Identifying Risks and Mitigating Deviations From Fundamentals in Investment Allocation Systems

by

Stavros Athans Valavanis

B.S. Electrical Science and Engineering
Massachusetts Institute of Technology, 2007

M.Eng. Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 2008

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2012

©2012 Massachusetts Institute of Technology. All rights reserved

Signature of Author ..................

Department of Electrical Engineering and Computer Science
January 31, 2012

Certified by .........................

Munther Dahleh
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Certified by ............................

Andrew W. Lo
Harris and Harris Group Professor of Finance
Thesis Supervisor

Accepted by ............................

Leslie Kolodziejski
Chairman, Department Committee on Graduate Theses
Identifying Risks and Mitigating Deviations From Fundamentals in Investment Allocation Systems

by

Stavros Athans Valavanis

Submitted to the Department of Electrical Engineering and Computer Science on January 31, 2012, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

The effects of the recent financial crisis have been devastating. Its causes are not well understood, but most people agree that incentive structures led to behaviors which are not captured by the standard theoretical paradigms of finance and economics. In this thesis, I analyze two models in which incentive structures cause deviations from such standard paradigms; one model focuses on the investment allocation process at the portfolio level and the other focuses on the investment allocation process at the financial system level. These two models are unified by the theme that incentive structures affect investment and risk in ways which are not captured by prevailing theoretical paradigms; my goal is to analyze these models so as to propose tools for identifying and/or mitigating their effects on risk and investment allocation. In analyzing the results of the asymmetric compensation incentive structure at the portfolio level, I propose a statistical inference tool which is able to decouple the information components of a portfolio allocation due to an equilibrium asset pricing model from the components due to the portfolio manager’s proprietary views. Such information is useful for risk management purposes as it allows one to examine whether the portfolio manager has implemented “outrageous” views in his portfolio. To explore the effects of incentive structure at the financial system level, I analyze an agent based model of the financial system inspired by recent empirical evidence of levered financial intermediary procyclical balance sheet management practices and propose an optimal rate setting rule for the Federal Reserve which mitigates some of the undesired effects on investment allocation and risk which arises from the endogenous financial system agent interactions.

Thesis Supervisor: Munther Dahleh
Title: Professor of Electrical Engineering and Computer Science
Thesis Supervisor: Andrew W. Lo
Title: Harris and Harris Group Professor of Finance
Acknowledgments

There do not exist words to describe the magnificent impact that my mother, Lena Valavani, has had on my life. I am extremely lucky to be her son as she is the best parent that someone could wish for. Her guidance and love have propelled me to where I am today. She has always supported me and has sacrificed a great deal to make me who I am. I would like to take this opportunity to thank her for all that she has been, done, and dreamed for me: I am eternally grateful!

I would also like to thank my two fantastic thesis supervisors: Professors Andrew Lo and Munther Dahleh. Professor Lo has been a tremendous inspiration in my academic and professional career and it is through him that I have come to love Finance. While at MIT, I was given the opportunity to get involved in many exciting projects with him; I am especially grateful that he gave me the opportunity to work with him on a key project with the US Treasury and in the pioneering efforts of the Office of Financial Research (OFR). I also cherish the experience of being his teaching assistant; he is truly a fantastic professor and I hope to be able to command an audience as well as he does one day. He has an original method of approaching research and teaching and I admire the fact that he is able to marry theory with practice as well as he does. I believe that Finance needs more leading figures like him, who are interested in addressing real world problems in an interdisciplinary fashion. Professor Dahleh was a pivotal figure in my PhD endeavor. He would always go in depth with me on questions that I had. He taught me how to organize my thoughts and analyze problems in an efficient manner. I will cherish the insightful discussions we had and his ability to delve right to the heart of a given problem. He is one of the smartest people I have met; many times, when explaining my work to him, I was in awe of how he was able to quickly assimilate new information from a different field and advise me appropriately. I also admire his passion for teaching his students and bonding with them. He is a very fun person and I thoroughly enjoyed spending time with him. In his new role of Associate Department Head, I believe that he will steer the department in the right direction; he is full of transformational ideas and EECS stands to benefit greatly from him. I also would like to thank Sanjoy Mitter for being a member of my Thesis Committee. His suggestions were very helpful and I thoroughly enjoyed working with him.

I would also like to thank my office mates at the Laboratory for Financial Engineering: Amir Khandani and Shawn Staker. I learned a great deal from both of them. I was very lucky to have met them early in my PhD career. I also would like to thank Michael Rinehart and Mardavij Roozbehani, who helped and advised me on my agent based model formulation. Furthermore, I would like to thank all of my friends for their help and support throughout my PhD.

Finally, I would like to thank MIT and EECS, for hosting me since 2002. Having obtained 3 degrees from this magnificent institution, I am ready to move on into the next stage of my life. But the wealth of knowledge and experience I have acquired here will serve me well for the rest of my life.
11 The Model Setup

11.1 Players and Preferences

11.2 The Risky Investment

11.3 The Solution

11.3.1 U's Allocations

11.3.2 L's Optimal Borrowing

11.3.3 The Equilibrium Solution

11.4 Solution Properties

11.4.1 Levered Balance Sheet Dynamics

11.4.2 Solution Dependence on Relative Size

11.4.3 Relative Size Growth Barrier

11.4.4 Leverage Versus Size

11.4.5 Solution Dependence on Outcome

11.4.6 Dynamics of \( P_t \)

11.5 Solution Simplification

11.5.1 The Linearization

11.5.2 Simplified Model Properties

11.5.2.1 The Dynamics of \( y_t \) and \( z_t \)

11.5.2.2 Relative Size Growth Barrier

11.5.2.3 Solution Dependence on Relative Size

11.5.2.4 Solution Dependence on Outcome

11.5.2.5 Simplified dynamics of \( P_t \)

11.6 Metrics of Interest

11.6.1 Investment Deviation From Fundamentals

11.6.2 Risk

12 Key Results

12.1 The Expected Growth of U and L

12.1.1 One Step Case

12.1.2 \( n \)-Step Case

12.2 Expected Future Investment and Deviation from Fundamentals

12.2.1 Deviation From Fundamentals Magnitude

12.2.2 One Step Expected Investment

12.2.3 One Step Expected Deviation from Fundamentals

12.2.4 \( n \)-Step Expected Investment

12.2.5 \( n \)-Step Expected Deviation From Fundamentals

12.3 Expected Risk

12.3.1 One Step Expected Risk

12.3.2 \( n \)-Step Expected Risk

12.4 Simulation Results for Full and Simplified Model

12.4.1 Simulation Overview

12.4.2 Additional Variables Definitions

12.4.3 Simulation Results
12.4.3.1 Simulation 1: Baseline ........................................... 100
12.4.3.2 Simulation 2: Lower Risk Free Rate ....................... 106
12.4.3.3 Simulation 3: Higher Variance .......................... 111

