have

$$\phi = \frac{\alpha_1 - \gamma_1}{\alpha_2 - \gamma_2}.$$

Plugging in $\alpha_1, \alpha_2, \gamma_1$ and γ_2 yields

$$\phi = \frac{\left(S_a + 1 - \frac{\phi}{\lambda\sqrt{\tau_x + \tau_y \phi^2}}\right) \frac{\tau_x}{\tau_a + \tau_x}}{\left(S_f + 1 - \frac{1}{\lambda\sqrt{\tau_x + \tau_y \phi^2}}\right) \frac{\tau_y}{\tau_f + \tau_y}}. \tag{2.3.10}$$

Analyzing this equation by further plugging in S_a and S_f, which are both functions of ϕ, we reach a unique cross-learning equilibrium, formally characterized by the following proposition.

Proposition **13.** *For a high enough noisy supply elasticity λ, there exists a cross-learning equilibrium in which the speculators put a positive weight $\phi > 0$ on the signal of the common productivity shock. For a high enough information precision τ_y (of the speculators' signal on the idiosyncratic shock), the equilibrium is unique.*

Establishing the existence of a unique equilibrium is essential for our further analysis regarding investment exuberance, as it allows us to investigate that how changes in economic environment affect investments and prices through the cross-learning mechanism. When ϕ is higher, the speculators put more weight on the information about the common shock in trading, encouraging all the capital providers to respond to the common shock more sensitively through cross learning, which in turn leads to an even higher ϕ. This new spiral gives rise to many implications in line with the empirical phenomena regarding industry-wide investment exuberance as we later explore.

The conditions to guarantee a unique cross-learning equilibrium are not only commonly seen in the feedback literature but also empirically plausible. A relatively high noisy supply elasticity λ implies that markets are liquid enough. A relatively high information precision τ_y of the speculators' signal on the idiosyncratic shock suggests that asset market participants understand their target firms better than the whole industry. These two conditions are appropriate in particular when we focus on the contexts leading to investment exuberance:

relatively liquid markets and relatively more uncertain macroeconomic conditions.[23]

2.3.3 Self-Feedback Benchmark

Having established the existence and uniqueness of a cross-learning equilibrium, we compare the cross-learning equilibrium to the corresponding self-feedback equilibrium in a comparable economy. This self-feedback benchmark helps understand how the presence of cross learning affects the capital providers' investment policy and the speculators' trading strategy, in contrast to the counterfactual where cross learning is absent. To show these effects, we again focus on the difference of the speculators' weight ϕ on the signal of the common productivity shock in the two respective equilibria, as all equilibrium outcomes are functions of this weight. We still consider unique symmetric equilibrium and denote by ϕ' the speculators' weight on the signal of the common productivity shock in the self-feedback benchmark.

Formally, the only difference of the benchmark economy is that each capital provider i observes its own asset price P_i but not other firms' asset prices $\{P_{-i}\}$. In other words, capital provider i's information set is $\Gamma_i = \{P_i, s_{a,i}, s_{f,i}\}$. We have

$$P_i = \exp\left(\frac{\phi'}{\lambda\sqrt{\tau_x^{-1}(\phi')^2+\tau_y^{-1}}}a + \frac{1}{\lambda\sqrt{\tau_x^{-1}(\phi')^2+\tau_y^{-1}}}f_i + \frac{\zeta+\xi_i}{\lambda} - \frac{\mu_i}{\lambda\sqrt{\tau_x^{-1}(\phi')^2+\tau_y^{-1}}}\right),$$

which is equivalent to the following two signals

$$z_a(P_i) = \frac{\lambda\sqrt{\tau_x^{-1}(\phi')^2+\tau_y^{-1}}\log P_i + \mu_i}{\phi'} = a + \frac{1}{\phi'}f_i + \frac{\sqrt{\tau_x^{-1}(\phi')^2+\tau_y^{-1}}}{\phi'}(\zeta+\xi_i)$$

in predicting the common shock a and

$$z_f(P_i) = \lambda\sqrt{\tau_x^{-1}(\phi')^2+\tau_y^{-1}}\log P_i + \mu_i = f_i + \phi' a + \sqrt{\tau_x^{-1}(\phi')^2+\tau_y^{-1}}(\zeta+\xi_i)$$

---

[23] We have numerically shown that these conditions are not restrictive. Even for reasonably small λ and τ_y, our model still features a unique cross-learning equilibrium. These numerical results are reported in Section 3.4.2. We have also explored other sufficient conditions that can guarantee a unique cross-learning equilibrium, which are less empirically irrelevant but the results are available upon request. The cross-learning mechanism itself is unaffected under other sets of sufficient conditions.

in predicting the corresponding idiosyncratic shock f_i. The precisions of $z_a(P_i)$ and $z_f(P_i)$ are denoted as τ_{pa} and τ_{pf} where

$$\tau_{pa} = \frac{1}{\frac{1}{(\phi')^2}\tau_f^{-1} + \frac{\tau_x^{-1}(\phi')^2+\tau_y^{-1}}{(\phi')^2}(\tau_\zeta^{-1} + \tau_\xi^{-1})},$$

and

$$\tau_{pf} = \frac{1}{(\phi')^2\tau_a^{-1} + (\tau_x^{-1}(\phi')^2 + \tau_y^{-1})(\tau_\zeta^{-1} + \tau_\xi^{-1})}.$$

Following the same definition of investment sensitivity and the same analysis for the capital providers' investment policy and the speculators' trading strategy, we have

$$\begin{aligned}
S_a' &= \frac{\tau_{sa} + \tau_{pa}}{\tau_a + \tau_{sa} + \tau_{pa}} + \frac{\tau_{pf}}{\tau_f + \tau_{sf} + \tau_{pf}}\phi', \\
S_f' &= \frac{\tau_{sf} + \tau_{pf}}{\tau_f + \tau_{sf} + \tau_{pf}} + \frac{\tau_{pa}}{\tau_a + \tau_{sa} + \tau_{pa}}\frac{1}{\phi'}, \\
\alpha_1' &= \left(S_a' + 1\right)\frac{\tau_x}{\tau_a + \tau_x}, \\
\alpha_2' &= \left(S_f' + 1\right)\frac{\tau_y}{\tau_f + \tau_y}, \\
\gamma_1' &= \frac{\phi'}{\lambda\sqrt{\tau_x + \tau_y(\phi')^2}}\frac{\tau_x}{\tau_a + \tau_x}, \\
\gamma_2' &= \frac{1}{\lambda\sqrt{\tau_x + \tau_y(\phi')^2}}\frac{\tau_y}{\tau_f + \tau_y}.
\end{aligned}$$

In the self-feedback equilibrium, we also have

$$\phi' = \frac{\alpha_1' - \gamma_1'}{\alpha_2' - \gamma_2'}$$

to pin down the speculators' weight on the information of the common shock. Plugging in $\alpha_1', \alpha_2', \gamma_1'$ and γ_2' yields

$$\phi' = \frac{\left(S_a' + 1 - \frac{\phi'}{\lambda\sqrt{\tau_x+\tau_y(\phi')^2}}\right)\frac{\tau_x}{\tau_a+\tau_x}}{\left(S_f' + 1 - \frac{1}{\lambda\sqrt{\tau_x+\tau_y(\phi')^2}}\right)\frac{\tau_y}{\tau_f+\tau_y}}. \tag{2.3.11}$$

Therefore, we have the following proposition regarding the comparison between the cross-learning equilibrium and the corresponding self-feedback benchmark. We focus on comparable cases in which a self-feedback equilibrium and its corresponding cross-learning equilibrium are both unique.

Proposition **14.** *For a high enough noisy supply elasticity λ, a low enough idiosyncratic noisy supply shock precision τ_{ξ}, and a high enough information precision τ_{y} (of the speculators' signal on the idiosyncratic shock), there exists a unique self-feedback equilibrium in which speculators put a positive weight $\phi' > 0$ on the signal of the common productivity shock. In particular, $\phi' < \phi$, where ϕ is the speculators' weight on the signal of the common productivity shock in the corresponding cross-learning equilibrium.*

The comparison between a cross-learning equilibrium and its corresponding self-feedback equilibrium implies that, the presence of cross-learning may encourage the speculators to put a higher weight ϕ on the information about the common productivity shock.

The results in Proposition 14 uncover the informational and real consequences of cross learning in equilibrium. Intuitively, when the capital providers are able to cross learn from each other's asset prices (in addition to their own firms' prices), they indeed do so in equilibrium as other firms' asset prices help them better infer the common shock. According to the results in Section 2.3.2, this makes firms' investments relatively more correlated with the common shock. Thus, the speculators find it more profitable to put more weight on the information about the common shock. This further makes asset prices becoming relatively more informative about the common shock in guiding investment decisions, and thus the capital providers respond to the common shock even more sensitively in investing. This spiral is absent in existing feedback models, and it indeed plays an important role in amplifying industry-wide investment exuberance as we fully explore in the next section.

2.4 Systematic Risks

The most important implications of cross learning are on the systematic risks in industry-wide investment exuberance. This derives from the endogenous spiral between the capital providers' investment sensitivity to the common shock and the speculators' weight on the information about the common shock, as shown in Section 2.3. The new spiral and strategic complementarities help generate and amplify systematic risks in many relevant economic environments, which seem jointly puzzling otherwise. These implications also suggest differing extent of cross learning in industries with different economic conditions.

2.4.1 Impacts of Speculators' Weight on Systematic Risks

It is instructive to first investigate the impact of the speculators' weight ϕ (on the information of the common shock) on systematic risks. We also introduce our measures of systematic risks in investment exuberance along the way.

Definition **5.** *The correlation coefficients between the investments of two firms and between the asset prices of two firms are defined as:*

$$\beta_I = \frac{Cov(\log I_i, \log I_j)}{\sqrt{Var(\log I_i)}\sqrt{Var(\log I_j)}}, \text{ and } \beta_P = \frac{Cov(\log P_i, \log P_j)}{\sqrt{Var(\log P_i)}\sqrt{Var(\log P_j)}},$$

respectively. We call β_I the investment beta and β_P the price beta henceforth.

We take the investment beta β_I and the price beta β_P as two major measures of systematic risks in investment exuberance, with both real and financial respects. We have the following intuitive results on the impact of speculators' weight ϕ on the two betas. When the speculators put more weight on the information of the common shock, the capital providers' investment sensitivity to the common shock increases, which makes their investments more correlated. This in turn encourages the speculators to put a higher weight on the common shock, resulting in a higher correlation between asset prices.

Lemma **13.** *Both the investment beta β_I and the price beta β_P are increasing in ϕ when $\phi > 0$.*

Similarly, we also look at the correlations between investment and the two productivity shocks, respectively. As a complement to Definition 4 of investment sensitivity and the associated Lemma 12, the following definition shoots a closer look at the equilibrium investments' correlation with the two shocks.

Definition **6.** *The correlation coefficient between investment and the common productivity shock,* *and between investment and the idiosyncratic productivity shock are defined as:*

$$\beta_A = \frac{Cov(\log I_i, \log A)}{\sqrt{Var(\log I_i)}\sqrt{Var(\log A)}}, \text{ and } \beta_F = \frac{Cov(\log I_i, \log F_i)}{\sqrt{Var(\log I_i)}\sqrt{Var(\log F_i)}},$$

respectively. We call β_A the common investment correlation and β_F the idiosyncratic investment correlation.

Intuitively, when the speculators put more weight on the information of the common shock, both investments and prices become more correlated with the common productivity shock instead of the idiosyncratic shocks. This is because the asset prices become more informative in predicting the common shock but less informative in predicting the idiosyncratic shocks.

Lemma **14.** *The common investment correlation β_A is increasing in ϕ while the idiosyncratic investment correlation β_F is decreasing in ϕ when $\phi > 0$.*

In what follows, we focus on the investment beta β_I and the price beta β_P in exploring the comparative statics, highlighting the speculators' endogenous weight and equilibrium systematic risks under cross learning. The investigation on the common investment correlation β_A and the idiosyncratic investment correlation β_F yields the same insights.

2.4.2 Endogenous Cross Learning and Systematic Risks

Having established the impacts of the speculators' weight ϕ (on the information about the common shock) on systematic risks, we turn to one of the most interesting parts of the paper, which investigates how the changes of economic environments affect equilibrium systematic risks through the cross-learning mechanism. This unifies several observations regarding shock amplifications that are otherwise hard to reconcile with fundamental shocks only. Mathematically, we perform formal comparative statics of the equilibrium betas with respect to exogenous parameters. We elaborate the first comparative statics (with respect to the common uncertainty) in more detail to explore the underlying mechanism, while the other comparative statics follow the same intuition.

Common Uncertainty

We first focus on the effects of common uncertainty, which is captured by the prior precision τ_a of the common productivity shock. We pay greater attention to common uncertainty, because many cases of industry-wide and economy-wide investment exuberance are associated with an increasing common uncertainty in the first place. The most typical driver for an increasing common uncertainty is the arrival of all-purpose technological or financial innovations, as discussed in Brunnermeier and Nagel (2004), Pastor and Veronesi (2006, 2009), and more broadly the literature of bubbles. Highlighting the cross-learning mechanism, our predictions help deliver a new perspective to look at the impacts of innovations and the accompanying increasing common uncertainty on the systematic risks in investment exuberance.

Lemma **15.** *Increasing the common uncertainty leads to a higher weight of the speculators on the information about the common shock. Specifically, the speculators' weight* ϕ *is decreasing in* τ_a.

Lemma 15 implies that an increase in the common uncertainty leads to a stronger cross-learning spiral towards the common shock. The following proposition further establishes the overall impact on the equilibrium systematic risks.

Proposition **15.** *Increasing the common uncertainty leads to both a higher investment beta and a higher price beta in equilibrium. Specifically,* β_I *and* β_P *are both decreasing in* τ_a. *We further decompose the effects into two negative components (the same for* β_I *and* β_P*):*

$$\frac{d\beta(\tau_a,\phi)}{d\tau_a} = \underbrace{\frac{\partial\beta(\tau_a,\phi)}{\partial\tau_a}}_{\text{Mechanical Effect } <0} + \underbrace{\frac{\partial\beta(\tau_a,\phi)}{\partial\phi}\frac{\partial\phi}{\partial\tau_a}}_{\text{Cross-Learning Effect } <0} < 0\,.$$

Proposition 15 indicates two effects contributing to the higher systematic risks associated with an increasing common uncertainty. The first is a mechanical effect that does not depend on the endogenous interaction between the capital providers and the speculators under cross learning. Intuitively, when the common uncertainty increases, the firms' fundamentals become more exposed to the common shock, and thus their capital providers' investment sensitivity to the common shock increases mechanically. This immediately results in a higher comovement among firms' investments and prices. Figure 2.3 illustrates this mechanical effect in a two-firm example. As pointed out in the literature, however, investment (and price) comovements during investment exuberance are hard to reconcile with fundamental shock comovements only, so a further amplification channel is needed.

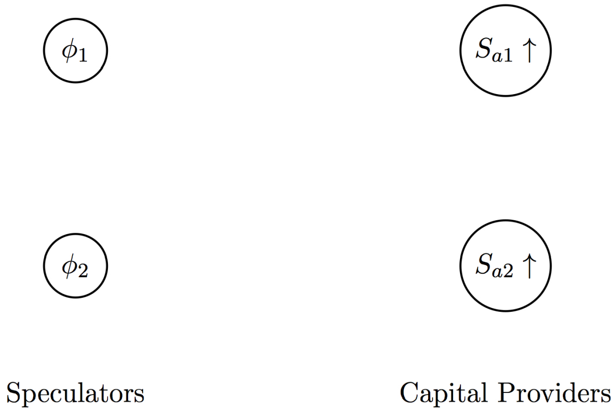


Figure 2.3: *Mechanical Effect on Systematic Risks*

The second effect, the cross-learning effect, provides a new perspective to understand the amplification from fundamental shocks to large investment (and price) comovements. This effect is novel and only at play in our multi-firm cross-learning framework with two types of shocks. It reflects the new spiral between the capital providers' investment sensitivity to the common shock and the speculators' weight on the signal of the common shock.

Figure 2.4 illustrates this cross-learning effect in a two-firm example. Interestingly, this effect takes place even when only some (not all) firms in the economy perceive the increasing common uncertainty. Suppose, without loss of generality, firm 1's capital provider perceives the increasing common uncertainty but firm 2's does not. As seen in the upper-left panel, firm 1's investment sensitivity to the common shock S_{a1} first increases (along with a decreasing investment sensitivity to its idiosyncratic shock), leading to a higher weight ϕ_1 on the information of the common shock by its speculators. Then, as seen in the upper-right panel, a higher ϕ_1 results in an even higher S_{a1} since firm 1 learns from its own price. More

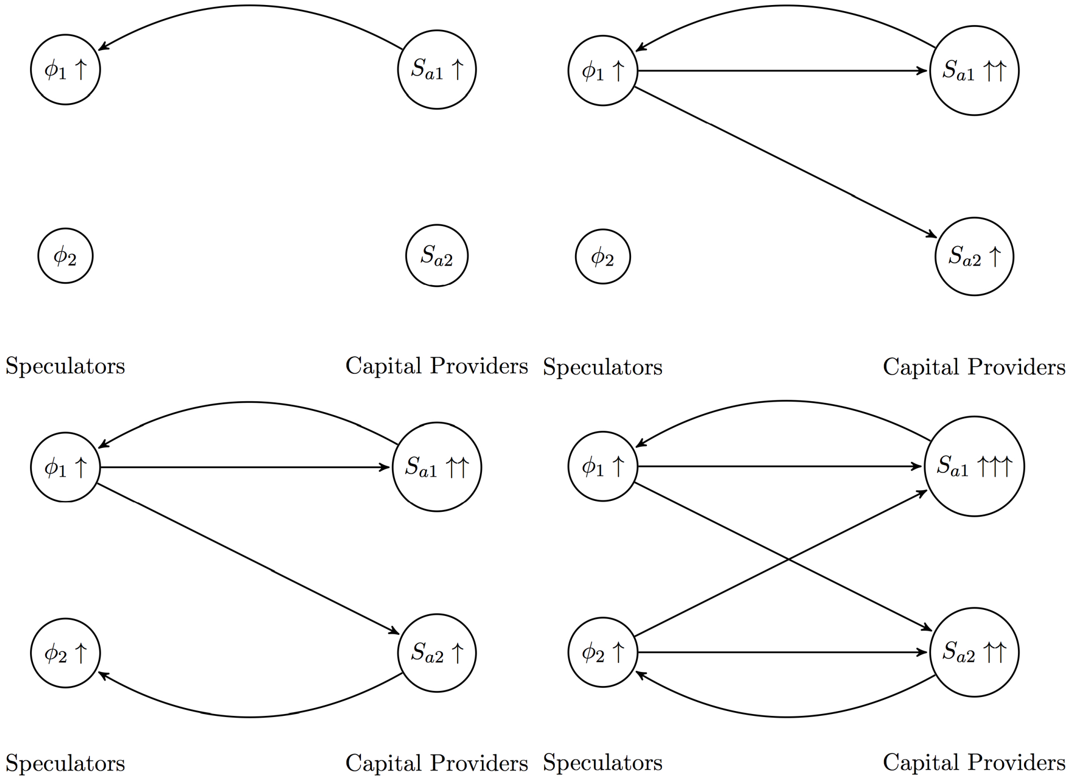

$\phi_1 \uparrow$
$S_{a1} \uparrow$
ϕ_2
S_{a2}
Speculators
Capital Providers
$\phi_1 \uparrow$
$S_{a1} \uparrow\uparrow$
ϕ_2
$S_{a2} \uparrow$
Speculators
Capital Providers
$\phi_1 \uparrow$
$S_{a1} \uparrow\uparrow$
$\phi_2 \uparrow$
$S_{a2} \uparrow$
Speculators
Capital Providers
$\phi_1 \uparrow$
$S_{a1} \uparrow\uparrow\uparrow$
$\phi_2 \uparrow$
$S_{a2} \uparrow\uparrow$
Speculators
Capital Providers

Figure 2.4: *Learning Effect on Systematic Risks*

Order: Upper-Left, Upper-Right, Lower-Left, Lower-Right

importantly, because of cross learning, firm 2's investment sensitivity to the common shock S_{a2} also increases, since firm 2 finds firm 1's price more informative about the common shock and thus better understands the common shock. It then naturally leads to a higher weight ϕ_2 on the information of the common shock by firm 2's speculators, as seen in the lower-left panel. Finally, the increase of ϕ_2 results in even higher S_{a1} and S_{a2} by cross learning, as seen in the lower-right panel. This spiral suggests two new strategic complementarities only under cross learning: the first is among speculators' weight on the information about the common shock in each market, and the second is among different firms' relative investment sensitivity to the common shock. With the two strategic complementarities, the spiral reinforces itself and eventually pushes the economy to a new equilibrium with much higher systematic risks.

Our predictions on systematic risks after an increasing common uncertainty are consistent with the literature (Brunnermeier and Nagel, 2004, Pastor and Veronesi, 2006, 2009) that documents the increasing systematic risks after major technological innovations, as these innovations often come with industry-wide uncertain market prospects. In particular, the cross-learning amplification effect helps understand the huge magnitude of systematic risks in these cases of investment exuberance, which is hard to be explained by fundamental shocks. These predictions also shed light on the empirical regularities on systematic risks in papers such as Rhodes-Kropf, Robinson and Viswanathan (2005) on merger and acquisitions, Hoberg and Phillips (2010) on investment waves, and Bhattacharyya and Purnanandam (2011) on financial institutions' collective risk-taking activities.

Capital Providers' Access to Information

We then turn to the capital providers' access to private information, captured by the two precisions τ_{sa} and τ_{sf} regarding the two productivity shocks, respectively. Again, we have the following lemma pertaining to the speculators' endogenous weight.

Lemma **16.** *Increasing the capital providers' information precision on the common shock leads to a higher speculator weight on the information about the common shock, while increasing the capital providers' information precision on the idiosyncratic shock leads to a lower weight. Specifically, the speculators' weight ϕ is increasing in τ_{sa} while decreasing in τ_{sf}.*

Lemma 16 prescribes that when the capital providers have better information on the common shock, they use that information more in investing, which in turn encourages the speculators to put more weight on the common shock. Hence, the equilibrium cross-learning spiral towards the common shock is stronger. In contrast, better information on the idiosyncratic shock pushes the cross-learning spiral towards the idiosyncratic shocks.

Moreover, similar to Proposition 15, when the overall impact of the mechanical effect and the cross-learning effect is considered, we have the following proposition.

Proposition **16.** *For the capital providers' access to private information, we have the following results.*

i) Increasing the precision on the common shock leads to a higher investment beta when the precision is not small, and always a higher price beta; specifically, β_I is increasing in τ_{sa} when $\tau_{sa} > \tau_a + \tau_x \tau_\zeta$ and β_P is always increasing in τ_{sa}.

ii) Increasing the precision on the idiosyncratic shock leads to both a lower investment beta and a lower price beta; specifically, β_I and β_P are always decreasing in τ_{sf}.

Liquidity Trading

We also investigate the effects of liquidity trading, captured by the market liquidity λ and the two precisions of noisy supplies τ_ζ and τ_ξ. Similarly, we have the following intuitive lemma on the speculators' endogenous weight.

Lemma **17.** *For liquidity trading, a higher weight of the speculators on the information about the common shock results from a lower market liquidity, a lower variance of common noisy supply, or a higher variance of idiosyncratic noisy supply. Specifically, the speculators' weight ϕ is decreasing in λ, increasing in τ_ζ, and decreasing in τ_ξ.*

The predictions along the three dimensions are all intuitive. When the market liquidity is higher, it is easier for the noisy traders to absorb speculators' demand, making asset prices harder to aggregate information so that the cross-learning spiral towards the common shock becomes weaker. When the variance of the common noisy supply is lower, speculators are more likely to trade upon the common productivity shock. In contrast, when the variance of the idiosyncratic noisy supply is lower, speculators are less likely to trade upon the common shock, which results in a weaker spiral towards the common shock.

These predictions are further reflected in the following proposition, speaking to the overall effects of liquidity trading on investment exuberance. Again, similar to Proposition 15, we have the mechanical effect and the cross-learning effect in the same direction.

Proposition **17.** *For liquidity trading, we have the following results.*

i) A higher investment beta β_I results from a lower market liquidity, a lower variance of common noisy supply, or a higher variance of idiosyncratic noisy supply. Specifically, β_I is decreasing in λ, increasing in τ_ζ, and decreasing in τ_ξ.

ii) A higher price beta β_P *results from a lower market liquidity, or a higher variance of idiosyncratic noisy supply. Specifically,* β_P *is decreasing in* λ *and decreasing in* τ_ξ*. The effect of the variance of common noisy supply* τ_ζ *on* β_P *is uncertain.*

2.5 Investment Inefficiency and Competition

An important question is how firm cross learning affects real investment efficiency. On the positive side, cross learning allows capital providers to take advantage of more information that would not be available if they were not able to observe their own and other firms' asset prices. However, the interests between capital providers and speculators in learning the two types of shocks are not perfectly aligned. More importantly, each firm's cross learning further creates a new pecuniary externality on other firms. These frictions result in investment inefficiency. In particular, the cross learning effect and the pecuniary externality associated with cross learning both increase in the number of firms, suggesting that more competitive industries may exhibit stronger and more inefficient investment exuberance. This industry-level heterogeneity is particularly helpful to separate aggregate effects specific to cross learning from other potential mechanisms which may also generate investment exuberance.

We proceed by two steps in evaluating how cross learning affects investment efficiency. First, we evaluate the overall investment efficiency and show that any cross-learning equilibrium always features investment inefficiency. Then we characterize the new pecuniary externality associated with cross learning for better understanding the origin of such inefficiency. By doing this, we in particular underscore the implications of competition on inefficient investment exuberance through the new pecuniary externality.

2.5.1 Overall Investment Efficiency

We follow Goldstein, Ozdenoren and Yuan (2013) to define investment efficiency by the ex-ante expected net benefit of the total investments by all of the firms,[24] given that capital providers may learn from all publicly available asset prices:

Definition 7. *The investment efficiency of the economy is defined as*

$$R = \int_0^1 R_i di \,,$$

where

$$R_i = \mathbb{E}\left[\mathbb{E}\left[AF_i I_i - \frac{c}{2} I_i^2 | \Gamma_i\right]\right]$$

denotes each firm i's ex-ante expected net benefit of investment, given its capital provider's information set under cross learning: $\Gamma_i = \{P_i, \{P_{-i}\}, s_{a,i}, s_{f,i}\}$.

We have the following proposition indicating the universal presence of investment inefficiency in a cross-learning equilibrium. We focus on unique cross-learning equilibria by assuming that the noisy supply elasticity λ and the information precision τ_y (of speculators' signal on the idiosyncratic shock) are high enough.

Proposition 18. *There always exists a unique optimal weight* $\phi^* \geqslant 0$ *of the speculators on the signal of the common shock that maximizes investment efficiency. In particular, for a high enough noisy supply elasticity* λ *and a high enough information precision* τ_y *(of the speculators' signal on the idiosyncratic shock), the optimal weight is always smaller than that in the corresponding cross-learning equilibrium, i.e.,* $\phi^* < \phi$.

Proposition 26 indicates that in a cross-learning equilibrium, the speculators tend to put an inefficient high weight on the signal about the common shock. This makes capital providers respond to the common shock too sensitively, leading to inefficient investment exuberance. This particular inefficiency fits quite in line with what we have observed in typical investment exuberance (for example, Rhodes-Kropf, Robinson and Viswanathan, 2005, Peng, Xiong, and Bollerslev, 2007, Hoberg and Phillips, 2010, Bhattacharyya and Purnanandam, 2011) that both primary and secondary market investors inefficiently pay too much attention to common shocks or noisy macroeconomics news while ignoring informative idiosyncratic news.

---

[24] Consistent with the literature, the private adjustment costs of the capital providers are included in the first-best welfare calculation, and they are also the only type of investment cost in the present framework. Thus, although the capital providers do not maximize firms' total expected values, this fact does not imply any inefficiency per se.

To better understand the impacts of cross learning on investment efficiency and potentially shed light on corrective policies, we perform comparative statistics of investment efficiency with respect to several economic parameters. Again, we focus on the cases in which a unique cross-learning equilibrium is guaranteed.

PROPOSITION **19.** *In a cross-learning equilibrium, investment efficiency is higher when the market liquidity is higher, the precision of idiosyncratic noisy supply is higher, or the capital providers' information precision on the idiosyncratic productivity shock is higher. Specifically, R is increasing in λ, τ_{ξ}, and τ_{sf}.*

The comparative statics regarding the investment efficiency are intuitive. First, a higher market liquidity has a corrective effect on the investment efficiency. In other words, in a deeper asset market, the speculators' trading positions can more easily be absorbed. Specifically, when an asset market is more liquid or deeper, it becomes harder for the same amount of informed trading to impact the asset price. This is beneficial in particular when cross learning is strong after the arrival of major innovations or other common news involving high uncertainty, because the inefficient impact from speculators' overuse of information about the common shock can be better absorbed.

Importantly, this corrective effect on real investment efficiency helps justify recent regulatory concerns and practices by the SEC in limiting informed speculators' trading positions but at the same time encouraging liquidity provision by less informed retail investors. These two are hard to be explained as approaches to correct investors' irrationality or to sidestep limits to arbitrage. In this sense, our cross-learning mechanism does a better job in delivering policy implications than typical models featuring bubbles.

Second, increasing investment efficiency in an economy with cross learning calls for a better use of information about the idiosyncratic shocks in the economy. Any policies on financial disclosure or government communication failing to keep this point in mind may end up crowding out the idiosyncratic news and resulting in investment inefficiency. This policy implication fits broadly in line with the recent studies that speak to the dark side of financial disclosures or central bank communications (Di Maggio and Pagano, 2013, Kurlat and Veldkamp, 2013, Goldstein and Yang, 2014b). Theoretically, the endogenous overuse of information on the common shock due to multi-firm cross learning results in an inefficient crowding-out effect on the use of information on the idiosyncratic shock. Thus, it also complements the idea on the crowding-out effect of public information provision on the use of private information (see Amador and Weill, 2010).

2.5.2 Competition and Cross Learning

To better understand the origin of investment inefficiency, we further identify a new pecuniary externality induced by cross learning. In doing so, we extend our baseline model to admit finite number of firms. This not only allows us to underscore the efficiency change associated with different extent of cross learning, but to investigate the relationship between competition and inefficient investment exuberance which has been a well documented puzzle in recent empirical literature (see Hoberg and Phillips, 2010, Greenwood and Hanson, 2014, among others). To flesh out the cross-learning channel, we use the number of firms as a proxy for competition[25] and intentionally abstract away from any product market competition as that in Peress (2010).[26] Doing this also helps empirically identify the aggregate effects specific to cross learning because it predicts industry-level heterogeneity, that is, differing extent of cross learning in different industries.

We first outline the extended cross-learning framework. A major challenge in identifying the externality associated with cross learning is to deal with an information endowment effect. Specifically, when the number of firms increase, the total amount of information in the economy also increases, leading to an efficiency gain to each firm. This information endowment effect confounds the identification of externalities and thus needs to be controlled properly. To achieve this goal, our extended cross-learning framework still features a continuum of 1 of firms being able to learn from all asset prices. However, we assume that the speculators believe that each firm only observes and learns from as many as $n \geqslant 1$ asset prices, including its own price. This setting delivers an equilibrium weight ϕ (of the speculators on the information about the common shock) identical to that in a corresponding economy with n finite firms operating and the speculators fully internalizing their cross learning, while keeping the total amount of information endowment invariant with n. Hence, we are able to distinguish the externality associated with cross learning as the number of firms increases.

---

[25] To use the number of firms to proxy competition is common in the literature, especially when the model focuses on information (see Vives, 2010, for a survey).

[26] Peress (2010) offers an interesting analysis on the impacts of monopolistic competition in product markets on stock market efficiency, but does not consider feedback to real investments or cross learning as we do. He does not consider the implications on investment exuberance as well. To incorporate product market competition into our framework is possible and could make our model richer, and we leave it for future research.

We rigorously formulate the idea above as follows. We divide all the firms into $n \geqslant 1$ groups, a continuum of $1/n$ of firms in each group. The firms still observe and learn from all the asset prices as in the baseline model, regardless of the grouping. However, the speculators do not fully internalize firms' cross learning. Specifically, let $i \in [0, 1/n)$ denote one firm in the first group. The speculators believe that for any i, the n firms in the set $\{i + k/n | 0 \leqslant k \leqslant n-1, k \in \mathbb{Z}\}$ only learn from the asset prices of each other but not from the asset prices of other firms outside the set. Figure 5 offers an illustration of the case when $n = 3$, in which the speculators believe that the three red firms (i, $i + 1/3$, and $i + 2/3$) cross learn only from each other and the three blue firms (i', $i' + 1/3$, and $i' + 2/3$) cross learn only from each other, similar for other firm triples.

This setting has several advantages, both economically and technically. First, it casts industry competition in a straightforward way. Since the speculators are risk neutral, it looks to them as if there are exactly n firms operating in the economy. Thus, the speculators' weight in equilibrium is identical to that in a corresponding economy with exactly n firms and the speculators fully internalizing their cross learning. Second, it helps identify

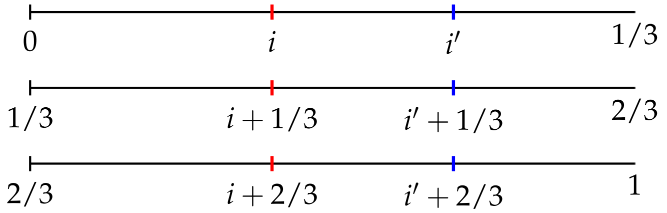


Figure 2.5: *Extended Cross-Learning Framework*

the pecuniary externality induced by cross learning while keeping the total information endowment fixed. Especially, the efficiency change associated with cross learning takes place only through the speculators' endogenous weight over the two types of shocks, making it possible to distinguish that from the change of firms' actual information endowment. Last, this setting offers a smooth transition between the baseline model with full cross learning (as n goes to infinity) and the self-feedback benchmark (as n equals to 1). This not only makes our analysis analytically tractable but also helps unify all of the results and intuitions.

We again acknowledge that we are abstracting away from any possible industrial organization of the firms' product markets. Rather, we make use of the number of firms as a proxy for competition, which we believe is the most straightforward and relevant measure. This also allows us to underscore the cross-learning mechanism by rendering it as the only form of interaction among firms. In this sense, our model serves as a benchmark for further research that may take into account more respects of industrial competition along with firm cross learning.

We proceed to characterize the equilibrium in the extended framework and the corresponding investment efficiency. We still consider symmetric equilibria, and denote by ϕ_n the speculators' weight on the signal of the common shock when the speculators believe that each firm only learns from as many as n asset prices, including its own. For convenience, we call the associated equilibrium an n-learning equilibrium.

Formally, each capital provider i still observes its own asset price P_i and all other firms' asset prices $\{P_{-i}\}$. Same as before, the information content of any asset price is characterized by

$$P_i = \exp\left(\frac{\phi_n}{\lambda\sqrt{\tau_x^{-1}\phi_n^2+\tau_y^{-1}}}a + \frac{1}{\lambda\sqrt{\tau_x^{-1}\phi_n^2+\tau_y^{-1}}}f_i + \frac{\zeta+\xi_i}{\lambda} - \frac{\mu_i}{\lambda\sqrt{\tau_x^{-1}\phi_n^2+\tau_y^{-1}}}\right),$$

equivalent to a signal

$$z_n\left(P_i\right) = \phi_n a + f_i + \sqrt{\tau_x^{-1}\phi_n^2+\tau_y^{-1}}\left(\zeta+\xi_i\right).$$

However, in an n-learning equilibrium, the speculators believe that each capital provider only learns from its own price as well as the other $n-1$ firms' asset prices. Specifically, from the speculators' perspective, due to symmetry, each capital provider i has four signals: her own private signals $s_{a,i}$ and $s_{f,i}$, the signal $z_n(P_i)$ from its own asset price, and another signal $z_n(P_{-i})$ coming from the other $n-1$ asset prices:[27]

$$z_n\left(P_{-i}\right) = \phi_n a + \frac{\sum_{l\neq i} f_l}{n-1} + \sqrt{\tau_x^{-1}\phi_n^2+\tau_y^{-1}}\left(\zeta + \frac{\sum_{l\neq i}\xi_l}{n-1}\right).$$

From the speculators' perspective, capital provider i uses these four signals to infer the sum of the two (log) productivity shocks, $a+f_i$, in making investment decisions. Concretely, the speculators believe that capital provider i updates her belief as

$$\mathbb{E}[a+f_i|\Gamma_i] = \mathbf{z}'\mathrm{Var}(\mathbf{z})^{-1}\mathrm{Cov}(a+f_i,\mathbf{z}), \tag{2.5.12}$$

where $\mathbf{z} = [s_{a,i}, s_{f,i}, z_n(P_i), z_n(P_{-i})]'$. As a consequence, the speculators' perceived investment sensitivity S_{an} to the common shock and S_{fn} to the idiosyncratic shocks are read off from the conditional expectation (2.5.12). Following the same approach as before in solving for the speculators' optimal weight in trading, we finally get

$$\phi_n = \frac{\left(S_{an}+1-\frac{\phi_n}{\lambda\sqrt{\tau_x+\tau_y\phi_n^2}}\right)\frac{\tau_x}{\tau_a+\tau_x}}{\left(S_{fn}+1-\frac{1}{\lambda\sqrt{\tau_x+\tau_y\phi_n^2}}\right)\frac{\tau_y}{\tau_f+\tau_y}}. \tag{2.5.13}$$

Clearly, the equilibrium condition (2.5.13), with the investment sensitivity prescribed by condition (2.5.12), is equivalent to that in a corresponding economy with n firms operating and speculators fully internalizing their cross learning, so is the equilibrium weight ϕ_n, while the expression of investment efficiency can be shown to be the same as that in Definition 9. Therefore, we have the following proposition regarding the equilibrium weight ϕ_n and investment efficiency R_n in an n-learning equilibrium. We still focus on comparable cases in which all n-learning equilibria are unique.

---

[27] When $n=1$, only the first three signals are relevant and the n-learning equilibrium degenerates to a self-feedback equilibrium.

Proposition **20.** *For a high enough noisy supply elasticity λ, a low enough idiosyncratic noisy supply shock precision τ_ξ, and a high enough information precision τ_y (of the speculators' signal on the idiosyncratic shock),*

i) for all $n \geqslant 1$, there exists a unique n-learning equilibrium in which the speculators put a positive weight $\phi_n > 0$ on the signal of the common productivity shock,

ii) for all $n > 1$, $\phi^* < \phi'(= \phi_1) < \phi_n < \phi$, in particular, ϕ_n is increasing in n, where ϕ, ϕ' are the equilibrium weights in the baseline cross-learning equilibrium and in the self-feedback equilibrium, respectively, and ϕ^* is the optimal weight that maximizes investment efficiency, and

iii) for all $n \geqslant 1$, $R < R_n < R^*$, in particular, R_n is decreasing in n, where R is the investment efficiency in the baseline cross-learning equilibrium and R^* is the optimal investment efficiency.

Proposition 20 offers a clear identification of the externality and efficiency loss associated with cross learning. Under the parameters guaranteeing a unique equilibrium, when the number of firms increases, cross learning makes the speculators put an increasing weight ϕ_n on the signal of the common shock. Along with the established results in Section 2.4.2, this implies investment exuberance with higher systematic risks. Moreover, since the information endowment is controlled, an increase in n leads to decreasing investment efficiency, associated with an increasing extent of cross learning. The key to understand this is a new externality through the speculators' weight over the two types of shocks in response to the capital providers' cross learning. When each capital provider learns from other firms' asset prices, she only cares about her own investment decision and wants to use other firms' asset prices for better inferring the common shock. This makes her investment more sensitive to the common shock, which in turn encourages the speculators to put more weight on the signal of the common shock. However, she does not internalize the endogenous cost on other firms' investment decisions because her cross learning makes asset prices endogenously less informative on other firms' idiosyncratic productivity shocks, through the speculators' endogenous response in terms of allocating weight over the two shocks. When there are more firms in the economy, the speculators respond more heavily to the capital providers' cross learning and each asset price is also used by more firms, which implies a stronger externality not being internalized by each capital provider in cross learning.

Along with the results in Section 2.4.2, the new externality associated with cross learning offers a new perspective to investigate the puzzling fact that more competitive industries exhibit more inefficient investment exuberance with higher systematic risks. This fact has recently been documented in Hoberg and Phillips (2010) and shown to be robust after many relevant controls. As they suggest, however, no single theory in the literature can accommodate their findings. More recently, Greenwood and Hanson (2014) find a similar pattern in the cargo ship industry that also applies to other industries. They estimate a behavioral theory in which firms over-extrapolate exogenous demand shocks and partially neglect the endogenous investment responses of their competitors. Our fully rational cross-learning framework helps reconcile these facts by explicitly identifying the externality associated with competition and its impacts on real investment efficiency. Relatedly, Ozoguz and Rebello (2013) have empirically identified that firms in more competitive industries adapt their investment more sensitively to stock prices of their peers, which supports our theory.

We highlight the externality that we have identified as a new pecuniary externality, taking effect through the informativeness of prices instead of price levels. In the classical pecuniary externality literature (see Stiglitz, 1982, Greenwald and Stiglitz, 1986, and Geanakoplos and Polemarchakis, 1985), agents do not internalize the impacts of their actions on equilibrium price levels, leading to a welfare loss under various frictions. In particular, the classical pecuniary externality generates welfare transfers across agents through the levels of prices.