13 The Optimal Risk Free Rate Rule ............................ 116
13.1 The Logic Behind the Rule ................................. 116
13.2 Risk Free Rate Rule: Full Model Case ..................... 118
13.3 Risk Free Rate Rule: Simplified Model Case .............. 123
  13.3.1 Derivation of \( y_{t-1} \) ................................... 124
  13.3.2 Derivation of Relationship Between \( F_t^* \) and \( V_t \) ........ 124
13.4 Sensitivity of Optimal Rate in Simplified Model ............. 125
  13.4.1 Derivation ............................................... 125
13.5 Risk Under The Optimal Risk Free Rate Rule ................. 127
  13.5.1 Derivation ............................................... 127
13.6 Optimal Risk Free Rate Discussion ........................ 130

14 Summary ............................................................. 131

15 Appendix for Part I of Thesis ................................. 134
  15.1 Proof of Equations (13) and (14) ......................... 134
  15.2 Proof of Theorem 1 .......................................... 134
  15.3 Proof of Theorem 2 .......................................... 136
  15.4 Proof of Theorem 3 .......................................... 140
  15.5 Proof of Theorem 4 .......................................... 140
  15.6 Proof of Theorem 5 .......................................... 141

16 Appendix for Part 2 of Thesis ................................. 145
  16.1 Proof of \( q_t = 1 \) .......................................... 145
  16.2 Proof that \( P(V_{t+1} \leq K) \leq c \) in non-VaR Constrained Mode .. 147
  16.3 The Simplified Model Approximation ..................... 147
    16.3.1 An Empirical Justification .......................... 147
    16.3.2 The Approximation .................................... 150
  16.4 Variance of Next Period Investment ...................... 152
# List of Figures

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Flow chart depiction of the PM’s BL allocation process (top panel) and the Inverse Problem logic (bottom panel).</td>
</tr>
<tr>
<td>2</td>
<td>The above figure gives a graphical depiction of the procedure of estimating $k$. We look at the absolute value of the entries in $x$ and count the number of distinct nonzero entries.</td>
</tr>
<tr>
<td>3</td>
<td>The above figure graphically depicts the procedure for generating the $P'$ matrix. We group the nonzero entries of $</td>
</tr>
<tr>
<td>4</td>
<td>Scatter plot of the elements in $</td>
</tr>
<tr>
<td>5</td>
<td>Mean and median estimates of the number of views for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 1.</td>
</tr>
<tr>
<td>6</td>
<td>Mean and median estimates of the number of the rows of the $P$ matrix that are estimated correctly for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 1.</td>
</tr>
<tr>
<td>7</td>
<td>Mean and median estimates of the distance between the estimated and true $v$ vectors for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 1.</td>
</tr>
<tr>
<td>8</td>
<td>Mean and median estimates of the number of views for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 2.</td>
</tr>
<tr>
<td>9</td>
<td>Mean and median estimates of the number of the rows of the $P$ matrix that are estimated correctly for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 2.</td>
</tr>
<tr>
<td>10</td>
<td>Mean and median estimates of the distance between the estimated and true $v$ vectors for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 2.</td>
</tr>
<tr>
<td>11</td>
<td>Mean and median estimates of the number of views for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 3.</td>
</tr>
<tr>
<td>12</td>
<td>Mean and median estimates of the number of the rows of the $P$ matrix that are estimated correctly for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 3.</td>
</tr>
<tr>
<td>13</td>
<td>Mean and median estimates of the distance between the estimated and true $v$ vectors for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 3.</td>
</tr>
</tbody>
</table>
Left: Graphical illustration of FF portfolio notation for the 5x5 case; the 2x3 case is analogous. Right: $\tau$ over time for the mutual funds that we analyze.

The PM’s estimated one asset equilibrium view deviations $v - P\pi$ over time. Note that the y-axis is in annualized %.  

The PM’s estimated two asset equilibrium view deviations $v - P\pi$ over time. Note that the y-axis is in annualized %.  

Scatter plot of the PM’s estimated view magnitudes on the next month’s returns for the assets involved in his views (y-axis) versus the realized returns of those asset combinations over the next month (x-axis). The points are color coded to reflect the number of assets that each view involves. Note that the numbers are monthly percentages and are not annualized.  

The PM’s estimated one asset equilibrium view deviations $v - P\pi$ over time. Note that the y-axis is in annualized %.  

The PM’s estimated two asset equilibrium view deviations $v - P\pi$ over time. Note that the y-axis is in annualized %.  

Scatter plot of the PM’s estimated view magnitudes on the next month’s returns for the assets involved in his views (y-axis) versus the realized returns of those asset combinations over the next month (x-axis). The data from August 2008 is excluded. The points are color coded to reflect the number of assets that each view involves. Note that the numbers are monthly percentages and are not annualized.  

The blue line plots the year end net funds invested in US money market accounts in US$ Billions. The red line plots the percentage of money market fund assets invested in the short term risky debt of levered players.  

This figure plots the difference between the the 30 day financial commercial paper rate and the 30 day risk free rate. The rates are not annualized; they reflect the effective return of an investment over 30 days.  

Graphical depiction of the investment opportunities available to each agent.  

Graphical depiction of L’s balance sheet dynamics in response to a “good” $V_{t+1}$.  

Graphical depiction of the system’s monetary inflows and outflows  

Graphical illustration of risk. The left panel system, where $y_t$ is small, has a lower potential systemic cost than the right panel system, where $y_t$ is large and L has attracted a larger proportion of U’s portfolio into his risky debt.  

Graphical depiction of the mechanism whereby investment deviates from fundamentals for the case where $y_t > y^c$ (top panel) and $y_t < y^c$ (bottom panel).  

Simulation 1: Plot of the simulated average investment (top panel) and default cost (bottom panel) for the simplified and full models.  

Simulation 1: Plot of the simulated average $a_D$ (top panel) and inverse risk (bottom panel) for the simplified and full models.  

Simulation 1: Plot of the simulated average $y$ (top panel) and leverage (bottom panel).
31 Simulation 1: Plot of the simulated average $X$ (top panel) for the simplified and full models as well as the simplified model epsilon approximations (bottom panel). .......................................................... 105
32 Simulation 2: Plot of the simulated average investment (top panel) and default cost (bottom panel) for the simplified and full models. ................................. 107
33 Simulation 2: Plot of the simulated average $aD$ (top panel) and inverse risk (bottom panel) for the simplified and full models. ........................................ 108
34 Simulation 2: Plot of the simulated average $y$ (top panel) and leverage (bottom panel). .......................................................... 109
35 Simulation 2: Plot of the simulated average $X$ (top panel) for the simplified and full models as well as the simplified model epsilon approximations (bottom panel). .......................................................... 110
36 Simulation 3: Plot of the simulated average investment (top panel) and default cost (bottom panel) for the simplified and full models. ................................. 112
37 Simulation 3: Plot of the simulated average $aD$ (top panel) and inverse risk (bottom panel) for the simplified and full models. ........................................ 113
38 Simulation 3: Plot of the simulated average $y$ (top panel) and leverage (bottom panel). .......................................................... 114
39 Simulation 3: Plot of the simulated average $X$ (top panel) for the simplified and full models as well as the simplified model epsilon approximations (bottom panel). .......................................................... 115
40 Graphical depiction of the situation where financial frictions do not affect real economy investment. .......................................................... 117
41 Graphical depiction of the optimal $F^*_t$ (top panel) and a zoomed in version (bottom panel). .......................................................... 121
42 Graphical depiction of the optimal $F^*_t$ response to $V_t > V_{t-1}$ for the case where $y_{t-1}$ is larger (top panel) and for the case where $y_{t-1}$ is smaller (bottom panel). Note that for both cases, all parameters are the same except $y_{t-1}$ . 122
43 Estimated percentage allocation of US mutual fund assets into levered financial intermediary debt. .......................................................... 152
List of Tables

1. The 5 steps of the inference algorithm and the data needed for each step. ... 34
2. Summary of the two mutual funds we analyzed. ... 55
3. This table depicts the four possible deviation from fundamentals scenarios. ... 91
4. This table depicts the parameter values used for the 3 simulations we present. All quoted rates are annualized. ... 100
5. This table depicts the possible $a_t^*$ and $y_t$ movements compared to $a_{t-1}^*$ and $y_{t-1}$ respectively. ... 119
6. This table depicts the four possible risk free rate adjustment scenarios of the simplified model. We refer to these cases as A through D as denoted in the table. ... 123
7. This table depicts the rate adjustment magnitude results for cases as A through D. ... 126
8. This table presents the total worldwide assets under management (AUM) of the major levered and unlevered institution types. ... 149
1 Introduction

The recent economic crisis started as a financial crisis; as is typical with such crises, it quickly spread to the real economy. While its causes are not well understood, most people agree that incentive structures led to behaviors which are not captured by the standard theoretical paradigms of finance and economics. One popular incentive structure problem involves financial decision makers, such as portfolio managers, who have adverse incentives: they have a high payoff when they make money yet have limited downside when they lose money; thus, they are incentivized to take large risks. Another popular incentive structure problem involves levered financial intermediaries, who play an active role in channeling capital from ultimate lenders to ultimate borrowers instead of their traditionally passive role; these intermediaries are governed by their desire to maximize returns but are regulated by a VaR constraint. Both cases described above have the common denominator of incentive structures resulting in agents taking actions that cause deviations from the prevailing standard academic paradigms. In the former case, a portfolio manager’s compensation scheme incentivizes him to try to “beat the market”; he thus adjusts his allocations away from the standard Markowitz efficient frontier in order to accommodate his personal views. In the latter case, a bankruptcy constraint imposed on levered financial intermediaries causes a series of unintended dynamics because it incentivizes them to act in a procyclical manner; this results in a behavior which can be summarized as “invest more following good investment returns and divest more following bad investment returns”; such behavior may have broader effects on the economy and it deviates from the paradigm set by existing standard macroeconomic models. In this thesis, my goal is to explore both of the paradigm deviating incentive structure cases described above in an analytic manner so as to understand their causes and propose tools for identifying and/or mitigating their effects.

In analyzing the results of the asymmetric compensation incentive structure at the portfolio level, I propose a statistical inference tool which is able to decouple the information components of a portfolio allocation due to an equilibrium asset pricing model from the components due to the PM’s proprietary views. PM’s are paid handsomely because of their investment ideas and input. Such PM’s have to combine publicly available information with their own ideas/information in order to make optimal investment decisions consistent with both information sources. Typically, PM’s bring to the table their expertise on certain sectors, inside/private information, and even their personal biases. Indeed, many PM’s portfolio allocations are governed to a great extent by their proprietary views because they are driven to add value to their portfolio and outperform the market. In this thesis, I propose an estimation algorithm that disentangles the proprietary views from the public information; given a set of portfolio weights and the public information, my algorithm estimates the number of views that the PM had, which assets each view involved, and the strength of each view. Being able to solve such an “Inverse Problem” allows us to exactly determine what the PM’s value added is to his portfolio and whether he has “outrageous” views. Such information can be useful to both the risk management committee of the firm as well as to the

\[^{1}\text{I group these manager “inputs” into the term “proprietary views” for brevity.}\]
PM himself. From here on, I term the above problem the *Inverse Problem*.

To explore the effects of incentive structure at the financial system level, I analyze an agent based model of the financial system inspired by recent empirical evidence of levered financial intermediary active balance sheet management practices and propose an optimal rate setting rule for the Federal Reserve (Fed) which mitigates some of the undesired effects arising from endogenous financial system agent interactions. The key players in my model are: the leveraged sector (e.g. hedge funds), the unleveraged sector (e.g. pension funds), and the Fed. These agents pursue local optimization strategies and their interaction dictates the amount of investment in the real economy at each point in time, the deviation of this investment amount from that which is justified by the fundamentals, and the buildup of risk in the system. In my analysis, I first explore the natural system dynamics under a constant Fed Funds Rate (FFR) so as to isolate the effects of the leveraged sector’s procyclical behavior.\(^2\) Subsequently, I propose an optimal FFR setting rule which mitigates the frictions to investment channeling caused by the endogenous interactions of these locally optimizing agents; furthermore, I demonstrate that there is a fundamental tradeoff as it is not generally possible to simultaneously have investment reflect fundamentals and have risk stay constant in the system by only controlling the FFR.

The two models discussed above which I analyze in this thesis are essentially two pieces of work which contribute to two different levels of applied finance, unified by the theme that incentive structures affect risk in ways which are not captured by prevailing theoretical paradigms: thus, this thesis is divided into two parts. The first part formulates and solves the *Inverse Problem*; the result is an algorithm which allows the user to identify the PM’s portfolio deviations from the efficient frontier implied by the prevailing public information; furthermore, the algorithm estimates exactly what the deviation implied PM’s proprietary views are. The second part formulates an agent based model of the financial system which incorporates an active levered financial intermediary who acts in a procyclical fashion; the effects of this behavior on aggregate investment and systemic risk are analyzed and a mitigating FFR setting rule is proposed.

\(^2\)A constant FFR is also a realistic assumption given that this rate stays approximately constant for long periods of time compared to the typical decision making frequency of the levered and unlevered sectors.
Part I
Extracting Portfolio Manager Alpha: An Inverse Black Litterman Approach
2 Motivation for Inverse Problem

The recent financial crisis started as a banking crisis yet quickly spread to the real economy. Indeed, one of the main culprits of the current crisis is widely believed to be the morally hazardous "bonus culture" of Wall Street which results in excessive risk taking. From a portfolio manager’s (PM for brevity) perspective, since his compensation schemes mainly take into account the short term profits he rakes in, he has an incentive to take huge risks because if they pay off, he gets rewarded handsomely; on the other hand, if the risky positions cause large losses for the portfolio, the PM does not lose his own money.