Instead in our framework, the capital providers do not fully internalize the impacts of cross learning on equilibrium price informativeness. This leads to a typical "tragedy of the commons" regarding the use of the price system as an information source under multi-firm cross learning. This tragedy-of-the-commons observation is absent in existing single-firm feedback models. In this sense, our pecuniary externality is also reminiscent of the notion of learning externality in the earlier dynamic learning and herding literature (for example, Vives, 1997) that an agent, when responding to private information, does not take into account the benefit of increased informativeness of public information in the future. This literature, however, does not explicitly consider the role of financial markets and in particular the feedback from market prices to investments as we do.

The new pecuniary externality associated with cross learning resembles and generalizes the "mismatch channel" of feedback identified in Goldstein and Yang (2014a). Due to cross learning, the capital providers already get a relatively better understanding about the common shock, so that they will want more information about the idiosyncratic shocks and want to see their own prices more informative about them. But this goes against the speculators' incentives in trading (and information aggregation), because the speculators still find it profitable to put considerable weight on the signal of the common shock in trading. This conflict of interest between capital providers (or firms) and speculators is generally present in the feedback literature in different forms (see the survey by Bond, Edmans and Goldstein, 2012), and Goldstein and Yang (2014a) formally identify it as the mismatch channel of feedback. Specifically, to uncover this channel, Goldstein and Yang (2014a) assume that capital providers can perfectly observe one shock but not the other and they do not consider the possibility of multi-firm cross learning. Thus, the contribution of our work is to endogenize capital providers' information advantage on the common shock through multi-firm cross learning, and then to show that the underlying conflict of interest becomes more severe as the number of firms get larger. Therefore, our cross-learning framework nicely resembles the classical notion of pecuniary externalities, in the sense that many capital providers endogenously ignore the full impacts of their cross learning actions on equilibrium prices.

Although our framework allows for an analytical characterization, we also offer numerical examples to help illustrate the pecuniary externality and efficiency loss associated with different extent of cross learning. We set $\tau_a = \tau_f = \tau_{sa} = \tau_{sf} = \tau_x = \tau_\zeta = 1$, $\tau_y = 10$, $\tau_\xi = 0.1$, $\lambda = 2$, $\kappa = 1$, and $c = 0.5$. The left panel of Figure 2.6 depicts the equilibrium weight ϕ_n as n increases as the blue solid line. When n becomes larger, the weight gradually approaches that in the baseline cross-learning equilibrium, as depicted by the red dashed line. The right panel of Figure 2.6 depicts the log of the efficiency loss due to cross learning, measured by $\log(R^*/R_n)$. As shown in the solid blue line, the efficiency loss associated with cross learning increases in n, suggesting a more severe pecuniary externality as competition becomes stronger. In the baseline cross-learning equilibrium, the pecuniary externality is the strongest, as depicted by the dashed red line. These results are robust to a very wide range of parameters when λ and τ_y are relatively large while τ_ξ is relatively small (and thus a unique equilibrium is guaranteed), which are empirically relevant as aforementioned.

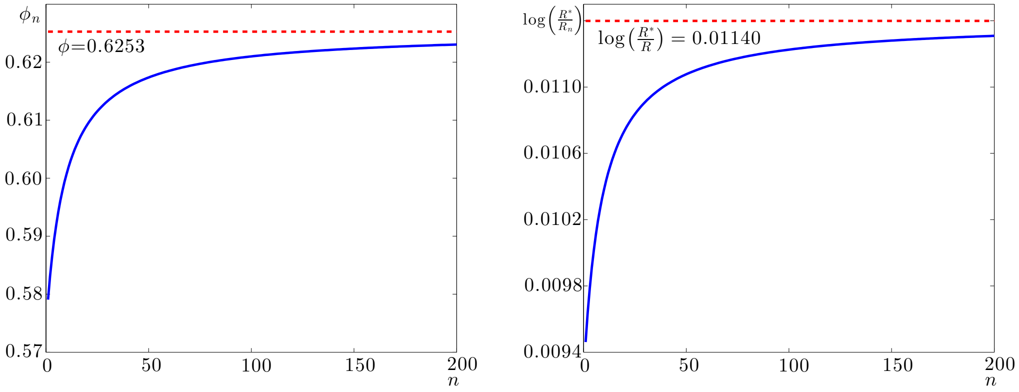


Figure 2.6: *Competition on Cross Learning and Efficiency Loss*

Admittedly, our identification of the pecuniary externality associated with cross learning does not attempt to offer a comprehensive evaluation of the merits of competition. Relatedly, the investment efficiency R_n in an n-learning equilibrium cannot be interpreted as a direct measure of the investment efficiency in an actual competitive industry with n firms. Our point is focused, however, to suggest a new perspective to look at the relationship between inefficient investment exuberance and competition, a puzzling fact recently documented and hard to be reconciled with existing theories. We admit that, despite the fact that competition increases the extent of cross learning, with new adverse implications for investment efficiency, it may well remain desirable when all other social benefits and costs of competition are taken into account.

2.6 Discussion and Extension

Our cross-learning framework focuses on the second moments and inefficiency of industry-wide investment exuberance, which are less understood in the literature. It is convenient to further extend the present model to generate more implications and to shed light on some other commonly observed phenomena, which offer additional empirical predictions to help differentiate cross learning from other well-understood mechanisms.

2.6.1 Growth Opportunities and Systematic Risks under Cross Learning

In a structural asset pricing framework, Berk, Green, and Naik (1999) first argues that firms with more growth opportunities are systematically more risky and their returns also comove more. Papanikolaou (2011) and Kogan and Papanikolaou (2014) further propose a concept of investment-specific technology shocks and argue that growth firms are more risky because they are more exposed to such shocks. The cross-learning framework can provide an alternative to look at the systematic risks associated with growth firms or industries. In the same spirit of the relationship between competition and cross learning, the theory predicts that industries with more growth opportunities may exhibit stronger extent of cross learning and higher systematic risks.

We could adapt our baseline model by incorporating both assets in place and growth opportunities while keeping all other settings. Specifically, each firm has an asset in place and a growth option, both subject to a common shock and an idiosyncratic shock. However, the capital provider can only invest in the growth option but not in the asset in place. In this setting, a larger growth option (relative to the asset in place) implies a stronger extent of cross learning and a higher weight of the speculators on the common shock. Consequently, the common beta becomes higher, implying higher systematic risks. Intuitively, as the size of growth option becomes higher, the capital providers tend to cross learn more, making the speculators over-value information about the common shock in trading. This in turn reinforce capital providers' cross learning, leading to a higher investment sensitivity to the common shock even if the shock to growth opportunities does not change per se.

Importantly, the cross-learning framework provides a new micro-foundation for growth firms' higher exposure to investment-specific shocks, without assuming that such shocks are larger for growth firms per se. Consistent with the thesis of Papanikolaou (2011) and Kogan and Papanikolaou (2014), as firms grow, the uncertainty of their assets in place decreases relative to growth opportunities, and investment-specific technology shocks become more positively correlated with their average q, so that they become systematically more risky. In this sense, our theory helps generate additional testable predictions regarding the aggregate effects specific to cross learning, highlighting the cross-sectional heterogeneity among industries with differing growth opportunities.

2.6.2 Over-investment under Cross Learning

Investment exuberance usually exhibits both high systematic risks (second moment) and over-investment (first moment). Although the latter has been well addressed in the literature, our framework is easily adaptable to generate so.

We could keep all the settings in our baseline model except for introducing two different investment technologies. Specifically, each firm now has two mutually exclusive projects, one only subject to a common shock while the other only subject to an idiosyncratic shock. We call the former common project and the latter idiosyncratic project. This adapted setting is in the similar spirit of Dow, Goldstein and Guembel (2011) but enriches it with both cross learning and the firm's debate between the common project and the idiosyncratic project.

Following the same equilibrium concept as our baseline model, a cross-learning equilibrium in this adapted setting features over-investment in the common project while under-investment in the idiosyncratic project, compared to the first best. The intuition is the same as before. When the capital providers are able to cross learn, the speculators again find it more profitable to put a higher weight on the information about the common shock. This makes the prices more informative about the common shock and thus encourages the capital providers to invest more on the common project while less on their idiosyncratic projects.

Complementary to the existing literature about over-investment, our cross-learning mechanism has two new implications. First, over-investment is more likely to happen in technologies or industries that are more sensitive to common shocks, which is reflected by the common project in our stylized model. This fits quite in line with the recent episodes of over-investment in the IT industry and housing markets. Second, which is perhaps more subtle and interesting, over-investment in the common project is always accompanied by under-investment in the idiosyncratic projects at the same time. This suggests that over-investment does not necessarily imply an inefficiently large economy scale but rather an inefficient composition of various economic activities.

2.6.3 Industry Momentum under Cross Learning

The contemporaneous study by Sockin and Xiong (2014b) uses a feedback model to generate return momentum in a housing cycle context. Although our model does not aim to provide a general dynamic account for boom-bust cycles, the introduction of multiple firms and the two types of shocks also help shed light on the understanding of momentum by further establishing a channel between cross learning and industry momentum.

Industry momentum, first identified by Moskowitz and Grinblatt (1999), suggests that industry portfolios also exhibit considerable short-term momentum, and it even accounts for much of the individual stock momentum. As discussed by Moskowitz and Grinblatt (1999), individual stock momentum may be explained by a number of behavioral theories focusing on investors' information barrier or risk appetite. But there have been no formal theories that directly speak to the existence and the magnitude of industry momentum. Our framework can potentially offer a consistent rational theory for both individual stock momentum and industry momentum, instead highlighting firms' investment activities and their cross learning.

In our benchmark three-period model, the standard definition of overall individual momentum is

$$M_i = \mathrm{Cov}\left(\log(AF_iI_i) - \log P_i, \log P_i\right) ,$$

and industry momentum can be defined as

$$\overline{M} = \mathrm{Cov}\left(\log \int_0^1 AF_iI_i \mathrm{d}i - \log \overline{P}, \log \overline{P}\right) .$$

It is straightforward to show that both individual stock momentum and industry momentum exist in equilibrium, and their magnitudes increase in the speculators' weight ϕ on the information about the common shock. Specifically, M_i and $\overline{M}$ are always positive when the noisy supply elasticity λ and the information precision τ_y (of speculators' signal on the idiosyncratic shock) are large enough, and they increase in ϕ. Intuitively, when the speculators put a higher weight on the common shock, the asset prices become more informative about the common shock and thus firm investment also becomes more sensitive to the common shock. Therefore, the common shock plays a more important role in determining both the asset prices and eventual cash flow of the firms, implying both a stronger individual stock momentum and a stronger industry momentum. Moreover, according to the results regarding the relationship between cross learning and competition in Section 3.4.2, both individual stock momentum and industry momentum may be stronger in more competitive industries, also due to a stronger cross learning effect.

It is worth highlighting that our mechanism to generate individual momentum and industry momentum is fundamentally different from the prevailing explanations that highlight overconfidence (Daniel, Hirshleifer, and Subrahmanyam, 1998), sentiment (Barberis, Shleifer, and Vishny, 1998), or slow information diffusion (Hong and Stein, 1999). In those models, investors generally ignore some information contents revealed by asset prices. In contrast, in our model, the capital providers' rational cross learning from all available asset prices plays a central role.

2.7 Conclusion

Firm cross learning behavior is not only empirically important but also theoretically relevant for commonly observed investment exuberance. We have developed a tractable model to admit cross learning and delivered a series of predictions regarding the second moments of investment exuberance. We have illustrated that the high systematic risks associated with investment exuberance come from a new spiral that coordinate capital providers' investment sensitivity and speculators' weight in trading towards the common productivity shock. Moreover, cross learning may lead to higher investment inefficiency, because capital providers do not internalize the new externality that other firms' asset prices become less informative on their idiosyncratic productivity shocks. In more competitive industries, cross learning tends to be stronger, potentially leading to more inefficient investment exuberance with higher systematic risks. Hence, appropriate policy interventions are called for to correct the inefficiency in industry-wide investment exuberance.

Bibliography

Aghion, Philippe and Jeremy Stein (2008). "Growth versus Margins: Destabilizing Consequences of Giving the Stock Market What It Wants." *Journal of Finance*, 63: 1025-1058.

Angeletos, George-Marios, Guido Lorenzoni and Alessandro Pavan (2012). "Wall Street and Silicon Valley: A Delicate Interaction." Working Paper.

Amador, Manuel and Pierre-Olivier Weill (2010). "Learning from Prices: Public Communication and Welfare." *Journal of Economic Theory*, 147: 910-940.

Bakke, Tor-Erik and Toni Whited (2010). "Which Firms Follow the Market? An Analysis of Corporate Investment Decisions." *Review of Financial Studies*, 23: 1941-1980.

Banerjee, Abhijit (1992). "A Simple Model of Herd Behavior." *Quarterly Journal of Economics*, 107: 797-817.

Barberis, Nicholas, Andrei Shleifer and Robert Vishny (1998). "A Model of Investor Sentiment." *Journal of Financial Economics*, 49: 307-343.

Berk, Jonathan, Richard Green and Vasant Naik (1998). "Optimal Investment, Growth Options, and Security Returns." *Journal of Finance*, 54: 1553-1607.

Bikhchandani, Sushil, David Hirshleifer and Ivo Welch (1992). "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy*, 100: 992-1026.

Bond, Philip, Alex Edmans and Itay Goldstein (2012). "The Real Effects of Financial Markets." *Annual Review of Finance and Economics*, 4: 1-22.

Bond, Philip, Itay Goldstein and Edward Prescott (2010). "Market-Based Corrective Actions." *Review of Financial Studies*, 23: 781-820.

Bhattacharyya, Sugato and Amiyatosh Purnanandam (2011). "Risk-Taking by Banks: What Did We Know and When Did We Know It?" Working paper.

Brunnermeier, Markus and Stefan Nagel (2004). "Hedge Funds and the Technology Bubble." *Journal of Finance*, 59: 2013-2040.

Brunnermeier, Markus, Thomas Eisenbach and Yuliy Sannikov (2013). "Macroeconomics with Financial Frictions: A Survey." in Daron Acemoglu, Manuel Arellano and Eddie Dekel eds., *Advances in Economics and Econometrics, Tenth World Congress of the Econometric Society*.

Brunnermeier, Markus and Martin Oehmke (2013). "Bubbles, Financial Crises, and Systemic Risk." in George Constantinides, Milton Harris, and Rene Stulz eds., *Handbook of the Economics of Finance*, 2B: 1221-1287. North-Holland.

Carlson, Murray, Adlai Fisher and Ron Giammarino (2004). "Corporate Investment and Asset Price Dynamics: Implications for the Cross-Section of Returns." *Journal of Finance*, 61: 171-194.

Cespa, Giovani and Thierry Foucault (2014). "Illiquidity Contagion and Liquidity Crashes." *Review of Financial Studies*, 27: 1615-1660..

Chen, Qi, Itay Goldstein and Wei Jiang (2007). "Price Informativeness and Investment Sensitivity to Stock Price." *Review of Financial Studies*, 20: 619-650.

Daniel, Kent, David Hirshleifer and Avanidhar Subrahmanyam (1998). "A Theory of Overconfidence, Self-attribution, and Security Market Under- and Overrations." *Journal of Finance*, 52: 1-34.

Dasgupta, Amil (2007). "Coordination and Delay in Global Games." *Journal of Economic Theory*, 134: 195-225.

DeMarzo, Peter, Ron Kaniel and Ilan Kremer (2007). "Technological Innovation and Real Investment Booms and Busts." *Journal of Financial Economics*, 85: 735-754.

Di Maggio, Marco and Marco Pagano (2013). "Financial Disclosure and Market Transparency with Costly Information Processing." Working paper.

Dow, James and Gary Gorton (1997). "Stock Market Efficiency and Economic Efficiency: Is There a Connection?" *Journal of Finance*, 52: 1087-1128.

Dow, James, Itay Goldstein and Alexander Guembel (2011). "Incentives for Information Production in Markets where Prices Affect Real Investment." Working Paper.

Edmans, Alex (2009). "Blockholder Trading, Market Efficiency, and Managerial Myopia." *Journal of Finance*, 64: 2481-2513.

Edmans, Alex, Itay Goldstein and Wei Jiang (2012). "The Real Effects of Financial Markets: the Impacts of Prices on Takeovers." *Journal of Finance*, 67: 933-971.

Foucault, Thierry and Laurent Fresard (2012). "Cross-Listing, Investment Sensitivity to Stock Price, and the Learning Hypothesis." *Review of Financial Studies*, 25: 3305-3350.

Foucault, Thierry and Laurent Fresard (2014). "Learning from Peers' Stock Prices and Corporate Investment." *Journal of Financial Economics*, 111: 554-577.

Foucault, Thierry and Thomas Gehrig (2008). "Stock Price Informativeness, Cross-Listings, and Investment Decisions." *Journal of Financial Economics*, 88: 146-168.

Fulghieri, Paolo and Dmitry Lukin (2001). "Information Production, Dilution Costs, and Optimal Security Design." *Journal of Financial Economics*, 61: 3-42.

Geanakoplos, John and Heracles Polemarchakis (1985). "Existence, Regularity, and Constrained Suboptimality of Competitive Allocations When the Asset Market Is Incomplete." Cowles Foundation for Research in Economics Discussion Papers 764.

Goldstein, Itay and Alexander Guembel (2008). "Manipulation and the Allocational Role of Prices." *Review of Economic Studies*, 75: 133-164.

Goldstein, Itay, Emre Ozdenoren and Kathy Yuan (2013). "Trading Frenzies and Their Impact on Real Investment." *Journal of Financial Economics*, 109: 566-582.

Goldstein, Itay and Liyan Yang (2014a). "Market Efficiency and Real Efficiency: The Connect and Disconnect via Feedback Effects." Working Paper.

Goldstein, Itay and Liyan Yang (2014b). "Good Disclosure, Bad Disclosure." Working Paper.

GOLDSTEIN, ITAY and LIYAN YANG (2014c). "Information Diversity and Complementarities in Trading and Information Acquisition." *Journal of Finance*, forthcoming.

GRAHAM, JOHN and CAMPBELL HARVEY (2001). "The Theory and Practice of Corporate Finance: Evidence from the Field." *Journal of Financial Economics*, 61: 187-243.

GREENWALD, BRUCE and JOSEPH STIGLITZ (1986). "Externalities in Economies with Imperfect Information and Incomplete Markets." *Quarterly Journal of Economics*, 101: 229-264.

GREENWOOD, ROBIN and SAMUEL HANSON (2014). "Waves in Ship Prices and Investment." *Quarterly Journal of Economics*, forthcoming.

HELLWIG, CHRISTIAN, ARIJIT MUKHERJI and ALEH TSYVINSKI (2006). "Self-fulfilling Currency Crises: The Role of Interest Rates." *American Economic Review*, 96: 1769-1987.

HIRSHLEIFER, DAVID, AVANIDHAR SUBRAHMANYAM and SHERIDAN TITMAN (2006). "Feedback and the Success of Irrational Investors." *Journal of Finance*, 81: 311-338.

HOBERG, GERARD and GORDON PHILLIPS (2010). "Real and Financial Industry Booms and Busts." *Journal of Finance*, 65: 45-86.

HONG, HARRISON and JEREMY STEIN (1999). "A Unified Theory of Underraction, Momentum Trading and Overreaction in Asset Markets." *Journal of Finance*, 54: 2143-2184.

KOGAN, LEONID and DIMITRIS PAPANIKOLAOU (2014). "Growth Opportunities, Technology Shocks, and Asset Prices." *Journal of Finance*, forthcoming.

KURLAT, PABLO and LAURA VELDKAMP (2013). "Should We Regulate Financial Information?" *Journal of Economic Theory*, forthcoming.

LIU, XUEWEN (2014). "Firm Disclosure, Financial Markets and the Macroeconomy." Working paper.

LUO, YUANZHI (2005). "Do Insiders Learn from Outsiders? Evidence From Mergers and Acquisitions." *Journal of Finance*, 60: 1951-1982.

MAKSIMOVIC, VOJISLAV, GORDON PHILLIPS and LIU YANG (2013). "Private and Public Merger Waves." *Journal of Finance*, 68: 2177-2217.

MENZLY, LIOR and OGUZHAN OZBAS (2010). "Market Segmentation and Cross-Predictability of Returns." *Journal of Finance*, 65: 1555-1580.

Moskowitz, Tobias and Mark Grinblatt (1999). "Do Industries Explain Momentum?" *Journal of Finance*, 54: 1249-1290.

Ozdenoren, Emre and Kathy Yuan (2008). "Feedback Effects and Asset Prices." *Journal of Finance*, 63: 1939-1975.

Ozdenoren, Emre and Kathy Yuan (2013). "Endogenous Contractual Externalities." Working paper.

Ozoguz, Arzu and Michael Rebello (2013). "Information, Competition, and Investment Sensitivity to Peer Stock Prices." Working paper.

Pastor, Lubos and Pietro Veronesi (2006). "Was there a Nasdaq bubble in the late 1990s?" *Journal of Financial Economics*, 81: 61-100.

Pastor, Lubos and Pietro Veronesi (2009). "Technological Revolutions and Stock Prices." *American Economic Review*, 99: 1451-1483.

Peng, Lin and Wei Xiong (2006). "Investor Attention, Overconfidence and Category Learning." *Journal of Financial Economics*, 80: 563-602.

Peng, Lin, Wei Xiong and Tim Bollerslev (2007). "Investor Attention and Time-Varying Comovements." *European Financial Management*, 13: 394-422.

Papanikolaou, Dimitris (2011). "Investment Shocks and Asset Prices." *Journal of Political Economy*, 119: 639-685.

Peress, Joel (2010). "Product Market Competition, Insider Trading, and Stock Market Efficiency." *Journal of Finance*, 65: 1-43.

Rhodes-Kropf, Matthew, David Robinson and S. Viswanathan (2005). "Valuation Waves and Merger Activity: the Empirical Evidence." *Journal of Financial Economics*, 77: 561-603.

Rhodes-Kropf, Matthew and S. Viswanathan (2004). "Market Valuation and Merger Waves." *Journal of Finance*, 59: 2685-2718.

Scharfstein, David and Jeremy Stein (1990). "Herd Behavior and Investment." *American Economic Review*, 80: 465-479.

Sockin, Michael and Wei Xiong (2014a). "Informational Frictions and Commodity Markets." *Journal of Finance*, forthcoming.

Sockin, Michael and Wei Xiong (2014b). “Learning about the Neighborhood: A Model of Housing Cycles.” Working Paper.

Stiglitz, Joseph (1982). “The Inefficiency of the Stock Market Equilibrium.” *Review of Economic Studies*, 49: 241-261.

Subrahmanyam, Avanidhar and Sheridan Titman (2001). “Feedback From Stock Prices to Cash Flows.” *Journal of Finance*, 56: 2389-2413.

Subrahmanyam, Avanidhar and Sheridan Titman (2013). “Financial Market Shocks and the Macroeconomy.” *Review of Financial Studies*, 26: 2687-2717..

Veldkamp, Laura (2006). “Information Markets and the Comovement of Asset Prices.” *Review of Economic Studies*, 73: 823-845.

Veldkamp, Laura (2011). *Information Choice in Macroeconomics and Finance*. Princeton Press.

Vives, Xavier (1997). “Learning from Others: A Welfare Analysis.” *Games and Economic Behavior*, 20: 177-200.

Vives, Xavier (2010). *Information and Learning in Markets: The Impact of Market Microstructure*. Princeton Press.

Vives, Xavier (2014). “Endogenous Public Information and Welfare in Market Games.” Working Paper.

Xiong, Wei (2013). “Bubbles, Crises, and Heterogeneous Beliefs.” in Jean-Pierre Fouque and Joe Langsam eds., *Handbook for Systemic Risk*, forthcoming.

Chapter 3

Financing Entrepreneurial Production: Security Design with Flexible Information Acquistion[1]

3.1 Introduction

Why are projects with different natures usually financed by different types of securities? Specifically, both debt and non-debt securities are commonly viewed in different real-world corporate finance contexts, but it is less clear why debt is optimal for financing some projects while non-debt securities are optimal for others. This paper provides a new answer to this question, under a single and natural premise that the investor can acquire costly information on the entrepreneur's project before making the financing decision.

The literature of security design usually postulates that an entrepreneur with a project but without financial resources proposes specific contracts to an investor to get finance. The entrepreneur is often modeled as an expert who is more informed about the project. However, this common approach misses a crucial point: in reality some investors are better able than the entrepreneur to acquire information and thus to assess a project's uncertain market prospects, drawing upon their industry experience. For instance, start-ups seek venture capital, and most venture capitalists are themselves former founders of successful start-ups, so they may be better able to determine whether new technologies match the market. Tirole (2006) points out that one shortcoming of the classical corporate finance literature is that it overlooks this informational advantage of investors. Our paper addresses this concern by uncovering the interaction between entrepreneurs' security design and investors' endogenous information acquisition and screening.[2] It enables us to provide a theory of debt and non-debt securities within a single framework, and in particular, to show under what conditions debt or non-debt securities will be optimal. These results are consistent with the empirical evidence regarding the finance of entrepreneurial businesses.

---

[1] Co-authored with Ming Yang.

In our model, an entrepreneur has the potential to produce but has no money to start the project, and thus the investor's endogenous information advantage over the entrepreneur[3] leads to a new informational friction. Specifically, in our model, the investor can flexibly (defined later) acquire costly information about the project's uncertain cash flow before making the financing decision. Only when the investor believes that the project is good enough, will she finance the project. Hence, the entrepreneur's real production depends on the investor's information acquisition, but these two are conducted separately, constituting the friction at the heart of our model.

Our model predicts standard debt and the combination of debt and equity[4] as optimal securities in different circumstances. When the project's ex-ante market prospects are already good and clear or the screening cost is high, the optimal security is debt, which does not induce information acquisition. This prediction is consistent with the evidence that conventional start-ups and mature private businesses rely heavily on plain-vanilla debt finance from investors who are not good at screening, such as relatives, friends, and banks (see Berger and Udell, 1998, Kerr and Nanda, 2009, Robb and Robinson, 2014). The intuition is clear: since the benefit of screening does not justify its cost, the entrepreneur finds it optimal to avoid costly information acquisition by using debt, the least information-sensitive security. The investor thus makes the investment decision based on her prior. This intuition for the optimality of debt is different from the conventional wisdom, as our mechanism does not feature adverse selection or signaling.[5]

---

[2] In our model, screening refers to the decision of whether or not to finance the project after acquiring information. As our model does not feature entrepreneurs' private information, our notion of screening is however different from the notion of separating (different types of entrepreneurs) commonly used in the literature.

[3] We do not attempt to deny that entrepreneurs in reality may have private information about their technologies, which has been discussed extensively in the previous literature. Rather, we highlight the overlooked fact that investors may acquire information and become more informed about the potential match between new technologies and the market.

[4] The formal mathematical definitions of debt, equity, and the combination of debt and equity in our framework are given in Sections 3.3.1 and 3.3.2. In defining them, we focus on the qualitative aspects of their cash flow rights but ignore the aspects of control rights. Specifically, debt means the security pays all the cash flow in low states but a constant face value in high states, while equity means the security payoff and its residual are both strictly increasing in the fundamental. Consistent with the reality, debt is also more senior than equity in our framework.

In contrast, when the project's ex-ante market prospects are obscure or the cost of screening is low, the optimal security is the combination of debt and equity that induces the investor to acquire information. Regarding cash flow rights only, this is equivalent to participating convertible preferred stock. This prediction fits well with the empirical facts (Sahlman, 1990, Gompers, 1999). In particular, convertible preferred stocks have been used in almost all the contracts between entrepreneurs and venture investors, and nearly half of them are participating, as documented in Kaplan and Stromberg (2003). Participating convertible preferred stock is popular in particular for the early rounds of investment (Kaplan and Stromberg, 2003), when the friction is more severe.

The optimality of the combination of debt and equity is subtle. First, the entrepreneur wants to induce the investor to screen only if the investor screens in a potentially good project and screens out bad ones.[6] That is, any project with a higher ex-post cash flow should have a better chance to be financed ex-ante. Only when the investor's payoff is high in good states while low in bad states, the investor has the right incentive to distinguish between these different states by developing such a screening rule, because she only wants to invest when the likelihood of high payment is high. Consequently, the entrepreneur can maximally benefit from this by ensuring that her own payoff is also high when the investor's is. Therefore, an equity component with payments that are strictly increasing in the project's cash flow is offered, encouraging the investor to acquire adequate information to distinguish between any different states. In addition, the investor's information after screening may still be imperfect, albeit perhaps with a better posterior. In other words, the investor might still end up financing a bad project after screening. Thus, downside protection is necessary to prevent the investor from rejecting the project without any information acquisition. This justifies the debt component. These intuitions also suggest that straight or leveraged equity alone is not optimal for financing entrepreneurial production, consistent with reality (Kaplan and Stromberg, 2003, Lerner, Leamon and Hardymon, 2012).

---

[5] Notable results regarding debt as the least information-sensitive security to mitigate adverse selection include Myers and Majluf (1984), Gorton and Pennacchi (1990) and DeMarzo and Duffie (1999).

[6] Our model features continuous state, but we use the notions of good and bad at times to help develop intuitions.

One methodological contribution of our work is to characterize the payment structure, or in visual terms, the "shape", of the optimal securities without either distributional assumptions or restrictions on the feasible security space. A new concept, flexible information acquisition, helps achieve this goal. For instance, debt, with its flatter shape, is less likely than equity to prompt screening. Moreover, in screening, a debt holder tends to pay attention to states with low cash flows, as the payments are constant over states with high cash flows so there is no point in differentiating the latter. In contrast, levered equity holders tend to pay attention to states with high cash flows, as they benefits from the upside payments. An arbitrary security determines the investor's incentives for screening and attention allocation in this state-contingent way, and these in turn affect the entrepreneur's incentives in designing the security. The traditional approach of exogenous information asymmetry does not capture these incentives. Recent models of endogenous information acquisition do not capture such flexibility of incentives adequately, since they only consider the amount or precision of information (see Veldkamp, 2011, for a review). Our approach of flexible information acquisition, following Yang (2013, 2015), is based on the literature of rational inattention (Sims, 2003, Woodford, 2008), but has a different focus. It captures not only how much but also what kind of information the investor acquires through state-contingent attention allocation. In our setting, when screening is desirable, the optimal security encourages the investor to allocate adequate attention to all states so as to effectively distinguish potentially good from bad projects, and thus delivers the highest possible ex-ante profit to the entrepreneur. This mechanism helps generate the exact shape of different optimal securities.

Our parsimonious framework accommodates a variety of theoretical corporate finance contexts and real-world scenarios of financing entrepreneurial production. On the one hand, we view the investor as a screening expert, which is natural but often overlooked.[7] As the cost of screening pertains both to the project's nature and to the investor's information expertise, it also allows us to cover various investors, including family and friends, banks, and venture capitalists. On the other hand, we highlight two particular aspects of the entrepreneur, capturing the nature of private businesses that account for most firms. First, the entrepreneur is financially constrained. Second, her human capital is inalienable, which means the investor cannot take over the project and the entrepreneur has bargaining power in designing the security. These settings fit the notion of entrepreneur-led financing proposed by Admati and Pfleiderer (1994) and the idea in Rajan (2012) that entrepreneurs' human capital is important in the early stages of firms' life cycles. These assumptions are also completely benign; even relaxing them does not affect our main results.[8] This paper, to the best of our knowledge, is the first to investigate the interplay between security design and an individual investor's screening in a corporate finance production setting.

---

[7] The acknowledgement of investors' screening dates back to Knight (1921) and Schumpeter (1942). Apart from extensive anecdotal evidence (see Kaplan and Lerner, 2010, Da Rin, Hellmann and Puri, 2011, for reviews), recent empirical literature (Chemmanur, Krishnan and Nandy, 2012, Kerr, Lerner and Schoar, 2014) has also identified direct screening by various types of investors. Theoretical developments in this direction have been surveyed in Bond, Edmans and Goldstein (2012). However, most of those papers focus on the role of competitive financial markets in soliciting or aggregating the information of investors or speculators (for instance, Boot and Thakor, 1993, Fulghieri and Lukin, 2001, Axelson, 2007, Garmaise, 2007, Hennessy, 2013, on security design) rather than the role of screening by individual investors. In reality, most firms are private and do not have easy access to a competitive financial market. A burgeoning security design literature highlights individual investors' endogenous information advantage directly (Dang, Gorton and Holmstrom, 2011, Yang, 2013), but these models are built to capture the asset-backed securities market as an exchange economy and not fit for our setting of financing entrepreneurial production.

[8] The results of optimal securities continue to hold if the project is transferrable or if the entrepreneur does not have full bargaining power in designing the security. See subsection 3.3.3 and appendix C.1.2.

Related Literature. In addition to the security design literature that identifies debt as the most information-insensitive form of finance (as mentioned in footnote 6), this paper is related to a series of theoretical papers that predict that non-debt securities (including equity and convertibles) can be optimal in various circumstances with asymmetric information (see Brennan and Kraus, 1987, Constantinides and Grundy, 1989, Stein, 1992, Nachman and Noe, 1994, Chemmanur and Fulghieri, 1997, Inderst and Mueller, 2006, Chakraborty and Yilmaz, 2011, Chakraborty, Gervais and Yilmaz, 2011, Fulghieri, Garcia and Hackbarth, 2013). Even closer to the present paper are Boot and Thakor (1993), Manove, Padilla and Pagano (2001), Fulghieri and Lukin (2001), Axelson (2007), Garmaise (2007), and Hennessy (2013), all of which highlight the competitive financial markets' role in soliciting or aggregating investors' private information. These papers, however, do not consider an individual investor's screening directly. Also different from these papers, our model allows for state-contingent decisions of information acquisition, which reflects the fact that the investor is able to acquire information in a very detailed, careful manner. This methodology helps to model arbitrary feasible securities over continuous states with arbitrary distributions and information structures, allowing us to characterize the matching between different projects and different optimal securities in an exhaustive way.

Our model also contributes to the venture contract design literature by highlighting screening. Security design is one focus of modern research in entrepreneurial finance and innovation, but the literature mostly focuses on control rights (Berglof, 1994, Hellmann, 1998, Kirilenko, 2001), monitoring (Ravid and Spiegel, 1997, Schmidt, 2003, Casamatta, 2003, Hellmann, 2006), and refinancing and staging (Admati and Pfleiderer, 1994, Bergemann and Hege, 1998, Cornelli and Yosha, 2003, Repullo and Suarez, 2004) and tends to ignore screening. Further, most of these models only focus on one class of optimal security. In contrast, our model unifies debt and non-debt-like securities in a general framework and provides a consistent mapping of their optimality to different real-world circumstances.

A new strand of literature on the real effects of rating agencies (see Kashyap and Kovrijnykh, 2013, Opp, Opp and Harris, 2013) is also relevant. On behalf of investors, the rating agency screens the firm, which does not know its own type. Information acquisition may improve social surplus through ratings and the resulting investment decisions. Unlike this literature, we study how different shapes of securities interact with the incentives to allocate attention in acquiring information and therefore the equilibrium financing choice.

The remaining of this paper proceeds as follows. Section 3.2 specifies the economy. Section 3.3 characterizes the optimal securities. Section 3.4 further characterizes under what conditions debt or non-debt securities will be optimal. Section 3.5 performs comparative statics on the optimal securities. Section 3.6 concludes.

3.2 The Model

We present a model focusing on the interplay between security design and flexible screening. We highlight one friction: the dependence of real production on information acquisition and the former's simultaneous separation from the latter.

3.2.1 Financing Entrepreneurial Production

Consider a production economy with two dates, $t = 0, 1$, and a single consumption good. There are two agents: an entrepreneur lacking financial resources and a deep-pocket investor, both risk-neutral. Their utility function is the sum of consumptions over the two dates: $u = c_0 + c_1$, where c_t denotes an agent's consumption at date t. In what follows we use subscripts E and I to indicate the entrepreneur and the investor, respectively.

The financing process of the entrepreneur's risky project is as follows. To initiate the project at date 0, the underlying technology requires an investment of $k > 0$. If financed, the project generates a non-negative verifiable random cash flow θ at date 1. The project cannot be initiated partially. Hence, the entrepreneur has to raise k, by selling a security to the investor at date 0. The payment of a security at date 1 is a mapping $s : \mathbb{R}_+ \to \mathbb{R}_+$ such that $s(\theta) \in [0, \theta]$ for any θ. We focus only on the cash flow of projects and securities rather than the control rights.

Security design and information acquisition both happen at date 0. The agents have a common prior Π on the potential project's future cash flow θ, and neither party has any private information ex-ante.[9] The entrepreneur designs the security, and then proposes a take-it-or-leave-it offer to the investor at price k. Facing the offer, the investor acquires information about θ in the manner of rational inattention (Sims, 2003, Woodford, 2008, Yang, 2013, 2015), updates beliefs on θ, and then decides whether to accept the offer. The information acquired is measured by reduction of entropy. The information cost per unit reduction of entropy is μ, defined as the cost of screening. We elaborate this information acquisition process in more detail in subsection 3.2.2.

The assumptions implicit in the setting reflect the key features of financing entrepreneurial production, in particular the role of screening. First, the entrepreneur cannot undertake the project except by external finance.[10] This is consistent with the empirical evidence that entrepreneurs and private firms are often financially constrained (Evans and Jovanovic, 1989, Holtz-Eakin, Joulfaian and Rosen, 1994). Even in mature firms, managers may seek outside finance where the internal capital market does not work well for risky projects (Stein, 1997, Scharfstein and Stein, 2000). Second, the investor can acquire information about the cash flow and thus screen the project through her financing decision. This point not only accounts for the empirical evidence but also sets this model apart from most of the previous security design literature, which features the entrepreneur's exogenous information advantage. These two points together lead to the dependence and separation of real production and information acquisition, which is the key friction in our model.

It is worth noting which aspects of finance in the production economy are abstracted away, and how much they affect our analysis. First, to focus on screening, we set aside moral hazard. To ignore moral hazard is common in the security design literature, especially when hidden information is important (see DeMarzo and Duffie, 1999, for a justification). Second, we do not focus on the bargaining process and the allocation of control rights. We assume that the entrepreneur's human capital is inalienable, so that direct project transfer is impossible and the entrepreneur has the bargaining power to design the security.[11] In subsection 3.3.3 we formally demonstrate that even if the project is transferrable, it is not optimal to transfer the project at any fixed price. Moreover, in appendix C.1.2 we discuss a general allocation of bargaining power between the two agents and we show that our main results are unaffected unless the investor's bargaining power is too strong. Third, consistent with the security design literature, we do not allow for partial financing or endogenous investment scale choice. Since our theory can admit any prior distribution, a fixed investment requirement in fact enables us to capture projects with differing natures in an exhaustive sense. Fourth, we do not model the staging of finance, and we accordingly interpret the cash flow θ as already incorporating the consequences of investors' exiting. Hence, each round of investment may be mapped to our model separately with a different prior. Last, we do not model competition among investors. The last two points pertain to the structure of financial markets, which is tangential to the friction we consider.

---

[9] We can interpret this setting as that the entrepreneur may still have some private information about the future cash flow, but she does not have any effective ways to signal that to the investor. Signaling has been extensively discussed in the literature and already well understood, so we leave it aside.

[10] As common in the corporate finance literature (see Tirole, 2006 for an overview), entrepreneurs are typically viewed as financially constrained. Alternatively, the entrepreneur may have her own capital but cannot acquire information, so that she wants to hire an information expert to improve the investment decision. This alternative situation boils down to a consulting problem. A large literature on the delegation of experimentation (for example, Manso, 2011) considers such consulting problems in corporate finance, which is beyond the scope of this paper.

3.2.2 Flexible Information Acquisition

We model the investor's screening by flexible information acquisition (Yang, 2015). This captures the nature of screening and allows us to work with arbitrary securities over continuous states and without distributional assumptions. Fundamentally, the entrepreneur can design the security's payoff structure arbitrarily, which may induce arbitrary incentive of attention allocation by the investor in screening the project. This therefore calls for an equally flexible account of screening to capture the interaction between the shape of the securities and the incentives to allocate attention. To map to the reality, flexible information acquisition also captures the fact that the investor can acquire information in a very detailed and careful manner.

---

[11] This notion of entrepreneur-led financing is also common in the literature (Brennan and Kraus, 1987, Constantinides and Grundy, 1989, Admati and Pfleiderer, 1994). Together with the differentiation argument in Rajan (2012), this assumption also broadly corresponds to the earlier incomplete contract literature, which suggests that ownership should go to the entrepreneur when firms are young (Aghion and Tirole, 1994).

The essence of flexible information acquisition is that it captures not only how much but also which aspects of information an agent acquires. Consider an agent who chooses a binary action $a \in \{0, 1\}$ and receives a payoff $u(a, \theta)$, where $\theta \in \mathbb{R}_+$ is the fundamental, distributed according to a continuous probability measure Π over $\mathbb{R}_+$. Before making a decision, the agent may acquire information through a set of binary-signal information structures, each signal corresponding to one optimal action.[12] Specifically, she may choose a measurable function $m : \mathbb{R}_+ \to [0, 1]$, the probability of observing signal 1 if the true state is θ, and acquire binary signals $x \in \{0, 1\}$ parameterized by $m(\theta)$; $m(\theta)$ is chosen to ensure that the agent's optimal action is 1 (or 0) when observing 1 (or 0). By choosing different functional forms of $m(\theta)$, the agent can make the signal correlate with the fundamental in any arbitrary way.[13] Intuitively, for instance, if the agent's payoff is sensitive to fluctuations of the state within some range $A \subset \mathbb{R}_+$, she would pay more attention to this range by making $m(\theta)$ co-vary more with θ in A. This gives us a desirable account to model an agent's incentive to acquire different aspects of information.

The conditional probability $m(\cdot)$ embodies a natural interpretation of screening. In our setting of financing entrepreneurial production, conditional on a cash flow θ, $m(\theta)$ is the probability of the project's being screened in and thus getting financed. It is state-contingent, capturing the investor's incentive to allocate attention in screening a project. In particular, the absolute value of the first order derivative $|dm(\theta)/d\theta|$ represents the screening intensity: when it is larger, the investor differentiates the states around θ better. Thus, in what follows we call $m(\cdot)$ a screening rule.

---

[12] In general, an agent can choose any information structure. But an agent always prefers binary-signal information structures in binary decision problems. See Woodford (2008) and Yang (2015).

[13] Technically, this allows agents to choose signals drawn from any conditional distribution of the fundamental.

We then characterize the cost of information acquisition. As in Woodford (2008) and Yang (2015), the amount of information conveyed by a screening rule $m(\cdot)$ is defined as the expected reduction of uncertainty through observation of the signal, where the uncertainty associated with a distribution is measured by Shannon's entropy $H(\cdot)$. This reduction from agents' prior entropy to expected posterior entropy can be calculated as:

$$I(m) = \mathbb{E}\left[g\left(m\left(\theta\right)\right)\right] - g\left(\mathbb{E}\left[m(\theta)\right]\right),$$

where $g(x) = x \cdot \ln x + (1-x) \cdot \ln(1-x)$, and the expectation operator $\mathbb{E}(\cdot)$ is with respect to θ under the probability measure Π.[14] Denote by $M = \{m \in L(\mathbb{R}_+, \Pi) : \theta \in \mathbb{R}_+, m(\theta) \in [0,1]\}$ the set of binary-signal information structures, and $c : M \to \mathbb{R}_+$ the cost of information. The cost is assumed to be proportional to the associated mutual information:

$$c(m) = \mu \cdot I(m),$$

where $\mu > 0$ is the cost of information acquisition per unit of reduction of entropy.[15,16]

Built upon flexible information acquisition, the agent's problem is to choose a functional form of $m(\theta)$ to maximize expected payoff less information cost. We characterize the optimal screening rule $m(\theta)$ in the following proposition. We denote $\Delta u(\theta) = u(1,\theta) - u(0,\theta)$, which is the payoff gain of taking action 1 over action 0. We also assume that $\Pr\left[\Delta u(\theta) \neq 0\right] > 0$ to exclude the trivial case where the agent is always indifferent between the two actions. The proof is in Yang (2013) (see also Woodford, 2008, for an earlier treatment).

---

[14] Formally, we have

$$I(m) = H(\Pi) - \int_x H(\Pi(\cdot|x))\Pi_x dx,$$

where Π denotes the prior, x the signal received, $\Pi(\cdot|x)$ the posterior distribution, and Π_x the marginal probability of signal x. Under binary-signal structure, standard calculation yields the result above.

[15] Although the cost $c(m)$ is linear in mutual information $I(m)$, it does not mean it is linear in information acquisition. Essentially, mutual information $I(m)$ is a non-linear functional of the screening rule $m(\cdot)$ and the prior Π, micro-founded by the information theory.

[16] The functional form of the information cost, following the literature of rational inattention, is not crucial in driving our qualitative results. See Woodford (2012) and Yang (2015) for discussions on related properties of this cost function.

PROPOSITION **21.** *Given u, Π, and $\mu > 0$, let $m^*(\theta) \in M$ be an optimal screening rule and*

$$\bar{\pi}^* = \mathbb{E}\left[m^*(\theta)\right]$$

be the corresponding unconditional probability of taking action 1. Then,

i) the optimal screening rule is unique;

ii) there are three cases for the optimal screening rule:

a) $\bar{\pi}^* = 1$, i.e., $Prob[m^*(\theta) = 1] = 1$ if and only if

$$\mathbb{E}\left[\exp\left(-\mu^{-1} \cdot \Delta u(\theta)\right)\right] \leqslant 1\,; \tag{3.2.1}$$

b) $\bar{\pi}^* = 0$, i.e., $Prob[m^*(\theta) = 0] = 1$ if and only if

$$\mathbb{E}\left[\exp\left(\mu^{-1} \cdot \Delta u(\theta)\right)\right] \leqslant 1\,;$$

c) $0 < \bar{\pi}^* < 1$ and $Prob[0 < m^*(\theta) < 1] = 1$ if and only if

$$\mathbb{E}\left[\exp\left(\mu^{-1} \cdot \Delta u(\theta)\right)\right] > 1 \text{ and } \mathbb{E}\left[\exp\left(-\mu^{-1} \cdot \Delta u(\theta)\right)\right] > 1\,; \tag{3.2.2}$$

in this case, the optimal screening rule $m^*(\theta)$ is determined by the equation

$$\Delta u(\theta) = \mu \cdot \left(g'\left(m^*(\theta)\right) - g'\left(\bar{\pi}^*\right)\right) \tag{3.2.3}$$

for all $\theta \in \mathbb{R}_+$, where

$$g'(x) = \ln\left(\frac{x}{1-x}\right)\,.$$

Proposition 21 fully characterizes the agent's possible optimal decisions of information acquisition. Case a) and Case b) correspond to the scenarios of optimal action 1 or 0. These two cases do not involve information acquisition. They correspond to the scenarios in which the prior is extreme or the cost of information acquisition is sufficiently high. But case c), the more interesting one, involves information acquisition. In particular, the optimal screening rule $m^*(\theta)$ is not constant in this case, and neither action 1 nor 0 is optimal ex-ante. This case corresponds to the scenario where the prior is not extreme, or the cost of information acquisition is sufficiently low. In Case c) where information acquisition is involved, the agent equates the marginal benefit of information with its marginal cost, as indicated by condition (3.2.3). So doing, the agent chooses the shape of $m^*(\theta)$ according to the shape of payoff gain $\Delta u(\theta)$ and her prior Π.[17] In the next section we will see that the shape of $m^*(\theta)$ is crucial in characterizing the way in which the investor screens a project.

3.3 Security Design

Now let us consider the entrepreneur's security design problem. Denote the optimal security of the entrepreneur by $s^*(\theta)$. The entrepreneur and the investor play a dynamic Bayesian game. Concretely, the entrepreneur designs the security, and then the investor screens the project given the security designed. Hence, we apply Proposition 21 to the investor's information acquisition problem, given the security, and then solve backwards for the entrepreneur's optimal security. To distinguish from the general decision problem in Section 3.2.2, we denote the investor's optimal screening rule as $m_s(\theta)$, given the security $s(\theta)$; hence the investor's optimal screening rule is now denoted by $m_s^*(\theta)$.

We formally define the equilibrium as follows.

Definition 8. *Given u, Π, k and $\mu > 0$, the sequential equilibrium is defined as a combination of the entrepreneur's optimal security $s^*(\theta)$ and the investor's optimal screening rule $m_s(\theta)$ for any generic security $s(\theta)$, such that*

i) the investor optimally acquires information given any generic security $s(\theta)$: $m_s(\theta)$ is prescribed by Proposition 21,[18] *and*

ii) the entrepreneur designs the optimal security:

$$s^*(\theta) \in \arg \max_{0 \leqslant s(\theta) \leqslant \theta} \mathbb{E}[m_s(\theta) \cdot (\theta - s(\theta))] \, .$$

According to Proposition 21, there are three possible investor behaviors, given the entrepreneur's optimal security. First, the investor may optimally choose not to acquire information and simply accept the security as proposed. This implies that the project is certainly financed. Second, the investor may optimally acquire some information, induced by the proposed security, and then accept the entrepreneur's optimal security with a positive probability. In this case, the project is financed with a probability that is positive but less than one. Third, the investor may simply reject the security without acquiring information, which implies that the project is certainly not financed. All the three cases can be accommodated by the equilibrium definition. This third case, however, represents the outside option of the entrepreneur, who can always offer nothing to the investor and drop the project. Accordingly, we focus on the first two cases. The following lemma helps distinguish the first two types of equilibrium from the third.

---

[17] See Woodford (2008), Yang (2013, 2015) for more examples on this decision problem.

[18] The specification of belief for the investor at any generic information set after information acquisition is implied by Proposition 21, provided the definition of $m_s(\theta)$.

Lemma 18. *The project can be financed with a positive probability if and only if*

$$\mathbb{E}\left[\exp(\mu^{-1}\cdot(\theta-k))\right] > 1\,. \tag{3.3.1}$$

Lemma 18 is an intuitive investment criterion. It implies that the project is more likely to get financed if the prior of the cash flow is better, if the initial investment k is smaller, or if the cost of screening μ is lower. When condition (3.3.1) is violated, the investor will reject the proposed security, whatever it is.

Condition (3.3.1) appears different from the ex-ante NPV criterion, which suggests that a project should be financed for sure when $\mathbb{E}[\theta] - k > 0$. In our model with screening, by Jensen's inequality, condition (3.3.1) implies that any project with positive ex-ante NPV will be financed with a positive probability. Moreover, some projects with negative ex-ante NPV may also be financed with a positive probability. This is consistent with our idea that real production depends on information acquisition. With the potential of screening, the ex-ante NPV criterion based on a fixed prior is generalized to a new information-adjusted one to admit belief updating.

The following Corollary 18 implies that the entrepreneur will never propose to concede the entire cash flow to the investor if the project is financed. This corollary is straightforward but worth stressing, in that it helps illustrate the key friction by showing that the interests of the entrepreneur and the investor are not perfectly aligned.

Corollary **4.** *When the project can be financed with a positive probability,* $s^*(\theta) = \theta$ *is not an optimal security.*

In what follows, we assume that condition (3.3.1) is satisfied, and characterize the entrepreneur's optimal security, focusing on the first two types of equilibria with a positive screening cost $\mu > 0$. As we will see, the entrepreneur's optimal security differs between the two cases, which implies that the investor screens the project differently. We further show that to transfer the project at a given price is always sub-optimal, which also justifies the security design approach. Finally, for additional intuitions, we consider two limiting cases, one with infinite and one with zero screening cost.

3.3.1 Optimal Security without Inducing Information Acquisition

In this subsection, we consider the case in which the entrepreneur's optimal security is accepted by the investor without information acquisition. In other words, the entrepreneur finds screening not worthwhile and wants to design a security to deter it. Concretely, this means $Pr\left[m_s^*(\theta) = 1\right] = 1$. We first consider the investor's problem of screening, given the entrepreneur's security, then characterize the optimal security.

Given a security $s(\theta)$, the investor's payoff gain from accepting rather than rejecting the security is

$$\Delta u_I(\theta) = u_I(1, \theta) - u_I(0, \theta) = s\left(\theta\right) - k\,. \tag{3.3.2}$$

According to Proposition 21 and conditions (3.2.1) and (3.3.2), any security $s(\theta)$ that is accepted by the investor without information acquisition must satisfy

$$\mathbb{E}\left[\exp\left(-\mu^{-1} \cdot \left(s\left(\theta\right) - k\right)\right)\right] \leqslant 1\,. \tag{3.3.3}$$

If the left-hand side of inequality (3.3.3) is strictly less than one, the entrepreneur could lower $s(\theta)$ to some extent to increase her expected payoff gain, without affecting the investor's incentives. Hence, condition (3.3.3) always holds as an equality in equilibrium.

By backward induction, the entrepreneur's problem is to choose a security $s(\theta)$ to maximize her expected payoff

$$u_E(s(\cdot)) = \mathbb{E}\left[\theta - s(\theta)\right]$$

subject to the investor's information acquisition constraint

$$\mathbb{E}\left[\exp\left(-\mu^{-1} \cdot \left(s\left(\theta\right) - k\right)\right)\right] = 1\,,$$

and the feasibility condition $0 \leqslant s(\theta) \leqslant \theta$.[19]

In this case, the entrepreneur's optimal security is a debt. We characterize this optimal security by the following proposition, along with its graphical illustration in Figure 3.1. It is easy to see that the face value of the debt is unique in this case.

PROPOSITION **22.** *If the entrepreneur's optimal security $s^*(\theta)$ induces the investor to accept the security without acquiring information, then it takes the form of a debt:*

$$s^*(\theta) = \min\left(\theta, D^*\right)$$

where the unique face value D^* is determined in equilibrium.

It is intuitive that debt is the optimal means of finance when the entrepreneur finds it optimal not to induce information acquisition. Since screening is not worthwhile, and the entrepreneur wants to design a security to deter it, debt is the least information-sensitive form that provides the entrepreneur's desired expected payoff. From another perspective, the optimal security enables the investor to break even between acquiring and not acquiring information. Specifically, it is the flat part of debt that mitigates the investor's incentive to acquire information to the extent at which she just gives up acquiring information, while delivering the highest possible expected payoff to the entrepreneur. This implies that the

---

[19] With this feasibility condition, the entrepreneur's individual rationality constraint $\mathbb{E}\left[\theta - s(\theta)\right] \geqslant 0$ is automatically satisfied, which is also true for the later case with information acquisition. This comes from the fact that the entrepreneur has no endowment. It also implies that the entrepreneur always prefers to undertake the project, which is consistent with real-world practices. However, it is not correct to interpret this as that the entrepreneur would like to contract with any investor, as we do not model the competition among different investors.

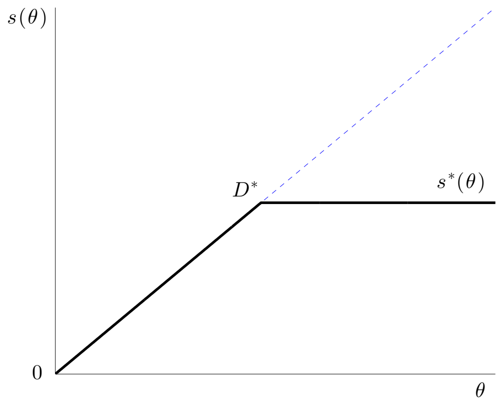


Figure 3.1: *The Unique Optimal Security without Information Acquisition*

optimal security should be as flat as possible when the limited liability constraint is not binding, which leads to debt.

Interestingly, although in this case the investor only provides material investment (rather than acquire costly information), the expected payment $\mathbb{E}[s^*(\theta)]$ exceeds the investment requirement k. The extra payment exceeding k works as a premium to make the investor comfortable with accepting the offer with certainty, as otherwise, without screening she might worry about financing a potentially bad project.

On the other hand, however, the prediction here implies that there is always a limit to the amount of optimal leverage, even without resorting to costs of financial distress which is typical in trade-off theories of debt. In our framework, intuitively, this derives from the separation of real production and information acquisition. As long as the face value is high enough to prevent costly and unnecessary information acquisition, the entrepreneur wants to retain as much as possible, inducing a cap on the optimal face value of debt.

The optimality of debt here accounts for the real-world scenarios in which new projects are financed by fixed-income securities. When a project's market prospects are good and thus not much extra information is needed, it is optimal to deter or mitigate investor's costly information acquisition by resorting to a debt security, which is the least information-sensitive. Interestingly, the rationale for debt in our model does not feature adverse selection, but rather a cost-benefit trade-off of screening. Empirical evidence suggests that many conventional businesses and less revolutionary start-ups relying heavily on plain vanilla debt finance from investors who are not good at screening, such as relatives, friends, and traditional banks (for example, Berger and Udell, 1998, Kerr and Nanda, 2009, Robb and Robinson, 2014), as opposed to more sophisticated financial contracts with venture capital or buyout funds.

The optimality of debt described here is also conceptually different from that in Yang (2013), who considers security design with flexible information acquisition in a comparable exchange economy. In that model, a seller has an asset in place and proposes a security to a more patient buyer to raise liquidity. The buyer can acquire information about the asset's cash flow before purchasing. In that model, debt is optimal because it offers the greatest mitigation of the buyer's adverse selection and hence helps achieve a higher selling price. In the present production economy, however, the investment requirement is fixed, and the optimality of debt derives from the aforementioned cost-benefit analysis of screening (i.e., information acquisition).

3.3.2 Optimal Security Inducing Information Acquisition

Here we characterize the entrepreneur's optimal security that does induce the investor to acquire information and to accept the security with positive probability but not certainty. In other words, the entrepreneur finds screening desirable in this case and designs a security to incentivize it. According to Proposition 21, this means $Prob\left[0 < m_s(\theta) < 1\right] = 1$.

Again, according to Proposition 21 and conditions (3.2.2) and (3.3.2), any generic security $s(\theta)$ that induces the investor to acquire information must satisfy

$$\mathbb{E}\left[\exp\left(\mu^{-1}\left(s\left(\theta\right)-k\right)\right)\right] > 1 \tag{3.3.4}$$

and

$$\mathbb{E}\left[\exp\left(-\mu^{-1}\left(s\left(\theta\right)-k\right)\right)\right] > 1\,, \tag{3.3.5}$$

Given such a security $s(\theta)$, Proposition 21 and condition (3.2.3) also prescribe that the investor's optimal screening rule $m_s(\theta)$ is uniquely characterized by

$$s(\theta) - k = \mu \cdot \left(g'(m_s(\theta)) - g'(\overline{\pi}_s)\right) \ , \tag{3.3.6}$$

where

$$\overline{\pi}_s = \mathbb{E}\left[m_s(\theta)\right]$$

is the investor's unconditional probability of accepting the security. In what follows, we denote by $\overline{\pi}_s^*$ the unconditional probability induced by the entrepreneur's optimal security $s^*(\theta)$.

We derive the entrepreneur's optimal security backwards. Taking account of investor's response $m_s(\theta)$, the entrepreneur chooses a security $s(\theta)$ to maximize

$$u_E(s(\cdot)) = \mathbb{E}\left[m_s(\theta) \cdot (\theta - s(\theta))\right] \tag{3.3.7}$$

subject to (3.3.4), (3.3.5), (3.3.6), and the feasibility condition $0 \leqslant s(\theta) \leqslant \theta$.[20]

We first offer an intuitive roadmap to investigate the optimal security and the associated screening rule, highlighting their key properties. Then we follow with a formal proposition to characterize the optimal security. The detailed derivation is in Appendix C.1.1.[21]

First, the investor's optimal screening rule $m_s^*(\theta)$, induced by the optimal security $s^*(\theta)$, must increase in θ. When the entrepreneur finds it optimal to induce information acquisition, screening by the investor benefits the entrepreneur. Effective screening makes sense only if the investor screens in a potentially good project and screens out bad ones; otherwise it lowers the total social surplus. Under flexible information acquisition, this implies that $m_s^*(\theta)$ should be more likely to generate a good signal and to result in a successful finance when the cash flow θ is higher, while more likely to generate a bad signal and a rejection when θ is lower. Therefore, $m_s^*(\theta)$ should be increasing in θ. As we will see, the monotonicity of

---

[20] Again, the entrepreneur's individual rationality constraint $\mathbb{E}\left[m_s(\theta) \cdot (\theta - s(\theta))\right] \geqslant 0$ is automatically satisfied.

[21] To facilitate understanding, the intuitive investigation of the optimal security is not organized in the same order as the derivation goes in the Appendix, but all the claims in the main text are guaranteed by the formal proofs.

$m_s^*(\theta)$ and the shape of $s^*(\theta)$ are closely interrelated.

To induce an increasing optimal screening rule $m_s^*(\theta)$, the optimal security $s^*(\theta)$ must be increasing in θ as well, according to the first order condition of information acquisition (3.3.6). Intuitively, this monotonicity reflects the dependence of real production on information acquisition: the entrepreneur is willing to compensate the investor more in the event of higher cash flow to encourage effective screening. Unlike the classical security design literature, which often restricts the feasible set to non-decreasing securities, our model uncovers the intrinsic force that drives the pervasiveness of increasing securities in reality.

We also argue that the non-negative constraint $s(\theta) \geqslant 0$ is not binding for the optimal security $s^*(\theta)$ for any $\theta > 0$. Suppose $s^*(\widetilde{\theta}) = 0$ for some $\widetilde{\theta} > 0$. Since $s^*(\theta)$ is increasing in θ, for all $0 \leqslant \theta \leqslant \widetilde{\theta}$ we must have $s^*(\theta) = 0$. This violates the foregoing argument that $s^*(\theta)$ must be increasing in θ. Intuitively, zero payment in states with low cash flows gives the investor too little incentive to acquire information, which is not optimal for the entrepreneur. The security with zero payment in states with low cash flows looks closest to levered common stock, which is the least commonly used security between entrepreneurs and investors in practice (Kaplan and Stromberg, 2003, Kaplan and Lerner, 2010, Lerner, Leamon and Hardymon, 2012).

For closer examination of the optimal security, a perturbation argument on the security design problem gives the entrepreneur's first order condition. Specifically, denote by $r^*(\theta)$ the marginal contribution to the entrepreneur's expected payoff $u_E(s(\cdot))$ of any feasible perturbation to the optimal security $s^*(\theta)$.[22] As $s^*(\theta) > 0$ for any $\theta > 0$, it is intuitive to show that for any $\theta > 0$:

$$r^*(\theta) \begin{cases} = 0 & \text{if} \quad 0 < s^*(\theta) < \theta \\ \geqslant 0 & \text{if} \quad\quad s^*(\theta) = \theta \end{cases},$$

---

[22] Formally, $r^*(\theta)$ is the Frechet derivative, the functional derivative used in the calculus of variations, of $u_E(s(\cdot))$ at $s^*(\theta)$. It is analogical to the commonly used derivative of a real-valued function of a single real variable but generalized to accommodate functions on Banach spaces.

which is further shown to be equivalent to

$$(1 - m_s^*(\theta)) \cdot (\theta - s^*(\theta) + w^*) \begin{cases} = \mu & \text{if} \quad 0 < s^*(\theta) < \theta \\ \geqslant \mu & \text{if} \quad \quad s^*(\theta) = \theta \end{cases}, \tag{3.3.8}$$

where w^* is a constant determined in equilibrium.

We argue that the optimal security $s^*(\theta)$ follows the 45° line in states with low cash flows and then increases in θ with some smaller slope in states with high cash flows. That is, the residual of the optimal security, $\theta - s^*(\theta)$, also increases in θ in states with high cash flows. According to the entrepreneur's first order condition (3.3.8) and the monotonicity of $m_s^*(\theta)$, if $s^*(\widehat{\theta}) = \widehat{\theta}$ for some $\widehat{\theta} > 0$, it must be $s^*(\theta) = \theta$ for any $0 < \theta < \widehat{\theta}$.[23] Similarly, if $s^*(\widehat{\theta}) < \widehat{\theta}$ for some $\widehat{\theta} > 0$, then for any $\theta > \widehat{\theta}$ it must be $s^*(\theta) < \theta$, again by condition (3.3.8) and the monotonicity of $m_s^*(\theta)$. In addition, Corollary 4 rules out $s^*(\theta) = \theta$ for all $\theta > 0$ as an optimal security. Thus, since $s^*(\theta)$ is increasing in θ, the limited liability constraint can only be binding in states with low cash flows.[24] Importantly, given condition (3.3.8) and, again the monotonicity of $m_s^*(\theta)$, when the limited liability constraint is not binding in states with high cash flows, not only $s^*(\theta)$ but also $\theta - s^*(\theta)$ are increasing in θ. In other words, $s^*(\theta)$ is dual monotone when it deviates from the 45° line in states with high cash flows.

The shape of the optimal security $s^*(\theta)$ reflects the friction of the economy. Recall that the monotonicity of $s^*(\theta)$ reflects the dependence of real production on information. The monotonicity of $\theta - s^*(\theta)$, however, reflects their separation: the entrepreneur wants to retain as much as possible while incentivizing the investor to screen the project. Specifically, the area between $s^*(\theta)$ and the 45° line not only captures the entrepreneur's retained benefit, but also reflects the degree to which the allocation of cash flow is inefficient when screening is desirable. This is intuitive: dependence implies that the investor should get all the cash flow, but separation precludes proposing such a deal, as shown in Corollary 4. The competition of the two forces is alleviated in a most efficient way: rewarding the investor more but also retaining more in better states. In this sense, again, our prediction of dual monotonicity derives endogenously from the friction of the economy, whereas in the previous literature it is commonly posited by assumptions.

---

[23] This argument can be seen by contradiction. Suppose $s^*(\theta) < \theta$ when $0 < \theta < \widehat{\theta}$. By the monotonicity of $m_s^*(\theta)$, we know that $1 - m_s^*(\theta) > 1 - m_s^*(\widehat{\theta})$. We also know that $\theta - s^*(\theta) > \widehat{\theta} - s^*(\widehat{\theta}) = 0$. Thus, we have $(1 - m_s^*(\theta))(\theta - s^*(\theta) + w^*) > (1 - m_s^*(\widehat{\theta}))(\widehat{\theta} - s^*(\widehat{\theta}) + w^*) \geqslant \mu$, the last inequality following the second row of condition (3.3.8) because $s^*(\widehat{\theta}) = \widehat{\theta}$. But this in turn implies that $(1 - m_s^*(\theta))(\theta - s^*(\theta) + w^*) > \mu$, which violates the first row of condition (3.3.8) because $s^*(\theta) < \theta$ requires $(1 - m_s^*(\theta))(\theta - s^*(\theta) + w^*) = \mu$, a contradiction.

[24] In the formal proofs we further show that the limited liability constraint must be binding for some states $(0, \widehat{\theta})$ with $\widehat{\theta} > 0$.

Formally, the following proposition characterizes the optimal security $s^*(\theta)$ that induces the investor to acquire information.

Proposition 23. *If the entrepreneur's optimal security $s^*(\theta)$ induces the investor to acquire information, then it takes the following form:*

$$s^*(\theta) = \begin{cases} \theta & \text{if} \quad 0 \leqslant \theta \leqslant \widehat{\theta} \\ \widehat{s}(\theta) & \text{if} \quad \theta > \widehat{\theta} \end{cases},$$

where $\widehat{\theta}$ is determined in equilibrium and the unconstrained part $\widehat{s}(\theta)$ satisfies:

i) $\widehat{\theta} < \widehat{s}(\theta) < \theta$;

ii) $0 < d\widehat{s}(\theta)/d\theta < 1$.[25]

Finally, the corresponding optimal screening rule satisfies $dm_s^*(\theta)/d\theta > 0$.

Proposition 23 offers a clear prediction on the entrepreneur's optimal security when screening is desirable. The form of this security most closely resembles participating convertible preferred stock, with $d\widehat{s}(\theta)/d\theta$ as the conversion ratio, which grants holders the right to receive both the face value and their equity participation as if it was converted, in the real-world event of a public offering or sale. The payoff structure shown in Figure 3.2 may

---

[25] In Appendix C.1.1, we provide the implicit function that determines $d\widehat{s}(\theta)/d\theta$ and further show that $d^2\widehat{s}(\theta)/d\theta^2 < 0$, which implies that the unconstrained part is concave, as illustrated in Figure 3.2. Specifically, we interpret this as a state-contingent conversion ratio, with which the entrepreneur retains more shares in better states. Kaplan and Stromberg (2003) have documented the frequent use of contingent contracts in venture finance and private equity buyouts, which is consistent with the state-contingent conversion ratio described here.

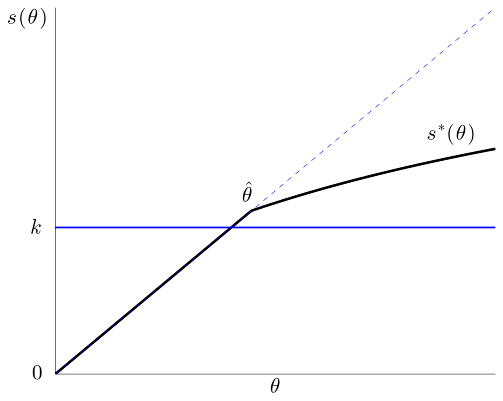


Figure 3.2: *The Unique Optimal Security with Information Acquisition*

be also interpreted as debt plus equity (common stock),[26] or participating convertible debt.[27] This prediction is consistent with the empirical evidence of venture contracts documented in Kaplan and Stromberg (2003), who find that 94.4% of all financing contracts are convertible preferred stock,[28] among which 40.8% are participating, and the participating feature is especially frequent in the earlier rounds of investment. Kaplan and Stromberg (2003) suggest that participating is preferable even to straight convertible preferred stock for screening purpose. Our model implies the same: straight convertible preferred stock is better than debt but still not optimal, because its flat payoffs in intermediate states do not provide enough incentive for the investor to differentiate these states. Our prediction also fits in line with earlier evidence (Sahlman, 1990, Bergemann and Hege, 1998, Gompers, 1999) on the popularity of participating convertible preferred stock and the combination of debt and equity in financing young firms and new projects. For brevity, in what follows we refer to the optimal security in this case as convertible preferred stock or the combination of debt and equity, and use the two terms interchangeably.

---

[26] The package of redeemable preferred stock (can be viewed as a form of debt) and common stock is used as equivalent to participating convertible preferred stock in practice. But the package is not popular, since it is harder to assign reasonable value to each component of the package. See Lerner, Leamon and Hardymon (2012) for more detailed discussion on this point. Also see Stein (1992) and Cornelli and Yosha (2003) for theoretical expositions on the advantage of convertible securities over the combination of debt and equity when characteristics other than the cash flow rights are taken into account.

[27] Compared to equity (common stock), debt and preferred stock are identical in our model, as the model only features two tranches and no dividends.

[28] If we include convertible debt and the combination of debt and equity, this number increases to 98.1%.

Flexible information acquisition plays an important role in predicting the shape of convertible preferred stock. When screening is desirable, on the one hand, a globally increasing security incentivizes the investor to pay sufficient attention to all states so as to discriminate between potentially good and bad projects. On the other hand, a higher conversion ratio $d\widehat{s}(\theta)/d\theta$ induces the investor to screen the underlying project more intensively, at a higher cost to the entrepreneur.[29] Hence, the entrepreneur weighs the benefit of screening against its cost by choosing the optimal conversion ratios on a state-contingent basis to ensure the highest possible ex-ante profit.

Our prediction of the multiple of convertible preferred stock, defined as the ratio of the face value $\widehat{\theta}$ to the investor's initial investment k, is also consistent with the empirical evidence (Kaplan and Stromberg, 2003, Lerner, Leamon and Hardymon, 2012). The multiple is a key characteristic of convertible preferred stock, sometimes taken as analogous to investment returns.

COROLLARY **5.** *The multiple of convertible preferred stock, as the optimal security $s^*(\theta)$ inducing information acquisition, is greater than one, that is, $\widehat{\theta} > k$.*

Like the debt case, this property again derives from the fact that the entrepreneur should offer the investor a premium for accepting the offer with positive probability. Even after screening, the investor's information is still imperfect. Thus, this premium makes the investor comfortable with not rejecting the offer, because even with information acquisition the investor may still end up financing a potentially bad project.

Finally, comparison of the production with an exchange economy helps show why our model can predict both debt and non-debt securities. In a production economy (a primary financial market), costly information contributes to the output, whereas in an exchange economy (a secondary financial market) it only helps reallocate existing resources.

---

[29] In the Appendix C.1.1, we formally show that the screening intensity $|d\widehat{m}(\theta)/d\theta|$ in the converting region is increasing in $d\widehat{s}(\theta)/d\theta$.

Specifically, in an exchange economy as modeled in Dang, Gorton and Holmstrom (2011) and Yang (2013), information is socially wasteful for the following two reasons. First, the information acquisition leads to endogenous adverse selection and results in illiquidity. Second, information is costly per se. As a result, to discourage information acquisition is desirable. In the present paper, however, the entrepreneur and the investor jointly tap the project's cash flow if the investor accepts the proposed security, not if the security is rejected. Thus, the present model features a production economy in which the social surplus may depend positively on costly information. In this case, adverse selection is no longer the focus. Instead, the entrepreneur may want to design a security that encourages the investor to acquire information favorable to the entrepreneur. That is, debt may no longer be optimal when information acquisition is desirable.[30]

3.3.3 Project Transfer

This subsection considers the potential for transferring the project at a fixed non-negative price and shows that it is not optimal even though agents are risk neutral. It can be seen as tantamount a scenario in which the entrepreneur works for the investor for a fixed wage, without moral hazard on the entrepreneur's side. We show that even if the project is transferrable, the entrepreneur still finds it not optimal.

The key to understand this idea is to posit project transfer as one of the feasible securities, as in Figure 3.3, and show that this security is not optimal. When the entrepreneur proposes a project transfer to the investor at a fixed price $p \geqslant 0$, it is equivalent to proposing a security $s(\theta) = \theta - p$ without the non-negative constraint $s(\theta) \geqslant 0$. To see why, notice that if the investor accepts the offer of transfer and undertakes the project, she gets the entire cash flow θ and pay p as an upfront cost. This interpretation allows us to analyze project transfer in our security design framework.

---

[30] This contrast is reminiscent of Hirshleifer (1971), which distinguishes between information value in an exchange and in a production economy. Earlier mechanism design literature on information gathering also hints at this difference, suggesting that the contribution of information provision on liquidity would differ accordingly (Cremer and Khalil, 1992, Cremer, Khalil and Rochet, 1998a,b).

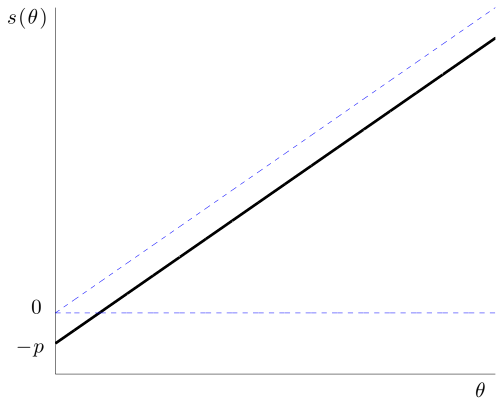


Figure 3.3: *Project Transfer as A Security*

To see why the equivalent security $s(\theta) = \theta - p$ is feasible but not optimal, observe that the non-negative constraint $s(\theta) \geqslant 0$ is not binding in either case of security design, as shown in Section 3.3.1 and Section 3.3.2. Hence, we may also consider a larger set of feasible securities, which is still restricted by the limited liability constraint $s(\theta) \leqslant \theta$ but allows negative payoffs to the investor. As debt and the combination of debt and equity are still the only two optimal securities in this generalized problem, and $s(\theta) = \theta - p$ (project transfer) is feasible, we conclude that project transfer is not optimal to the entrepreneur at any transfer price p. Intuitively, project transfer is not optimal because it does not follow the least costly way to compensate the investor, whether or not information acquisition is induced.

Proposition **24.** *Transferring a project at a fixed price $p \geqslant 0$ is always sub-optimal for the entrepreneur.*

The timeline and sequence of moves in the economy is important for Proposition 24. Consider an alternative sequence in which the investor acquires information only after the transfer. In this economy friction is no longer present, because real production and information acquisition are both performed by the investor after the transfer. With this timeline, project transfer is optimal to the entrepreneur through setting a price that is equivalent to the expected profit of the investor. In this case, the entrepreneur's bargaining power is too strong in the sense that she can prevent the investor from acquiring information when proposing the transfer deal, which essentially removes the friction from the economy. In practice, however, it is common (and reasonable) for investors to have the option of acquiring information about the project before the transfer. This justifies our sequence of moves and suggests that transfer is not optimal when screening is inevitable.

3.4 Optimal Securities in Different Circumstances

A natural question is: in our production economy, when is debt optimal, and when is the combination of debt and equity (convertible preferred stock) optimal? Having characterized the optimal securities with and without inducing screening, we take them together and determine the optimal security given the characteristics of the production economy. This implies that the entrepreneur chooses different optimal securities and thus different capital structures in order to finance projects with differing natures. We focus on cases where the project can be financed with a positive probability and thus security design makes sense, that is, when condition (3.3.1) is satisfied.

3.4.1 NPV Dimension

We first investigate how the optimal security varies when the ex-ante NPV of the project is different.

PROPOSITION **25.** *Consider the ex-ante NPV (i.e.,* $\mathbb{E}[\theta] - k$*) of the project:*

i). if $\mathbb{E}[\theta] - k \leqslant 0$*, the optimal security* $s^*(\theta)$ *is convertible preferred stock; and*

ii). if $\mathbb{E}[\theta] - k > 0$*,* $s^*(\theta)$ *may be either convertible preferred stock or debt.*

Intuitively, a negative NPV project can only be financed by convertible preferred stock: only through screening could the investor's belief be updated. This is consistent with the conventional wisdom that a negative NPV project can never be financed by debt with a given, fixed belief.

Interestingly, convertible preferred stock may be optimal for financing both negative- and positive-NPV projects, but the underlying mechanisms of screening are subtly different. In both cases, the dependence of real production on information acquisition is strong.

When the project has a zero or negative NPV, convertible preferred stock is used to encourage the investor to screen in a potentially good project. The investor will never finance the project without screening it, because it incurs an expected loss even if the entrepreneur promises the entire cash flow. Thus, if it is to be financed, the only way is to use convertible preferred stock to encourage screening. This implies that the dependence of real production on information acquisition is strong due to the relatively poor prior, and thus the friction is accordingly severe. When the investor acquires information, she may expect either a good signal, which leads to a deal, or a bad signal, which results in a rejection, but the ex-ante probability of financing the project becomes positive since a potentially good project can be screened in. Hence, the entrepreneur is better off by proposing convertible preferred stock.

In contrast, when the project has a positive NPV, convertible preferred stock may still be used, but now the aim is to encourage the investor to screen out a potentially bad project. Here, the dependence of real production on information acquisition is still strong due to a relatively mediocre prior. In the status quo where the investor is unable to screen the project, the entrepreneur can finance the positive-NPV project certainly (i.e., with probability one) by proposing debt with a sufficiently high face value. However, when the investor can acquire information, such certain financing (with probability one) could be too expensive, because it leaves too little for the entrepreneur. Instead, the entrepreneur could retain more by offering convertible preferred stock, less generous, and invite the investor to screen the project. Although this results in less than certain financing for the project, the entrepreneur's total expected profit could be higher, since a potentially bad project may be screened out, which justifies convertible preferred stock as optimal.

Finally, debt may be optimal for some positive-NPV projects. When the prior is sufficiently good, the dependence of real production on information acquisition is weak, and thus the benefit from screening does not justify the cost. In this case, it is optimal for the entrepreneur to propose debt to deter costly screening while still retaining enough profit.

3.4.2 The Efficiency Dimension

We then go deeper to investigate how the optimal security varies when the extent of the informational friction in the economy changes. This helps reveal the mechanism of this model at a more fundamental level.

To understand how the optimal security evolves with the severity of friction, we consider a frictionless centralized economy in which real production and information acquisition are aligned. We define a new efficiency dimension with help of this centralized economy. If and only if the friction in the decentralized economy is not severe in the sense that an optimal security can achieve efficiency, the optimal security is debt and screening is not induced in equilibrium. If and only if the friction is severe in the sense that an optimal security cannot achieve efficiency, the optimal security is convertible preferred stock and screening is induced. This dichotomy again highlights the close interconnection of the shape of the optimal security, the role of screening, and the extent of friction in the production economy.

First let us define the expected social surplus and the efficiency dimension. In the decentralized economy, the expected surplus is the difference between the expected profit of the project and the cost of screening, both of which are functions of the screening rule. Thus, an optimal security achieves efficiency if the induced optimal screening rule maximizes expected social surplus in equilibrium.

Definition 9. *An optimal security in the decentralized economy achieves efficiency if and only if the associated optimal screening rule* $m_s^*(\theta)$ *satisfies:*

$$m_s^*(\theta) \in \arg\max_{0 \leqslant m(\theta) \leqslant 1} \mathbb{E}[m(\theta) \cdot (\theta - k)] - \mu \cdot I(m(\theta)) \,. \tag{3.4.9}$$

To facilitate discussion, we characterize a frictionless centralized economy to help benchmark the friction in the corresponding decentralized economy. In the centralized economy, u, Θ, Π, k and μ are given as the same. However, we assume that the entrepreneur has sufficient initial wealth and can also screen the project. Thus, real production still depends on information acquisition, but the two are aligned. In this economy, security design is irrelevant. The entrepreneur's problem is to decide whether to undertake the project directly, to screen it, or to abandon it. The entrepreneur's payoff gain from undertaking the project rather than abandoning it is

$$\Delta u_I(\theta) = u_I(1, \theta) - u_I(0, \theta) = \theta - k\,.$$

We denote an arbitrary screening rule in the centralized economy by $m_c(\theta)$ and the optimal screening rule by $m_c^*(\theta)$. Thus, the entrepreneur's problem in the centralized economy is

$$\max_{0 \leqslant m_c(\theta) \leqslant 1} \mathbb{E}[m_c(\theta) \cdot (\theta - k)] - \mu \cdot I(m_c(\theta))\,. \tag{3.4.10}$$

By construction, the entrepreneur's objective (3.4.10) in the centralized economy is exactly the expected social surplus in the decentralized economy. Hence, we can examine the centralized economy to analyze the efficiency of equilibria in the corresponding decentralized economy.