In this paper, we concern ourselves with portfolio managers; for intuition, you can think of them as managing equities although the type of asset really doesn’t matter in terms of the techniques we present. Typically, PM’s at large hedge funds, mutual funds, and investment banks are paid handsomely because of their investment ideas and input. Such PM’s have to combine publicly available information with their own proprietary views. If PM’s used only public information for their investment decisions, they would be obsolete. In fact, it is plausible that many PM’s portfolio allocations are governed to a great extent by their proprietary views because they are driven to add value to their portfolio and beat the market; a manager’s ego often adds to this phenomenon.

When observing the PM’s portfolio allocations, an interesting question to ask is what part of those allocations are due to the PM’s proprietary views as opposed to the public information available: in other words, can the proprietary views be disentangled from the public information? Being able to solve such an Inverse Problem would allow us to exactly determine what the PM’s value added is to his portfolio. Such information can be useful to both the risk management committee of the firm as well as to the manager himself. For example, a PM may consistently have outrageous views (given market circumstances) on some of the assets that he manages; knowing that the PM is consistently inputting outrageous views could be very useful for the risk committee. PM’s rarely explicitly state their outrageous views and senior management rarely intervenes when PM’s are profitable, no matter how the PM’s make their investment decisions. Clearly, having an inference algorithm which can pick out PM’s who make wild bets can be used in mitigating company-wide risks.

Such an inference algorithm could also prove useful for the PM himself. For example, an equity PM may have the proprietary view that the German DAX index will outperform the US S&P 500 by 10% next year and change positions in his portfolio to reflect his belief; however, due to the complex interaction of the assets in his portfolio through factors such as correlations and currency risks, it may well be the case that his positions do not adjust to consistently reflect both the public information and his proprietary view; in other words, the PM may be incapable of updating his portfolio correctly to consistently incorporate his subjective beliefs. An inference algorithm which would output the proprietary views implied...
by his portfolio allocations would be a useful “self check” tool.

A third application of such an inference algorithm could be “insider trading” detection. Typically, when endowed with inside information on a particular stock, a sophisticated PM will try to mask his view; he may adjust his position in multiple stocks so as to conceal his information. With an inference algorithm that can separate proprietary views from public information, a regulator can track a PM and see if he often inputs proprietary views on a particular asset that uncannily turn out to be “right” most of the time. Tracking a PM’s proprietary views could also be used to predict his future proprietary views: for example, if a PM often has a positive proprietary view on a stock, one can “frontrun” him and make a profit.5

A fourth application of such an inference algorithm involves the aggregation of the proprietary views of a set of PM’s. For example, at the firm level, one can aggregate the proprietary views of each PM in the firm and generate an aggregate firm consensus or anti-consensus. Thus, instead of just trusting the investment advice that a firm offers its clients, one can examine whether the firm is implementing the advice that it gives to its clients. In theory, there should be an overall proprietary view consistency result for the market as a whole: if one could back out the proprietary views of every market player, these views should net out to zero in some sense. For every positive proprietary view on an asset, there must be an equally negative proprietary view from the rest of the market; else, the positive proprietary view could not have been implemented in the first place.

The consistent way of combining public information with proprietary views is through Bayesian updating: the public information results in a prior and the proprietary views “update” the prior to form the posterior. In the portfolio management literature, much emphasis is given on the expected returns. The Black-Litterman (BL) asset allocation model ([16] and [17]) is essentially a Bayesian updating algorithm which updates the prior on the expected asset returns by incorporating the PM’s proprietary views. The BL model is quite popular in the portfolio management industry due to its intuitiveness, flexibility and ease of implementation: the PM can express an arbitrary number of proprietary views on subsets of the assets that he manages and obtain the updated market returns distribution that reflects these views in a consistent way. Given that a PM is using the BL model framework, we seek to develop an inference algorithm which will take as input the PM’s portfolio allocations and the public information available and will detect how many proprietary views the PM has, which assets these proprietary views are on, and what these proprietary views are. We call this problem the Inverse Problem. In what follows, the relevant previous literature is presented in Section 3, the description of the BL model is presented in Section 4, some key theorems needed to solve the Inverse Problem are presented in Section 5, the inference algorithm to solve the Inverse Problem is presented in Section 6, simulation results are presented in Section 7, and an empirical application of the inference algorithm on mutual fund data along with

---

5 Frontrunning is the act of taking a position in an asset in advance of an action which will move the asset’s price in a predictable fashion. For example, if we know that a portfolio manager will have a positive view on a stock in the next period, we know that he will buy more of it thus raising its price. We can take advantage of this by buying the stock today and selling it to tomorrow since its price will be driven up by the manager’s action.
discussion are presented in Section 8.
3 Literature Review for Inverse Problem

There are two main versions of the BL model: the original model, which was created by Black and Litterman ([16] and [17]), and the “simplified” model, which assumes that the prior on the mean of the returns is a point estimate instead of a distribution. Relatively good presentations of the original BL model can be found in [40], [41], and [28]. The simplified model is presented in [56] and Meucci ([49] and [51]). Although there is no shortage of literature discussing the BL model, the presentations are largely mediocre: mathematical derivations are often questionable and/or missing and explanations are disappointing. It is our opinion that much of the BL literature copies previous literature resulting in mistakes and omissions being carried through. In fact, we believe that the “simplified” model came as a result of a shallow understanding of the original BL papers ([16] and [17]). Since the BL model related publications are mainly from investment management practitioners who either don’t want to give explicit details regarding their ideas/methods or are not trained in rigorous academic technical writing, the aforementioned problems should not be a surprise.

[40] compares the performance of the BL model with that of traditional mean-variance optimization through an application to a portfolio consisting of the major stock index of the G7. They demonstrate that the BL model gives more stable portfolios than the traditional mean-variance approach. [36] modifies the classical BL model by applying different priors for asset returns (the t-student distribution and alpha-stable distributions). [48] proposes using a four moment CAPM risk model in the estimation of the prior distribution of returns; however, they seem to ignore the posterior variance of the asset returns in their portfolio optimization step. [33] illustrates how to incorporate a trading strategy in the BL framework: they use the BL model to combine the output of a cross sectional momentum strategy (this plays the role of the proprietary views) with the prior distribution resulting from public information. [51] modifies the BL model so that the PM inputs views on the returns instead of the expected returns: however, he simplifies the prior by assuming no estimation error. [27] presents an extension of the BL model in which proprietary views can be on both individual assets as well as on Fama-French type factor portfolios (see [34]). [63] presents an extension of the BL model which allows for Bayesian learning from the data in addition to the PM’s proprietary views and the public information.