Since the optimal screening rules are unique for both the centralized and the decentralized economy, efficiency is achieved if and only if information is acquired in the same manner in both.

Lemma **19.** *An optimal security in the decentralized economy achieves efficiency if and only if the associated optimal screening rule $m_s^*(\theta)$ satisfies*

$$Prob[m_s^*(\theta) = m_c^*(\theta)] = 1\,,$$

where $m_c^*(\theta)$ is the optimal screening rule in the corresponding centralized economy.

The efficiency concept in Lemma 19 demonstrates the role of screening in the production economy and provides a natural measure of the severity of friction in the decentralized economy. Fundamentally, we may view the expected social surplus (3.4.10) or (3.4.9) in our production economy as a production function, with information characterized by the screening rule $m_c(\theta)$ or $m(\theta)$ as the sole input. This again fits with the idea that real production depends on information acquisition. Consequently, efficiency is achieved if and only if the optimal security in the decentralized economy delivers the same equilibrium allocation of input, information acquisition, as the centralized economy. If the optimal security does this, friction in the decentralized economy is not severe, as it can be effectively removed by optimal security design. Otherwise, friction is severe, in that it cannot be completely removed even by an optimal security.

By the efficiency concept in Definition 9 and Lemma 19, we can characterize the optimal securities over the efficiency dimension, as follows.

Proposition **26.** *In the decentralized production economy:*

i) the optimal security $s^*(\theta)$ is debt if and only if friction in the decentralized economy is not severe, i.e., an optimal security achieves efficiency; and

ii) $s^*(\theta)$ is convertible preferred stock if and only if the friction is severe, i.e., an optimal security cannot achieve efficiency.

These results are important because they answer the question why different projects should be financed by different types of securities by pointing to its origin: the friction between real production and information acquisition. In the decentralized economy, real production is performed by the entrepreneur while information acquisition by the investor. This separation is always present, unchanged despite the different exogenous characteristics of the economy. Hence, the severity of the friction is reflected in the extent to which real production depends on information acquisition. If the friction is severe, the dependence is strong, which makes screening worthwhile and makes convertible preferred stock optimal. Similarly, if the friction is not severe, the dependence is weak, screening does not justify its cost, and debt is optimal.

Our predictions help unify the empirical evidence. They are particularly suited for entrepreneurial businesses. Debt financing is popular for conventional projects and for investors who have less expertise in screening, i.e., when the informational friction is not severe. Instead, financing with convertible preferred stock (or the combination of debt and equity) is common for innovative projects, especially in the early rounds, which benefit more from screening, that is, when the friction is severe.

3.5 Comparative Statics of the Optimal Security

For additional intuitions, we look at numerical comparative statics on the shape of the optimal securities with respect to two empirical dimensions: the profitability of the project and its uncertainty.[31] When the environment varies, the role of screening changes, and the way in which the entrepreneur incentivizes screening changes accordingly, producing different shapes of optimal securities.

3.5.1 Project Profitability

First, we consider the effects of variations in the project's profitability on the shape of the optimal security $s^*(\theta)$, holding constant the project's market prospects (i.e., the prior distribution of the cash flow θ), and the cost of screening, μ. Thus, a decrease in the investment requirement k implies that the project is more profitable ex-ante.

We show the results in Figure 3.4. The investment k takes three increasing values: 0.4, 0.475, and 0.525. When $k = 0.4$, the optimal security is debt; for two other projects with larger k, one with positive and one with negative ex-ante NPV, it is convertible preferred stock. In particular, the face value $\widehat{\theta}$ and the conversion ratios $d\widehat{s}(\theta)/d\theta$ of the convertible preferred stock are both increasing in k. For the prior of the cash flow θ, we take a normal distribution with mean 0.5 and standard deviation 0.125, and then truncate and normalize this distribution to the interval $[0, 1]$. The screening cost μ is fixed at 0.2.

The comparative statics with respect to the profitability of the project serve as a detailed illustration of Proposition 25. When the project is sufficiently profitable ex-ante ($k = 0.4$), the friction is not severe and the project will be financed by debt without inducing screening. When the project looks mediocre in terms of its profitability but still has a positive ex-ante NPV ($k = 0.475$), the friction becomes severe, and information acquisition becomes

---

[31] We are not aware of any analytical comparative statics pertaining to functionals. An analytical comparative statics requires a total order, which is not applicable for our security space. Even for some ordered characteristics of the optimal security, for instance, the face value, analytical comparative statics are not achievable. Thus, we rely on numerical results to deliver intuitions and leave analytical work to future research. Numerical analysis in our framework is tractable but already fairly technical. The algorithm and codes are available upon request.

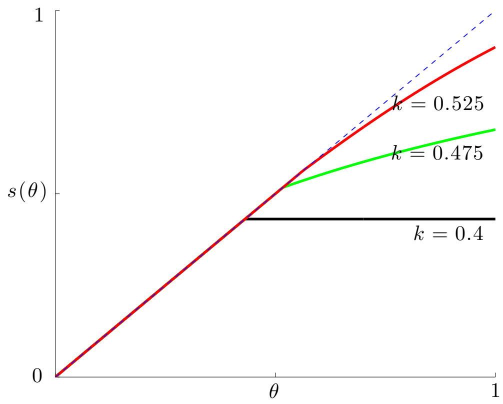


Figure 3.4: *Change of Investment with* $\mathbb{E}[\theta] = 0.5, \mu = 0.2$

worthwhile to screen bad projects out, so that convertible preferred stock becomes optimal. When the project is not profitable in the sense that its NPV is negative ($k = 0.525$), the friction is more severe, and the only way for the entrepreneur to obtain financing is to propose convertible preferred stock and expect a potentially good project to be screened in. For this type of project, in particular, screening is more valuable, and hence the entrepreneur is willing to compensate the investor more generously to induce more effective screening, as seen in Figure 3.4.

3.5.2 Project Uncertainty

We then consider how varying the degree of the project's uncertainty affects the optimal security $s^*(\theta)$. Concretely, we consider different prior distributions of the cash flow θ with the same mean, ranked by second order stochastic dominance.[32] We also hold constant the investment requirement k and the cost of screening μ. Note that, the effect of varying uncertainty cannot be accounted for by any argument involving risks, because both the entrepreneur and the investor are risk-neutral. Instead, we still focus on friction and the role of screening to explain these effects.

---

[32] There are other ways to measure the project's uncertainty. For comparative statics, our idea is to find a partial order of uncertainty over the space of distributions, while to keep the project's ex-ante NPV constant. Thus, second order stochastic dominance is a natural choice.

Interestingly, the comparative statics with respect to uncertainty depend on the sign of the project's ex-ante NPV. As implied by Proposition 25, the role of screening differs when these signs differ. This further leads to different patterns of comparative statics when the degree of uncertainty varies.

First, we consider projects with positive ex-ante NPV and increasing uncertainty. The results are shown in Figure 3.5, the left panel illustrating the priors of the cash flow θ and the right panel the evolution of the optimal security. When the project is the least uncertain, the optimal security is debt. For more uncertain projects convertible preferred stock becomes optimal, while the face value $\widehat{\theta}$ and the conversion ratios $d\widehat{s}(\theta)/d\theta$ are both increasing in uncertainty. For the priors, we take normal distributions with mean 0.5 and standard deviations 0.125 and 0.25, and then truncate and normalize them to the interval $[0, 1]$. We also construct a third distribution, in which the project is so uncertain that the cash flow has a greater probability of taking extreme values in $[0, 1]$. The investment is $k = 0.4$, and the cost of screening is $\mu = 0.2$.

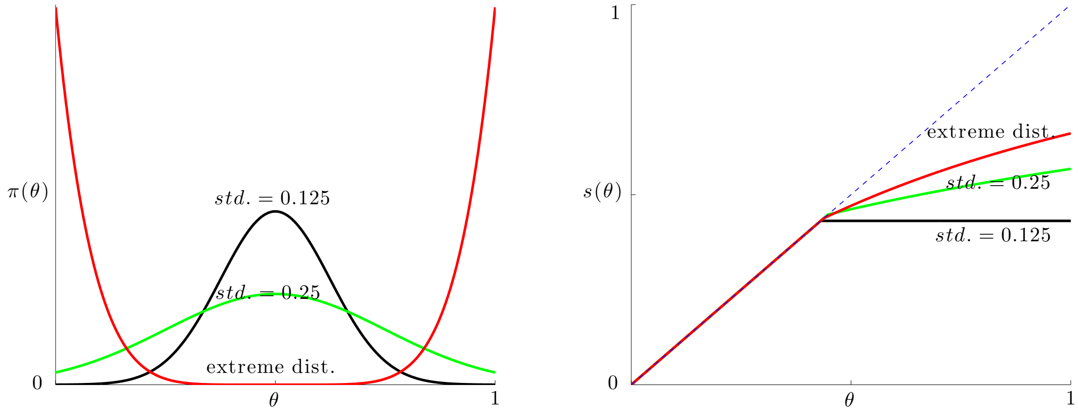


Figure 3.5: *Change of Uncertainty:* $k = 0.4 < \mathbb{E}[\theta] = 0.5$, $\mu = 0.2$

The comparative statics in this case demonstrate how varying uncertainty affects the screening-out of bad projects, given positive ex-ante NPV. When the project is least uncertain, it is least likely to be bad, which implies that screening-out is least relevant and debt financing is accordingly optimal. When uncertainty increases, the project is more likely to be bad, and screening-out becomes more valuable. Hence, the entrepreneur finds it optimal to propose a more generous convertible preferred stock to induce screening-out.

Next we consider projects with negative ex-ante NPV, focusing on those that may be financed with a positive probability due to screening-in through convertible preferred stock. The results are shown in Figure 3.6: both the face value $\widehat{\theta}$ and the conversion ratio $d\widehat{s}(\theta)/d\theta$ of the convertible preferred stock are decreasing in uncertainty. The priors are generated as we did in Figure 3.5. The investment is $k = 0.525$ and the cost of screening is $\mu = 0.2$.

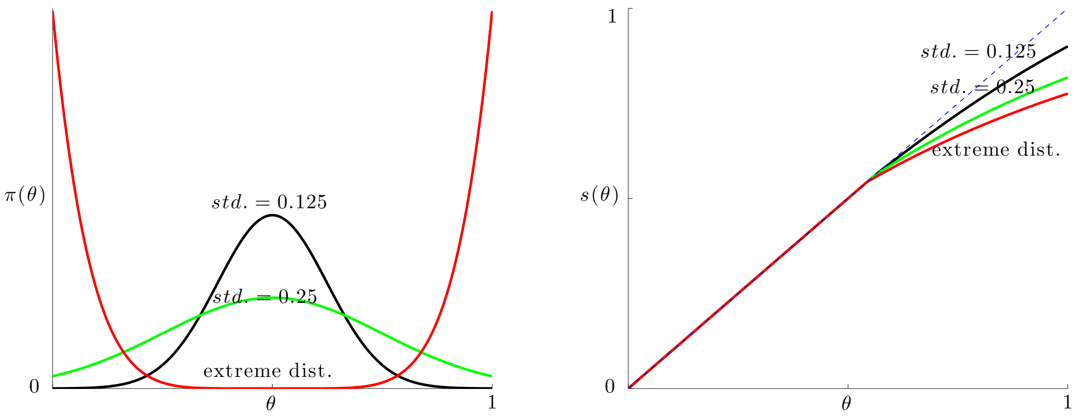


Figure 3.6: *Change of Uncertainty:* $k = 0.525 > \mathbb{E}[\theta] = 0.5$, $\mu = 0.2$

The comparative statics in this case are also intuitive, according to the role of screening-in. Given negative ex-ante NPV, the investor screens in potentially good projects. In contrast to the positive-NPV case, here the increase in uncertainty means that the ex-ante negative-NPV project is more likely to be good. Thus, to acquire costly information to screen in a potentially good project becomes less necessary. Therefore, the entrepreneur wants to propose a less generous convertible preferred stock for less costly screening. Not surprisingly, the resulting convertible preferred stock moves away from the 45° line when the project is more uncertain.

3.6 Conclusion

This paper posits a new type of informational friction to investigate security design. Real production depends on information acquisition, but these two functions are performed separately by entrepreneur and investor. We predict that debt, which does not induce screening, is optimal when the dependence is weak and the friction is therefore not severe, whereas the combination of debt and equity (participating convertible preferred stock), which does induce screening, is optimal when the dependence is strong and the friction accordingly severe. These predictions are supported by the empirical evidence.

This paper contributes to the security design literature in several respects, as well as to the broader corporate finance and contract design literature. By flexible information acquisition, we can work with arbitrary securities over continuous states while dispensing with usual distributional assumptions, thus offering an exhaustive examination at the question why projects with different natures should be financed by different types of securities.

Bibliography

Admati, Anat and Paul Pfleiderer (1994). "Robust Financial Contracting and the Role of Venture Capitalists." *Journal of Finance*, 49: 371-402.

Aghion, Philippe and Jean Tirole (1992). "On The Management of Innovation." *Quarterly Journal of Economics*, 109: 1185-1207.

Axelson, Ulf (1994). "Security Design with Investor Private Information." *Journal of Finance*, 62: 2587-632.

Bergemann, Dirk and Ulrich Hege (1998). "Venture Capital Financing, Moral Hazard, and Learning." *Journal of Banking and Finance*, 22: 703-735.

Berger, Allen and Gregory Udell (1998). "The Economics of Small Business Finance: The Roles of Private Equity and Debt Markets in the Financial Growth Cycle." *Journal of Banking and Finance*, 22: 613-673.

Berglof, Erik (1994). "A Control Theory of Venture Capital Finance." *Journal of Law, Economics and Organization*, 10: 247-67.

Biais, Bruno and Thomas Mariotti (2005). "Strategic Liquidity Supply and Security Design." *Review of Economic Studies*, 72: 617-49.

Bond, Philip, Alex Edmans and Itay Goldstein (2012). "The Real Effects of Financial Markets." *Annual Review of Finance and Economics*, 4(2): 1-22.

Boot, Arnoud and Anjan Thakor (1993). "Security Design." *Journal of Finance*, 48: 1349-1378.

Brennan, Michael and Alan Kraus (1987). "Efficient Financing Under Asymmetric Information." *Journal of Finance*, 42: 1225-1243.

CASAMATTA, CATHERINE (2003). "Financing and Advising: Optimal Financial Contracts with Venture Capitalists." *Journal of Finance*, 58(5): 2059-2085.

CHAKRABORTY, ARCHISHMAN and BILGE YILMAZ (2011). "Adverse Selection and Covertible Bonds." *Review of Economic Studies*, 78: 148-175.

CHAKRABORTY, ARCHISHMAN, SIMON GERVAIS and BILGE YILMAZ (2012). "Security Design in Initial Public Offerings." *Review of Finance*, 15: 327-357.

CHEMMANUR, THOMAS and PAOLO FULGHIERI (1997). "Why Include Warrants in New Equity Issues? A Theory of Unit IPOs." *Journal of Financial and Quantitative Analysis*, 32: 1-24.

CHEMMANUR, THOMAS, KARTHIK KRISHNAN and DEBARSHI NANDY (2012). "How Does Venture Capital Financing Improve Efficiency in Private Firms? A Look Beneath the Surface." *Review of Financial Studies*, forthcoming.

CONSTANTINIDES, GEORGE and BRUCE GRUNDY (1989). "Optimal Investment with Stock Repurchase and Financing as Signals." *Review of Financial Studies*, 2: 445-465.

CREMER, JACQUES and FAHAD KHALIL (1992). "Gathering Information Before Signing a Contract." *American Economic Review*, 82: 566-578.

CREMER, JACQUES, FAHAD KHALIL and JEAN-CHARLES ROCHET (1998a). "Strategic Information Gathering before a Contract Is Offered." *Journal of Economic Theory*, 81: 163-200.

CREMER, JACQUES, FAHAD KHALIL and JEAN-CHARLES ROCHET (1998b). "Contracts and Productive Information Gathering." *Games and Economic Behavior*, 25: 174-193.

CORNELLI, FRANCESCA and OVED YOSHA (2003). "Stage Financing and the Role of Convertible Securities." *Review of Economic Studies*, 70: 1-32.

DA RIN, MARCO, THOMAS HELLMANN and MANJU PURI (2011). "A Survey of Venture Capital Research." Forthcoming in George Constantinides, Milton Harris, and Rene Stulz (eds), *Handbook of the Economics of Finance*, 2.

DANG, TRI VI, GARY GORTON and BENGT HOLMSTROM (2011). "Ignorance and the Optimality of Debt for Liquidity Provision." Working paper.

DeMarzo, Peter and Darrell Duffie (1999). “A Liquidity-Based Model of Security Design.” *Econometrica*, 67: 65-99.

DeMarzo, Peter (2005). “The Pooling and Tranching of Securities: A Model of Informed Intermediation.” *Review of Financial Studies*, 18: 1-35.

Evans, David and Boyan Jovanovic (1989). “An Estimated Model of Entrepreneurial Choice under Liquidity Constraints.” *Journal of Political Economy*, 97: 808-827.

Fulghieri, Paolo and Dmitry Lukin (2001). “Information Production, Dilution Costs, and Optimal Security Design.” *Journal of Financial Economics*, 61: 3-42.

Fulghieri, Paolo, Diego Garcia and Dirk Hackbarth (2013). “Asymmetric Information and the Pecking (Dis)Order.” Working Paper.

Garmaise, Mark (2007). “Informed Investors and the Financing of Entrepreneurial Projects.” Working Paper.
(2013).

Gompers, Paul (1999). “Ownership and Control in Entrepreneurial Firms: An Examination of Convertible Securities in Venture Capital Investment.” Working Paper.

Gorton, Gary and George Pennacchi (1990). “Financial Intermediaries and Liquidity Creation.” *Journal of Finance*, 45: 49-72.

Hellmann, Thomas (1998). “The Allocation of Control Rights in Venture Capital Contracts.” *RAND Journal of Economics*, 29: 57-76.

Hellmann, Thomas (2006). “IPOs, Acquisitions, and the Use of Convertible Securities in Venture Capital.” *Journal of Financial Economics*, 81: 649-679.

Hennessy, Christopher (2013). “A Theory of ABS Design Based on Rational Noise-Traders.” Working Paper.

Hirshleifer, Jack (1971). “The Private and Social Value of Information and the Reward to Inventive Activity.” *American Economic Review*, 61: 561-574.

Holtz-Eakin, Douglas, David Joulfaian and Harvey Rosen (1994). “Sticking It Out: Entrepreneurial Survival and Liquidity Constraints.” *Journal of Political Economy*, 102: 53-75.

INDERST, ROMAN and HOLGER MUELLER (2006). "Informed Lending and Security Design." *Journal of Finance*, 61: 2137-2162.

INNES, ROBERT (1990). "Limited Liability and Incentive Contracting with Ex-ante Action Choices." *Journal of Economic Theory*, 52: 45-67.

KAPLAN, STEVEN and JOSH LERNER (2010). "It Ain't Broke: The Past, Present, and Future of Venture Capital." *Journal of Applied Corporate Finance*, 22: 36-47.

KAPLAN, STEVEN and PER STROMBERG (2003). "Financial Contracting Theory Meets the Real World: An Empirical Analysis of Venture Capital Contracts." *Review of Economic Studies*, 70: 281-315.

KASHYAP, ANIL and NATALIA KOVRIJNYKH (2013). "Who Should Pay for Credit Ratings and How?" Working paper.

KERR, WILLIAM, JOSH LERNER and ANTOINETTE SCHOAR (2014). "The Consequences of Entrepreneurial Finance: Evidence from Angel Financings." *Review of Financial Studies*, 27: 20-55.

KERR, WILLIAM and RAMANA NANDA (2009). "Democratizing Entry: Banking Deregulations, Financing constraints, and Entrepreneurship." *Journal of Financial Economics*, 94: 124-149.

KIRILENKO, ANDREI (2001). "Valuation and Control in Venture Finance." *Journal of Finance* 56: 565-587.

KNIGHT, FRANK (1921). *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin.

LERNER, JOSH, ANN LEAMON and FELDA HARDYMON (2012). *Venture Capital, Private Equity, and the Financing of Entrepreneurship*. John Wiley & Sons, Inc.

MANOVE, MICHAEL, A. JORGE PADILLA and MARCO PAGANO (2001). *Collateral versus Project Screening: A Model of Lazy Banks*. *The RAND Journal of Economics*, 32: 726-744.

MANSO, GUSTAVO (2011). "Motivating Innovation." *Journal of Finance* 66: 1823-1860.

MYERS, STEWART and NICHOLAS MAJLUF (1984). "Corporate Financing and Investment Decisions when Firms Have Information that Investors Do Not Have." *Journal of Financial Economics*, 13: 187-221.

NACHMAN, DAVID and THOMAS NOE (1994). "Optimal Design of Securities under Asymmetric Information." *Review of Financial Studies*, 7:1-44.

Opp, Christian, Marcus Opp and Milton Harris (2013). "Rating Agencies in the Face of Regulation." *Journal of Financial Economics*, 108: 46-61.

Rajan, Raghu (2012). "Presidential Address: The Corporation in Finance." *Journal of Finance*, 67: 1173-1217.

Ravid, Abraham and Matthew Spiegel (1997). "Optimal Financial Contracts for a Start-Up with Unlimited Operating Discretion." *Journal of Financial and Quantitative Analysis*, 32: 269-286.

Repullo, Rafael and Javier Suarez (2004). "Venture Capital Finance: A Security Design Approach." *Review of Finance*, 8: 75-108.

Robb, Alicia and David Robinson (2014). "The Capital Structure Decisions of New Firms." *Review of Financial Studies*, 27: 153-179.

Sahlman, William (1990). "The Structure and Governance of Venture-Capital Organizations." *Journal of Financial Economics*, 27: 473-521.

Scharfstein, David and Jeremy Stein (2000). "The Dark Side of Internal Capital Markets: Divisional Rent-Seeking and Inefficient Investment." *Journal of Finance,* 55: 2537-2564.

Schmidt, Klaus (2003). "Convertible Securities and Venture Capital Finance." *Journal of Finance*, 58: 1139-1166.

Schumpeter, Joseph (1942). *Capitalism, Socialism, and Democracy*. New York: Harper Brothers.

Sims, Christopher (2003). "Implications of Rational Inattention." *Journal of Monetary Economics*, 50: 665-690.

Stein, Jeremy (1992). "Convertible bonds as backdoor equity financing." *Journal of Financial Economics*, 32: 3-21.

Stein, Jeremy (1997). "Internal Capital Markets and the Competition for Corporate Resources." *Journal of Finance*, 52: 111-133.

Tirole, Jean (2006). *The Theory of Corporate Finance*. Princeton University Press.

Veldkamp, Laura (2011). *Information Choice in Macroeconomics and Finance*. Princeton Press.

WOODFORD, MICHAEL (2008). "Inattention as A Source of Randomized Discrete Adjustment." Working paper.

WOODFORD, MICHAEL (2012). "Inattentive Valuation and Reference-Dependent Choice." Working paper.

YANG, MING (2013). "Optimality of Debt with Flexible Information Acquisition." Working paper.

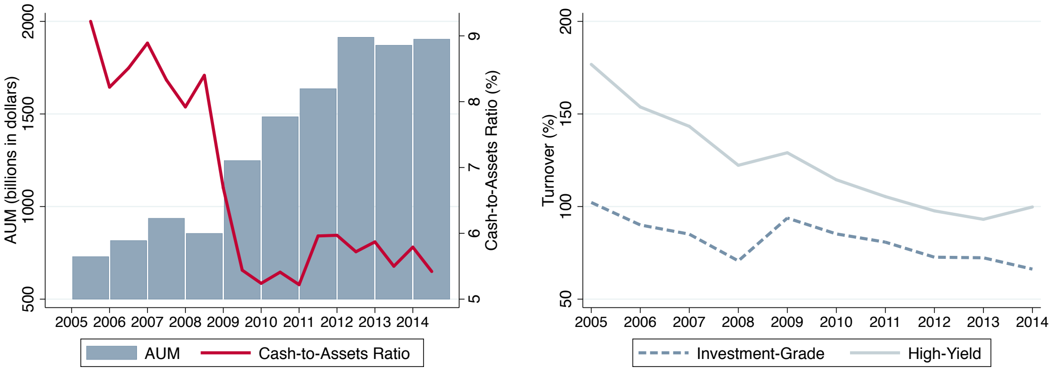
YANG, MING (2015). "Coordination with Flexible Information Acquisition." *Journal of Economic Theory*, 158: 721-738.

Appendix A

Appendix to Chapter 1

A.1 Institutional Background

In this appendix, I depict the current trends of U.S. corporate bond mutual funds. These trends show that these funds are growing rapidly in size, investing in more illiquid assets, and holding less cash relative to their assets. The settings in my model are consistent with these trends.



The left panel plots total assets under management (data from the Investment Company Institute) and weighted-average cash-to-assets ratios (data from the SEC form N-SAR) of corporate bond funds. Cash equivalents such as repurchase agreements and other short-term debt securities are included in the SEC cash holdings data. The right panel plots annual turnovers of the underlying corporate bond markets (data from Barclays Research).

Figure A.1: *Trends in Corporate Bond Mutual Funds*

It is worth noting that, although the trends illustrated here suggest increasing financial stability risks of mutual funds and potentially large fire sale losses when runs occur, the mechanism of this paper is general and does not rely on the trends.

A.2 The Analysis of the Intermediate Cash-to-Assets Ratio Region G_m

In this appendix, I provide a complete equilibrium analysis of the stage game when the fund starts at the intermediate cash-to-assets ratio region G_m.

Again, I first characterize how shareholder runs affect the fund's forced fire sales on dates $2t$ and $2t+1$, respectively. For convenience, I define

$$\widehat{\lambda}_{2t} \equiv \frac{x_{2t} - \mu_E(Ra_{2t} + x_{2t})}{\mu_L(Ra_{2t} + x_{2t})} .$$

By construction, within the intermediate region, there is always $\widehat{\lambda}_{2t} \in [0, 1)$. The economic meaning of $\widehat{\lambda}_{2t}$ will become clear shortly.

Lemma 20. *When $\eta_{2t} \in G_m$, there are:*

i) if $\lambda_{2t} \in [0, \widehat{\lambda}_{2t}]$, then

$$\begin{aligned} q_{2t}(\lambda_{2t}) &= 0 , \\ q_{2t+1}(\lambda_{2t}) &= \frac{(\mu_E + \mu_L)(Ra_{2t} + x_{2t}) - x_{2t}}{\widehat{p}_L(\lambda_{2t})} , \end{aligned} \tag{A.2.1}$$

where q_{2t+1} is increasing in λ_{2t}, and,

ii) if $\lambda_{2t} \in (\widehat{\lambda}_{2t}, 1]$, then

$$q_{2t}(\lambda_{2t}) = \frac{(\mu_E + \lambda_{2t}\mu_L)(Ra_{2t} + x_{2t}) - x_{2t}}{\widehat{p}_E(\lambda_{2t})} , \tag{A.2.2}$$

$$q_{2t+1}(\lambda_{2t}) = \frac{(1 - \lambda_{2t})\mu_L \cdot \dfrac{R(a_{2t} - q_{2t})}{1 - \mu_E - \lambda_{2t}\mu_L}}{\widehat{p}_L(\lambda_{2t})} , \tag{A.2.3}$$

where q_{2t} is increasing in λ_{2t} but q_{2t+1} is decreasing in λ_{2t}.

Moreover, $q_{2t} + q_{2t+1}$ *is increasing in* λ_{2t} *for all* $\lambda_{2t} \in [0, 1]$.

I first discuss the intuition behind the results when $\lambda_{2t} \leqslant \widehat{\lambda}_{2t}$. In this case, the fund has enough cash to satisfy all the $\mu_E + \lambda_{2t}\mu_L$ redeeming shareholders on date $2t$ at the initial NAV. Thus, no illiquid assets are forced to sell on date $2t$, that is, $q_{2t}(\lambda_{2t}) = 0$. However, in the intermediate region the fund does not have enough cash to satisfy all the late shareholders on the odd date. Specifically, the cash gap at the beginning of date $2t + 1$ is indicated by the numerator of (A.2.1). Following the same intuition of Lemma 4, the fund manager will close the gap by selling at the effective price $\widehat{p}_L(\lambda_{2t})$. As λ_{2t} increases, the effective selling price $\widehat{p}_L(\lambda_{2t})$ becomes lower, suggesting that the fund will be forced to sell more on date $2t + 1$, that is, q_{2t+1} becomes larger.

The situation becomes different when $\lambda_{2t} > \widehat{\lambda}_{2t}$. Compared to Lemma 4, the two conditions (A.2.2) and (A.2.3) are exactly the same as conditions (1.3.6) and (1.3.7) there. This is because when $\lambda_{2t} > \widehat{\lambda}_{2t}$ the cash position x_{2t} becomes inadequate to satisfy the $\mu_E + \lambda_{2t}\mu_L$ redeeming shareholders on date $2t$ at the initial NAV, so that the stage game effectively jumps into the low cash-to-assets ratio region. The monotonicity of q_{2t}, q_{2t+1}, and $(q_{2t} + q_{2t+1})$ all follows the same intuition there.

It is worth noting that, regardless of whether $\lambda_{2t} \leqslant \widehat{\lambda}_{2t}$ or $\lambda_{2t} > \widehat{\lambda}_{2t}$, more late shareholder runs always lead to unambiguously higher forced fire sales within the entire stage (including both date $2t$ and $2t + 1$).

Similarly, I can characterize the NAVs in the intermediate region. When $\lambda_{2t} \in [0, \widehat{\lambda}_{2t}]$, by Lemma 20 there is

$$NAV_{2t}(\lambda_{2t}) = Ra_{2t} + x_{2t}\,, \tag{A.2.4}$$

and

$$\begin{aligned} NAV_{2t+1}(\lambda_{2t}) &= \frac{\overbrace{R\left(a_{2t} - q_{2t+1}(\lambda_{2t}) - s_{2t+1}\right)}^{\text{illiquid assets remained}} + \overbrace{x_{2t} - (\mu_E + \lambda_{2t}\mu_L)(Ra_{2t} + x_{2t})}^{\text{cash remained}} + \overbrace{\delta_L R(q_{2t+1}(\lambda_{2t}) + s_{2t+1})}^{\text{cash raised and rebuilt}}}{\underbrace{1 - (\mu_E + \lambda_{2t}\mu_L)}_{\text{shareholders remained on date } 2t+1}} \\ &= NAV_{2t} - \frac{(1 - \delta_L)R\left(q_{2t+1}(\lambda_{2t}) + s_{2t+1}\right)}{1 - \mu_E - \lambda_{2t}\mu_L}\,. \end{aligned} \tag{A.2.5}$$

where $q_{2t+1}(\lambda_{2t})$ is given in (A.2.1). Clearly, the NAV on date $2t$ as in (A.2.4) is also constant and the same as (1.3.1) in the high region. The NAV on date $2t+1$ as in (A.2.5) also features the same expression as (1.3.3) in the high region. These suggest that shareholders' strategic interaction in this sub-region is the same as that in the high region.

When $\lambda_{2t} \in (\widehat{\lambda}_{2t}, 1]$, by Lemma 20 there are

$$NAV_{2t}(\lambda_{2t}) = Ra_{2t} + x_{2t} - (1 - \delta_E)Rq_{2t}(\lambda_{2t}) \,, \tag{A.2.6}$$

and

$$NAV_{2t+1}(\lambda_{2t}) = \frac{R\left(a_{2t} - q_{2t}(\lambda_{2t}) - q_{2t+1}(\lambda_{2t}) - s_{2t+1}\right) + \delta_L R\left(q_{2t+1}(\lambda_{2t}) + s_{2t+1}\right)}{1 - \mu_E - \lambda_{2t}\mu_L} \,, \tag{A.2.7}$$

where $q_{2t}(\lambda_{2t})$ and $q_{2t+1}(\lambda_{2t})$ are given in (A.2.2) and (A.2.3). Note that, the NAVs as in (A.2.6) and (A.2.7) are exactly the same as (1.3.9) and (1.3.10) in the low region, suggesting that shareholders' strategic interaction in this sub-region is the same as that in the low region.

Formally, the following lemma tells us that the stage game in the intermediate region indeed features a switch from strategic complementarity to substitutability.

Lemma 21. *When $\eta_{2t} \in G_m$, there are:*

i) if $\lambda_{2t} \in [0, \widehat{\lambda}_{2t}]$, late shareholders' run decision λ_{2t} exhibits strategic complementarity for any feasible $s_{2t+1} \in [0, a_{2t} - q_{2t+1}(\lambda_{2t})]$, and the strategic complementarity becomes stronger as s_{2t+1} increases, and,

ii) if $\lambda_{2t} \in (\widehat{\lambda}_{2t}, 1]$, late shareholders' run decision λ_{2t} exhibits strategic substitutability for any λ_{2t} satisfying $\theta NAV_{2t}(\lambda_{2t}) \geqslant NAV_{2t+1}(\lambda_{2t})$ and any feasible $s_{2t+1} \in [0, a_{2t} - q_{2t}(\lambda_{2t}) - q_{2t+1}(\lambda_{2t})]$, and the strategic substitutability becomes weaker as s_{2t+1} increases.

Not surprisingly, Lemma 21 can be understood in view of Lemma 3 (in the analysis for the high region) and Lemma 5 (in the analysis for the low region). In the first sub-region $[0, \widehat{\lambda}_{2t}]$, shareholders who run can get the endogenously fixed NAV on date $2t$ at the expense of shareholders who wait. More running shareholders or a more rapid cash rebuilding policy implies a larger magnitude of predictable decline in the NAV on date

$2t+1$, leading to a stronger strategic complementarity. In the second sub-region, however, running shareholders have to accept an endogenously lower NAV themselves because the fund is forced to sell its illiquid assets on date $2t$, when the fire sale price is extremely low. The resulting higher fire sale losses suggest that more shareholder runs make the other shareholders who plan to wait less likely to run. But again, more rapid cash rebuilding still gives rise to a larger magnitude of predictable decline in the NAV on date $2t+1$ and thus reinforces the run incentive.

Because of the switch of strategic interaction, shareholders' equilibrium run behaviors exhibit a richer pattern. Despite the complicated equilibrium construction in the intermediate region, it still indicates that fund cash rebuilding leads to runs and more rapid cash rebuilding triggers more severe runs in equilibrium.

Proposition 27. *When $\eta_{2t} \in G_m$, late shareholders' equilibrium run behaviors are given by the following five cases:*

i) none of the late shareholders runs, that is, $\lambda_{2t} = 0$, if

$$s_{2t+1} < \underline{s}_m \equiv \frac{Ra_{2t} - \theta(1-\mu_E - \widehat{\lambda}_{2t}\mu_L)(Ra_{2t}+x_{2t})}{(1-\delta_L)R} - q_{2t+1}(\widehat{\lambda}_{2t})\,,$$

ii) if

$$s_{2t+1} > \bar{s}_m \equiv \frac{Ra_{2t} - \theta(1-\mu_E)(Ra_{2t}+x_{2t})}{(1-\delta_L)R} - q_{2t+1}(0)\,,\ \textit{then}\,,$$

a) all of the late shareholders run, that is, $\lambda_{2t} = 1$, if

$$s_{2t+1} > \bar{\bar{s}}_m \equiv \frac{Ra_{2t} - \theta(1-\mu_E-\mu_l)(Ra_{2t}+x_{2t}) - (1-\theta(1-\delta_E)(1-\mu_E-\mu_l))Rq_{2t}(1)}{(1-\delta_L)R}\,,$$

b) some of the late shareholder runs, that is, $\lambda_{2t} = \tilde{\tilde{\lambda}}_{2t} \in [\widehat{\lambda}_{2t}, 1)$, if

$$s_{2t+1} \leqslant \bar{\bar{s}}_m\,,$$

where $\tilde{\tilde{\lambda}}_{2t}$ is the solution to

$$s_{2t+1} = \frac{Ra_{2t} - \theta(1-\mu_E-\tilde{\tilde{\lambda}}_{2t}\mu_l)(Ra_{2t}+x_{2t}) - (1-\theta(1-\delta_E)(1-\mu_E-\tilde{\tilde{\lambda}}_{2t}\mu_l))Rq_{2t}(\tilde{\tilde{\lambda}}_{2t})}{(1-\delta_L)R} - q_{2t+1}(\tilde{\tilde{\lambda}}_{2t})\,,$$

iii) if $\underline{s}_m \leqslant s_{2t+1} \leqslant \bar{s}_m$, then, *c)* $\lambda_{2t} \in \{0, \widetilde{\lambda}_{2t}, 1\}$*, if*

$$s_{2t+1} > \overline{\overline{s}}_m \, ,$$

where $\widetilde{\lambda}_{2t}$ *is the solution to*

$$s_{2t+1} = \frac{(1-\theta)(1-\mu_E-\widetilde{\lambda}_{2t}\mu_L)(Ra_{2t}+x_{2t})}{(1-\delta_L)R} - q_{2t+1}(\widetilde{\lambda}_{2t}) \, ,$$

d) $\lambda_{2t} \in \{0, \widetilde{\lambda}_{2t}, \widetilde{\widetilde{\lambda}}_{2t}\}$*, if*

$$s_{2t+1} \leqslant \overline{\overline{s}}_m \, ,$$

where $\widetilde{\lambda}_{2t}$ *is given in Case c) and* $\widetilde{\widetilde{\lambda}}_{2t}$ *is given in Case b).*

All of the $q_{2t}(\lambda_{2t})$ *and* $q_{2t+1}(\lambda_{2t})$ *are given in Lemma 20. Moreover,* $\overline{s}_m \geqslant 0$ *and* $\overline{s}_m > \underline{s}_m$*.*

The intuition behind Proposition 27 is clear in view of Propositions 1 and 2. By Lemma 21, the stage game in the intermediate region starts with strategic complementarity when only a small fraction of late shareholders decides to run. Hence, it is the strategic complementarity in the first sub-region $[0, \widehat{\lambda}_{2t}]$ that determines whether any late shareholder will run at all. As in Case i), when $\widehat{\lambda}_{2t}$ of the late shareholders decide to run, if the utility gain of running over waiting is still not positive, none of the late shareholders will ever run. In Case ii), the utility gain of running over waiting is already positive even if no one runs, so that at least $\widehat{\lambda}_{2t}$ of the late shareholders will run due to the strategic complementarity in the sub-region $[0, \widehat{\lambda}_{2t}]$. However, as the stage game switches to the second sub-region $(\widehat{\lambda}_{2t}, 1]$, there can be strategic substitutability. In sub-case a), the fund uses a rapid cash rebuilding policy so that all of the late shareholders run despite the strategic substitutability, while in sub-case b) the substitutability is strong so that $\lambda_{2t} = \widetilde{\widetilde{\lambda}}_{2t} \in [\widehat{\lambda}_{2t}, 1)$ of the late shareholders are going to run. Finally, in Case iii), the strategic complementarity in the first sub-region $[0, \widehat{\lambda}_{2t}]$ is moderate. When this happens, the worst equilibrium will be determined by the magnitude of strategic substitutability in the sub-region $(\widehat{\lambda}_{2t}, 1]$, as shown in Case c) and Case d).

Interestingly, Proposition 2 suggests that the risk of shareholder runs can be higher in the intermediate region than that in the low region (with the same set of cash rebuilding policy s_{2t+1} and other model parameters). This is because the run threshold $\underline{s}_m$ in Proposition 27 can be smaller than $\underline{s}_l$ in Proposition 2. This predication may be surprising given that the fund has more cash in the intermediate region. But it is intuitive in view of the strategic complementarity in the intermediate region. Concretely, when the fund starts from either the low or the intermediate region, there are forced fire sales on date $2t+1$. But only when starting from the intermediate region, can some of the running shareholders get an endogenously fixed NAV on date $2t$. As a result, shareholders may be more willing to run to get the higher NAV on date $2t$ when the fund starts from the intermediate region, compared to the case when if the fund starts from the low region where they would always have to accept a lower NAV on date $2t$ if they run.

As usual, I show how shareholder runs increase the risk of forced fire sales by exploring the laws of motions in the intermediate region.

Corollary **6.** *When $\eta_{2t} \in G_m$, the law of motions of (a_{2t}, x_{2t}) is given by*

$$a_{2t+2} = a_{2t} - (q_{2t}(\lambda_{2t}) + q_{2t+1}(\lambda_{2t})) - s_{2t+1}\,, \text{ and}\,, \tag{A.2.8}$$

$$x_{2t+2} = \widehat{p}_L(\lambda_{2t}) s_{2t+1}\,, \tag{A.2.9}$$

where λ_{2t} is the equilibrium run probability induced by (a_{2t}, x_{2t}) and s_{2t+1}, as characterized in Proposition 27.

Similarly, Corollary 6 is easier to understand in view of Corollary 2; the two laws of motions (A.2.8) and (A.2.9) share the same mathematical expressions with (1.3.11) and (1.3.12) in the low region.[1] They again suggest two different costs of shareholder runs: more forced fire sales in the current stage and higher risk of future-stage fire sales.

A.3 A Micro-foundation for the Pattern of Selling Prices

In this appendix, I show that the pattern of selling prices in the baseline model can emerge endogenously by modeling the slow-moving of liquidity providers in the spirit of Grossman and Miller (1988) and Duffie (2010). It shows that the reduced-form assumption can be rationalized as the outcome of a full-fledged equilibrium model with both liquidity demanders and providers. To make the idea more transparent, I set the micro-foundation in continuous time.[2] I follow the building blocks in Duffie, Garleanu and Pedersen (2005, 2007), Weill (2007) and Lagos, Rocheteau and Weill (2011) to model the gradual entry of liquidity providers and focus on the equilibrium price implications.