In terms of previous work related to the Inverse Problem, the only relevant paper seems to be [35], which attempts to attack a somewhat related inference problem: in their paper, given a set of portfolio weights and benchmark index returns, their goal was to derive the implied expected returns that would make the given portfolio optimal; note that in their case, the implied expected returns distribution was the “blended” return distribution resulting from both the PM’s proprietary views distribution and the prior returns distribution resulting from a risk model such as the CAPM. Our Inverse Problem goal is much deeper than just estimating what the PM’s posterior returns distribution should be to make his portfolio optimal: we wish to determine how many proprietary views the PM has, which assets the proprietary views are on, and what these proprietary views are. It should be noted that the inference method we present to solve the Inverse Problem can also be used (with slight modifications) for the modified BL models of [48],[33], [51], and [63]. Furthermore, it can
also be used for the “simplified” BL model setup.
4 The Black-Litterman Model Setup and Intuition

In this paper, we use the original BL model because it is more complex, realistic, and rich. Furthermore, the Inverse Problem solution is more complex and thus more interesting. The assumptions underlying the BL model are the same as those underlying Modern Portfolio Theory (MPT). In the BL setup, the PM is rational and faces a one period asset allocation problem. He is risk averse and his risk-reward tradeoff can be described via a CARA (constant absolute risk aversion) utility function, which reduces to a quadratic utility function when returns are normal. In order to arrive at a portfolio allocation, he needs to have a returns distribution on the assets. Given the returns distribution and his utility function, the PM runs an optimization routine so as to generate "optimal" (with respect to his risk-return preferences) portfolio weights.

The BL model is a Bayesian updating algorithm which updates a prior returns distribution by incorporating the PM's proprietary views. More specifically, assume that there are n assets in the market (which may include equities, bonds, and other assets); the returns distribution of these assets is assumed normal:

\[ x \sim N(\mu, \Sigma) \]  

(1)

where \( x \) represents the \( nx1 \) returns vector of the assets and the covariance matrix \( \Sigma \) is known. With the Bayesian approach, the expected returns \( \mu \) are random variables themselves: they are not observable and one can only infer their probability distribution. This inference starts with a prior belief on \( \mu \) which comes from public information: this prior is updated by incorporating the PM's proprietary views, resulting in a posterior distribution of the returns. Below, we discuss the prior and proprietary views in detail.

4.1 The BL Prior

The prior distribution on \( \mu \) is formed through an asset pricing equilibrium model, such as the Capital Asset Pricing Model (CAPM). The main assumptions needed are that all investors have access to the same public information \( G \) (e.g. historical market data, announced company-specific news, etc.) and that all investors maximize their expected utility (which is CARA); furthermore, the returns distribution is assumed normal and there are no market frictions. Thus, under the asset pricing equilibrium, all investors hold the same optimal (with respect to their objective function) portfolio of risky assets; they differ only on the amount of the risk free asset that they hold, which depends on their risk aversion. Since all investors hold the same optimal portfolio, it must be the portfolio that the market holds as a whole; this is called the market portfolio. By observing the market portfolio, one can back out what the expected returns \( \pi \) implied by the asset pricing equilibrium are: these are the

---

6For a discussion of MPT and its assumptions, see Markowitz ([47]).
7Unless otherwise noted, all returns refer to excess returns; an asset's excess return is defined as its return minus the risk-free rate.
8The CAPM was originally proposed by Sharpe [57] and Lintner [45].
9Unless otherwise noted, the term "portfolio" will be used to refer to the portfolio of risky assets.
expected returns for which the demand of each asset equals its supply [15]. Mathematically, each investor decides today on the portfolio allocations to maximize his expected utility in time 1 conditional on the public information G available at time 0. Without loss of generality, assume that at time 0, his portfolio is worth $1: we denote by $\omega$ the vector of weights on the risky assets, $w_0$ the weight on the risk free rate, $r$ the random vector of total returns, $r_f$ the risk free rate, and by $M_1$ the value of his portfolio at time 1. The PM’s optimization problem is:

$$\max_{\omega} E_0 (U_1 | G) = \max_{\omega} E_0 \left( -\frac{1}{\Delta} e^{-\Delta M_1 | G} \right) = \max_{\omega} E_0 \left( -e^{-\Delta (1+\omega' r + (1-\omega') r_f) | G} \right)$$

$$= \max_{\omega} E_0 \left( -e^{-\Delta (1+r_f)} \right) = \max_{\omega} \left( -e^{-\Delta \omega' \Delta} \right)$$

where $\Delta$ denotes his risk aversion and $\pi = E_0 (x | G) = E_0 (r - r_f | G)$ is the vector of expected excess equilibrium returns. For the market representative agent, who by definition holds only the market portfolio $\omega_{eq}$ (implying $1'\omega_{eq} = 1$), the first order condition (FOC) of the maximization in equation (2) is:

$$\omega_{eq} = \frac{1}{\Delta_{mkt}} \Sigma^{-1} \pi \Rightarrow \pi = \Delta_{mkt} \Sigma \omega_{eq}$$

where $\Delta_{mkt}$ is the market risk aversion.\(^{10}\) Thus, the equilibrium expected excess returns can be backed out if one is given the market equilibrium weights $\omega_{eq}$ and $\Delta_{mkt}$. The market equilibrium weights $\omega_{eq}$ are observable: for an asset, its market equilibrium weight is defined as the ratio of its market capitalization with the total market capitalization of all assets in the investment universe.\(^{11}\) Furthermore, $\Delta_{mkt}$ can be estimated by noting that the expected market portfolio return is given by $\omega_{eq}' \pi$: thus, multiplying both sides of the market representative agent’s FOC in equation (3) by $\omega_{eq}'$ leads to the following expression for $\Delta_{mkt}$:

$$\omega_{eq}' \pi = \Delta_{mkt} \Rightarrow E_0 (x_{mkt}) = \Delta_{mkt} Var_0 (x_{mkt}) \Rightarrow \Delta_{mkt} = \frac{E_0 (x_{mkt})}{Var_0 (x_{mkt})}$$

\(^{10}\)Note that if the risk-free rate asset is in zero net supply, $\Delta_{mkt}$ is a "weighted average" of all market players’ $\Delta$.

\(^{11}\)Of course, calculating the market capitalization of all possible investable assets is unrealistic in practice. Depending on the type of assets that a portfolio manager invests in, the investment universe is approximated accordingly. For example, a US equity portfolio manager may use as his investment universe all stocks listed on the NYSE, NASDAQ, and AMEX when calculating the market equilibrium weights. A Global Tactical Asset Allocation (GTAA) manager who invests in G8 world markets may use as his investment universe all stocks and bonds traded in the countries which he invests in.
Thus, an estimate of $\Delta_{mkt}$ can be obtained through equation (4) where historical data is used to estimate $E_0(x_{mkt})$ and $Var_0(x_{mkt})$ by their sample counterparts.