---

[1] However, it is worth noting that the equilibrium determination and the endogenous functions of $q_{2t}(\lambda_{2t})$ and $q_{2t+1}(\lambda_{2t})$ are still different between the two regions. So that the intermediate region still features a different law of motions of (a_{2t}, x_{2t}).

Time is continuous and infinite. A probability space $(\Omega, \mathcal{F}, P)$ is fixed with an information filtration $\{\mathcal{F}_t, t \geqslant 0\}$ satisfying the usual measurability conditions (Sun, 2006). There is a common discount rate $r > 0$. There is a continuum of 1 of risk-neutral, infinitely lived, and competitive investors. There is a centralized market with many different assets. The total supply of all assets is $S \in [0, 1)$. Investors can hold at most one unit of assets and cannot short sell the assets. There is also a riskfree saving account with return r, which can be interpreted as cash equivalents. Under usual non-arbitrage conditions, this implies that the fundamental value of the assets is $1/r$.

There are two types of investors: liquidity providers and liquidity demanders. Liquidity providers enjoy a high utility flow per time by holding one unit of assets, which is normalized to 1, while liquidity demanders enjoy a low utility flow $\delta \in (0, 1)$.

At the beginning, the economy is hit by an unanticipated liquidity shock that makes all investors liquidity demanders. However, as time goes by, they will randomly and pairwise independently switch to liquidity providers.[3] Specifically, the times at which investors switch to liquidity providers are i.i.d. exponentially distributed with a parameter α. Denote the endogenous population of liquidity providers by $\rho(t)$. By the exact law of larger numbers (Sun, 2006, Theorem 2.16), there is

$$\rho(t) = 1 - \exp(-\alpha t) . \tag{A.3.10}$$

---

[2] This baseline model is set in discrete time to highlight the discrete nature of daily redemptions and the end-of-day NAV. But in the micro-foundation, the discrete nature is no longer important. As a result, setting a continuous-time model incurs no loss of generality but makes the derivation mathematically more convenient.

[3] This dynamic process is in the spirit of Grossman and Miller (1988), in which liquidity providers only enter the market one period after the initial liquidity shock.

Intuitively, this implies that there is no liquidity provider available right at the shock time (i.e., $t = 0$), while there will be more and more liquidity providers stepping into the market after the shock.

In this simple framework, the following proposition shows the pattern of asset selling price over time:

PROPOSITION **28.** *The asset selling price at time t is characterized by*

$$p(t) = \frac{\delta + (1-\delta)\exp(-r(t_S - t))}{r},$$

where t_S satisfies $\rho(t_S) = S$ and $\rho(\cdot)$ is given by (A.3.10).

Intuitively, the selling price drops discontinuously at $t = 0$ from the fundamental value, but rebounds gradually over time (as more liquidity providers become available) until it gets back to the fundamental value at time t_S. When the next shock comes, this process repeats itself, giving rise to the price pattern in the baseline model.

It is instructive to provide the proof here to help build intuition. First of all, I show that there is a time at which the selling positions can be completely absorbed by liquidity providers so that the price goes back to the fundamental. Specifically, condition (A.3.10) shows that more liquidity providers step into the market as time goes by after the shock. Denote the endogenous time by which liquidity providers can absorb all the asset supply by t_S, which implies that $\rho(t_S) = S$. Since $\rho(t_S)$ is monotone, this uniquely determines t_S. This corresponds to the baseline model that if the game ends (i.e., there are no future shocks), the asset selling price will ultimately reflect the fundamental value.

Then I show that, between the shock time 0 and the full recovery time t_S (before the next possible shock), the asset selling price first drops and then rebounds gradually, as that in the baseline model. Note that, at any time t between 0 and t_S, there are no enough liquidity providers in the market, so that the marginal investor is a liquidity demander who has a low valuation of the assets. Since this liquidity demander is infinitely lived, the

Hamilton-Jacobi-Bellman equation leads to:

$$rp(t)dt = \delta dt + p(t) \tag{A.3.11}$$

This condition has an intuitive interpretation. At any time t between 0 and t_S, the left hand side of (A.3.11) denotes the return of selling the unit of assets at t and investing the proceeds in cash equivalents in the time interval $[t, t + dt)$, while the right hand side denotes the valuation flow by holding one unit of assets in the time interval $[t, t + dt)$ plus the proceeds from selling it after that. In any equilibrium path, the liquidity demander should be indifferent between these two options of selling earlier or later. Therefore, solving the differential equation implied by (A.3.11) with the boundary conditions yields the equilibrium selling price. This concludes the proof.

Fundamentally, this micro-foundation follows the spirit of Grossman and Miller (1988) and Duffie (2010), but differs in an important way. Specifically, liquidity providers in their models share risks with liquidity demanders, while in both my baseline model and the micro-foundation, all the investors are risk neutral.[4] However, similar selling price pattern emerges. This is because liquidity providers in my model have higher valuation of the underlying assets, which more resembles the notion of natural buyers in Shleifer and Vishny (1992, 1997) and thus is closer to the fire sale interpretation in the baseline model. Like that in Grossman and Miller (1988) and Duffie (2010), liquidity providers step into the market only gradually after the shock, implying that only a few liquidity providers are present in the market right after the shock. Hence, investors who want to sell the assets right after the shock have to incur an extremely low fire sale price. As time goes by (but before the next possible shock comes), more liquidity providers with high valuation of the underlying assets step into the market, implying that it becomes increasingly easier for the liquidity demanders to find a better selling price.

This micro-foundation has other nice properties, which are also consistent with other ingredients of the baseline model. First, different from Grossman and Miller (1988), all investors in the micro-foundation are infinitely lived and perfectly forward-looking. This implies that the resulting selling price pattern does not come from any myopia of the investors. Second, as suggested by Duffie, Garleanu and Pedersen (2005, 2007) and many follow-up models, the setting of low and high valuations parsimoniously captures many potential benefits and costs of holding the assets and the resulting trading motives, which are consistent with the baseline model. When a mutual fund experiences a large redemption, holding the illiquid assets would incur more costs and thus imply a lower valuation.

---

[4] This assumption of risk neutrality also appears in other search-based models (see Duffie, Garleanu and Pedersen, 2005, Weill, 2007, Lagos, Rocheteau and Weill, 2011, among many others).

A.4 Proofs

In this appendix, I provide proofs for all the results in the main text.

Proof of Lemma 1. First, in the high cash-to-assets ratio region, the fund needs to sell no illiquid assets on either date $2t$ or $2t+1$. Since no fire sale losses are incurred in this region, both early and late shareholders are able to get the same NAV as that at the beginning of date $2t$, that is,

$$NAV_{2t} = NAV_{2t+1} = \frac{Ra_{2t} + x_{2t}}{n_{2t}}.$$

Moreover, the initial cash position should be large enough to meet the redemption needs of all shareholders on dates $2t$ and $2t+1$ at such a constant NAV:

$$x_{2t} \geqslant (\mu_E + \mu_L) n_{2t} \cdot \frac{Ra_{2t} + x_{2t}}{n_{2t}},$$

yielding

$$\eta_{2t} \geqslant \frac{(\mu_E + \mu_L)R}{1 - \mu_E - \mu_L}, \tag{A.4.12}$$

the criterion for the high region.

Then, in the intermediate region, as no fire sale is incurred on date $2t$, the initial cash position is high enough to meet the redemption needs of early shareholders at the initial

NAV but insufficient to meet late shareholders' redemption needs:

$$\mu_E n_{2t} \cdot \frac{Ra_{2t} + x_{2t}}{n_{2t}} \leqslant x_{2t} < (\mu_E + \mu_L) n_{2t} \cdot \frac{Ra_{2t} + x_{2t}}{n_{2t}},$$

which leads to

$$\frac{\mu_E R}{1 - \mu_E} \leqslant \eta_{2t} < \frac{(\mu_E + \mu_L) R}{1 - \mu_E - \mu_L}, \tag{A.4.13}$$

the criterion for the intermediate region.

Finally, in the low region, the cash position is even inadequate to meet early shareholders' redemption needs at the initial NAV. This means

$$x_{2t} < \mu_E n_{2t} \cdot \frac{Ra_{2t} + x_{2t}}{n_{2t}},$$

which yields

$$\eta_{2t} < \frac{\mu_E R}{1 - \mu_E}, \tag{A.4.14}$$

the criterion for the intermediate region.

It is straightforward to check that (A.4.12), (A.4.13), and (A.4.14) are also sufficient conditions. This concludes the proof. □

Proof of Lemma 2. Suppose $\lambda_{2t}\mu_L$ late shareholders decide to run. This situation is equivalent to a counterfactual in which there are initially $\mu'_E = \mu_E + \lambda_{2t}\mu_L$ early shareholders and $\mu'_L = (1 - \lambda_{2t})\mu_L$ late shareholders but no late shareholder runs. Since $\mu'_E + \mu'_L = \mu_E + \mu_L$, by Lemma 1, $q_{2t} = q_{2t+1} = 0$ is true in the counterfactual situation and so is true in the original situation with $\lambda_{2t}\mu_L$ late shareholders running. This concludes the proof. □

Proof of Lemma 3. By Lemma 2 and the definition of $\Delta u_L(\lambda_{2t})$:

$$\Delta u_L(\lambda_{2t}) = \theta NAV_{2t} - NAV_{2t+1} = (\theta - 1)(Ra_{2t} + x_{2t}) + \frac{(1 - \delta_L) R s_{2t+1}}{1 - \mu_E - \lambda_{2t}\mu_L}.$$

Taking derivatives yields:

$$\frac{\partial \Delta u_L(\lambda_{2t})}{\partial \lambda_{2t}} = \frac{(1 - \delta_L)\mu_L R s_{2t+1}}{(1 - \mu_E - \lambda_{2t}\mu_L)^2} > 0,$$

which takes value 0 when $s_{2t+1} = 0$, and

$$\frac{\partial^2 \Delta u_L(\lambda_{2t})}{\partial \lambda_{2t} \partial s_{2t+1}} = \frac{(1-\delta_L)\mu_L R}{(1-\mu_E-\lambda_{2t}\mu_L)^2} > 0\,.$$

This concludes the proof. □

PROOF OF PROPOSITION 1. By Lemma 3, the stage game exhibits strategic complementarity when $s_{2t+1} > 0$. Also notice that any shareholder runs only if $\theta NAV_{2t} \geqslant NAV_{2t+1}$. Thus, in Case i), none of the late shareholders runs if

$$\theta NAV_{2t} < NAV_{2t+1}(1)\,, \tag{A.4.15}$$

in which $NAV_{2t+1}(\lambda_{2t})$ is a function of λ_{2t}. Solving inequality (A.4.15) leads to

$$s_{2t+1} < \frac{(1-\theta)(1-\mu_E-\mu_L)(Ra_{2t}+x_{2t})}{(1-\delta_L)R} \equiv \underline{s}_h\,.$$

Alternatively, in Case ii), all of the late shareholders run if

$$\theta NAV_{2t} > NAV_{2t+1}(0)\,, \tag{A.4.16}$$

the solution of which is

$$s_{2t+1} > \frac{(1-\theta)(1-\mu_E)(Ra_{2t}+x_{2t})}{(1-\delta_L)R} \equiv \overline{s}_h\,.$$

Finally, in Case iii), if neither (A.4.15) nor (A.4.16) holds, there exists a $\widetilde{\lambda}_{2t} \in [0,1]$ that solves

$$\theta NAV_{2t} = NAV_{2t+1}(\widetilde{\lambda}_{2t})\,.$$

Note that, $\widetilde{\lambda}_{2t}$ constructs an equilibrium because by definition $\Delta u_L(\widetilde{\lambda}_{2t}) = 0$ and thus no shareholder would have an incentive to deviate from it. In addition, in this case, again by Lemma 3, there are $\theta NAV_{2t} \geqslant NAV_{2t+1}(1)$ and $\theta NAV_{2t} \leqslant NAV_{2t+1}(0)$, which means $\lambda_{2t} = 1$ and $\lambda_{2t} = 0$ are also two equilibria when (A.4.15) and (A.4.16) are both violated. This concludes the proof. □

PROOF OF COROLLARY 1. By Lemma 2, $q_{2t}(\lambda_{2t}) = q_{2t+1}(\lambda_{2t}) = 0$ for any arbitrary $\lambda_{2t} \in$

$[0,1]$. Thus, the evolution of the asset position directly follows:

$$a_{2t+2} = a_{2t} - q_{2t} - q_{2t+1} - s_{2t+1} = a_{2t} - s_{2t+1} .$$

For the evolution of the cash position, the fund pays all the redeeming shareholders by cash at the respective end-of-day NAVs on date $2t$ and $2t+1$, and rebuilds its cash buffer on date $2t+1$. Note that there will be no cash raised by forced fire sales. Thus:

$$\begin{aligned} x_{2t+2} &= x_{2t} - (\mu_E + \lambda_{2t}\mu_L)NAV_{2t} - (1-\lambda_{2t})\mu_L NAV_{2t+1} + p_L s_{2t+1} \\ &= x_{2t} - (\mu_E + \mu_L)(Ra_{2t} + x_{2t}) + \delta_L R s_{2t+1} + \frac{(1-\lambda_{2t})\mu_L(1-\delta_L)Rs_{2t+1}}{1-\mu_E-\lambda_{2t}\mu_L} . \end{aligned}$$

This concludes the proof. □

Proof of Lemma 4. Recall that, when forced fire sales occur, the fund sells up to a point at which it can satisfy the redemptions at the end-of-day NAV, which will take into account the losses from forced fire sales. On the one hand, on date $2t$ the fund starts with a cash position x_{2t}. Hence, on date $2t$, q_{2t} solves

$$x_{2t} + p_E q_{2t} = (\mu_E + \lambda_{2t}\mu_L)[(a_{2t} - q_{2t})R + x_{2t} + p_E q_{2t}] ,$$

yielding

$$q_{2t}(\lambda_{2t}) = \frac{(\mu_E + \lambda_{2t}\mu_L)(Ra_{2t} + x_{2t}) - x_{2t}}{[\delta_E + (1-\delta_E)(\mu_E + \lambda_{2t}\mu_L)]\, R} . \tag{A.4.17}$$

On the other hand, on date $2t+1$, by construction, the fund has no cash at all at the beginning. Hence, q_{2t+1} solves

$$p_L q_{2t+1} = (1-\lambda_{2t})\mu_L \frac{(a_{2t} - q_{2t} - q_{2t+1})R + p_L q_{2t+1}}{1-\mu_E-\lambda_{2t}\mu_L} ,$$

yielding

$$q_{2t+1} = \frac{(1-\lambda_{2t})\mu_L(a_{2t} - q_{2t})}{(1-\mu_E-\lambda_{2t}\mu_L)\delta_L + (1-\lambda_{2t})\mu_L(1-\delta_L)} . \tag{A.4.18}$$

Plugging (A.4.17) into (A.4.18) leads to

$$q_{2t+1}(\lambda_{2t}) = \frac{(1-\lambda_{2t})\mu_L \cdot \dfrac{R(a_{2t}-q_{2t})}{1-\mu_E-\lambda_{2t}\mu_L}}{\left[\delta_L + \dfrac{(1-\lambda_{2t})\mu_L(1-\delta_L)}{1-\mu_E-\lambda_{2t}\mu_L}\right] R} \,. \tag{A.4.19}$$

For the monotonicity of $q_{2t}(\lambda_{2t})$, taking derivative of (A.4.17) leads to

$$\frac{\partial q_{2t}(\lambda_{2t})}{\lambda_{2t}} = \frac{\mu_L(\delta_E R a_{2t} + x_{2t})}{[\mu_E(1-\delta_E) + (1-\lambda_{2t}\mu_L)\delta_E + \lambda_{2t}\mu_L]^2} > 0\,,$$

implying that $q_{2t}(\lambda_{2t})$ is increasing in λ_{2t}. Similar procedures based on (A.4.17) and (A.4.19) show that $q_{2t+1}(\lambda_{2t})$ is decreasing in λ_{2t} while $q_{2t}(\lambda_{2t}) + q_{2t+1}(\lambda_{2t})$ is increasing in λ_{2t}. This concludes the proof. □

PROOF OF LEMMA 5. By Lemma 4 and the definition of $\Delta u_L(\lambda_{2t})$:

$$\begin{aligned} \Delta u_L(\lambda_{2t}) &= \theta NAV_{2t} - NAV_{2t+1} \\ &= \theta[(Ra_{2t}+x_{2t}) - (1-\delta_E)Rq_{2t}] - \frac{(Ra_{2t}+x_{2t}) - Rq_{2t} - (1-\delta_L)R(q_{2t+1}+s_{2t+1})}{1-\mu_E-\lambda_{2t}\mu_L} \end{aligned} \tag{A.4.20}$$

in which q_{2t} and q_{2t+1} are functions of λ_{2t} by Lemma 4. It is straightforward that $\Delta u_L(\lambda_{2t})$ is larger when s_{2t+1} increases. To focus on the value of λ_{2t} that satisfies $\theta NAV_{2t} \geqslant NAV_{2t+1}$, there is no loss of generality to consider $\theta = 1$, and the analysis for a general θ naturally follows by considering subsets of λ_{2t}. Now plug (A.4.17) and (A.4.18) into (A.4.20) and then take derivative with respect to λ_{2t}. After rearrangement, this yields:

$$\frac{\partial \Delta u_L(\lambda_{2t})}{\partial \lambda_{2t}} = -\frac{(1-\delta_L)\mu_L\left(x_{2t}C_1 - \dfrac{R\left(s_{2t+1}C_2 - a_{2t}\delta_E(1-\mu_E-\lambda_{2t}\mu_L)^2 C_1\right)}{(1-\mu_E-\lambda_{2t}\mu_L)^2}\right)}{((1-\lambda_{2t})\mu_L + \delta_L(1-\mu_E-\mu_L))^2(\mu_E+\lambda_{2t}\mu_L+\delta_E(1-\mu_E-\lambda_{2t}\mu_L))^2}\,, \tag{A.4.21}$$

where

$$\begin{aligned} C_1 &= (1-\delta_E)(1-\lambda_{2t})^2\mu_L^2 + \delta_L(1-\mu_E-\mu_L)(\mu_E+\mu_L) + \delta_E\delta_L(1-\mu_E-\mu_L)^2 > 0\,, \\ C_2 &= ((1-\lambda_{2t})\mu_L + \delta_L(1-\mu_E-\mu_L))^2\left(\mu_E+\lambda_{2t}\mu_L+\delta_E(1-\mu_E-\lambda_{2t}\mu_L)\right)^2 > 0\,. \end{aligned}$$

Consider

$$C = x_{2t}C_1 - \frac{R\left(s_{2t+1}C_2 - a_{2t}\delta_E(1-\mu_E-\lambda_{2t}\mu_L)^2C_1\right)}{(1-\mu_E-\lambda_{2t}\mu_L)^2}$$

for any $0 \leqslant x_{2t} < \mu_E n_{2t}(Ra_{2t}+x_{2t})$ and any $0 \leqslant s_{2t+1} \leqslant a_{2t} - q_{2t}(\lambda_{2t}) - q_{2t+1}(\lambda_{2t})$. Since $C_1 > 0$, there is $x_{2t}C_1 > 0$ and thus

$$C > \frac{R\left(a_{2t}\left(\delta_E(1-\mu_E-\lambda_{2t}\mu_L)^2C_1 - C_2\right)\right)}{(1-\mu_E-\lambda_{2t}\mu_L)^2}\,. \tag{A.4.22}$$

Notice that C_1 and C_2 are only functions of λ_{2t}, μ_E, μ_L, δ_E, δ_L, and are independent of a_{2t} and x_{2t}. By construction,

$$\delta_E(1-\mu_E-\lambda_{2t}\mu_L)^2C_1 - C_2 \geqslant 0$$

for any $\lambda_{2t} \in [0,1]$.

As a result, since $a_{2t} > 0$ and $R > 0$, inequality (A.4.22) implies that $C > 0$. Plugging back to (A.4.21) finally yields

$$\frac{\partial \Delta u_L(\lambda_{2t})}{\partial \lambda_{2t}} < 0\,,$$

implying strategic substitutability.

It is straightforward that $\Delta u_L(\lambda_{2t})$ is larger when s_{2t+1} increases. Also, by definition, C is decreasing in s_{2t+1} when $0 \leqslant s_{2t+1} \leqslant a_{2t} - q_{2t}(\lambda_{2t}) - q_{2t+1}(\lambda_{2t})$. By (A.4.21) and the derivation above, the strategic substitutability becomes weaker when s_{2t+1} increases. This concludes the proof. □

PROOF OF PROPOSITION 2. Notice that any shareholder runs only if $\theta NAV_{2t} \geqslant NAV_{2t+1}$. Also by Lemma 5, the stage game exhibits strategic substitutability whenever an incentive to redeem earlier exists. Thus, in Case i), none of the late shareholders runs if

$$\theta NAV_{2t}(0) < NAV_{2t+1}(0)\,, \tag{A.4.23}$$

which implies that $\theta NAV_{2t}(\lambda_{2t}) < NAV_{2t+1}(\lambda_{2t})$ for any λ_{2t} by using the expressions in Lemma 4. Thus, solving inequality (A.4.23) leads to

$$s_{2t+1} < \frac{Ra_{2t} - \theta(1-\mu_E)(Ra_{2t}+x_{2t}) - (1-\theta(1-\delta_E)(1-\mu_E))Rq_{2t}(0)}{(1-\delta_L)R} - q_{2t+1}(0) \equiv \underline{s}_l \, .$$

Alternatively, in Case ii), all of the late shareholders run if

$$\theta NAV_{2t}(1) > NAV_{2t+1}(1) \, , \tag{A.4.24}$$

which implies that $\theta NAV_{2t}(\lambda_{2t}) > NAV_{2t+1}(\lambda_{2t})$ for any λ_{2t} despite the underlying strategic substitutability suggested by Lemma 5. Solving inequality (A.4.24) leads to

$$s_{2t+1} > \frac{Ra_{2t} - \theta(1-\mu_E-\mu_L)(Ra_{2t}+x_{2t}) - (1-\theta(1-\delta_E)(1-\mu_E-\mu_L))Rq_{2t}(1)}{(1-\delta_L)R} \equiv \bar{s}_l \, .$$

Using the expressions in Lemma 4, plugging $q_{2t}(0)$, $q_{2t+1}(0)$ and $q_{2t}(1)$ into the definition of $\underline{s}_l$ and $\bar{s}_l$ directly yields $\bar{s}_l \geqslant 0$ and $\bar{s}_l > \underline{s}_l$.

Finally, in Case iii), there exists some $\widetilde{\lambda}_{2t} \in [0,1]$ that solves

$$\theta NAV_{2t}(\widetilde{\lambda}_{2t}) = NAV_{2t+1}(\widetilde{\lambda}_{2t}) \, ,$$

where $\widetilde{\lambda}_{2t}$ constructs an equilibrium because by definition $\Delta u_L(\widetilde{\lambda}_{2t}) = 0$ and thus no shareholder would have an incentive to deviate from it. This leads to

$$s_{2t+1} = \frac{Ra_{2t} - \theta(1-\mu_E-\widetilde{\lambda}_{2t}\mu_l)(Ra_{2t}+x_{2t}) - (1-\theta(1-\delta_E)(1-\mu_E-\widetilde{\lambda}_{2t}\mu_l))Rq_{2t}(\widetilde{\lambda}_{2t})}{(1-\delta_L)R} - q_{2t+1}(\widetilde{\lambda}_{2t}) \, .$$

This concludes the proof. □

Proof of Corollary 2. By Lemma 4, $q_{2t}(\lambda_{2t}) > 0$ for any arbitrary $\lambda_{2t} \in [0,1]$ and $q_{2t+1}(\lambda_{2t}) > 0$ for any arbitrary $\lambda_{2t} \in [0,1)$. Thus, the evolution of the asset position directly follows:

$$a_{2t+2} = a_{2t} - q_{2t}(\lambda_{2t}) - q_{2t+1}(\lambda_{2t}) - s_{2t+1} \, .$$

For the evolution of the cash position, notice that all the proceeds from forced fire sales

$q_{2t}(\lambda_{2t})$ and $q_{2t}(\lambda_{2t+1})$ go to the redeeming shareholders. Also, by definition of the low cash-to-assets region, the fund starts with no cash on date $2t+1$. Thus:

$$\begin{aligned} x_{2t+2} &= p_L(q_{2t+1}(\lambda_{2t}) + s_{2t+1}) - (1-\lambda_{2t})\mu_L NAV_{2t+1} \\ &= \delta_L R s_{2t+1} + \frac{(1-\lambda_{2t})\mu_L(1-\delta_L)Rs_{2t+1}}{1-\mu_E-\lambda_{2t}\mu_L}. \end{aligned}$$

This concludes the proof. □

Proof of Proposition 3. The existence of a Markov equilibrium of the fully dynamic game follows a special case of Theorem 2 and Corollary 6 in Khan and Sun (2002). The key is to find a measurable selection of Nash equilibria in each stage game determined by the state variables (a_{2t}, x_{2t}). The Arsenin-Kunugui Theorem (see Kechris, 1995 for a textbook treatment) guarantees that any usual equilibrium selection mechanism such as selecting the best, the worst or the one based on the global game approach is measurable.

Under any Markov strategy profile, by definition, the strategies of both the fund manager and all the shareholders are functions of the two state variables (a_{2t}, x_{2t}), and their strategies are mutually best responses as well. In other words, strategies played in the past stages influence current-stage strategies only through the two state variables. For convenience, in what follows I call a stage game (a_{2t}, x_{2t}) when the fund starts from the portfolio position (a_{2t}, x_{2t}) on date $2t$.

Consider any arbitrary $\phi \in (0,1)$. Define $a'_{2t} = \phi a_{2t}$, $x'_{2t} = \phi x_{2t}$, and $s'_{2t+1} = \phi a_{2t+1}$. By Lemma 1, game (a_{2t}, x_{2t}) and game (a'_{2t}, x'_{2t}) start from the same cash-to-assets ratio region. By Propositions 1, 2, and 27, if λ_{2t} constructs a run equilibrium in game (a_{2t}, x_{2t}) under the cash rebuilding policy s_{2t+1}, it must also construct a run equilibrium in game (a'_{2t}, x'_{2t}) under the cash rebuilding policy s'_{2t+1}. Hence, by Lemmas 2, 4, and 20, the equilibrium amounts of forced fire sales in game (a'_{2t}, x'_{2t}) must be $q'_{2t} = \phi q_{2t}$, $q'_{2t+1} = \phi q_{2t+1}$, where q_{2t} and q_{2t+1} are the equilibrium amounts of forced fire sales in game (a_{2t}, x_{2t}).

Then consider the dynamics. Fix a consistent equilibrium selection mechanism if multiple equilibria occur. Let (a_{2t+2}, x_{2t+2}) be the next stage game when game (a_{2t}, x_{2t}) is played under the cash rebuilding policy s_{2t+1}. By Corollaries 1, 2 and 6, the next stage game must be (a'_{2t+2}, x'_{2t+2}), where $a'_{2t+2} = \phi a_{2t+2}$ and $x'_{2t+2} = \phi x_{2t+2}$, if the current stage game (a'_{2t}, x'_{2t}) is played under the cash rebuilding policy s'_{2t+1}. Therefore, if $s_{2t+1}(a_{2t}, x_{2t})$ is the optimal cash rebuilding policy in stage t for game (a_{2t}, x_{2t}), $s'_{2t+1}(a'_{2t}, x'_{2t}) = \phi s_{2t+1}(a_{2t}, x_{2t})$ must be the optimal cash rebuilding policy in stage t for game (a'_{2t}, x'_{2t}). Hence, $V(a'_{2t}, x'_{2t}) = \phi V(a_{2t}, x_{2t})$ is indeed the value function for the dynamic game with a starting position (a'_{2t}, x'_{2t}).

Finally, it is straightforward to see that $V(0,0) = 0$. This concludes the proof. □

Proof of Lemma 6. If $a_{2t} = 0$, it is trivial that $s^*_{2t+1}(a_{2t}, x_{2t}) = 0$. So it is only worth considering a strictly positive a_{2t}.

On the one hand, consider a perturbation $\varepsilon > 0$ of cash rebuilding around $s^*_{2t+1}(a_{2t}, x_{2t}) = 0$. On date $2t+1$ (in stage t), regardless of the starting portfolio position (a_{2t}, x_{2t}), the effective selling price on $2t+1$ is at most $\widehat{p}_L(0) > 0$. Thus, the fire sale loss in stage t is at least

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(0)} > 0\,.$$

On other other hand, consider an initial cash gap ε on date $2t+2$ (in stage $t+1$). Regardless of the starting portfolio position (a_{2t+2}, x_{2t+2}), the effective selling price on $2t+2$ is at least $\widehat{p}_E(0) > 0$, and the physical selling price in stage $t+1$ is at least $\delta_E R$. Thus, the expected fire sale loss in stage $t+1$ due to this cash gap is at most

$$\frac{\varepsilon(1-\delta_E)R}{\widehat{p}_E(0)} > 0\,.$$

Therefore, for any π satisfying

$$\pi > 1 - \frac{(1-\delta_L)\widehat{p}_E(0)}{(1-\delta_E)\widehat{p}_L(0)} \in (0,1)\,,$$

it is optimal to choose $s^*_{2t+1}(a_{2t}, x_{2t}) = 0$. This concludes the proof. □

Proof of Lemma 7. This directly follows late shareholders' utility function. □

Proof of Lemma 8. First consider the case of $\eta_{2t} \in G_h$. By Proposition 1, $\theta = 1$ implies that $\bar{s}_h = 0$. Again by Proposition 1, there is $\lambda_{2t} = 1$ for any $s_{2t+1} > 0$ regardless of (a_{2t}, x_{2t}).

Then consider the case of $\eta_{2t} \in G_l$. Similarly, by Proposition 2, $\theta = 1$ implies that $\bar{s}_l = 0$. Again by Proposition 2, there is $\lambda_{2t} = 1$ for any $s_{2t+1} > 0$ regardless of (a_{2t}, x_{2t}).

Finally, consider the case of $\eta_{2t} \in G_m$. By Proposition 27, $\theta = 1$ implies that $\bar{s}_m = \bar{\bar{s}}_m = 0$. Again by Proposition 27, there is $\lambda_{2t} = 1$ for any $s_{2t+1} > 0$ regardless of (a_{2t}, x_{2t}). This concludes the proof. □

Proof of Proposition 4. I consider two cases according to the starting cash-to-assets ratio on date $2t$.

Case 1. $\eta_{2t} \in G_l \cup G_m \cup G_{hl}$. First, consider a perturbation $-\varepsilon < 0$ of cash rebuilding around σ^*_{2t+1} that satisfies $\eta^*_{2t+2} = \mu_E R/(1-\mu_E)$. On date $2t+1$ (in stage t), since there are no runs (by Lemma 7), the effective selling price on $2t+1$ is $\widehat{p}_L(0)$. Thus, the fire sale loss saved in stage t is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(0)} > 0\,.$$

Now consider the same cash gap ε on date $2t+2$ (under the perturbed cash rebuilding policy $(\sigma^*_{2t+1}, -\varepsilon)$). This implies that $\eta_{2t+2} \in G_l$. Since there are no runs, the fund has to fire sell its assets on date $2t+2$ at the effective selling price $\widehat{p}_E(0) > 0$. Hence, the expected increase of fire sale loss in stage $t+1$ due to this cash gap ε is

$$\frac{\varepsilon(1-\delta_E)R}{\widehat{p}_E(0)} > 0\,.$$

Since $\delta_E < \delta_L$ and $\widehat{p}_E(0) < \widehat{p}_L(0)$, there is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(0)} < (1-\pi)\frac{\varepsilon(1-\delta_E)R}{\widehat{p}_E(0)}$$

for a sufficiently small but positive π, implying that the perturbation $-\varepsilon$ is not profitable.

Next, consider another perturbation $\varepsilon > 0$ of cash rebuilding around σ^*_{2t+1} that satisfies $\eta^*_{2t+2} = \mu_E R/(1-\mu_E)$. On date $2t+1$ (in stage t), similarly, the fire sale loss increased in stage t is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(0)} > 0\,.$$

Under the perturbed cash rebuilding policy $(\sigma^*_{2t+1}, \varepsilon)$ the fund gets ε more cash in stage

$t+1$. This implies that $\eta_{2t+2} \in G_m$. Since there are no runs, the fund does not have to fire sell its assets on date $2t+2$. Rather, the marginal cash saves the fund's active asset sales on date $2t+3$ at the effective selling price $\widehat{p}_L(0) > 0$. Hence, the expected fire sale saved in stage $t+1$ due to this marginal cash ε is also

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(0)} > 0\,.$$

Since $\pi \in (0,1)$, this perturbation ε is also not profitable. This verifies the optimality of $\eta^*_{2t+2} = \mu_E R/(1-\mu_E)$ when $\eta_{2t} \in G_l \cup G_m \cup G_{hl}$.

Case 2. $\eta_{2t} \in G_{hm} \cup G_{hh}$. Consider a perturbation $\varepsilon > 0$ of cash rebuilding around $\sigma^*_{2t+1} = 0$. On date $2t+1$ (in stage t), similarly, the fire sale loss increased in stage t is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(0)} > 0\,.$$

Similarly, under the perturbed cash rebuilding policy $(\sigma^*_{2t+1}, \varepsilon)$ the fund gets ε more cash in stage $t+1$. The expected fire sale loss saved in stage $t+1$ due to this marginal cash ε is also

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(0)} > 0\,.$$

Since $\pi \in (0,1)$, this perturbation ε is again not profitable. This verifies the optimality of $\sigma^*_{2t+1} = 0$ when $\eta_{2t} \in G_{hm} \cup G_{hh}$. This finally concludes the proof. □

Proof of Proposition 5. I consider two cases according to the starting cash-to-assets ratio on date $2t$.

Case 1. $\eta_{2t} \in G_l \cup G_m \cup G_{hl} \cup G_{hm}$. First, consider a perturbation $-\varepsilon < 0$ of cash rebuilding around σ^*_{2t+1} that satisfies $\eta^*_{2t+2} = (\mu_E + \mu_L)R/(1-\mu_E-\mu_L)$. On date $2t+1$ (in stage t), since $\lambda_{2t} = 1$ (by Lemma 8), the effective selling price on $2t+1$ is $\widehat{p}_L(1)$. Thus, the fire sale loss saved in stage t is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(1)} > 0\,.$$

Now consider the same cash gap ε on date $2t+2$ (under the perturbed cash rebuilding policy $(\sigma^*_{2t+1}, -\varepsilon)$. This implies that $\eta_{2t+2} \in G_l \cup G_m$. Since $\lambda_{2t+2} = 1$, by Lemmas 4 and 20 the fund always has to fire sell its assets on date $2t+2$ at the effective selling price $\widehat{p}_E(1) > 0$, even if $\eta_{2t+2} \in G_m$. Hence, the expected increase of fire sale loss in stage $t+1$ due to this cash gap ε is

$$\frac{\varepsilon(1-\delta_E)R}{\widehat{p}_E(1)} > 0\,.$$

Since $\delta_E < \delta_L$ and $\widehat{p}_E(1) < \widehat{p}_L(1)$, there is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(1)} < (1-\pi)\frac{\varepsilon(1-\delta_E)R}{\widehat{p}_E(1)}$$

for a sufficiently small but positive π. Hence, the perturbation $-\varepsilon$ is not profitable.

Next, consider another perturbation $\varepsilon > 0$ of cash rebuilding around σ^*_{2t+1} that satisfies $\eta^*_{2t+2} = (\mu_E + \mu_L)R/(1-\mu_E-\mu_L)$. On date $2t+1$ (in stage t), similarly, the fire sale loss increased in stage t is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(1)} > 0\,.$$

Under the perturbed cash rebuilding policy $(\sigma^*_{2t+1}, \varepsilon)$ the fund gets ε more cash in stage $t+1$. This implies that $\eta_{2t+2} \in G_h$. Hence, by Lemma 2, regardless of runs the fund does not have to fire sell its assets on date $2t+2$. Rather, the marginal cash saves the fund's active asset sales on date $2t+3$ at the effective selling price $\widehat{p}_L(1) > 0$. Hence, the expected fire sale saved in stage $t+1$ due to this marginal cash ε is also

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(1)} > 0\,.$$

Since $\pi \in (0,1)$, this perturbation ε is also not profitable. This verifies the optimality of $\eta^*_{2t+2} = (\mu_E + \mu_L)R/(1-\mu_E-\mu_L)$ when $\eta_{2t} \in G_l \cup G_m \cup G_{hl} \cup G_{hm}$.

Case 2. $\eta_{2t} \in G_{hh}$. Consider a perturbation $\varepsilon > 0$ of cash rebuilding around $\sigma^*_{2t+1} = 0$. On date $2t+1$ (in stage t), similarly, the fire sale loss increased in stage t is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(1)} > 0\,.$$

Similarly, under the perturbed cash rebuilding policy $(\sigma^*_{2t+1}, \varepsilon)$ the fund gets ε more cash in stage $t+1$. The expected fire sale loss saved in stage $t+1$ due to this marginal cash ε is also

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(1)} > 0\,.$$

Since $\pi \in (0,1)$, this perturbation ε is again not profitable. This verifies the optimality of $\sigma^*_{2t+1} = 0$ when $\eta_{2t} \in G_{hh}$. This finally concludes the proof. □

PROOF OF PROPOSITION 6. This proof proceeds in three steps. First, I show that when θ is sufficiently small, the equilibrium is the same as that characterized by Proposition 4. Second, I characterize the equilibrium when θ takes an intermediate value. Lastly, I show that when θ is sufficiently large, the equilibrium is the same as that characterized by Proposition 5.

STEP 1. Recall that when $\theta = 0$, the equilibrium cash rebuilding policy is characterized by Proposition 4. By Propositions 1, 2, and 27, $\underline{s}_h$, $\underline{s}_l$, and $\underline{s}_m$ are all continuous in θ. Hence, there exists a $\underline{\theta} > 0$ (explicit expression will be calculated in the next step) such that when $\theta \in (0, \underline{\theta}]$, none of the late shareholders chooses to run in any region if the fund still follows the cash rebuilding policy as described in Proposition 4. In addition, the proof of Proposition 4 only relies on the fact that there are no shareholder runs. This confirms that the cash rebuilding policy as described in Proposition 4 is still optimal when $\theta \in (0, \underline{\theta}]$, which in turn confirms the late shareholders' run decision $\lambda_{2t} = 0$.

STEP 2. By the definition of $\underline{\theta}$, when $\theta > \underline{\theta}$ there exists a non-zero-measure set G_{run} in which at least some of the late shareholders will run given the cash rebuilding policy described in Proposition 4. I first show that G_{run} takes the form of

$$G_{run} = G_l \cup \underline{G_m}\,, \tag{A.4.25}$$

where $\underline{G_m} \subseteq G_m$ is connected and

$$\inf \underline{G_m} = \frac{\mu_E R}{1-\mu_E}\,.$$

To see this, first recall the definition of $\underline{s}_l$:

$$\underline{s}_l = \frac{Ra_{2t} - \theta(1-\mu_E)(Ra_{2t}+x_{2t}) - (1-\theta(1-\delta_E)(1-\mu_E))Rq_{2t}(0)}{(1-\delta_L)R} - q_{2t+1}(0)\,.$$

Note that, for every pair of (a_{2t}, x_{2t}) and η_{2t+2}, there is an implied s_{2t+1}. Using that as the threshold $\underline{s}_l$ and solving for θ backward yields that, under the cash rebuilding policy $\eta_{2t+2} = \mu_E R/(1-\mu_E)$, when

$$\theta > \frac{\delta_L}{\mu_E + \mu_L + \delta_L(1-\mu_E-\mu_L)} \equiv \underline{\theta} \in (0,1)$$

there must be $\lambda_{2t} > 0$ for $\eta_{2t} \in G_l$.

Similarly, consider the definitions of $\underline{s}_m$ and $\underline{s}_h$. Also under the cash rebuilding policy $\eta_{2t+2} = \mu_E R/(1-\mu_E)$, solving for the threshold θ backward yields that, when

$$\theta > \frac{\delta_L}{\mu_E + \mu_L + \delta_L(1-\mu_E-\mu_L)} = \underline{\theta}$$

there must be $\lambda_{2t} > \widehat{\lambda}_{2t}$ for $\eta_{2t} \in G_m$, while when

$$\theta > \frac{\delta_L + \mu_L - \delta_L\mu_L}{\mu_E + \mu_L + \delta_L(1-\mu_E-\mu_L)} \equiv \bar{\bar{\theta}} \in (0,1)$$

there must be $\lambda_{2t} > 0$ for $\eta_{2t} \in G_h$.

Notice that

$$\underline{\theta} < \bar{\bar{\theta}}\,.$$

Thus, under the cash rebuilding policy $\eta_{2t+2} = \mu_E R/(1-\mu_E)$, when $\theta \in (\underline{\theta}, \bar{\bar{\theta}})$, there is $\lambda_{2t} = 0$ when $\eta_{2t} \in G_h$. This confirms the claim in (A.4.25).