The equilibrium expected excess returns vector $\pi$ is the mean of the prior distribution on $\mu$; however, there is uncertainty attached to the estimate because the market may not necessarily be in equilibrium and/or the equilibrium weight estimation may suffer from noise. Thus, the BL model models the prior on $\mu$ as a normal distribution with a mean of $\pi$ and a covariance matrix of $\tau \Sigma$:

$$\mu_G \sim N(\pi, \tau \Sigma)$$  \hspace{1cm} (5)

where the subscript $G$ indicates that the prior is based on the public information and the scalar $\tau$ represents the confidence that the PM has in the market equilibrium excess return estimation. The fact that the covariance matrix of the prior distribution on $\mu$ is proportional to $\Sigma$ is based on the assumption that the asset pricing equilibrium model for $\pi$ is a linear factor model (see [28]). As for $\tau$, it is smaller than 1 since the uncertainty in the mean returns should be less than the uncertainty in the actual returns. There will be a further discussion of $\tau$ later on.

### 4.2 The PM Proprietary Views

The PM may input proprietary views on various subsets of the assets that he manages: for our purposes, it does not matter how and why these proprietary views are formed. The proprietary views are statements on the vector $\mu$ and can be of varying degrees of confidence. They may be absolute views of the form “I expect stock a to return 5% with an uncertainty of $x\%$” or portfolio views of the form “I expect the portfolio consisting of 1 unit of stock b and 1 unit of stock c to outperform the portfolio consisting of 1 unit of stock d, 1 unit of stock e, and 1 unit of stock f by 10% with an uncertainty of $y\%$”. Denoting by $k$ the number of views that the PM has, we define the $(k \times n)$ matrix $P$ as the pick matrix, where $n$ is the number of assets in the PM’s investment universe: each view occupies one row of $P$. Furthermore, we define the $(k \times 1)$ views vector $v$, which contains the PM’s views. A main assumption of the BL model is that the PM’s views $v$ conditional on $\mu$ can be modeled as a normal distribution centered at $P \mu$:

$$v = P \mu + \epsilon_2, \quad \epsilon_2 \sim N(0, \Omega)$$  \hspace{1cm} (6)

where $\Omega$ is a matrix containing the PM’s uncertainties on his views. For illustration purposes, if the PM has the two views specified in the first paragraph, then $P$ and $v$ are populated as:

$$P = \begin{bmatrix} a & b & c & d & e & f \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & -1 & -1 & -1 \end{bmatrix}, \quad v = \begin{bmatrix} 0.05 \\ 0.1 \end{bmatrix}$$  \hspace{1cm} (7)
As for the uncertainty in each view, this is really just the standard deviation of the view according to the model in equation (6): thus, in the example in the first paragraph, the $\Omega$ would be a diagonal matrix with $x$ and $y$ as elements. However $\Omega$ need not be diagonal: the model allows for correlation between the PM’s views. The only requirement on $\Omega$ is that it be positive definite. As an example of correlated PM views, consider the following: the PM believes that Apple will outperform Walmart by 3% and that Google will outperform CVS by 7%. These views may both stem from a belief that financials will be stronger than consumer staples. Thus, allowing for view correlation enriches the model.12,13

4.3 The BL Posterior

Given the prior on $\mu$ and the PM’s views, we can determine the posterior distribution on $\mu$ and thus the posterior distribution of $x$. Given the setup, the natural way to proceed is through Bayes rule. Mathematically:

$$pdf(\mu|v) = \frac{pdf(v|\mu)pdf(\mu)}{pdf(v)} \propto pdf(v|\mu)pdf(\mu)$$

where

$$pdf(\mu|v) \sim N(\mu_{BL}, \Sigma_{BL})$$

Equations (10) and (11) can also be rewritten in a more intuitive manner as:

$$\mu_{BL} = \Sigma_{BL} [(\tau \Sigma)^{-1} \pi + P'\Omega^{-1}v]$$

$$\Sigma_{BL} = [(\tau \Sigma)^{-1} + P'\Omega^{-1}P]^{-1}$$

Equations (10) and (11) can also be derived through a mixed estimation approach.14 Notice that $\mu_{BL}$ can also be written as:

$$\mu_{BL} = \Sigma_{BL} [(\tau \Sigma)^{-1} \pi + P'\Omega^{-1}v]$$

Equations (13) and (14) can also be rewritten in a more intuitive manner as:

$$\mu_{BL} = \pi + \tau \Sigma P(\tau \Sigma P' + \Omega)^{-1}(v - P\pi)$$

$$\Sigma_{BL} = \tau \Sigma - \tau^2 \Sigma P(\tau \Sigma P' + \Omega)^{-1}P \Sigma$$
The variance of the posterior mean distribution is lower than that of the prior because more information is added thus reducing the uncertainty of the model. The proof of equations (13) and (14) can be found in the Appendix. Given the posterior distribution of $\mu$, the posterior distribution of the returns is:

$$x|v \sim N(\mu|v, \Sigma|v) = N(\mu_{BL}, \Sigma + \Sigma_{BL})$$  \hspace{1cm} (15)$$

Since the posterior mean of the returns is a distribution itself, the variance of the posterior returns needs to reflect this added uncertainty. Thus, the covariance matrix of the posterior returns distribution is that of the prior plus the uncertainty in the posterior mean estimate $\Sigma_{BL}$. Armed with the posterior distribution, the PM can now proceed to optimize according to equation (2), resulting in the BL portfolio weights:

$$\omega_{BL} = \frac{1}{\Delta}(\Sigma + \Sigma_{BL})^{-1}\mu_{BL}$$  \hspace{1cm} (16)$$

Before continuing, a few notes about $\Omega$ and $\tau$. As is evident from equations (10) and (11), $\tau\Sigma$ and $\Omega$ weight the prior expected excess returns and the PM views respectively. Thus, higher uncertainty in an information source leads to a lower weight on that information. It is clear that $\mu_{BL}$ can be written as a function of the “uncertainty ratio” $\Omega/\tau$. However, this is not the case for $\Sigma_{BL}$ which can only be written as $\tau$ times a function of $\Omega/\tau$. Thus, if both $\tau$ and $\Omega$ go to zero (meaning full confidence in both the prior and the views on $\mu$) yet $\Omega/\tau$ goes to a non-zero finite matrix, then the posterior $\mu$ becomes deterministic since $\Sigma_{BL}$ goes to 0.