Now define

$$\eta(\check{\lambda}) \equiv \frac{(\mu_E + \check{\lambda}\mu_L)R}{1-\mu_E-\check{\lambda}\mu_L} \in G_m\,.$$

For any $\check{\lambda} \in (0,1)$, consider the following cash rebuilding policy:

$$\eta_{2t+2} = \eta(\check{\lambda})\,.$$

Since

$$\frac{\mu_E R}{1-\mu_E} < \eta(\check{\lambda}) < \frac{(\mu_E+\mu_L)R}{1-\mu_E-\mu_L}\,,$$

there exists a $\check{\theta} \in (\underline{\theta}, \bar{\bar{\theta}})$ such that when $\theta = \check{\theta}$, there is

$$\begin{cases} \lambda_{2t} > 0 & \text{iff} \quad \eta_{2t} < \eta(\check{\lambda})\,, \\ \lambda_{2t} = 0 & \text{iff} \quad \eta_{2t} \geqslant \eta(\check{\lambda})\,. \end{cases}$$

Thus, it is natural to define that

$$\underline{G_m} \equiv \left\{ \eta_{2t} | \frac{\mu_E R}{1-\mu_E} \leqslant \eta_{2t} < \eta(\check{\lambda}) \right\} ,$$

and

$$\underline{G_{hm}} \equiv \left\{ \eta_{2t} | \eta_{2t} \geqslant \frac{(\mu_E + \mu_L)R}{1-\mu_E-\mu_L} \text{ and } \frac{\mu_E R}{1-\mu_E} \leqslant \eta_{2t+2} < \eta(\check{\lambda}) \text{ for } \sigma_{2t+1} = 0 \right\} .$$

Now I confirm that $\eta^*_{2t+2} = \eta(\check{\lambda})$ is the optimal cash rebuilding policy when $\theta = \check{\theta}$ and $\eta_{2t} \in \underline{G_m}$. First, consider a perturbation $-\varepsilon < 0$ of cash rebuilding around σ^*_{2t+1} that satisfies $\eta^*_{2t+2} = \eta(\check{\lambda})$. On date $2t+1$ (in stage t), since $\lambda_{2t} = 1$ (by Lemma 8), the effective selling price on $2t+1$ is $\widehat{p}_L(\check{\lambda})$. Thus, the fire sale loss saved in stage t is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(\check{\lambda})} > 0 .$$

Now consider the same cash gap ε on date $2t+2$ (under the perturbed cash rebuilding policy $(\sigma^*_{2t+1}, -\varepsilon)$. This implies that $\eta_{2t+2} \in \underline{G_{hm}}$. Since $\lambda_{2t+2} \geqslant \check{\lambda}$, the fund always has to fire sell its assets on date $2t+2$ at most at the effective selling price $\widehat{p}_E(\check{\lambda}) > 0$. Hence, the expected increase of fire sale loss in stage $t+1$ due to this cash gap ε is at least

$$\frac{\varepsilon(1-\delta_E)R}{\widehat{p}_E(\check{\lambda})} > 0 .$$

Since $\widehat{p}_E(\check{\lambda}) < \widehat{p}_L(\check{\lambda})$, there is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(\check{\lambda})} < (1-\pi)\frac{\varepsilon(1-\delta_E)R}{\widehat{p}_E(\check{\lambda})}$$

for a sufficiently small but positive π. Hence, the perturbation $-\varepsilon$ is not profitable.

Next, consider another perturbation $\varepsilon > 0$ of cash rebuilding around σ^*_{2t+1} that satisfies $\eta^*_{2t+2} = \eta(\check{\lambda})$. On date $2t+1$ (in stage t), similarly, the fire sale loss increased in stage t is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(\check{\lambda})} > 0 .$$

Under the perturbed cash rebuilding policy $(\sigma^*_{2t+1}, \varepsilon)$ the fund gets ε more cash in stage $t+1$. This implies that $\eta_{2t+2} \in G_h$. Since there will be no runs on date $2t+2$, the marginal cash saves the fund's active asset sales on date $2t+3$ at the effective selling price $\widehat{p}_L(1) > 0$. Hence, the expected fire sale saved in stage $t+1$ due to this marginal cash ε is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(1)} > 0\,.$$

Since $\widehat{p}_L(\check{\lambda}) < \widehat{p}_L(1)$ (and also $\pi \in (0,1)$), this perturbation ε is also not profitable. This verifies the optimality of $\eta^*_{2t+2} = (\mu_E + \mu_L)R/(1-\mu_E-\mu_L)$ when $\eta_{2t} \in \underline{G_m}$. This analysis can be readily extended to other subset of $G_l \cup G_m \cup G_{hl} \cup \underline{G_{hm}}$ as well as $\overline{G_{hm}} \cup G_{hh}$ following the same argument.

Finally, define

$$\bar{\theta} \equiv \check{\theta}(\check{\lambda} = 1) \in (\underline{\theta}, \bar{\bar{\theta}})\,.$$

By construction, when $\theta = \bar{\theta}$, there are $\lambda_{2t} > 0$ for $\eta_{2t} \in G_l \cup G_m$ while $\lambda_{2t} = 0$ for $\eta_{2t} \in G_h$ under the corresponding optimal cash rebuilding policy $\eta^*_{2t+2} = \eta(1)$.

Step 3. This step shows that when $\theta > \bar{\theta}$ there can not be equilibria other than that descried by Proposition 5. In this step, I use Figure A.2 to help illustrate the idea. I first show that, when $\theta > \bar{\theta}$, there must be $G_{run} = G_l \cup G_m \cup G_h$. Note that, by Step 2, there must be $G_l \cup G_m \subseteq G_{run}$ when $\theta > \bar{\theta}$, and thus it suffices to show that it cannot be that

$$\sup G_{run} < \sup G_h\,.$$

I prove this by contradiction. First suppose $\sup G_{run} \in G_{hl} \cup G_{hm}$. Define

$$G_{h1} \equiv G_{run}/(G_l \cup G_m)\,.$$

By the argument in the proof of Proposition 5, when $\eta_{2t} \in G_{run}$ the equilibrium cash rebuilding policy still features

$$\eta^*_{2t+2} = \frac{(\mu_E+\mu_L)R}{1-\mu_E-\mu_L}\,. \tag{A.4.26}$$

However, because $\sup G_{run} \in G_{hl} \cup G_{hm}$, one can now find another non-zero-measure connected set $G_{h2} \subseteq G_{hl} \cup G_{hm}$ that satisfies

$$\inf G_{h2} = \sup G_{h1}\,,$$

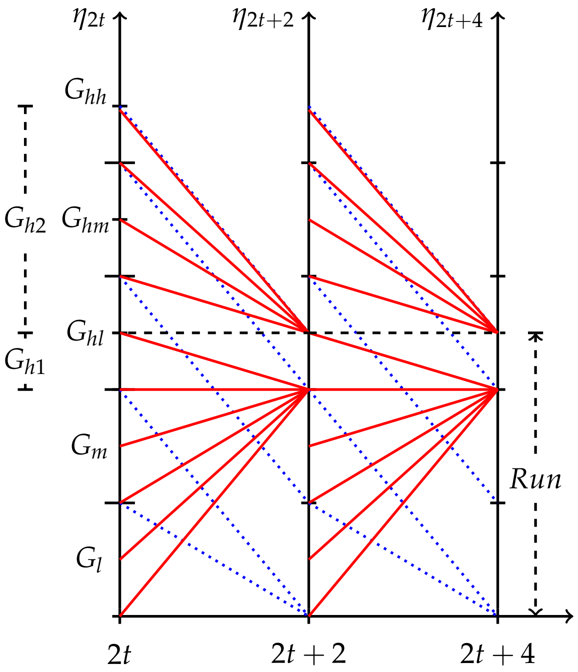


Figure A.2: *Hypothetical Equilibrium Cash Rebuilding Policy*

in which shareholders will not run under the cash rebuilding policy (A.4.26).

By construction, the optimal cash rebuilding policy when $\eta_{2t} \in G_{h2}$ should be

$$\eta^*_{2t+2} \geqslant \sup G_{h1} > \frac{(\mu_E + \mu_L)R}{1 - \mu_E - \mu_L} \, . \tag{A.4.27}$$

To see this, consider a perturbation $-\varepsilon$ of cash rebuilding around this cash rebuilding policy. On date $2t+1$ (in stage t), since there are no runs when $\eta_{2t} \in G_{h2}$, the effective selling price on $2t+1$ is $\widehat{p}_L(0)$. Thus, the fire sale loss saved in stage t is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(0)} > 0 \, .$$

Now consider the same cash gap ε on date $2t+2$ (under the perturbed cash rebuilding policy $(\sigma^*_{2t+1}, -\varepsilon)$. This implies that $\eta_{2t+2} \in G_{h1}$. Because of shareholder runs on date $2t+2$, the fund will sell its assets at the effective selling price $\widehat{p}_L(1) > 0$. Hence, the expected increase of fire sale loss in stage $t+1$ due to this cash gap ε is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(1)} > 0\,.$$

Since $\widehat{p}_L(1) < \widehat{p}_L(0)$, there is

$$\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(0)} < (1-\pi)\frac{\varepsilon(1-\delta_L)R}{\widehat{p}_L(1)}$$

for a sufficiently small but positive π, implying that the perturbation $-\varepsilon$ is not profitable.

However, under the new, more rapid cash rebuilding policy (A.4.27), by the definition of $\underline{s}_h$, there must be a subset of G_{h2} in which late shareholders are going to run. This violates the definition of G_{h2}: a contradiction.

Now instead suppose $\inf G_{hh} \leqslant \sup G_{run} < \sup G_h$. Again by the monotonicity and continuity of $\underline{s}_h$ in a_{2t} and x_{2t}, there is no loss of generality to assume that $\sup G_{run} = \inf G_{hh} + \epsilon$, where $\epsilon > 0$ is arbitrarily small. Similarly, the optimal cash rebuilding policy when $\eta_{2t} = \inf G_{hh} + \epsilon$ should be

$$\eta^*_{2t+2} = \inf G_{hh} > \frac{(\mu_E + \mu_L)R}{1-\mu_E-\mu_L}\,. \tag{A.4.28}$$

However, when $\eta_{2t} = \sup G_{hm} = \inf G_{hh}$, the optimal cash rebuilding policy already leads to runs. By the definition of $\underline{s}_h$, there must be shareholder runs when $\eta_{2t} = \inf G_{hh} + \epsilon$ under the cash rebuilding policy (A.4.28). This is again a contraction. As a result, there must be $G_{run} = G_l \cup G_m \cup G_h$.

Finally, by Proposition 5, the optimal cash rebuilding policy must be the same as described there because the pattern of shareholder runs is the same as described by Lemma 8. This ultimately concludes the proof. □

Proof of Proposition 7. It directly follows Propositions 4, 5, and 6 that for any θ, the proposed cash rebuilding policy

$$\eta^*_{2t+2} < \frac{\mu_E R}{1-\mu_E}$$

is not optimal when there is no commitment device. Here I provide a sufficient condition that it can be optimal if a commitment device is introduced.

Consider $\underline{\theta}$ as defined in Proposition 6. By definition, when $\theta = \underline{\theta} + \epsilon$, where $\epsilon > 0$ is arbitrarily small, and $\eta_{2t+2} = \mu_E R/(1-\mu_E)$, there is $\lambda_{2t} > 1$ for any $\eta_{2t} \in G_l$. Consider a perturbation $-\varepsilon < 0$ of cash rebuilding around $\eta_{2t+2} = \mu_E R/(1-\mu_E)$ when $\eta_{2t} = \mu_E R/(1-\mu_E)$, where the perturbation is chosen such that there is $\lambda_{2t} = 0$ for any $\eta_{2t} \in G_l$. On the one hand, since $\lambda_{2t+2} = 0$, the cash gap resulted from this perturbation on date $2t+2$ leads to the following expected increase of fire sale loss in stage $t+1$:

$$\frac{\varepsilon(1-\delta_E)R}{\widehat{p}_E(0)}.$$

On the other hand, when a commitment device is introduced, the determination of η_{2t+2} on $2t$ directly affects q_{2t} and q_{2t+1} through λ_{2t}. Thus, under the proposed perturbation $-\varepsilon < 0$, there are no runs on date $2t$, and thus the fire sale loss saved in stage t is

$$\Delta q_{2t}(1-\delta_E)R + \left(\frac{\varepsilon}{\widehat{p}_L(\lambda_{2t})} + \Delta q_{2t+1}\right)(1-\delta_L)R\,,$$

where λ_{2t} solves

$$\frac{\varepsilon}{\widehat{p}_L(\lambda_{2t})} = \frac{Ra_{2t} - \theta(1-\mu_E-\lambda_{2t}\mu_l)(Ra_{2t}+x_{2t}) - (1-\theta(1-\delta_E)(1-\mu_E-\lambda_{2t}\mu_l))Rq_{2t}(\lambda_{2t})}{(1-\delta_L)R} - q_{2t+1}(\lambda_{2t})\,,$$

and $\Delta q_{2t} = q_{2t}(\lambda_{2t}) - q_{2t}(0)$ and $\Delta q_{2t+1} = q_{2t+1}(0) - q_{2t+1}(\lambda_{2t})$.

Note that $\Delta q_{2t}(1-\delta_E)R + \Delta q_{2t+1}(1-\delta_L)R > 0$. Thus, if

$$\Delta q_{2t}(1-\delta_E) + \Delta q_{2t+1}(1-\delta_L) > (1-\pi)\frac{\varepsilon(1-\delta_E)}{\widehat{p}_E(0)} - \frac{\varepsilon(1-\delta_L)}{\widehat{p}_L(\lambda_{2t})}$$

satisfies, it is optimal for the fund to choose a less rapid cash rebuilding policy $\eta_{2t+2} < \mu_E R/(1-\mu_E)$. This concludes the proof. □

PROOF OF PROPOSITION 8. This directly follows Propositions 4, 5, and 6. Under a Markov strategy profile, because the equilibrium is stationary, it suffices to show that a higher θ leads to higher total fire sale losses within stage t for any given positive (a_{2t}, x_{2t}). I then consider two cases.

CASE 1. $\eta_{2t} \in G_l \cup G_m$. By Propositions 4, 5, and 6, the optimal cash rebuilding policy is always $\eta^*_{2t+2}(\theta) > 0$ where $\eta^*_{2t+2}(\theta)$ is increasing in θ. By definition, $\eta_{2t+2} = 0$ if $s_{2t+1} = 0$.

Thus, the total fire sale losses in stage t is given by

$$L_t(\theta) = (1-\delta_E)Rq_{2t} + (1-\delta_L)Rq_{2t+1} + \frac{\eta^*_{2t+2}}{\widehat{p}_L + \eta^*_{2t+2}}(1-\delta_L)R(a_{2t} - q_{2t} - q_{2t+1}),$$

where q_{2t}, q_{2t+1}, and $\widehat{p}$ are both functions of λ_0 and in turn functions of θ. Propositions 4, 5, and 6 also imply that λ_{2t} is increasing in θ for any given positive (a_{2t}, x_{2t}). Hence, it follows Lemmas 4 and 20 that $L(\theta)$ is increasing in θ.

CASE 2. $\eta_{2t} \in G_h$. By Lemma 2, $q_{2t} = q_{2t+1} = 0$ regardless of λ_{2t} or θ. Thus, the total fire sale losses in stage t is given by

$$L_t(\theta) = (1-\delta_L)Rs^*_{2t+1}.$$

Define $\underline{\eta_{2t+2}}$ as the target cash-to-assets ratio if $s_{2t+1} = 0$ and $\eta_{2t} \in G_h$. By Propositions 4, 5, and 6, the difference $\eta^*_{2t+2}(\theta) - \underline{\eta_{2t+2}}$ is increasing in θ. Moreover, $\widehat{p}_L$ is decreasing in λ_{2t} and thus decreasing in θ. This implies that s^*_{2t+1} is increasing in θ and so is $L(\theta)$ in this case. This finally concludes the proof. □

PROOF OF PROPOSITION 9. Recall that, the Bellman equation for the non-commitment case is:

$$V(a_{2t}, x_{2t}) = -(1-\delta_E)Rq_{2t} - (1-\delta_L)Rq_{2t+1} + \max_{s^*_{2t+1}}[-(1-\delta_L)Rs_{2t+1} + (1-\pi)V(a_{2t+2}, x_{2t+2})]. \tag{A.4.29}$$

When a commitment device is introduced, the Bellman equation instead becomes:

$$V(a_{2t}, x_{2t}) = \max_{s^*_{2t+1}}[-(1-\delta_E)Rq_{2t} - (1-\delta_L)R(q_{2t+1} + s_{2t+1}) + (1-\pi)V(a_{2t+2}, x_{2t+2})]. \tag{A.4.30}$$

Also, the fund manager's objective function, which is to minimize the total expected fire sale losses, can be re-written as

$$\max_{\{s^*_{2\tau+1}\}_{\tau=t}^{\infty}} \mathbb{E}_t \sum_{\tau=t}^{T-1} [-(1-\delta_E)Rq_{2t} - (1-\delta_L)R(q_{2t+1} + s_{2t+1})],$$

where the expectation is taken over the random variable T,[5] which is in turn govern by probability π. By the Principle of Optimality, the solution to (A.4.30) maximizes the fund manager's objective function, while the solution to (A.4.29) is feasible for the sequential problem associated with (A.4.30). This concludes the proof. □

PROOF OF PROPOSITION 10 AND COROLLARY 3. First, according to the starting cash-to-assets ratio η_{2t}, I still divide the stage game into three different regions. Without loss of generality, I consider $n_{2t} = 1$ as in the baseline model. Suppose the fund does not rebuild its cash buffer and no late shareholder is going to run, that is, $s_{2t+1} = 0$ and $\lambda_{2t} = 0$. Then there are three regions of the cash-to-assets ratio η_{2t} in the stage-t game. In these three regions, the amounts of illiquid assets that the fund has to sell passively on dates t and $t+1$ are characterized by:

$$\begin{cases} \text{High Region } G_h^\kappa\text{:} & q_{2t}^\kappa = 0, q_{2t+1}^\kappa = 0\,, \\ \\ \text{Intermediate Region } G_m^\kappa\text{:} & q_{2t}^\kappa = 0, q_{2t+1}^\kappa > 0\,, \\ \\ \text{Low Region } G_l^\kappa\text{:} & q_{2t}^\kappa > 0, q_{2t+1}^\kappa > 0\,. \end{cases}$$

I use the superscript κ to indicate the existence of the redemption fees. Note that, if a starting position (a_{2t}, x_{2t}) falls into a region G_j, $j \in \{h, m, l\}$, it does not necessarily falls into the same region G_j^κ when redemption fees are introduced. But by construction, there is

$$G_h^\kappa \cup G_m^\kappa \cup G_l^\kappa = G_h \cup G_m \cup G_l\,,$$

and

$$G_j^\kappa \cap G_k^\kappa = \varnothing, j \neq k\,.$$

Thus it suffices to consider the three regions G_h^κ, G_m^κ, and G_l^κ separately. Here I provide a complete analysis of the high region G_h^κ and the derivation for the other two regions directly follows.

---

[5] To be precise, the random variable T here denotes the stage (rather than the date) before which the game ends.

In the high region G_h^κ, when $q_{2t}^\kappa = 0$ and $\lambda_{2t} = 0$, there is

$$NAV_{2t}^\kappa = Ra_{2t} + x_{2t} \,,$$

and

$$NAV_{2t+1}^\kappa = \frac{1-\kappa\mu_E}{1-\mu_E}(Ra_{2t} + x_{2t}) \,.$$

Thus, $q_{2t}^\kappa = 0$ and $q_{2t+1}^\kappa = 0$ imply

$$\eta_{2t} \geqslant \frac{\left(\kappa\mu_E + \kappa\mu_L \dfrac{1-\kappa\mu_E}{1-\mu_E}\right) R}{1 - \kappa\mu_E - \kappa\mu_L \dfrac{1-\kappa\mu_E}{1-\mu_E}} \,.$$

This suggests that $G_h \subseteq G_h^\kappa$. This also suggests that Lemma 2 still holds. That means, for any λ_{2t}:

$$NAV_{2t}^\kappa(\lambda_{2t}) = Ra_{2t} + x_{2t} \,.$$

Meanwhile, when shareholder runs and cash rebuilding are introduced, there is

$$NAV_{2t+1}^\kappa(\lambda_{2t}) = \frac{R(a_{2t} - s_{2t+1}) + x_{2t} - \kappa(\mu_E + \lambda_{2t}\mu_L)(Ra_{2t} + x_{2t}) + \delta_L R s_{2t+1}}{1 - \mu_E - \lambda_{2t}\mu_L} \,.$$

Therefore, when $\theta = 1$, late shareholders' run incentives are governed by

$$\Delta NAV^\kappa(\lambda_{2t}) = \frac{\delta_L R s_{2t+1}}{1 - \mu_E - \lambda_{2t}\mu_L} - \frac{(1-\kappa)(\mu_E + \lambda_{2t}\mu_L)(Ra_{2t} + x_{2t})}{1 - \mu_E - \lambda_{2t}\mu_L} \,. \qquad \text{(A.4.31)}$$

Clearly, when there are no redemption fees, that is, when $\kappa = 1$, this goes back to wedge (1.3.3) in the baseline model. For any $\kappa \in (0, 1)$ and any $\lambda_{2t} \in [0, 1]$, the second term in (A.4.31) is strictly positive. This directly implies that for any feasible s_{2t+1}, there is $\lambda_{2t}^\kappa \leqslant \lambda_{2t}$, where λ_{2t}^κ is the equilibrium run probability in the game with the redemption fee while λ_{2t} is that in the game without redemption fees, leading to the results in Proposition 10.

Also, for any (a_{2t}, x_{2t}) and any $\kappa \in (0, 1)$, define

$$\bar{s} = \inf_{\lambda_{2t} \in [0,1]} \inf_{s_{2t+1}} \{ s_{2t+1} | \Delta NAV^\kappa(s_{2t+1}; \lambda_{2t}) \geqslant 0 \} \,.$$

By construction, there is $\bar{s} > 0$. Then the result follows because $\Delta NAV^\kappa(\lambda_{2t})$ is strictly increasing in s_{2t+1}. This leads to the results in Corollary 3 and finally concludes the proof. $\square$

Proof of Proposition 11. Under in-kind redemptions, any shareholder who redeems on date t will get a_t/n_t unit of assets and x_t/n_t unit of cash. Since there will be no forced fire sales at the fund level, the fund will no longer manage its cash buffer. This implies $\eta_t = \eta_0$ for any date t, where η_0 is the initial cash-to-assets ratio.

Consider any late shareholder on any odd date $2t+1$. If she redeems and consumes on date $2t+1$, she gets $\delta_L R a_0/n_0 + x_0/n_0$, while if she redeemed and consumed on date $2t$, she would get $\delta_E R a_0/n_0 + x_0/n_0$. Since $\delta_L > \delta_E$, no late shareholder will ever run in an equilibrium.

Now I consider total fire sale losses when $\theta = 0$. There is no loss of generality to consider $\eta_{2t} = \mu_E R/(1-\mu_E)$, which is the steady-state cash-to-assets ratio in the baseline model. Again due to the scale-invariance of the dynamic game, it suffices to consider an arbitrary state t. In the baseline model, by Proposition 4, the fire sale losses in stage t under the optimal cash rebuilding policy are:

$$\begin{aligned} L_t &= (1-\delta_L)R(q_{2t+1} + s^*_{2t+1}) \\ &= \frac{(1-\delta_L)R\mu_E\left(a_{2t} - \frac{\mu_L(Ra_{2t}+x_{2t})}{\left(\delta_L + \frac{\mu_L(1-\delta_L)}{1-\mu_E}\right)}\right)}{(1-\mu_E)\delta_L + \mu_L(1-\delta_L) + \mu_E} + \frac{(1-\delta_L)R\mu_L(Ra_{2t}+x_{2t})}{\left(\delta_L + \frac{\mu_L(1-\delta_L)}{1-\mu_E}\right)}, \end{aligned} \tag{A.4.32}$$

while the fire sale losses in stage t under in-kind redemptions are

$$L_t^{in-kind} = (1-\delta_E)R\mu_E a_{2t} + (1-\delta_L)R\mu_L a_{2t}\,. \tag{A.4.33}$$

Note that, when $\mu_L = 0$, (A.4.32) reduces to

$$L_t = \frac{(1-\delta_L)R\mu_E a_{2t}}{(1-\mu_E)\delta_L + \mu_E}\,,$$

while (A.4.33) reduces to

$$L_t^{in-kind} = (1-\delta_E)R\mu_E a_{2t}\,.$$

Clearly, when δ_L is sufficiently larger than δ_E such that

$$1 - \delta_E > \frac{1 - \delta_L}{(1 - \mu_E)\delta_L + \mu_E},$$

there is $L_t^{in-kind} > L_t$. Since fire sale losses are continuous in μ_L, and by Proposition 6 they are also continuous in θ, this concludes the proof. $\square$

PROOF OF PROPOSITION 12. This directly follows the proof of Proposition 10. The only difference is that under redemption fees any individual redeeming shareholder gets κNAV_t on date t while she gets NAV_t (if not being denied) under redemption restrictions. In expectation, any shareholder gets ζNAV_t when she redeems her shares under redemption restrictions. As such, this proof is identical to that of Proposition 10. $\square$

Appendix B

Appendix to Chapter 2

B.1 Proofs

This appendix provides all proofs omitted above with auxiliary results.

Proof of Lemma 9. According to the equilibrium definition, any speculator (i, j) longs one share of asset i when $\phi_i x_{ij} + y_{ij} > \mu_i$ and shorts one share otherwise. Equivalently, speculator (i, j) longs one share of asset i when

$$\frac{\phi_i x_{ij} + y_{ij}}{\sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}} > \frac{\mu_i - (\phi_i a + f_i)}{\sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}},$$

and shorts one share otherwise. Thus, in asset market i, all speculators' aggregate demand is

$$D_i = 1 - 2\Phi\left(\frac{\mu_i - (\phi_i a + f_i)}{\sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}}\right).$$

Hence, in equilibrium, market clearing implies

$$1 - 2\Phi\left(\frac{\mu_i - (\phi_i a + f_i)}{\sqrt{\tau_x^{-1}\phi_i^2 + \tau_y^{-1}}}\right) = 1 - 2\Phi(\zeta + \xi_i - \lambda \log P_i),$$

which further implies that the equilibrium price in asset market i is

$$P_i = \exp\left(\frac{\phi_i}{\lambda\sqrt{\tau_x^{-1}\phi_i^2+\tau_y^{-1}}}a + \frac{1}{\lambda\sqrt{\tau_x^{-1}\phi_i^2+\tau_y^{-1}}}f_i + \frac{\zeta+\xi_i}{\lambda} - \frac{\mu_i}{\lambda\sqrt{\tau_x^{-1}\phi_i^2+\tau_y^{-1}}}\right).$$

This concludes the proof. □

Proof of Lemma 10. In a symmetric equilibrium, the capital providers put a same weight ϕ on the information of the common shock in any asset market i. Thus, by Lemma 9, for asset price P_i, its equivalent signal in predicting the common shock a becomes

$$z_a(P_i) = a + \frac{1}{\phi}f_i + \frac{\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}{\phi}(\zeta+\xi_i).$$

Since f_i and ξ_i are both i.i.d. and have zero means, the aggregate price $\overline{P}$ is equivalent to the following signal

$$z_a(\overline{P}) = \int_0^1\left(a + \frac{1}{\phi}f_i + \frac{\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}{\phi}(\zeta+\xi_i)\right)\mathrm{d}i = a + \frac{\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}{\phi}\zeta$$

in predicting the common shock a. It immediately follows the construction of the other signal $z_{f,i}(\overline{P})$ in predicting the idiosyncratic shock.

Finally, it is easy to verify that any combination of the asset prices $\{P_i, i \in [0,1]\}$ cannot be more informative in predicting the two productivity shocks. This concludes the proof. □

Proof of Lemma 11. From the capital providers' problem (2.2.1), the first order condition is

$$\begin{aligned} I_i &= \frac{\kappa}{c}\mathbb{E}[\exp(a+f_i)|\Gamma_i] \\ &= \frac{\kappa}{c}\exp\left(\mathbb{E}[a+f_i|\Gamma_i] + \frac{1}{2}\mathrm{Var}[a+f_i|\Gamma_i]\right). \end{aligned}$$

By Lemma 10, we know that $s_{a,i}$ and $z_a(\overline{P})$ are only informative about the common shock a and $s_{f,i}$ and $z_{f,i}(\overline{P})$ are only informative about the idiosyncratic shock f_i. Applying Bayesian updating immediately leads to the following optimal investment policy

$$I_i = \frac{\kappa}{c}\exp\left[\frac{\tau_{sa}s_{a,i}+\tau_{\overline{pa}}z_a(\overline{P})}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}} + \frac{1}{2(\tau_a+\tau_{sa}+\tau_{\overline{pa}})} + \frac{\tau_f s_{f,i}+\tau_{\overline{pf}}z_{f,i}(\overline{P})}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}} + \frac{1}{2(\tau_f+\tau_{sf}+\tau_{\overline{pf}})}\right].$$

This concludes the proof. ☐

Proof of Lemma 12. This is a direct application of Lemma 10 to Definition 4. ☐

Proof of Proposition 13. We proceed step by step.

Step 1: Proof of the existence of the solution.

Following the equilibrium condition (2.3.10) for the cross-learning case, let

$$g(\phi) = \phi - \frac{\alpha_1 - \gamma_1}{\alpha_2 - \gamma_2} = \phi - \frac{\frac{\tau_{sa}+\tau_{\overline{pa}}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}} + 1 - \frac{\phi}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}}{\frac{\tau_{sf}+\tau_{\overline{pf}}}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}} + 1 - \frac{1}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}} \frac{\frac{\tau_x}{\tau_a+\tau_x}}{\frac{\tau_y}{\tau_f+\tau_y}},$$

where $\tau_{\overline{pa}}$ is given by (2.3.5) and $\tau_{\overline{pf}}$ is given by (2.3.7), both being function of ϕ. It is easy to check that $\lim_{\phi\to-\infty} g(\phi) < 0$ and $\lim_{\phi\to+\infty} g(\phi) > 0$ by the following two equations:

$$\lim_{\phi\to-\infty} \frac{\frac{\tau_{sa}+\tau_{\overline{pa}}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}} + 1 - \frac{\phi}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}}{\frac{\tau_{sf}+\tau_{\overline{pf}}}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}} + 1 - \frac{1}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}} \frac{\frac{\tau_x}{\tau_a+\tau_x}}{\frac{\tau_y}{\tau_f+\tau_y}} = \frac{\frac{\tau_x}{\tau_a+\tau_x}}{\frac{\tau_y}{\tau_f+\tau_y}} \frac{\frac{\tau_{sa}+\tau_\zeta\tau_x}{\tau_a+\tau_{sa}+\tau_\zeta\tau_x} + 1 - \frac{\sqrt{\tau_x}}{\lambda}}{\frac{\tau_{sf}}{\tau_f+\tau_{sf}} + 1},$$

and

$$\lim_{\phi\to+\infty} \frac{\frac{\tau_{sa}+\tau_{\overline{pa}}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}} + 1 - \frac{\phi}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}}{\frac{\tau_{sf}+\tau_{\overline{pf}}}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}} + 1 - \frac{1}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}} \frac{\frac{\tau_x}{\tau_a+\tau_x}}{\frac{\tau_y}{\tau_f+\tau_y}} = \frac{\frac{\tau_x}{\tau_a+\tau_x}}{\frac{\tau_y}{\tau_f+\tau_y}} \frac{\frac{\tau_{sa}+\tau_\zeta\tau_x}{\tau_a+\tau_{sa}+\tau_\zeta\tau_x} + 1 - \frac{\sqrt{\tau_x}}{\lambda}}{\frac{\tau_{sf}}{\tau_f+\tau_{sf}} + 1}.$$

The analysis above indicates that there always exists a solution of ϕ to the equilibrium condition (2.3.10), i.e., $g(\phi) = 0$, by the intermediate value theorem. Especially, when $\lambda > 1/\sqrt{\tau_x^{-1}}$, we know that

$$f(0) = -\frac{\frac{\tau_{sa}+\tau_{\overline{pa}}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}} + 1}{\frac{\tau_{sf}+\tau_{\overline{pf}}}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}} + 1 - \frac{1}{\lambda\sqrt{\tau_x^{-1}}}} \frac{\frac{\tau_x}{\tau_a+\tau_x}}{\frac{\tau_y}{\tau_f+\tau_y}} < 0.$$

We conclude that there always exists a positive solution $\phi > 0$ as long as λ is large enough.

Step 2: Proof of the uniqueness of the solution when τ_f is large enough.

By simple algebra, the equilibrium condition (2.3.10) is re-expressed as

$$\left(\frac{\tau_{sa}+\tau_{\overline{pa}}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}} + 1 - \frac{\phi}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}\right) \frac{\tau_x}{\tau_a+\tau_x} = \phi \left(\frac{\tau_{sf}+\tau_{\overline{pf}}}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}} + 1 - \frac{1}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}\right) \frac{\tau_y}{\tau_f+\tau_y}. \tag{B.1.1}$$

Applying Taylor expansion to the terms in equation (B.1.1) with respect to τ_y^{-1} yields:

$$\frac{\tau_{sa}+\tau_{\overline{pa}}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}} = -\frac{\tau_a}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}} - \frac{\tau_a}{(\tau_a+\tau_{sa}+\tau_{\overline{pa}})^2}\frac{\tau_\zeta}{\tau_x^{-2}\phi^2}\,\tau_y^{-1} + o(\tau_y^{-1})\,,$$

$$\frac{\tau_{sf}+\tau_{\overline{pf}}}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}} = 1-\frac{\tau_f}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}} - \frac{\tau_f}{(\tau_f+\tau_{sf}+\tau_{\overline{pf}})^2}\frac{\tau_\xi}{\tau_x^{-2}\phi^4}\,\tau_y^{-1} + o(\tau_y^{-1})\,,$$

$$\frac{\phi}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}} = \frac{1}{\lambda\sqrt{\tau_x^{-1}}} - \frac{1}{2\lambda\tau_x^{-\frac{3}{2}}\phi^2}\tau_y^{-1} + o(\tau_y^{-1})\,,$$

and

$$\frac{1}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}} = \frac{1}{\lambda\phi\sqrt{\tau_x^{-1}}} - \frac{1}{2\lambda\tau_x^{-\frac{3}{2}}\phi^3}\tau_y^{-1} + o(\tau_y^{-1})\,.$$

Plugging them back into equation (B.1.1), we have:

$$\begin{aligned} &\tfrac{\tau_x}{\tau_a+\tau_x}\tfrac{\tau_f+\tau_y}{\tau_y}\left[2-\tfrac{\tau_a}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}}-\tfrac{\tau_a}{(\tau_a+\tau_{sa}+\tau_{\overline{pa}})^2}\tfrac{\tau_\zeta}{\tau_x^{-2}\phi^2}\,\tau_y^{-1}-\tfrac{1}{\lambda\sqrt{\tau_x^{-1}}}+\tfrac{1}{2\lambda\tau_x^{-\frac{3}{2}}\phi^2}\tau_y^{-1}+o(\tau_y^{-1})\right] \\ = \quad &\phi\left[2-\tfrac{\tau_f}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}}-\tfrac{\tau_f}{(\tau_f+\tau_{sf}+\tau_{\overline{pf}})^2}\tfrac{\tau_\xi}{\tau_x^{-2}\phi^4}\,\tau_y^{-1}-\tfrac{1}{\lambda\phi\sqrt{\tau_x^{-1}}}+\tfrac{1}{2\lambda\tau_x^{-\frac{3}{2}}\phi^3}\tau_y^{-1}+o(\tau_y^{-1})\right], \end{aligned}$$

which becomes a cubic equation of ϕ when τ_y goes to infinity:

$$\left(\frac{\tau_{sa}+\tau_x\tau_\zeta}{\tau_a+\tau_{sa}+\tau_x\tau_\zeta}+1-\frac{\sqrt{\tau_x}}{\lambda}\right)\frac{\tau_x}{\tau_a+\tau_x} = \frac{\tau_{sf}\phi^3+\tau_x\tau_\xi\phi}{\tau_f\phi^2+\tau_{sf}\phi^2+\tau_x\tau_\xi}+\phi-\frac{\sqrt{\tau_x}}{\lambda}\,. \tag{B.1.2}$$

Note that, the left hand side of equation (B.1.2) does not depends on ϕ. Denote by $h(\phi)$ the right hand side of (B.1.2), and its first order derivative with respect to ϕ is given by

$$\frac{\partial h(\phi)}{\partial\phi} = 1-\frac{\tau_f\phi^2}{\tau_f\phi^2+\tau_{sf}\phi^2+\tau_x\tau_\xi}+1-\frac{2\tau_f\tau_x\tau_\xi\phi^2}{(\tau_f\phi^2+\tau_{sf}\phi^2+\tau_x\tau_\xi)^2} > 0\,,$$

which indicates that the right hand side of equation (B.1.2) is increasing in ϕ and thus we have a unique solution to equation (B.1.2). Therefore, since $g(\phi)$ is a continuous function of τ_y, there always exists one unique solution to $g(\phi)=0$, i.e., equation (B.1.1), when τ_y is large enough. This concludes the proof. □

PROOF OF PROPOSITION 14. This proof is similar to the proof of Proposition 13. In the benchmark case without cross learning, the equilibrium condition (2.3.11) is re-expressed as

$$\left(\frac{\tau_{sa}+\tau_{pa}}{\tau_a+\tau_{sa}+\tau_{pa}}+\frac{\tau_{pf}}{\tau_f+\tau_{sf}+\tau_{pf}}\phi'+1-\frac{\phi}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}\right)\frac{\tau_x}{\tau_a+\tau_x}$$
$$=\quad\phi\left(\frac{\tau_{sf}+\tau_{pf}}{\tau_f+\tau_{sf}+\tau_{pf}}+\frac{\tau_{pa}}{\tau_a+\tau_{sa}+\tau_{pa}}\frac{1}{\phi'}+1-\frac{1}{\lambda\sqrt{\tau_x^{-1}\phi^2+\tau_y^{-1}}}\right)\frac{\tau_y}{\tau_f+\tau_y},$$

which further reduces to

$$\left(\frac{\tau_{sa}}{\tau_a+\tau_{sa}}+1-\frac{\sqrt{\tau_x}}{\lambda}\right)\frac{\tau_x}{\tau_a+\tau_x}=\phi\left(\frac{\tau_{sf}}{\tau_f+\tau_{sf}}+1-\frac{\sqrt{\tau_x}}{\lambda\phi}\right), \tag{B.1.3}$$

when τ_y goes to infinity, τ_ξ goes to zero and $\lambda > \sqrt{\tau_x}$. Since the right hand side of equation (B.1.3) is increasing inf ϕ, we know that there must exist one unique solution to equation (B.1.3).

On the other hand, when τ_ξ goes to zero (and when τ_y goes to infinity and $\lambda > \sqrt{\tau_x}$), equation (B.1.2) in the case with cross learning becomes

$$\left(\frac{\tau_{sa}+\tau_{\overline{pa}}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}}+1-\frac{\sqrt{\tau_x}}{\lambda}\right)\frac{\tau_x}{\tau_a+\tau_x}=\phi\left(\frac{\tau_{sf}}{\tau_f+\tau_{sf}}+1-\frac{\sqrt{\tau_x}}{\lambda\phi}\right). \tag{B.1.4}$$

We compare between the two equations (B.1.3) and (B.1.4). It is clear that their right hand sides are the same, while the left hand side of (B.1.3) (for the benchmark case without cross learning) is smaller than the left hand side of (B.1.4) (for the case with cross learning). By the continuity with respect to τ_y and τ_ξ of the two equilibrium conditions (2.3.11) and (2.3.10) in the two cases, we conclude that the equilibrium ϕ' in the benchmark case without cross learning is always lower than the equilibrium ϕ is the problem with cross learning, as long as $\lambda > \sqrt{\tau_x}$, τ_y is large enough, and τ_ξ is small enough. □

PROOF OF LEMMA 13. We first consider the investment beta β_I. Recall the investment policy (2.3.8), we have

$$\log I_i=\frac{\tau_{sa}s_{a,i}+\tau_{\overline{pa}}z_a(\overline{P})}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}}+\frac{1}{2(\tau_a+\tau_{sa}+\tau_{\overline{pa}})}+\frac{\tau_f s_{f,i}+\tau_{\overline{pf}}z_{f,i}(\overline{P})}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}}+\frac{1}{2(\tau_f+\tau_{sf}+\tau_{\overline{pf}})}.$$

Following the definition of β_I and after some algebra, we reach

$$\beta_I = \frac{S_a/\tau_a - \frac{\tau_{sa}}{(\tau_a+\tau_{sa}+\tau_{\overline{pa}})^2}}{S_a/\tau_a + S_f/\tau_f}.$$

To simplify the analysis, let $g_1 = S_a/\tau_a$, $g_2 = \tau_{sa}/\left(\tau_a + \tau_{sa} + \tau_{\overline{pa}}\right)^2$, and $g_3 = S_f/\tau_f$. By Lemma 12, it is straightforward that g_1 is increasing in ϕ and both g_2 and g_3 are decreasing in ϕ. Thus, as $\phi > 0$, we also have that g_1 is increasing in ϕ^2 and both g_2 and g_3 are decreasing in ϕ^2. Furthermore, we have

$$\begin{aligned}\frac{\partial \beta_I}{\partial \phi^2} &= \frac{g_1' - g_2'}{g_1 + g_3} - \frac{(g_1 - g_2)(g_1' + g_3')}{(g_1 + g_3)^2} \\ &= \frac{[(g_1' - g_2')g_1 - (g_1 - g_2)g_1'] + (g_1' - g_2')g_3 - (g_1 - g_2)g_3'}{(g_1 + g_3)^2},\end{aligned}$$

where g_1', g_2' and g_3' stands for the first order derivative with respect to ϕ^2.

Since we know that $g_1' - g_2' > g_1'$ (due to the fact that g_2 is decreasing in ϕ^2) and $g_1 > g_1 - g_2$, we have $(g_1' - g_2')g_1 - (g_1 - g_2)g_1' > 0$. Meanwhile, we have $g_1' > 0$, $g_2' < 0$, and $g_3 > 0$, so that $(g_1' - g_2')g_3 > 0$. Lastly, since $g_1 > g_2$ and $g_3' < 0$, we also know that $(g_1 - g_2)g_3' < 0$. Therefore, we conclude that $\partial\beta_I/\partial\phi^2 > 0$, i.e., β_I is an increasing function of ϕ when $\phi > 0$.

We then consider the price beta β_P. Recall the pricing function (2.3.1), we have

$$\log P_i = \frac{\phi}{\lambda\sqrt{\tau_x^{-1}\phi^2 + \tau_y^{-1}}}a + \frac{1}{\lambda\sqrt{\tau_x^{-1}\phi^2 + \tau_y^{-1}}}f_i + \frac{\zeta + \xi_i}{\lambda} - \frac{\mu}{\lambda\sqrt{\tau_x^{-1}\phi^2 + \tau_y^{-1}}}.$$

Following the definition of β_P and after some algebra, we reach that

$$\beta_P = \frac{\frac{\phi^2}{\tau_x^{-1}\phi^2+\tau_y^{-1}}\frac{1}{\tau_a} + \frac{1}{\tau_\zeta}}{\frac{\phi^2}{\tau_x^{-1}\phi^2+\tau_y^{-1}}\frac{1}{\tau_a} + \frac{1}{\tau_\zeta} + \frac{1}{\tau_x^{-1}\phi^2+\tau_y^{-1}}\frac{1}{\tau_f} + \frac{1}{\tau_\xi}}.$$

To simplify, let

$$h_1 = \frac{\phi^2}{\tau_x^{-1}\phi^2 + \tau_y^{-1}}\frac{1}{\tau_a} + \frac{1}{\tau_\zeta} \tag{B.1.5}$$

and

$$h_2 = \frac{1}{\tau_x^{-1}\phi^2 + \tau_y^{-1}}\frac{1}{\tau_f} + \frac{1}{\tau_\xi}. \tag{B.1.6}$$

It is straightforward that h_1 is increasing in ϕ and h_2 is decreasing in ϕ. Hence, we have

$$\frac{\partial \beta_P}{\partial \phi^2} = \frac{-g_1 \frac{\partial g_2}{\partial \phi^2}}{(g_1 + g_2)^2} > 0\,,$$

which indicates that β_P is an increasing function of ϕ when $\phi > 0$. $\square$

Proof of Lemma 14. We first consider common investment correlation β_A. Recall the investment policy (2.3.8) and the definition of β_A, we have

$$\begin{aligned}
\beta_A &= \frac{\operatorname{Cov}(\log I_i, \log A)}{\sqrt{\operatorname{Var}(\log I_i)}\sqrt{\operatorname{Var}(\log A)}} \\
&= \frac{\frac{\tau_{sa}+\tau_{\overline{pa}}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}}}{\sqrt{\tau_a}\sqrt{\frac{\tau_{sa}+\tau_{\overline{pa}}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}}\frac{1}{\tau_a} + \frac{\tau_f+\tau_{\overline{pf}}}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}}\frac{1}{\tau_f}}} = \frac{S_a}{\sqrt{\tau_a}\sqrt{S_a/\tau_a + S_f/\tau_f}}\,.
\end{aligned}$$

It is convenient for us to consider instead

$$\frac{1}{\tau_a \beta_A^2} = \frac{S_a/\tau_a + S_f/\tau_f}{S_a^2} = \frac{1}{\tau_a S_a} + \frac{S_f}{\tau_f S_a^2}\,.$$

By Lemma 12, since S_a is increasing in ϕ^2 and S_f is decreasing in ϕ^2 when $\phi > 0$, it is straightforward that $1/\tau_a \beta_A^2$ is decreasing in ϕ^2. This indicates that β_A is an increasing function of ϕ when $\phi > 0$.

We then consider the idiosyncratic investment correlation β_F. Again, recall the investment policy (2.3.8) and the definition of β_F, we have

$$\begin{aligned}
\beta_F &= \frac{\operatorname{Cov}(\log I_i, \log F_i)}{\sqrt{\operatorname{Var}(\log I_i)}\sqrt{\operatorname{Var}(\log F_i)}} \\
&= \frac{\frac{\tau_f+\tau_{\overline{pf}}}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}}}{\sqrt{\tau_f}\sqrt{\frac{\tau_{sa}+\tau_{\overline{pa}}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}}\frac{1}{\tau_a} + \frac{\tau_f+\tau_{\overline{pf}}}{\tau_f+\tau_{sf}+\tau_{\overline{pf}}}\frac{1}{\tau_f}}} = \frac{S_f}{\sqrt{\tau_f}\sqrt{S_a/\tau_a + S_f/\tau_f}}\,.
\end{aligned}$$

Similarly, it is convenient for us to consider instead

$$\frac{1}{\tau_f \beta_F^2} = \frac{S_a/\tau_a + S_f/\tau_f}{S_f^2} = \frac{S_a}{\tau_a S_f^2} + \frac{1}{\tau_f S_f}\,,$$

which is decreasing in ϕ^2, again by Lemma 12. Hence, we conclude that β_F is an increasing function of ϕ as well, when $\phi > 0$. $\square$

Proof of Lemma 15. We focus on the case when τ_y and λ are large enough so that a unique solution of ϕ is guaranteed. Recall that the reduced equilibrium condition (B.1.2) (in the proof of Proposition 13):

$$\left(\frac{\tau_{sa}+\tau_x\tau_\zeta}{\tau_a+\tau_{sa}+\tau_x\tau_\zeta}+1-\frac{\sqrt{\tau_x}}{\lambda}\right)\frac{\tau_x}{\tau_a+\tau_x}=\frac{\tau_{sf}\phi^3+\tau_x\tau_\xi\phi}{\tau_f\phi^2+\tau_{sf}\phi^2+\tau_x\tau_\xi}+\phi-\frac{\sqrt{\tau_x}}{\lambda}\,.$$

It is clear that the left hand side of (B.1.2) is decreasing in τ_a. Hence, when τ_a increase, the right hand side of (B.1.2) decreases as well. On the other hand, we have already known that the right hand side of (B.1.2) is increasing in ϕ. Hence, in equilibrium, ϕ decreases. This indicates that ϕ is a decreasing function of τ_a. □

Proof of Proposition 15. We still focus on the case when τ_y and λ are large enough so that a unique solution of ϕ is guaranteed.

We first consider the investment beta

$$\beta_I=\frac{S_a-g}{S_a+\tau_a S_f/\tau_f}\,,$$

where

$$g=\left(\frac{\tau_{sa}}{\tau_a+\tau_{sa}+\tau_{\overline{pa}}}\right)^2\frac{\tau_a}{\tau_{sa}}\,. \tag{B.1.7}$$

It is instructive for us to decompose the total effects of the changing of τ_a on β_I into two parts: the mechanical effect and the cross-learning effect:

$$\frac{d\beta_I(\tau_a,\phi)}{d\tau_a}=\frac{\partial\beta_I(\tau_a,\phi)}{\partial\tau_a}+\frac{\partial\beta_I(\tau_a,\phi)}{\partial\phi}\frac{\partial\phi}{\partial\tau_a}\,.$$

The sign of the second term, the cross-learning effect, is straightforward by Lemma 13 and Lemma 15. Specifically, Lemma 13 indicates that $\partial\beta_I(\tau_a,\phi)/\partial\phi>0$ and Lemma 15 indicates that $\partial\phi/\partial\tau_a<0$, so that the cross-learning effect is negative in this case.

For the first term, the mechanical effect, by the facts that $\tau_{\overline{pa}}=\tau_x\tau_\zeta$ when τ_y goes to infinity and that S_f is independent of τ_a, we know that $\partial\beta_I(\tau_a,\phi)/\partial\tau_a<0$, i.e., the mechanical effect is negative as well.

Taking the two effects together, we know that the total effect is also negative, i.e., $d\beta_I(\tau_a,\phi)/d\tau_a<0$.

We then consider the price beta

$$\beta_P = \frac{\frac{\phi^2}{\tau_x^{-1}\phi^2+\tau_y^{-1}}\frac{1}{\tau_a}+\frac{1}{\tau_\zeta}}{\frac{\phi^2}{\tau_x^{-1}\phi^2+\tau_y^{-1}}\frac{1}{\tau_a}+\frac{1}{\tau_\zeta}+\frac{1}{\tau_x^{-1}\phi^2+\tau_y^{-1}}\frac{1}{\tau_f}+\frac{1}{\tau_\xi}} = \frac{h_1}{h_1+h_2},$$

where h_1 and h_2 are already defined in (B.1.5) and (B.1.6) (in the proof of Lemma 13).

Again, we decompose the total effects of the changing of τ_a on β_P into two parts: the mechanical effect and the cross-learning effect:

$$\frac{d\beta_P(\tau_a,\phi)}{d\tau_a} = \frac{\partial\beta_P(\tau_a,\phi)}{\partial\tau_a} + \frac{\partial\beta_P(\tau_a,\phi)}{\partial\phi}\frac{\partial\phi}{\partial\tau_a}.$$

First, it is straightforward to see that the mechanical effect $\partial\beta_P(\tau_a,\phi)/\partial\tau_a$ is negative, because β_P is an increasing function of h_1 that is in turn decreasing in τ_a at the same time. Furthermore, Lemma 13 indicates that $\partial\beta_P(\tau_a,\phi)/\partial\phi > 0$ and Lemma 15 indicates that $\partial\phi/\partial\tau_a < 0$, which together imply that the cross-learning effect is negative as well. Therefore, we conclude that the total effect is also negative, i.e., $d\beta_P(\tau_a,\phi)/d\tau_a < 0$. □

Proof of Lemma 16. The proof is similar to the proof of Lemma 15. We again focus on the case when τ_y and λ are large enough so that a unique solution of ϕ is guaranteed. In this case, we recall the reduced equilibrium condition (B.1.2) (in the proof of Proposition 13):

$$\left(\frac{\tau_{sa}+\tau_x\tau_\zeta}{\tau_a+\tau_{sa}+\tau_x\tau_\zeta}+1-\frac{\sqrt{\tau_x}}{\lambda}\right)\frac{\tau_x}{\tau_a+\tau_x} = \frac{\tau_{sf}\phi^3+\tau_x\tau_\xi\phi}{\tau_f\phi^2+\tau_{sf}\phi^2+\tau_x\tau_\xi}+\phi-\frac{\sqrt{\tau_x}}{\lambda}.$$

It is clear that the left hand side of (B.1.2) is increasing in τ_{sa}. Hence, when τ_{sa} increase, the right hand side of (B.1.2) increases as well. On the other hand, we have already known that the right hand side of (B.1.2) is increasing in ϕ. Hence, in equilibrium, ϕ increases. This indicates that ϕ is an increasing function of τ_{sa}.

The analysis is similar for τ_{sf}. The right hand side of (B.1.2) is increasing in τ_{sf}, while the left hand side is independent of τ_{sf}. Thus, when τ_{sf} increase, ϕ decreases to ensure a constant right hand side of (B.1.2). This indicates that ϕ is a decreasing function of τ_{sf}. □

Proof of Proposition 16. We first consider the comparative statics with respect to τ_{sa}. For the investment beta β_I, we have

$$\beta_I = 1 - \frac{S_f/\tau_f + g/\tau_a}{S_a/\tau_a + S_f/\tau_f},$$

where g is already defined in (B.1.7) (in the proof of Proposition 15).

We also decompose the total effects of the changing of τ_{sa} on β_I into two parts: the mechanical effect and the cross-learning effect:

$$\frac{d\beta_I(\tau_{sa}, \phi)}{d\tau_{sa}} = \frac{\partial \beta_I(\tau_{sa}, \phi)}{\partial \tau_{sa}} + \frac{\partial \beta_I(\tau_{sa}, \phi)}{\partial \phi} \frac{\partial \phi}{\partial \tau_{sa}}.$$

Clearly, Lemma 13 and Lemma 16 indicate that the second term, the cross-learning effect, is positive. For the first term, the mechanical effect, we first know that S_a is increasing in τ_{sa}, given ϕ fixed. Also, it is easy to show that g/τ_a is increasing in τ_{sa} (given ϕ fixed) when $\tau_{sa} < \tau_a + \tau_x\tau_\zeta$, and decreasing in τ_{sa} (also given ϕ fixed) when $\tau_{sa} > \tau_a + \tau_x\tau_\zeta$. Hence, we conclude that when $\tau_{sa} > \tau_a + \tau_x\tau_\zeta$, the mechanical effect is positive, and thus the total effect $d\beta_I(\tau_{sa}, \phi)/d\tau_{sa}$ is positive as well. When $\tau_{sa} < \tau_a + \tau_x\tau_\zeta$, the mechanical effect is ambiguous and so is the total effect.

For the price beta β_P, there is only cross-learning effect but no mechanical effect. Hence, by Lemma 13 and Lemma 16 we have that

$$\frac{d\beta_P(\tau_{sa}, \phi)}{d\tau_{sa}} = \frac{\partial \beta_P(\tau_{sa}, \phi)}{\partial \phi} \frac{\partial \phi}{\partial \tau_{sa}} > 0.$$

We then consider the comparative statics with respect to τ_{sf} in a similar manner. For the investment beta β_I, we have

$$\beta_I = \frac{S_a/\tau_a - g/\tau_a}{S_a/\tau_a + S_f/\tau_f}.$$

By the similar decomposition and again by Lemma 13 and Lemma 16, we know that both the mechanical effect and the cross-learning effect in this case are negative. So that the total effect is also negative:

$$\frac{d\beta_I(\tau_{sf}, \phi)}{d\tau_{sf}} = \frac{\partial \beta_I(\tau_{sf}, \phi)}{\partial \tau_{sf}} + \frac{\partial \beta_I(\tau_{sf}, \phi)}{\partial \phi} \frac{\partial \phi}{\partial \tau_{sf}} < 0.$$

For the price β_P, again, there is only cross-learning effect but no mechanical effect.

Hence, by Lemma 13 and Lemma 16 we have that

$$\frac{d\beta_P(\tau_{sf},\phi)}{d\tau_{sf}} = \frac{\partial\beta_P(\tau_{sf},\phi)}{\partial\phi}\frac{\partial\phi}{\partial\tau_{sf}} < 0\,.$$

This concludes the proof. □

Proof of Lemma 17. We focus on the case when τ_y and λ are large enough so that a unique solution of ϕ is guaranteed. We first consider the effect of λ. Re-express the reduced equilibrium condition (B.1.2) (in the proof of Proposition 13):

$$\left(\frac{\tau_{sa}+\tau_x\tau_\zeta}{\tau_a+\tau_{sa}+\tau_x\tau_\zeta}+1\right)\frac{\tau_x}{\tau_a+\tau_x}+\frac{\sqrt{\tau_x}}{\lambda}\frac{\tau_a}{\tau_a+\tau_x} = \frac{\tau_{sf}\phi^3+\tau_x\tau_\xi\phi}{\tau_f\phi^2+\tau_{sf}\phi^2+\tau_x\tau_\xi}+\phi\,. \tag{B.1.8}$$

It is clear that the left hand side of (B.1.8) is decreasing in λ while the right hand side is independent of λ. Hence, when λ increases, ϕ decreases in equilibrium.

We then consider the effects of τ_ζ and τ_ξ. Again recall the reduced equilibrium condition (B.1.2):

$$\left(\frac{\tau_{sa}+\tau_x\tau_\zeta}{\tau_a+\tau_{sa}+\tau_x\tau_\zeta}+1-\frac{\sqrt{\tau_x}}{\lambda}\right)\frac{\tau_x}{\tau_a+\tau_x} = \frac{\tau_{sf}\phi^3+\tau_x\tau_\xi\phi}{\tau_f\phi^2+\tau_{sf}\phi^2+\tau_x\tau_\xi}+\phi-\frac{\sqrt{\tau_x}}{\lambda}\,.$$

On the one hand, the left hand side of (B.1.2) is increasing in τ_ζ, while the right hand side is independent of τ_ζ, so that ϕ is increasing in τ_ζ in equilibrium. On the other hand, the right hand side of (B.1.2) is increasing in τ_ξ while the left hand side is independent of τ_ξ, so that ϕ is decreasing in τ_ξ in equilibrium. □

Proof of Proposition 17. We first consider the comparative statics with respect to λ. Since λ has no mechanical effect on ether β_I or β_P, we focus on the cross-learning effect along. By Lemma 13 and Lemma 17, we know that both β_P and β_I are decreasing in λ.

We then consider the comparative statics with respect to τ_ζ. For the investment beta β_I, we have

$$\beta_I = \frac{S_a - g}{S_a + \tau_a S_f/\tau_f}\,,$$

where g is defined in (B.1.7) (in the proof of Proposition 15).

Again, we decompose the total effects of the changing of τ_ζ on β_I into two parts: the mechanical effect and the cross-learning effect:

$$\frac{d\beta_I(\tau_\zeta, \phi)}{d\tau_\zeta} = \frac{\partial\beta_I(\tau_\zeta, \phi)}{\partial\tau_\zeta} + \frac{\partial\beta_I(\tau_\zeta, \phi)}{\partial\phi} \frac{\partial\phi}{\partial\tau_\zeta}.$$

By Lemma 13 and Lemma 17, we know that the cross-learning effect is positive. For the mechanical effect, when ϕ is fixed, it is easy to show that $\partial\beta_I(\tau_\zeta, \phi)/\partial\tau_{\overline{pa}} > 0$. Since we know that $\partial\tau_{\overline{pa}}/\partial\tau_\zeta > 0$, we get that the mechanical effect is also positive. Hence, the total effect $d\beta_I(\tau_\zeta, \phi)/d\tau_\zeta$ is positive.

However, the total effect on the price beta β_P is ambiguous in this case. We have

$$\beta_P = \frac{h_1}{h_1 + h_2},$$

where h_1 and h_2 are already defined in (B.1.5) and (B.1.6) (in the proof of Lemma 13). Decomposition gives

$$\frac{d\beta_P(\tau_\zeta, \phi)}{d\tau_\zeta} = \frac{\partial\beta_P(\tau_\zeta, \phi)}{\partial\tau_\zeta} + \frac{\partial\beta_P(\tau_\zeta, \phi)}{\partial\phi} \frac{\partial\phi}{\partial\tau_\zeta}.$$

Lemma 13 and Lemma 17 give a positive cross-learning effect, i.e., the second term. However, it is easy to show that the first term, the mechanical effect, is negative. Hence, the total effect is ambiguous and is determined by other model parameters.

We finally consider the comparative statics with respect to τ_ξ. Similarly, we follow the decomposition above. For the investment beta β_I, by Lemma 13 and Lemma 17, we know that the cross-learning effect is negative. For the mechanical effect, when ϕ is fixed, it is easy to show that $\partial\beta_I(\tau_\xi, \phi)/\partial\tau_{\overline{pf}} > 0$. Since we know that $\partial\tau_{\overline{pf}}/\partial\tau_\xi > 0$, we get that the mechanical effect is also negative. Hence, the total effect $d\beta_I(\tau_\xi, \phi)/d\tau_\xi$ is negative. Following similar arguments and again by Lemma 13 and Lemma 17, we know that both the mechanical effect and the cross-learning effect on the price beta β_P are also negative, so that the total effect on β_P is negative as well. $\square$

PROOF OF PROPOSITION 26. Following the definition of real investment efficiency, we know that

$$R = \int_0^1 R_i di,$$

where

$$R_i = \mathbb{E}\left[AF_i I_i - \frac{c}{2} I_i^2\right] = \frac{\kappa(2-\kappa)}{2c}\mathbb{E}\left[AF_i \mathbb{E}\left[AF_i|\Gamma_i\right]\right] ,$$

and

$$\mathbb{E}[AF_i|\Gamma_i] = \exp\left[\frac{\tau_{sa}s_{a,i} + \tau_{\overline{pa}} z_a(\overline{P})}{\tau_a + \tau_{sa} + \tau_{\overline{pa}}} + \frac{1}{2(\tau_a + \tau_{sa} + \tau_{\overline{pa}})} + \frac{\tau_f s_{f,i} + \tau_{\overline{pf}} z_{f,i}(\overline{P})}{\tau_f + \tau_{sf} + \tau_{\overline{pf}}} + \frac{1}{2(\tau_f + \tau_{sf} + \tau_{\overline{pf}})}\right] .$$

Since κ and c are constant, without loss of generality, we set $\kappa = 1$ and $c = 0.5$ to ease the exposition. After some tedious algebra, the investment efficiency R is re-expressed in a much simpler and more intuitive form:

$$R = \exp\left(\frac{1+S_a}{\tau_a} + \frac{1+S_f}{\tau_f}\right) . \tag{B.1.9}$$

We solve for the socially optimal ϕ^* that maximizes R. Taking the first order condition gives

$$\frac{\partial \log R}{\partial(\phi^2)} = \frac{\tau_x^2 \tau_\zeta \tau_y}{(\tau_a + \tau_{sa} + \tau_{\overline{pa}})^2} \frac{1}{(\tau_x + \tau_y \phi^2)^2} - \frac{\tau_x \tau_y^2 \tau_\xi}{(\tau_f + \tau_{sf} + \tau_{\overline{pf}})^2} \frac{1}{(\tau_x + \tau_y \phi^2)^2} = 0 , \tag{B.1.10}$$

which reduces to

$$\frac{(\tau_a + \tau_{sa} + \tau_{\overline{pa}})^2}{(\tau_f + \tau_{sf} + \tau_{\overline{pf}})^2} = \frac{\tau_x \tau_\zeta}{\tau_y \tau_\xi} . \tag{B.1.11}$$

Since $\tau_{\overline{pa}}$ is increasing in ϕ^2 while $\tau_{\overline{pf}}$ is decreasing in ϕ^2, we know that the left hand side of (B.1.11) is increasing in ϕ^2. Therefore, there is a unique non-negative solution of ϕ^*.

We further compare between the socially optimal weight ϕ^* and the weight ϕ in the cross-learning equilibrium, focusing on the case in which τ_y and λ are large enough so that there is always a unique positive solution of ϕ. We re-express the first order condition (B.1.10) as

$$\frac{\partial \log R}{\partial(\phi^2)} = \frac{\tau_x \tau_y^2}{(\tau_a + \tau_{sa} + \tau_{\overline{pa}})^2}\left[\frac{\tau_x \tau_\zeta}{\tau_y} - \frac{(\tau_a + \tau_{sa} + \tau_{\overline{pa}})^2 \tau_\xi}{(\tau_f + \tau_{sf} + \tau_{\overline{pf}})^2}\right] = 0 . \tag{B.1.12}$$

When τ_y goes to infinity, $\tau_x \tau_\zeta / \tau_y$ goes to 0, and we also have

$$\left(\frac{\tau_a + \tau_{sa} + \tau_{\overline{pa}}}{\tau_f + \tau_{sf} + \tau_{\overline{pf}}}\right)^2 = \left(\frac{\tau_a + \tau_{sa} + \tau_x \tau_\zeta}{\tau_f + \tau_{sf} + \tau_x \tau_\zeta / \phi^2}\right)^2 > 0 .$$

Hence, when τ_y and λ are large enough, the left hand side of (B.1.12) is always negative.

Therefore, we conclude that the cross-learning equilibrium ϕ is always larger than the socially optimal ϕ^* when τ_y and λ are large enough. □

Proof of Proposition 19. Recall the expression of investment efficiency (B.1.9):

$$R = \exp\left(\frac{1+S_a}{\tau_a} + \frac{1+S_f}{\tau_f}\right).$$

We again focus on the case when τ_y and λ are large enough so that a unique solution of ϕ is guaranteed. We first consider the comparative statics with respect to λ. Lemma 12 implies that $\partial S_f/\partial\phi < 0$ and Lemma 17 implies that $\partial\phi/\partial\lambda < 0$. Since there is no direct effect of λ on S_f, we know that S_f is increasing in λ in equilibrium. Moreover, because the effect of ϕ on S_a is negligible when τ_y is sufficiently large, we eventually know that that R is an increasing function of λ.

We then consider τ_ξ. Similarly, Lemma 12 implies that $\partial S_f/\partial\phi < 0$ and Lemma 17 implies that $\partial\phi/\partial\tau_\xi < 0$, so that S_f, and thus R is increasing in τ_ξ in equilibrium.

We finally consider τ_{sf}. It is clear that $\partial S_f/\partial\tau_{sf} > 0$, i.e., the mechanical effect is positive. For the cross-learning effect, Lemma 12 implies that $\partial S_f/\partial\phi < 0$ and Lemma 16 implies that $\partial\phi/\partial\tau_\xi < 0$. Hence, the total effect is positive as well, i.e., S_f is increasing in τ_{sf} in equilibrium. Since the effect of ϕ on S_a is negligible when τ_y is sufficiently large, we eventually know that that R is an increasing function of τ_{sf}. □

Proof of Proposition 20. Part i) is straightforward following the proofs of Proposition 13 and Proposition 14. For part ii), we make use of the conditional expectation (2.5.12). Specifically, we have

$$\mathrm{Var}(\mathbf{z}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix},$$

where

$$
\begin{aligned}
\sigma_{11} &= \tau_a^{-1} + \tau_{sa}^{-1}\,, \\
\sigma_{12} &= 0\,, \\
\sigma_{13} &= \phi_n \tau_a^{-1}\,, \\
\sigma_{14} &= \phi_n \tau_a^{-1}\,, \\
\sigma_{22} &= \tau_f^{-1} + \tau_{sf}^{-1}\,, \\
\sigma_{23} &= \tau_f^{-1}\,, \\
\sigma_{24} &= 0\,,
\end{aligned}
$$

$$
\begin{aligned}
\sigma_{33} &= \tau_f^{-1} + \phi_n^2 \tau_a^{-1} + (\tau_x^{-1}\phi_n^2 + \tau_y^{-1})(\tau_\zeta^{-1} + \tau_\xi^{-1})\,, \\
\sigma_{34} &= \phi_n^2 \tau_a^{-1} + (\tau_x^{-1}\phi_n^2 + \tau_y^{-1})\tau_\zeta^{-1}\,, \\
\sigma_{44} &= \frac{\tau_f^{-1}}{n-1} + \phi_n^2 \tau_a^{-1} + (\tau_x^{-1}\phi_n^2 + \tau_y^{-1})\left(\tau_\zeta^{-1} + \frac{\tau_\xi^{-1}}{n-1}\right)\,,
\end{aligned}
$$

and

$$\mathrm{Cov}(\mathbf{z}, a + f_i) = [\tau_a^{-1}, \tau_f^{-1}, \phi_n \tau_a^{-1} + \tau_f^{-1}, \phi_n \tau_a^{-1}]\,.$$

By condition (2.5.12), we get the expressions of S_{an} and S_{fn} after some tedious algebra and plug them into the equilibrium condition. Denote by RHS the right hand side of the equilibrium condition (2.5.13) and we get

$$\frac{\partial \lim_{\tau_y \to \infty} RHS(\phi_n, n)}{\partial n} = C_1 C_2 C_3 C_4 C_5 (C_6 + C_7 + C_8)^2\,, \qquad \text{(B.1.13)}$$

where

$$\begin{aligned}
C_1 &= \phi_n^2(\tau_f+\tau_{sf})\tau_\zeta+\tau_x\tau_\zeta\tau_\xi-\phi_n(\tau_a+\tau_{sa}+\tau_x\tau_\zeta)\tau_\xi\,,\\
C_2 &= \phi_n^2\tau_a(\tau_f+2\tau_{sf})\tau_\zeta+2\tau_a\tau_x\tau_\zeta\tau_\xi+\phi_n\tau_f(\tau_a+2(\tau_{sa}+\tau_x\tau_\zeta))\tau_\xi\,,\\
C_3 &= \phi_n^2\tau_f+\tau_x\tau_\xi\,,\\
C_4 &= \phi_n^2\tau_f\tau_x^2\tau_\xi\,,\\
C_5 &= \tau_a+\tau_x\,,\\
C_6 &= \phi_n^3\tau_f^2\tau_x\tau_\zeta\tau_\xi+\phi_n\tau_f\tau_x^2\tau_\zeta\tau_\xi^2+2(\tau_a+\tau_{sa})\tau_x^2\tau_\zeta\tau_\xi^2\,,\\
C_7 &= \phi_n^4\tau_f(\tau_f+2\tau_{sf})((\tau_a+\tau_{sa})\tau_\zeta+n(\tau_a+\tau_{sa}+\tau_x\tau_\zeta)\tau_\xi)\,,\\
C_8 &= \phi_n^2\tau_x\tau_\xi((\tau_a+\tau_{sa})(3\tau_f+2\tau_{sf})\tau_\zeta+((2n-1)\tau_f+2\tau_{sf})(\tau_a+\tau_{sa}+\tau_x\tau_\zeta)\tau_\xi)\,.
\end{aligned}$$

Note that, only the first term C_1 has a negative component. However, when τ_ξ is small enough, C_1 is always strictly positive, so is the entire derivative (B.1.13). It implies that when τ_y is large enough and τ_ξ is small enough, the equilibrium ϕ_n is increasing in n. Also, the proof of Proposition 26 directly implies that $\phi'(=\phi_1)>\phi^*$, so that $\phi_n>\phi^*$ for all $n\geqslant 1$.

Finally, for part iii), by the proof of Proposition 26, in particular condition (B.1.9), we know that

$$R_n=\exp\left(\frac{1+S_a(\phi_n)}{\tau_a}+\frac{1+S_f(\phi_n)}{\tau_f}\right),$$

where S_a and S_f are the capital providers' investment sensitivities in the baseline cross-learning case pinned down by however the equilibrium weight in the corresponding n-learning equilibrium. By Lemma 12, it follows that $R<R_n<R^*$ for all $n\geqslant 1$. This concludes the proof. □

Appendix C

Appendix to Chapter 3

C.1 Derivation, Extensions, and Proofs

C.1.1 Derivation of Convertible Preferred Stock as the Optimal Security

This appendix derives the optimal security $s^*(\theta)$ when it induces information acquisition. We proceed by two steps.

First, we solve for an "unconstrained" optimal security without the feasibility condition $0 \leqslant s(\theta) \leqslant \theta$. We denote the solution by $\widehat{s}(\theta)$. We also denote the corresponding screening rule by $\widehat{m}_s(\theta)$. The unconstrained optimal security recovers the unconstrained part $\widehat{s}(\theta)$ of the eventual optimal security in Proposition 23. After that, we resume the feasibility condition and characterize the optimal security $s^*(\theta)$.

Lemma **22.** *In an equilibrium with information acquisition, the unconstrained optimal security $\widehat{s}(\theta)$ and its corresponding screening rule $\widehat{m}_s(\theta)$ are determined by*

$$\widehat{s}\left(\theta\right) - k = \mu \cdot \left(g'\left(\widehat{m}_s\left(\theta\right)\right) - g'\left(\overline{\pi}_s^*\right)\right) \,, \tag{C.1.1}$$

where

$$\overline{\pi}_s^* = \mathbb{E}\left[m_s^*(\theta)\right] \,,$$

and

$$\left(1-\widehat{m}_s(\theta)\right)\cdot\left(\theta-\widehat{s}(\theta)+w^*\right)=\mu\,, \tag{C.1.2}$$

where

$$w^*=\mathbb{E}\left[\left(\theta-s^*\left(\theta\right)\right)\frac{g''\left(\overline{\pi}_s^*\right)}{g''\left(m_s^*\left(\theta\right)\right)}\right]\left(1-\mathbb{E}\left[\frac{g''\left(\overline{\pi}_s^*\right)}{g''\left(m_s^*\left(\theta\right)\right)}\right]\right)^{-1}\,,$$

in which $\overline{p}_s^*$ and w^* are two constants determined in equilibrium, and $s^*(\theta)$ and $m_s^*(\theta)$ are the solutions of the original constrained problem.

Lemma 22 exhibits the relationship between the unconstrained optimal security $\widehat{s}(\theta)$ and the corresponding screening rule $\widehat{m}_s(\theta)$. Condition (C.1.1) specifies how the investor responses to the unconstrained optimal security by adjusting her screening rule. On the other hand, condition (C.1.2) is derived from the entrepreneur's optimization problem. It indicates the entrepreneur's optimal choices of payments across states, given the investor's screening rule. In equilibrium, $\widehat{s}(\theta)$ and $\widehat{m}_s(\theta)$ are jointly determined.

Although it is not tractable to fully solve the system of equations (C.1.1) and (C.1.2), we are able to deliver important analytical characteristics of the unconstrained optimal security $\widehat{s}(\theta)$ and the corresponding screening rule $\widehat{m}_s(\theta)$.

Lemma 23. *In an equilibrium with information acquisition, the unconstrained optimal security $\widehat{s}(\theta)$ and the corresponding screening rule $\widehat{m}_s(\theta)$ satisfy*

$$\frac{\partial\widehat{m}_s\left(\theta\right)}{\partial\theta}=\mu^{-1}\cdot\widehat{m}_s\left(\theta\right)\cdot\left(1-\widehat{m}_s\left(\theta\right)\right)^2>0\,, \tag{C.1.3}$$

and

$$\frac{\partial\widehat{s}\left(\theta\right)}{\partial\theta}=1-\widehat{m}_s\left(\theta\right)\in\left(0,1\right). \tag{C.1.4}$$

We have several interesting observations from Lemma 23. First, condition (C.1.3) implies that the unconstrained optimal screening rule $\widehat{m}_s\left(\theta\right)$ is strictly increasing. Second, condition (C.1.4) implies that the unconstrained optimal security $\widehat{s}\left(\theta\right)$ is also strictly increasing. These are because, according to Proposition 21, we have $Prob[0<\widehat{m}_s\left(\theta\right)<1]=1$ in this case, and thus the right hand sides of (C.1.3) and (C.1.4) are positive. It follows immediately that the residual of the unconstrained optimal security, $\theta-\widehat{s}\left(\theta\right)$, is also strictly increasing. Last, the unconstrained optimal security $\widehat{s}\,(\theta)$ is strictly concave. This is because conditions (C.1.3) and (C.1.4) imply that

$$\frac{\partial^2 \widehat{s}(\theta)}{\partial \theta^2} = -\mu^{-1} \cdot \widehat{m}_s\,(\theta) \cdot (1 - \widehat{m}_s\,(\theta))^2 < 0\ .$$

Therefore, the unconstrained optimal security $\widehat{s}(\theta)$ is an increasing concave function of θ.

Given Lemma 23, it is instructive to have the following lemma to illustrate the possible relative positions between the unconstrained optimal security and the feasibility constraints.

Lemma 24. *Three possible relative positions between the unconstrained optimal security* $\widehat{s}(\theta)$ *and the feasibility constraints* $0 \leqslant s(\theta) \leqslant \theta$ *may occur in equilibrium, in the* $\theta \sim s$ *space:*

i) $\widehat{s}(\theta)$ *intersects with the* 45° *line* $s = \theta$ *at* $(\widehat{\theta}, \widehat{\theta})$, $\widehat{\theta} > 0$, *and does not intersect with the horizontal axis* $s = 0$;

ii) $\widehat{s}(\theta)$ *goes through the origin* $(0, 0)$, *and does not intersect with either the* 45° *line* $s = \theta$ *or the horizontal axis* $s = 0$ *for any* $\theta \neq 0$;

iii) $\widehat{s}(\theta)$ *intersects with the horizontal axis* $s = 0$ *at* $(\widetilde{\theta}, 0)$, $\widetilde{\theta} > 0$, *and does not intersect with the* 45° *line* $s = \theta$.

In the three different cases, the actual optimal security $s^*(\theta)$ will be constrained by the feasibility condition in different ways. For example, $s^*(\theta)$ will be constrained by the 45° line $s = \theta$ in Case i) while by the horizontal axis $s = 0$ in Case iii). By imposing the feasibility conditions, we have the following characterization for $s^*(\theta)$:

Lemma 25. *In an equilibrium with information acquisition, the corresponding optimal security* $s^*(\theta)$ *satisfies*

$$s^*(\theta) = \begin{cases} \theta & \textit{if} \quad \widehat{s}(\theta) > \theta \\ \widehat{s}(\theta) & \textit{if} \quad 0 \leqslant \widehat{s}(\theta) \leqslant \theta \\ 0 & \textit{if} \quad \widehat{s}(\theta) < 0 \end{cases},$$

where $\widehat{s}(\theta)$ *is the corresponding unconstrained optimal security.*

Lemma 25 is helpful because it tells us how to construct an optimal security $s^*(\theta)$ from its corresponding unconstrained optimal security $\widehat{s}(\theta)$. Concretely, $s^*(\theta)$ will follow $\widehat{s}(\theta)$

when the latter is within the feasible region $0 \leqslant s \leqslant \theta$. When $\widehat{s}(\theta)$ goes out of the feasible region, the resulting optimal security will follow one of the feasibility constraints that is binding.

We apply Lemma 25 to the three cases of the unconstrained optimal security $\widehat{s}(\theta)$ described in Lemma 24. This gives the three potential cases of the optimal security $s^*(\theta)$, respectively.

Lemma **26.** *In an equilibrium with information acquisition, the optimal security $s^*(\theta)$ may take one of the following three forms:*

i) when the corresponding unconstrained optimal security $\widehat{s}(\theta)$ intersects with the $45°$ line $s = \theta$ at $(\widehat{\theta}, \widehat{\theta})$, $\widehat{\theta} > 0$, we have

$$s^*(\theta) = \begin{cases} \theta & \text{if} \quad 0 \leqslant \theta < \widehat{\theta} \\ \widehat{s}(\theta) & \text{if} \quad \theta \geqslant \widehat{\theta} \end{cases};$$

ii) when the corresponding unconstrained optimal security $\widehat{s}(\theta)$ goes through the origin $(0, 0)$, we have $s^*(\theta) = \widehat{s}(\theta)$ for $\theta \in \mathbb{R}_+$;

iii) when the corresponding unconstrained optimal security $\widehat{s}(\theta)$ intersects with the horizontal axis $s = 0$ at $(\widetilde{\theta}, 0)$, $\widetilde{\theta} > 0$, we have

$$s^*(\theta) = \begin{cases} 0 & \text{if} \quad 0 \leqslant \theta < \widetilde{\theta} \\ \widehat{s}(\theta) & \text{if} \quad \theta \geqslant \widetilde{\theta} \end{cases}.$$

The optimal security $s^*(\theta)$ takes different shapes in the three potential cases. In Case i), $s^*(\theta)$ follows a debt in states with low cash flows but increases in states with high cash flows. In Case iii), $s^*(\theta)$ has zero payment in states with low cash flows, while is an increasing function in states with high cash flows. Case ii) lies in between as a knife-edge case.

We proceed by determining whether these three potential cases are valid solutions to the entrepreneur's problem in an equilibrium with information acquisition. Interestingly, not all the three cases can occur in equilibrium.

Lemma **27.** *If the entrepreneur's optimal security $s^*(\theta)$ induces the investor to acquire information in equilibrium, then it must follow Case i) in Lemma 26, which corresponds to a participating* *convertible preferred stock with a face value* $\widehat{\theta} > 0$.

Together with the lemmas already established, Lemma 27 immediately leads to Proposition 23. Intuitively, Case ii) and Case iii) in Lemma 26 cannot sustain an equilibrium with information acquisition because the investor is underpaid. Recall that the investor provides two types of inputs. The first is the investment required to initiate the project, and the second is the costly information to screen the project. As a result, the entrepreneur wants to make sure that the investor is sufficiently compensated for both inputs to be willing to accept the security. This argument is further strengthened by Corollary 5, which suggested that $\widehat{\theta}$ should be larger than the investment requirement k.

C.1.2 General Allocation of Bargaining Power

This appendix extends our baseline model to a more general setting that allows for the arbitrary allocation of bargaining power between the entrepreneur and the investor. It demonstrates that our framework and qualitative results are robust to the allocation of bargain power.

Without loss of generality, let the entrepreneur's bargaining power in security design be $1 - \alpha$ and the investor's α. Suppose a third party in the economy knows α, designs the security and proposes it to the investor. The investor acquires information according to the security and decides whether or not to accept this offer. The third party's objective function is an average of the entrepreneur's and the investor's utilities, weighted according to the bargaining power of each. When $\alpha = 0$, this reduces to our baseline model. The derivations for the results are the same as in the baseline model.