### 4.4 The Intuition Behind BL Portfolios

The basis of the BL model rests on the fact that in the absence of proprietary views, the PM should hold the market equilibrium portfolio and possibly the risk-free asset, depending on his risk aversion relative to the market risk aversion. Essentially, the BL weights reflect a weighted average of two sources of information on $\mu$: below, we present 4 intuition building observations, which are straightforward applications of the BL formulae:

1. If the PM has complete confidence in the equilibrium market prior (i.e. $\tau = 0$) then:

$$\omega_{BL,\tau=0} = \frac{1}{\Delta}(\Sigma^{-1}\tau = \frac{\Delta_{mkt}}{\Delta}\omega_{eq}$$  \hspace{1cm} (17)$$

Thus, if the PM has complete confidence in the prior, he will chose to hold the market equilibrium portfolio. If his risk aversion is higher than the market risk aversion, then his BL weights sum to $\Delta_{mkt}/\Delta < 1$, implying that he will also hold some of the risk-free asset. If the PM is less risk averse than the market, his BL weights sum to $\Delta_{mkt}/\Delta > 1$ implying that he will be short some of the risk-free asset in order to further invest in the market portfolio.
2. If the PM has no confidence in his views (i.e., $\Omega \to \infty$) then:

$$\omega_{BL,\Omega=\infty} = \frac{1}{\Delta} \frac{1}{1+\tau} \pi = \frac{\Delta_{mkt}}{\Delta} \frac{\omega_{eq}}{1 + \tau}$$

(18)

Thus, if the manager has no confidence in his views, he will allocate his assets according to the market equilibrium weights, scaled by $1/(1 + \tau)$, which reflects his uncertainty on the prior. As his uncertainty $\tau$ about the prior increases, he allocates less of his wealth to risky assets. His risk-free asset allocation also depends on $\Delta_{mkt}/\Delta$ as was the case above.

3. If the PM has no views (i.e., $P=0$), then:

$$\omega_{BL,P=0} = \frac{1}{\Delta} \frac{1}{1+\tau} \pi = \frac{\Delta_{mkt}}{\Delta} \frac{\omega_{eq}}{1 + \tau}$$

(19)

Thus, if the manager has no views, he will allocate his assets in exactly the same way as in the case where he has no confidence in his views (case 2 above).

4. If the PM is completely confident in his views (i.e., $\Omega = 0$), then the entries in $\mu_{BL}$ corresponding to the assets on which the manager has views on will depend only on the views and not on the market equilibrium. For illustration purposes, assume that the manager has absolute views on all $n$ assets in his investment universe. Then, $P = I_n$ and the resulting BL weights are:

$$\omega_{BL,\Omega=0,P=I} = \frac{1}{\Delta} \Sigma^{-1} v$$

(20)

Note that equation (20) is exactly analogous to the standard FOC of equation (3), where $\pi$ has been replaced by $v$ and $\Delta_{mkt}$ has been replaced by $\Delta$. Thus, the PM is allocating assets based solely on his views vector $v$. 

26
5 Key Theorems and Definitions Needed to Solve the Inverse Problem

Given the intuitiveness of these extreme case results presented in the previous section, one can conjecture that the BL weights are a weighted average of the two sets of weights that would be implemented under each of the two information sets. Clearly, the posterior distribution on \( \mu \) is a blend of the two information sets: this leads to a posterior returns distribution, of which the PM passes the first two moments through a mean-variance optimizer in order to arrive at his BL allocation. Thus, it is not obvious what the relationship is between the BL weights and the two sets of weights that would have been implemented under each of the two information sets separately. However, if such a relationship did exist, then a solution to the Inverse Problem may be possible.

**Figure 1:** Flow chart depiction of the PM’s BL allocation process (top panel) and the Inverse Problem logic (bottom panel).

The top panel of Figure 1 depicts the flow chart of the PM’s BL allocation process; the bottom panel depicts the logical flow of the Inverse Problem. The Inverse Problem uses the PM’s observed allocation in conjunction with the public information available in order to estimate three things: the number of views \( k \) that the manager input, the assets that each view involves (i.e. the \( P \) matrix), and the magnitude of each view (i.e. the \( v \) vector). By definition, the Inverse Problem is ill defined in the sense that the PM may have views on only a subset of the \( n \) assets that he manages (i.e. \( k < n \)); furthermore, these views may be
complicated portfolio views involving multiple assets. Moreover, we do not know how many views the PM has. Thus, given the public information and PM’s allocation on \( n \) assets, the “system” is intuitively ill defined: loosely speaking, even if we had the information about how many views the PM has and which assets each of these views involves, we would still have \( n \) data points for estimating the magnitude of each of these views (i.e. the \( k \times 1 \) views vector \( v \)).

The beauty of the inference algorithm we present is that under a set of loose and realistic assumptions, the Inverse Problem can be transformed into an exactly identified problem thus allowing us to obtain a unique and exact solution. In order to develop an inference algorithm to solve the Inverse Problem, we first present the necessary definitions needed and derive the key theorems which lay the groundwork: these are presented in this section.

We denote the PM’s proprietary views with the triplet \((P, \Omega, v)\). Below, we provide the definition of the concepts of internal consistency and informational equivalence of the PM’s proprietary views.

**Definition 1 Internal Consistency:** Assume that the PM expresses a \((k \times m)\) matrix \( P \) of rank \((l \leq k)\), a \((k \times 1)\) views vector \( v \), and a \((k \times k)\) confidence matrix \( \Omega \). Furthermore, define as \( B \) the \((k \times k)\) matrix which transforms \( P \) to its row reduced echelon form \( P_{\text{ref}} \); in other words, \( P_{\text{ref}} = BP \). The PM’s proprietary views \((P, \Omega, v)\) are internally consistent if all three of the following hold:

1. There exists at least one solution vector \( \mu \) which satisfies the condition:

   \[
   v = P \mu \quad (21)
   \]

2. The last \((k-l)\) rows and the last \((k-l)\) columns of the matrix \( \Omega_{\text{ref}} = B \Omega B' \) are zero vectors.

3. The “upper left” \((l \times l)\) submatrix of \( \Omega_{\text{ref}} = B \Omega B' \) is positive definite.

The intuition of the first condition in Definition 1 relies on the fact that each PM’s view is a statement on the expected return of some combination of assets. In other words, he is making statements on the realization of the random vector \( P \mu \). A violation of the first condition in Definition 1 would mean that there exists at least one asset on which the PM has conflicting views. Note that if \( P \) has full row rank \((l = k)\), then condition 1 above holds for any \( v \).

The intuition of the second and third statements comes from the main assumption of the BL model that the manager’s views can be modeled as:

\[
 v = P \mu + \epsilon_2, \quad \epsilon_2 \sim N(0, \Omega) \quad (22)
\]

where \( \Omega \) is a matrix containing the PM’s uncertainties on his views. If we apply the \( B \) matrix to both sides of equation (22):
where $v_{rref} \equiv Bv$. From equation (23):

$$v_{rref} \sim N(P_{rref}\mu, B\Omega B')$$

Note that since the matrix $B$ performs the row operations on $P$ required to get it in its row reduced echelon form $P_{rref}$, $B$ is invertible. Since $P$ has rank $l$, the last $(k-l)$ rows of $P_{rref}$ are zero vectors. Thus, since the first condition in Definition 1 must hold, the last $(k-l)$ entries of the vector $v_{rref}$ must be zero. Therefore, the condition that requires the last $(k-l)$ rows and the last $(k-l)$ columns of $\Omega = B\Omega B'$ to be zero makes intuitive sense. As for the “upper left” $(lxl)$ submatrix of $\Omega = B\Omega B'$, we require it to be positive definite because we assume that no linear combination of the $l$ “independent” views has full certainty. Note that for the remainder of this section, we assume that the confidence matrix $\Omega$ is positive definite (as is the case in the classic BL setup).