In this setting, the third-party's objective function, that is, the payoff gain, is

$$u_T(s(\theta)) = \alpha \cdot (\mathbb{E}[(s(\theta) - k) \cdot m(\theta)] - \mu \cdot I(m)) + (1 - \alpha) \cdot \mathbb{E}[(\theta - s(\theta)) \cdot m(\theta)] \, .$$

We can show that, with information acquisition, the equation that governs information acquisition is still the same as condition (3.3.6):

$$s(\theta) - k = \mu \cdot \left(g'(m_s(\theta)) - g'(\overline{\pi}_s)\right) ,$$

while the equation that characterizes the optimality of the unconstrained optimal security becomes

$$r(\theta) = (2\alpha - 1) \cdot m(\theta) + (1 - \alpha) \cdot \mu^{-1} \cdot m(\theta) \cdot (1 - m(\theta)) \cdot (\theta - s(\theta) + w) .$$

The following two propositions characterize the optimal security in the general setting.[1]

Proposition 29. *When $0 \leqslant \alpha < 1/2$ and information acquisition happens in equilibrium, the unconstrained optimal security $\widehat{s}(\theta)$ and the corresponding screening rule $\widehat{m}_s(\theta)$ satisfy*

$$\frac{d\widehat{s}(\theta)}{d\theta} = \frac{1 - \widehat{m}_s(\theta)}{1 - \frac{\alpha}{1-\alpha}\widehat{m}_s(\theta)} \in (0, 1)$$

and

$$\frac{d\widehat{m}_s(\theta)}{d\theta} = \frac{\mu^{-1} \cdot \widehat{m}_s(\theta) \cdot (1 - \widehat{m}_s(\theta))^2}{1 - \frac{\alpha}{1-\alpha}\widehat{m}_s(\theta)} > 0 .$$

Also, all the results from Lemma 24 to Lemma 27 and from Proposition 21 to Proposition 26 still hold.

Proposition 30. *When $1/2 \leqslant \alpha \leqslant 1$, the optimal security features $s^*(\theta) = \theta$.*

Our generalized results show that when the investor has some bargaining power, but not too much, all the qualitative results remain unchanged. But if the investor's bargaining power is strong, the optimal security is most favorable to the investor: a complete takeover. This is intuitive considering the situation from the standpoint of friction. If the entrepreneur dominates, she will still plays a considerable role in real production, which depends on the investor's information acquisition. Thus, the presence of friction still calls for a meaningful security design that follows our interaction between the shape of the securities and the incentive for screening. On the contrary, if the investor dominates, she may take over the project and effectively eliminate the friction. In this case, real production and information acquisition are joined and security design becomes less relevant. This corresponds to the empirical fact that buyouts and takeovers are common for mature companies, where the role of entrepreneurs and founders is no longer inalienable, a point also highlighted in Rajan (2012) and Lerner, Leamon and Hardymon (2012).

---

[1] The proofs for the extended model follow those for the benchmark model closely, so we do not repeat them in the appendix.

C.1.3 Proofs

This appendix provides all proofs omitted above.

PROOF OF LEMMA 18. We first prove the "only if" part. Suppose that

$$\mathbb{E}\left[\exp(\mu^{-1}(\theta - k))\right] \leqslant 1\,.$$

According to Proposition 21, even if the entrepreneur proposes all the future cash flow to the investor, the investor will still reject the offer without acquiring information. Since $s(\theta) \leqslant \theta$, the project cannot be initiated in this case.

Then we prove the "if" part. Let $t \in (0, 1)$. Since $\mathbb{E}\left[\exp(\mu^{-1}(t \cdot \theta - k))\right]$ is continuous in t, there exists $t < 1$ such that

$$\mathbb{E}\left[\exp(\mu^{-1}(t \cdot \theta - k))\right] > 1\,.$$

Hence, according to Proposition 21, the security $s_t(\theta) = t \cdot \theta$ would be accepted by the investor with a positive probability. Moreover, let $m_t(\theta)$ be the corresponding screening rule. As $s_t(\theta)$ would be accepted with a positive probability, $m_t(\theta)$ cannot be always zero. Hence, the entrepreneur's expected payoff is $\mathbb{E}[(1 - t) \cdot \theta \cdot m_t(\theta)]$, which is strictly positive.

Note that the security $s_t(\theta)$ is a feasible security. Hence, the optimal security $s^*(\theta)$ will also be accepted with a positive probability and deliver a positive expected payoff to the entrepreneur. This concludes the proof. □

PROOF OF COROLLARY 4. The proof is straightforward following the above proof of Lemma 18. Proposing $s^*(\theta) = \theta$ gives the entrepreneur a zero payoff, while proposing $s_t(\theta) = t \cdot \theta$

constructed in the proof of Lemma 18 gives her a strictly positive expected payoff. This suggests that $s^*(\theta) = \theta$ is not optimal. $\square$

Proof of Proposition 22. The Lagrangian of the entrepreneur's problem is

$$\mathscr{L} = \mathbb{E}\left[\theta - s(\theta) + \lambda \cdot \left(1 - \exp\left(\mu^{-1} \cdot (k - s(\theta))\right)\right) + \eta_1(\theta) \cdot s(\theta) + \eta_2(\theta) \cdot (\theta - s(\theta))\right],$$

where λ, $\eta_1(\theta)$ and $\eta_2(\theta)$ are multipliers.

The first order condition is

$$\frac{d\mathscr{L}}{ds(\theta)} = -1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot (q - s(\theta))\right) + \eta_1(\theta) - \eta_2(\theta) = 0\,. \tag{C.1.5}$$

We first consider a special case that is helpful for us to solve the optimal security. If $0 < s(\theta) < \theta$, the two feasibility conditions are not binding. Thus $\eta_1(\theta) = \eta_2(\theta) = 0$, and the first order condition is simplified as

$$-1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot (k - s(\theta))\right) = 0\,.$$

By rearrangement, we get

$$s(\theta) = k - \mu \cdot \ln(\lambda^{-1} \cdot \mu)\,. \tag{C.1.6}$$

We denote the right hand side of (C.1.6), which is irrelevant of θ, as D^*. By definition, we have $D^* > 0$. Also, it is straightforward to have

$$-1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot (k - D^*)\right) = 0\,. \tag{C.1.7}$$

In what follows, we characterize the optimal solution $s^*(\theta)$ for different regions of θ.

First, we consider the region of $\theta > D^*$. We show that $0 < s^*(\theta) < \theta$ in this region by contradiction.

If $s^*(\theta) = \theta > D^*$, we have $\eta_1(\theta) = 0$ and $\eta_2(\theta) \geqslant 0$. From the first order condition (C.1.5) we obtain

$$-1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot (k - \theta)\right) = \eta_2(\theta) \geqslant 0\,. \tag{C.1.8}$$

On the other hand, as $\theta > D^*$, we have

$$-1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot (k - D^*)\right) > -1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot (k - \theta)\right). \tag{C.1.9}$$

Conditions (C.1.7), (C.1.8), and (C.1.9) construct a contradiction. So we must have $s^*(\theta) < \theta$ if $\theta > D^*$.

Similarly, if $s^*(\theta) = 0$, we have $\eta_1(\theta) \geqslant 0$ and $\eta_2(\theta) = 0$. Again from the first order condition (C.1.5) we obtain

$$-1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot k\right) = -\eta_1(\theta) \leqslant 0. \tag{C.1.10}$$

On the other hand, as $D^* > 0$, we have

$$-1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot (k - D^*)\right) < -1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot k\right). \tag{C.1.11}$$

Conditions (C.1.7), (C.1.10), and (C.1.11) construct another contradiction. So we must have $s^*(\theta) > 0$ if $\theta > D^*$.

Therefore, we have shown that $0 < s^*(\theta) < \theta$ for $\theta > D^*$. From the discussion above for this specific case, we conclude that $s^*(\theta) = D^*$ for $\theta > D^*$.

We then consider the region of $\theta < D^*$. We show that $s^*(\theta) = \theta$ in this region.

Since $s^*(\theta) \leqslant \theta < D^*$, we have

$$-1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot (k - s^*(\theta))\right) > -1 + \lambda \cdot \mu^{-1} \cdot \exp\left(\mu^{-1} \cdot (k - D^*)\right). \tag{C.1.12}$$

From condition (C.1.7), the right hand side of this inequality (C.1.12) is zero. Together with the first order condition (C.1.5), the inequality (C.1.12) implies that $\eta_1(\theta) = 0$ and $\eta_2(\theta) > 0$. Therefore, we have $s^*(\theta) = \theta$ in this region.

Also, from the first order condition (C.1.5) and condition (C.1.7), it is obvious that $s^*(D^*) = D^*$.

To sum up, the entrepreneur's optimal security without inducing the investor to acquire information features a debt with face value D^* determined by condition (C.1.6).

We need to check that there exists $D^* > 0$ and the corresponding multiplier $\lambda > 0$ such that

$$\mathbb{E}\left[\exp\left(-\mu^{-1}\cdot(\min(\theta,D^{*})-k)\right)\right]=1\,, \tag{C.1.13}$$

where D^{*} is determined by condition (C.1.6).

Consider the left hand side of condition (C.1.13). Clearly, it is continuous and monotonically decreasing in D^{*}. When D^{*} is sufficiently large, the left hand side of (C.1.13) approaches $\mathbb{E}\left[\exp\left(-\mu^{-1}\cdot(\theta-k)\right)\right]$, a number less than one, which is guaranteed by condition (3.3.3) as well as the feasibility condition $s(\theta)\leqslant\theta$. On the other hand, when $D^{*}=0$, it approaches $\exp\left(\mu^{-1}\cdot k\right)$, which is strictly greater than one. Hence, there exists $D^{*}>0$ such that condition (C.1.13) holds.

Moreover, from condition (C.1.6), we also know that D^{*} is continuous and monotonically increasing in λ. When λ approaches zero, D^{*} approaches negative infinity, while when λ approaches positive infinity, D^{*} approaches positive infinity as well. Hence, for any $D^{*}>0$ there exists a corresponding multiplier $\lambda>0$.

Last, suppose $D^{*}\leqslant k$. It is easy to see that this debt would be rejected by the investor due to Proposition 21, a contradiction.

Finally, by condition (3.3.3) again, since the optimal security $s^{*}(\theta)$ satisfies

$$\mathbb{E}\left[\exp\left(-\mu^{-1}\cdot\left(s^{*}\left(\theta\right)-k\right)\right)\right]=1\,,$$

Jensen's inequality implies that $\mathbb{E}[s^{*}(\theta)]>k$ given $\mu>0$. This concludes the proof. □

PROOF OF LEMMA 22. We derive the entrepreneur's optimal security $s^{*}\left(\theta\right)$ and the corresponding unconstrained optimal security $\widehat{s}(\theta)$ through variational methods. Specifically, we characterize how the entrepreneur's expected payoff responds to the perturbation of her optimal security.

Let $s\left(\theta\right)=s^{*}\left(\theta\right)+\alpha\cdot\varepsilon\left(\theta\right)$ be an arbitrary perturbation of the optimal security $s^{*}\left(\theta\right)$. Note that the investor's optimal screening rule $m_{s}\left(\theta\right)$ appears in the entrepreneur's expected payoff $u_{E}(s(\cdot))$, according to condition (3.3.7), and it is implicitly determined by the proposed security $s\left(\theta\right)$ through the functional equation (3.3.6). Hence, we need to first characterize how $m_{s}\left(\theta\right)$ varies with respect to the perturbation of $s^{*}\left(\theta\right)$. Taking derivative with respect to α at $\alpha = 0$ for both sides of (3.3.6) leads to

$$\mu^{-1}\varepsilon(\theta) = g''(m_s^*(\theta)) \cdot \left.\frac{\partial m_s(\theta)}{\partial \alpha}\right|_{\alpha=0} - g''(\overline{\pi}_s^*) \cdot \mathbb{E} \left.\frac{\partial m_s(\theta)}{\partial \alpha}\right|_{\alpha=0} .$$

Take expectation of both sides and we get

$$\mathbb{E}\left[\left.\frac{\partial m_s(\theta)}{\partial \alpha}\right|_{\alpha=0}\right] = \mu^{-1} \cdot \left(1 - \mathbb{E}\left[\left(g''(m_s^*(\theta))\right)^{-1}\right] \cdot g''(\overline{\pi}_s^*)\right)^{-1} \cdot \mathbb{E}\left[\left(g''(m_s^*(\theta))\right)^{-1}\varepsilon(\theta)\right] .$$

Combining the above two equations, for any perturbation $s(\theta) = s^*(\theta) + \alpha \cdot \varepsilon(\theta)$, the investor's screening rule $m_s(\cdot)$ is characterized by

$$\begin{aligned} \left.\frac{\partial m_s(\theta)}{\partial \alpha}\right|_{\alpha=0} &= \mu^{-1} \cdot \left(g''(m_s^*(\theta))\right)^{-1}\varepsilon(\theta) \\ &+ \frac{\mu^{-1} \cdot \left(g''(m_s^*(\theta))\right)^{-1} \cdot \mathbb{E}\left[\left(g''(m_s^*(\theta))\right)^{-1}\varepsilon(\theta)\right]}{\left(g''(\overline{\pi}_s^*)\right)^{-1} - \mathbb{E}\left[\left(g''(m_s^*(\theta))\right)^{-1}\right]} . \end{aligned} \tag{C.1.14}$$

We interpret condition (C.1.14). The first term of the right hand side of (C.1.14) is the investor's local response to $\varepsilon(\theta)$. It is of the same sign as the perturbation $\varepsilon(\theta)$. When the payment of the security increases at state θ, the investor is more likely to accept the security at this state. The second term measures the investor's average response to perturbation $\varepsilon(\theta)$ over all states. It is straightforward to verify that the denominator of the second term is positive due to Jensen's inequality. As a result, if the perturbation increases the investor's payment on average over all states, she is more likely to accept the security as well.

Now we can calculate the variation of the entrepreneur's expected payoff $u_E(s(\cdot))$, according to condition (3.3.7). Taking derivative of $u_E(s(\cdot))$ with respect to α at $\alpha = 0$ leads to

$$\left.\frac{\partial u_E(s(\cdot))}{\partial \alpha}\right|_{\alpha=0} = \mathbb{E}\left[\left.\frac{\partial m_s(\theta)}{\partial \alpha}\right|_{\alpha=0} (\theta - s(\theta))\right] - \mathbb{E}\left[m_s^*(\theta) \cdot \varepsilon(\theta)\right] . \tag{C.1.15}$$

Substitute (C.1.14) into (C.1.15) and we get

$$\left.\frac{\partial u_E(s(\cdot))}{\partial \alpha}\right|_{\alpha=0} = \mathbb{E}\left[r(\theta) \cdot \varepsilon(\theta)\right] , \tag{C.1.16}$$

where

$$r(\theta) = -m_s^*(\theta) + \mu^{-1} \cdot \left(g''(m_s^*(\theta))\right)^{-1} \cdot (\theta - s^*(\theta) + w^*) \tag{C.1.17}$$

and

$$w^* = \mathbb{E}\left[(\theta - s^*(\theta))\frac{g''(\overline{\pi}_s^*)}{g''(m_s^*(\theta))}\right]\left(1 - \mathbb{E}\left[\frac{g''(\overline{\pi}_s^*)}{g''(m_s^*(\theta))}\right]\right)^{-1}.$$

Note that w^* is a constant that does not depend on θ and will be endogenously determined in the equilibrium. Besides, $r(\theta)$ is the Frechet derivative of the entrepreneur's expected payoff $u_E(s(\cdot))$ at $s^*(\theta)$, which measures the marginal contribution of any perturbation to the entrepreneur's expected payoff when the security is optimal. Specifically, the first term of (C.1.17) is the direct contribution of perturbing $s^*(\theta)$ disregarding the variation of $m_s^*(\theta)$, and the second term measures the indirect contribution through the variation of $m_s^*(\theta)$. This Frechet derivative $r(\theta)$ plays an important role in shaping the entrepreneur's optimal security.

To further characterize the optimal security, we discuss the Frechet derivative $r(\theta)$ in detail. Recall that the optimal security would be restricted by the feasibility condition $0 \leqslant s^*(\theta) \leqslant \theta$. Let

$$A_0 = \{\theta \in \Theta : \theta \neq 0, s^*(\theta) = 0\},$$

$$A_1 = \{\theta \in \Theta : \theta \neq 0, 0 < s^*(\theta) < \theta\},$$

$$A_2 = \{\theta \in \Theta : \theta \neq 0, s^*(\theta) = \theta\}.$$

Clearly, $\{A_0, A_1, A_2\}$ is a partition of $\Theta \backslash \{0\}$. Since $s^*(\theta)$ is the optimal security, we have

$$\left.\frac{\partial u_E(s(\cdot))}{\partial \alpha}\right|_{\alpha=0} \leqslant 0$$

for any feasible perturbation $\varepsilon(\theta)$.[2] Hence, condition (C.1.16) implies

$$r(\theta)\begin{cases} \leqslant 0 & \text{if } \theta \in A_0 \\ = 0 & \text{if } \theta \in A_1 \\ \geqslant 0 & \text{if } \theta \in A_2 \end{cases}. \tag{C.1.18}$$

According to Proposition 21, when the optimal security $s^*(\theta)$ induces the investor to acquire information, we have $0 < m_s^*(\theta) < 1$ for all $\theta \in \Theta$. Hence, condition (C.1.18) can be rearranged as

---

[2] A perturbation $\varepsilon(\theta)$ is feasible with respect to $s^*(\theta)$ if there exists $\alpha > 0$ such that for any $\theta \in \Theta$, $s^*(\theta) + \alpha \cdot \varepsilon(\theta) \in [0, \theta]$.

$$\frac{r(\theta)}{m_s^*(\theta)} = -1 + \mu^{-1} \cdot (1 - m_s^*(\theta)) \cdot (\theta - s^*(\theta) + w^*) \begin{cases} \leqslant 0 & \text{if } \theta \in A_0 \\ = 0 & \text{if } \theta \in A_1 \\ \geqslant 0 & \text{if } \theta \in A_2 \end{cases} . \tag{C.1.19}$$

Recall condition (3.3.6), given the optimal security $s^*(\theta)$, the investor's optimal screening rule $m_s^*(\theta)$ is

$$s^*(\theta) - k = \mu \cdot \left(g'(m_s^*(\theta)) - g'(\overline{\pi}_s^*)\right) , \tag{C.1.20}$$

where

$$\overline{\pi}_s^* = \mathbb{E}\left[m_s^*(\theta)\right]$$

is the investor's unconditional probability of accepting the optimal security $s^*(\theta)$. Conditions (C.1.19) and (C.1.20) as a system of functional equations jointly determine the optimal security $s^*(\theta)$ when it induces the investor's information acquisition.

Finally, when we focus on the unconstrained optimal security $\widehat{s}(\theta)$, note that is would not be restricted by the feasibility condition. Hence, the corresponding Frechet derivative $r(\theta)$ would be always zero at the optimum. On the other hand, the investor's optimal screening rule would not be affected. As a result, conditions (C.1.20) and (C.1.19) become

$$\widehat{s}(\theta) - k = \mu \cdot \left(g'(\widehat{m}_s(\theta)) - g'(\overline{\pi}_s^*)\right) ,$$

where

$$\overline{p}_s^* = \mathbb{E}\left[m_s^*(\theta)\right] ,$$

and

$$(1 - \widehat{m}_s(\theta)) \cdot (\theta - \widehat{s}(\theta) + w^*) = \mu ,$$

where

$$w^* = \mathbb{E}\left[(\theta - s^*(\theta)) \frac{g''(\overline{\pi}_s^*)}{g''(m_s^*(\theta))}\right] \left(1 - \mathbb{E}\left[\frac{g''(\overline{\pi}_s^*)}{g''(m_s^*(\theta))}\right]\right)^{-1} ,$$

in which $\overline{p}_s^*$ and w^* are two constants that do not depend on θ. This concludes the proof. □

PROOF OF LEMMA 23. From Lemma 22, $(\widehat{s}(\theta), \widehat{m}_s(\theta))$ satisfies the two equations (C.1.1) and

(C.1.2). By condition (C.1.2), we get

$$\widehat{m}_s(\theta) = 1 - \frac{\mu}{\theta - \widehat{s}(\theta) + w^*} \, . \tag{C.1.21}$$

Substituting (C.1.21) into (C.1.1) leads to

$$\mu^{-1}\left(\widehat{s}(\theta) - k\right) = g'\left(\frac{\mu}{\theta - \widehat{s}(\theta) + w^*}\right) - g'\left(\overline{\pi}_s^*\right) \, .$$

Taking derivatives of both sides of the above equation with respect to θ leads to

$$\begin{aligned}
\mu^{-1} \cdot \frac{d\widehat{s}(\theta)}{d\theta} &= g''\left(\widehat{m}_s(\theta)\right) \cdot \frac{d\widehat{m}_s(\theta)}{d\theta} \\
&= g''\left(\widehat{m}_s(\theta)\right) \cdot \frac{\mu \cdot \left(1 - \frac{d\widehat{s}(\theta)}{d\theta}\right)}{\left(\theta - \widehat{s}(\theta) + w^*\right)^2} \\
&= \frac{1 - \frac{d\widehat{s}(\theta)}{d\theta}}{\theta - \widehat{s}(\theta) + w^* - \mu} \, ,
\end{aligned}$$

where we use

$$g''(x) = \frac{1}{x(1-x)}$$

while deriving the third equality. Rearrange the above equation, and we get

$$\begin{aligned}
\frac{d\widehat{s}(\theta)}{d\theta} &= \frac{\mu}{\theta - \widehat{s}(\theta) + w^*} \\
&= 1 - \widehat{m}_s(\theta) \, ,
\end{aligned}$$

where the last equality follows (C.1.21).

Again, taking derivatives of both sides of the above equation with respect to θ leads to

$$\begin{aligned}
\mu^{-1} \cdot \frac{d\widehat{s}(\theta)}{d\theta} &= g''\left(\widehat{m}_s(\theta)\right) \cdot \frac{d\widehat{m}_s(\theta)}{d\theta} \\
&= \frac{1}{\widehat{m}_s(\theta)\left(1 - \widehat{m}_s(\theta)\right)} \cdot \frac{d\widehat{m}_s(\theta)}{d\theta} \, .
\end{aligned}$$

Hence

$$\begin{aligned}
\frac{d\widehat{m}_s(\theta)}{d\theta} &= \mu^{-1} \cdot \widehat{m}_s(\theta) \cdot \left(1 - \widehat{m}_s(\theta)\right) \cdot \frac{d\widehat{s}(\theta)}{d\theta} \\
&= \mu^{-1} \cdot \widehat{m}_s(\theta) \cdot \left(1 - \widehat{m}_s(\theta)\right)^2 .
\end{aligned}$$

This completes the proof. □

Proof of Lemma 24. From Lemma 23, it is easy to see that the slope of $\widehat{s}(\theta)$ is always less than one. Hence, Lemma 24 is straightforward. □

Proof of Lemma 25. We proceed by discussing three cases.

Case 1: We show that $\widehat{s}(\theta) > \theta$ implies $s^*(\theta) = \theta$.

Suppose $s^*(\theta) < \theta$. Then we have $s^*(\theta) < \widehat{s}(\theta)$. Since both $(s^*(\theta), m_s^*(\theta))$ and $(\widehat{s}(\theta), \widehat{m}_s(\theta))$ satisfy condition (3.3.6), we must have $m_s^*(\theta) < \widehat{m}_s(\theta)$. Therefore,

$$\begin{aligned}
\frac{r(\theta)}{m_s^*(\theta)} &= -1 + \mu^{-1} \cdot (1 - m_s^*(\theta)) \cdot (\theta - s^*(\theta) + w^*) \\
&> -1 + \mu^{-1} \cdot (1 - \widehat{m}_s(\theta)) \cdot (\theta - \widehat{s}(\theta) + w^*) \\
&= 0\,,
\end{aligned}$$

which implies $s^*(\theta) = \theta$, a contradiction.

Note that, the logic for the inequality above is as follows. Since $(\widehat{s}(\theta), \widehat{m}_s(\theta))$ satisfies condition (C.1.2), we must have $\theta - \widehat{\theta} + w^* > 0$. Hence, $\widehat{s}(\theta) > s^*(\theta)$ implies that

$$\theta - s^*(\theta) + w^* > \theta - \widehat{s}(\theta) + w^* > 0\,.$$

Also by noting that

$$1 - m_s^*(\theta) > 1 - \widehat{m}_s(\theta) > 0\,,$$

we get the inequality above.

Hence, we have $s^*(\theta) = \theta$ in this case.

Case 2: We show that $\widehat{s}(\theta) < 0$ implies $s^*(\theta) = 0$.

Suppose $s^*(\theta) > 0$. Then we have $s^*(\theta) > \widehat{s}(\theta)$. By similar argument we know that $m_s^*(\theta) > \widehat{m}_s(\theta)$. Therefore,

$$\begin{aligned}
\frac{r(\theta)}{m_s^*(\theta)} &= -1 + \mu^{-1} \cdot (1 - m_s^*(\theta)) \cdot (\theta - s^*(\theta) + w^*) \\
&< -1 + \mu^{-1} \cdot (1 - \widehat{m}_s(\theta)) \cdot (\theta - \widehat{s}(\theta) + w^*) \\
&= 0\,,
\end{aligned}$$

which implies $s^*(\theta) = 0$. This is a contradiction. Hence, we have $s^*(\theta) = 0$ in this case.

Case 3: We show that $0 \leqslant \widehat{s}(\theta) \leqslant \theta$ implies $s^*(\theta) = \widehat{s}(\theta)$.

Suppose $\widehat{s}(\theta) < s^*(\theta)$. Then similar argument suggests $r(\theta)/m_s^*(\theta) < 0$, which implies $s^*(\theta) = 0 < \widehat{s}(\theta)$. This is a contradiction.

Similarly, suppose $s^*(\theta) < \widehat{s}(\theta)$. Similar argument suggests that $r(\theta)/m_s^*(\theta) > 0$, which implies $s^*(\theta) = \theta > \widehat{s}(\theta)$. This is, again, a contradiction. Hence, we have $s^*(\theta) = \widehat{s}(\theta)$ in this case.

This concludes the proof. □

Proof of Lemma 26. Apply Lemma 24 to Lemma 25, then Lemma 26 is straightforward. □

Proof of Lemma 27. We prove by contradiction. Suppose that the last two cases in Lemma 26 can occur in equilibrium. Hence, there exists a $\widetilde{\theta} \geqslant 0$, such that $s^*(\theta) = 0$ when $0 \leqslant \theta \leqslant \widetilde{\theta}$ and $s^*(\theta) = \widehat{s}(\theta)$ when $\theta > \widetilde{\theta}$.

Note that, $s^*(\theta)$ is strictly increasing when $\theta > \widetilde{\theta}$. Also, since we focus on the equilibrium with information acquisition, there must exist a θ'' such that $s^*(\theta'') > k$; otherwise the optimal security would be rejected without information acquisition. Therefore, there exists a $\theta' > \widetilde{\theta}$ such that $s^*(\theta') = \widehat{s}(\theta') = k$. Recall condition (C.1.1), we have

$$m_s^*(\theta') = \bar{\pi}_s^* .$$

Moreover, notice that we have $s^*(\theta') \in (0, \theta')$, we have

$$\begin{aligned} 0 = r(\theta') &= -m_s^*(\theta') + \mu^{-1} \cdot m_s^*(\theta') \cdot (1 - m_s^*(\theta')) \cdot (\theta' - s^*(\theta') + w^*) \\ &= -\bar{\pi}_s^* + \mu^{-1} \cdot \bar{\pi}_s^* \cdot (1 - \bar{\pi}_s^*) \cdot (\theta' - k + w^*) \\ &= \mu^{-1} \cdot \bar{\pi}_s^* \cdot (1 - \bar{\pi}_s^*) \cdot (\theta' - k) + \mathbb{E}[r(\theta)] , \end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}[r(\theta)] &= -\bar{\pi}_s^* + \mu^{-1}\left(\mathbb{E}\left[\frac{(\theta - s(\theta))\cdot g''(\bar{\pi}_s^*))}{g''(m(\theta))}\right] / g''(\bar{\pi}_s^*) + w^*\mathbb{E}\left[\frac{1}{g''(m(\theta))}\right]\right) \\
&= -\bar{\pi}_s^* + \mu^{-1}\left(w^*\cdot\left(1 - \mathbb{E}\left[\frac{g''(\bar{\pi}_s^*)}{g''(m(\theta))}\right]\right) / g''(\bar{\pi}_s^*) + w^*\mathbb{E}\left[\frac{1}{g''(m(\theta))}\right]\right) \\
&= -\bar{\pi}_s^* + \frac{\mu^{-1}w^*}{g''(\bar{\pi}_s^*)} \\
&= -\bar{\pi}_s^* + \mu^{-1}\cdot\bar{\pi}_s^*\cdot(1-\bar{\pi}_s^*)\cdot w^* \,.
\end{aligned}$$

We can express the expectation term $\mathbb{E}[r(\theta)]$ in another way. Note that, for any $\theta \in [0, \widetilde{\theta}]$, by definition we have

$$\begin{aligned}
r(\theta) &= -m_s^*(\theta) + \mu^{-1}\cdot m_s^*(\theta)\cdot(1-m_s^*(\theta))\cdot(\theta - s^*(\theta) + w^*) \\
&= -\widehat{m}_s(\widetilde{\theta}) + \mu^{-1}\cdot\widehat{m}_s(\widetilde{\theta})\cdot(1-\widehat{m}_s(\widetilde{\theta}))\cdot(\theta - 0 - \widetilde{\theta} + \widetilde{\theta} + w^*) \\
&= r(\widetilde{\theta}) - \mu^{-1}\cdot\widehat{m}_s(\widetilde{\theta})\cdot(1-\widehat{m}_s(\widetilde{\theta}))\cdot(\widetilde{\theta} - \theta) \\
&= -\mu^{-1}\cdot\widehat{m}_s(\widetilde{\theta})\cdot(1-\widehat{m}_s(\widetilde{\theta}))\cdot(\widetilde{\theta} - \theta) \,.
\end{aligned}$$

Also, as $s^*(\theta) = \widehat{s}(\theta)$ for any $\theta > \widetilde{\theta}$, we have $r(\theta) = 0$ for all $\theta > \widetilde{\theta}$. Hence,

$$\mathbb{E}[r(\theta)] = -\mu^{-1}\cdot\widehat{m}_s(\widetilde{\theta})\cdot(1-\widehat{m}_s(\widetilde{\theta}))\int_0^{\widetilde{\theta}}(\widetilde{\theta} - \theta)d\Pi(\theta) \,.$$

Therefore, we have

$$\begin{aligned}
\mu^{-1}\cdot\bar{\pi}_s^*\cdot(1-\bar{\pi}_s^*)\cdot(\theta' - k) &= -\mathbb{E}[r(\theta)] && \text{(C.1.22)} \\
&= \mu^{-1}\cdot\widehat{m}_s(\widetilde{\theta})\cdot(1-\widehat{m}_s(\widetilde{\theta}))\int_0^{\widetilde{\theta}}(\widetilde{\theta} - \theta)d\Pi(\theta) \,. && \text{(C.1.23)}
\end{aligned}$$

Now we take the tangent line of $s^*(\theta)$ at $\theta = \widetilde{\theta}$. The tangent line intersects $s = k$ at $\widetilde{\theta}'$, which is given by

$$\frac{k}{\widetilde{\theta}' - \widetilde{\theta}} = \left.\frac{ds^*(\theta)}{d\theta}\right|_{\widetilde{\theta}} = 1 - \widehat{m}_s(\widetilde{\theta}) \,.$$

Hence, we have

$$\widetilde{\theta}' = \widetilde{\theta} + \frac{k}{1 - \widehat{m}_s(\widetilde{\theta})} \,. \tag{C.1.24}$$

Also, note that we have shown that for any $\theta \geqslant \widetilde{\theta}$, we have

$$\frac{ds^*(\theta)}{d\theta} = \frac{d\widehat{s}(\theta)}{d\theta} = 1 - \widehat{m}_s(\theta) = 1 - m_s^*(\theta)\,.$$

Hence,

$$\frac{d^2 s^*(\theta)}{d\theta^2} = -\mu^{-1} \cdot m_s^*(\theta) \cdot (1 - m_s^*(\theta))^2 < 0\,.$$

Therefore, $s^*(\theta)$ is strictly concave for $\theta \geqslant \widetilde{\theta}$, and we also have $\widetilde{\theta}' < \theta'$. Consequently, by condition (C.1.24) and then conditions (C.1.22) and (C.1.23), we have

$$\begin{aligned}
\bar{\pi}_s^* \cdot (1 - \bar{\pi}_s^*) \cdot \left(\widetilde{\theta} + \frac{\widehat{m}_s(\widetilde{\theta})}{1 - \widehat{m}_s(\widetilde{\theta})} \cdot k\right) &= \bar{\pi}_s^* \cdot (1 - \bar{\pi}_s^*) \cdot (\widetilde{\theta}' - k) \\
&< \bar{\pi}_s^* \cdot (1 - \bar{\pi}_s^*) \cdot (\theta' - k) \\
&= \widehat{m}_s(\widetilde{\theta}) \cdot (1 - \widehat{m}_s(\widetilde{\theta})) \int_0^{\widetilde{\theta}} (\widetilde{\theta} - \theta) d\Pi(\theta)\,.
\end{aligned}$$

On the other hand, by Jensen's inequality, we know that

$$\bar{\pi}_s^* \cdot (1 - \bar{\pi}_s^*) > \mathbb{E}\left[m_s^*(\theta) \cdot (1 - m^*(\theta))\right]\,.$$

Therefore, we have

$$\begin{aligned}
\widehat{m}_s(\widetilde{\theta}) \cdot (1 - \widehat{m}_s(\widetilde{\theta})) \int_0^{\widetilde{\theta}} (\widetilde{\theta} - \theta) d\Pi(\theta) &> \bar{\pi}_s^* \cdot (1 - \bar{\pi}_s^*) \cdot \left(\widetilde{\theta} + \frac{\widehat{m}_s(\widetilde{\theta})}{1 - \widehat{m}_s(\widetilde{\theta})} \cdot k\right) \\
&> \mathbb{E}\left[m_s^*(\theta) \cdot (1 - m^*(\theta))\right] \cdot \left(\widetilde{\theta} + \frac{\widehat{m}_s(\widetilde{\theta})}{1 - \widehat{m}_s(\widetilde{\theta})} \cdot k\right)\,.
\end{aligned}$$

Expand the expectation term above and rearrange, we get

$$\begin{aligned}
& \widehat{m}_s(\widetilde{\theta})^2 \cdot k \cdot Prob[\theta \leqslant \widetilde{\theta}] + \int_{\widetilde{\theta}}^{+\infty} m_s^*(\theta) \cdot (1 - m_s^*(\theta)) d\Pi(\theta) \cdot \left(\widetilde{\theta} + \frac{\widehat{m}_s(\widetilde{\theta})}{1 - \widehat{m}_s(\widetilde{\theta})} \cdot k\right) \\
< \; & \widehat{m}_s(\widetilde{\theta}) \cdot (1 - \widehat{m}_s(\widetilde{\theta})) \int_0^{\widetilde{\theta}} (-\theta) d\Pi(\theta) \\
\leqslant \; & 0\,.
\end{aligned}$$

Nevertheless, the left hand side of the above inequality should be positive, which is a contradiction. This concludes the proof. □

Proof of Proposition 24. We first consider the case with a positive transfer price $p > 0$.

Suppose the corresponding security $s(\theta) = \theta - p$ is optimal in a generalized security design problem without the non-negative constraint. However, this security can be accommodated by neither Proposition 22 or Proposition 23, which exclusively characterize the optimal security in the generalized security design problem, a contradiction.

By Corollary 4, we also know that the security $s(\theta) = \theta$ that represents transfer with a zero price is not optimal. This concludes the proof. $\square$

Proof of Corollary 5. First, note that $s^*(\theta)$ is strictly increasing and continuous. Also, note that there exists a θ'' such that $s^*(\theta'') > k$; otherwise, the offer will be rejected without information acquisition.

Therefore, there exists a unique θ' such that $s^*(\theta') = k$, which ensures that $m_s^*(\theta') = \bar{\pi}_s^*$, and

$$\begin{aligned} r(\theta') &= -\bar{\pi}_s^* + \mu^{-1} \cdot \bar{\pi}_s^* \cdot (1 - \bar{\pi}_s^*) \cdot (\theta' - s^*(\theta') + w^*) \\ &= \mu^{-1} \cdot \bar{\pi}_s^* \cdot (1 - \bar{\pi}_s^*) \cdot (\theta' - s^*(\theta')) + \mathbb{E}[r(\theta)] \,. \end{aligned}$$

Note that $\mathbb{E}[r(\theta)] > 0$ and $\theta' - s^*(\theta') \geqslant 0$, we have $\theta' < \widehat{\theta}$. As $\theta' = s^*(\theta') = k$, it follows that $\widehat{\theta} > \theta' = k$. This concludes the proof. $\square$

Proof of Proposition 25. When we have $\mathbb{E}[\theta] \leqslant k$ and

$$\mathbb{E}\left[\exp(\mu^{-1}(t \cdot \theta - k))\right] > 1 \,,$$

according to Proposition 21, even if the entrepreneur proposes all the future cash flow to the investor, the security would induce the investor to acquire information and accept it with a positive (but less than one) probability. The only optimal security for this case is convertible preferred stock. This concludes the proof. $\square$

Proof of Lemma 19. The "if" part is straightforward, following the definition of efficiency. The "only if" part is ensured by the fact that the optimal screening rule is always unique given an arbitrary security, established in Proposition 21. $\square$

PROOF OF PROPOSITION 26. We state a useful lemma to begin. It allow us to focus on the first two types of equilibria for welfare analysis.

LEMMA **28.** *A project is initiated with a positive probability in the decentralized economy if and only if it is initiated with a positive probability in the corresponding centralized economy.*

PROOF OF LEMMA 28. With the objective function (3.4.10) in the centralized economy, the entrepreneur's optimal screening rule $m_c^*(\theta)$ is characterized by Proposition 21. Specifically, the investor will initiate the project without information acquisition, i.e., $Prob[m_c^*(\theta) = 1] = 1$ if and only if

$$\mathbb{E}[\exp(-\mu^{-1} \cdot (\theta - k))] \leqslant 1\,,$$

will skip the project without information acquisition, i.e., $Prob[m_c^*(\theta) = 0] = 1$ if and only if

$$\mathbb{E}[\exp(\mu^{-1} \cdot (\theta - k))] \leqslant 1\,,$$

and will initiate the project with probability $0 < \bar{\pi}_c^* < 1$, $\bar{\pi}_c^* = \mathbb{E}[m_c^*(\theta)]$, if and only if

$$\mathbb{E}[\exp(-\mu^{-1} \cdot (\theta - k))] > 1 \text{ and } \mathbb{E}[\exp(\mu^{-1} \cdot (\theta - k))] > 1\,,$$

in which $m_c^*(\theta)$ is determined by

$$\theta - k = \mu \cdot (g'(m_c^*(\theta)) - g'(\bar{\pi}_c^*))\,.$$

It is straightforward to observe that, the project is initiated with a positive probability in the frictionless centralized economy if and only if

$$\mathbb{E}[\exp(\mu^{-1} \cdot (\theta - k))] > 1\,. \tag{C.1.25}$$

Note that, condition (C.1.25) is the same as condition (3.3.1) in Lemma 18 that gives the investment criterion in the corresponding decentralized economy. This concludes the proof. □

We continue the proof of Proposition 26. By Lemma 28, once we prove the "only if" parts of both cases of debt and convertible preferred stock, the "if" parts will get proved simultaneously.

First, consider the case when $s^*(\theta)$ is debt. In this case, we have $Prob[m_s^*(\theta) = 1] = 1$, and

$$\mathbb{E}[\exp\left(-\mu^{-1} \cdot (s^*(\theta) - k)\right)] \leqslant 1,$$

both from Proposition 21. Since $s^*(\theta) < \theta$ when $\theta > D^*$, it follows that

$$\mathbb{E}[\exp\left(-\mu^{-1} \cdot (\theta - k)\right)] \leqslant 1,$$

which implies that $Prob[m_c^*(\theta) = 1] = 1$, also by Proposition 21. Hence, we know that

$$Prob[m_s^*(\theta) = m_c^*(\theta)] = 1,$$

which suggests that $s^*(\theta)$, as debt, achieves efficiency, according to Lemma 19.

Second, consider the case when $s^*(\theta)$ is convertible preferred stock that induces information acquisition. In this case, we have $Prob[0 < m_s^*(\theta) < 1] = 1$, and

$$\mathbb{E}[\exp\left(-\mu^{-1} \cdot (s^*(\theta) - k)\right)] > 1,$$

again both from Proposition 21. Since $s^*(\theta) < \theta$ when $\theta > \widehat{\theta}$, the relationship between $\mathbb{E}[\exp\left(-\mu^{-1} \cdot (\theta - k)\right)]$ and 1 is ambiguous. If

$$\mathbb{E}[\exp\left(-\mu^{-1} \cdot (\theta - k)\right)] \leqslant 1,$$

we have $Prob[m_c^*(\theta) = 1] = 1$, and information acquisition is not induced in the centralized economy. It follows that

$$Prob[m_s^*(\theta) = m_c^*(\theta)] \neq 1.$$

Otherwise, if

$$\mathbb{E}[\exp\left(-\mu^{-1} \cdot (\theta - k)\right)] > 1,$$

suppose we also have $Prob[m_s^*(\theta) = m_c^*(\theta)] = 1$, then according to condition (C.1.20), we have

$$Prob[s^*(\theta) = \theta] = 1,$$

which violates Corollary 4. A contradiction. As a result, from Lemma 19, we know that $s^*(\theta)$, as convertible preferred stock, cannot achieve efficiency. This concludes the proof. $\square$