**Definition 2 Informational Equivalence:** Assume that the PM expresses a $(kxn)$ matrix $P$, a $(kx1)$ views vector $v$, and a $(kxk)$ confidence matrix $\Omega$. Furthermore, define $B$ as some $(kxk)$ invertible matrix. For a given public information set, all triplets $(P_*, \Omega_*, v_*)$ of the form $(P_*, \Omega_*, v_*) = (BP, B\Omega B', Bv)$ for some $B$ are said to be informationally equivalent: they all carry the same information about the manager’s beliefs.

The above definition is best illustrated through and example: the following three internally consistent statements are informationally equivalent:

1. “asset 1 is expected to outperform asset 2 by 5% and asset 1 is expected to return 8%”
2. “asset 1 is expected to outperform asset 2 by 5% and asset 2 is expected to return 3%”
3. “asset 1 is expected to return 8% and asset 2 is expected to return 3%”

In terms of our notation, the above three views correspond to the following three triplets:

$$
P_1 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad v_1 = \begin{bmatrix} 5\% \\ 8\% \end{bmatrix}, \quad \Omega_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
$$

$$
P_2 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad v_2 = \begin{bmatrix} 5\% \\ 3\% \end{bmatrix}, \quad \Omega_2 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}
$$

$$
P_{rref} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad v_{rref} = \begin{bmatrix} 8\% \\ 3\% \end{bmatrix}, \quad \Omega_{rref} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}
$$

The following theorem shows that for any $(P, \Omega, v)$ triplet in a family which results in a given $\omega_{BL}$, the row reduced echelon form of the triplet is also in the family; the proof is presented in the Appendix.
**Theorem 1** Assume that the PM has internally consistent views. Assume that based on his \((P, \Omega, v)\) triplet, the PM’s BL allocation is \(\omega_{BL}\). Define \(B\) as some \((k \times k)\) invertible matrix. If \(P\) has full row rank, all allocations using the informationally equivalent triplet \((P_*, \Omega_*, v_*)\) of the form \((P_*, \Omega_*, v_*) = (BP, B\Omega B', Bv)\) lead to the same BL allocation \(\omega_{BL}\). A special case of such an informationally equivalent triplet is the row reduced echelon form triplet \((P_{\text{rref}}, \Omega_{\text{rref}}, v_{\text{rref}})\).

For the *Inverse Problem*, we seek to find an element of the triplet family which led to \(\omega_{BL}\) as our goal is to find the PM’s proprietary views, not the specific \((P, \Omega, v)\) triplet which he input. Thus, the above theorem allows us the flexibility of assuming a relatively simple structure on \(P\), namely its row reduced echelon form. We will discuss the merit of this later on in this section. Given the complicated expressions for \(\mu_{BL}\) and \(\Sigma_{BL}\) involved in the FOC in equation (16), the task of proving that a tractable closed form expression which disentangles the public information from the PM’s proprietary views exists seems daunting. [40] conjectures that such an expression exists and propose the one given below, yet they do not present a proof. Furthermore, to our knowledge, no other paper in the relevant literature mentions or proves this result. In what follows, for simplicity, we assume that the PM’s risk aversion \(\Delta\) is the same as the aggregate market risk aversion \(\Delta_{\text{mkt}}\); however, this assumption is not critical and a directly analogous result exists when the two risk aversions differ. Theorem 2 below presents a key decomposition of the BL weights; it is proved in the Appendix.

**Theorem 2** The BL weights given in equation (16) can be factored into the following form:

\[
\omega_{BL} = \frac{\omega_{eq} + P'm}{1+\tau}
\]

where the \((k \times 1)\) vector \(m\) is given by:

\[
m = \frac{1}{\Delta} \Sigma^{-1} (v - \frac{P\pi}{1+\tau}), \quad \Delta = \frac{\Omega}{\tau} + \frac{P\Sigma P'}{1+\tau} \tag{27}
\]

The \((k \times 1)\) vector \(m\) is the "tilt" vector. Each view ultimately results in an element in \(m\): stronger views result in larger (in absolute value) \(m\) elements and thus cause a larger deviation from equilibrium. The beauty of the factored form of the BL weights given above is that it is intuitive and forms the basis for our attempt to solve the *Inverse Problem*, in which \(P, v, \Omega, \) and \(\tau\) are unknown. The next section presents the inference algorithm created for solving the *Inverse Problem*. 

30
6 Solving the Inverse Problem

6.1 The Key Idea

Utilizing the factorization presented in Theorem 2 is the stepping stone for solving the Inverse Problem. However, it is a seemingly impossible task: even if we knew $\tau$, the system in equation (26) boils down to factoring an $(n \times 1)$ vector into the product of a $(n \times k)$ matrix and a $(k \times 1)$ vector:

$$x \equiv (1 + \tau) \left( \omega_{BL} - \frac{1}{1 + \tau} \omega_{eq} \right) = \mathbf{P}^t \mathbf{m} \quad (28)$$

To make matters worse, we don’t even know $k$: thus, we are factoring $x$ into a product of a rectangular matrix with another vector without even knowing what their dimensions are. Clearly, solving such a system is impossible as there are an infinite number of $(\mathbf{P}, \mathbf{m})$ solution pairs. Thus, we proceed by placing some relatively general and intuitive restrictions on the $\mathbf{P}$ matrix. We will discuss how to get a reasonable estimate of $\tau$ later on; for now, assume it is given and thus that $x$ is observed.

**Definition 3 Allowable $\mathbf{P}$ Matrices:** A $(k \times n)$ pick matrix is **allowable** if the following two conditions hold:

1. Each asset in the PM’s investment universe may be involved in at most one proprietary view. A proprietary view may involve an asset in a “long” or in a “short” position. Thus, if asset $j$ is involved in view $k$, then all entries in the $j$th column are zero except for the $k$th one.

2. If an asset is involved in a view, it must be involved as a whole unit. Thus, if asset $j$ is involved in view $k$, then the $(k, j)$ entry of the $\mathbf{P}$ matrix must be $+1$ or $-1$.

The following three $\mathbf{P}$ matrices illustrate the two definition conditions.

$$\mathbf{P} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \quad \text{allowable}$$

$$\mathbf{P} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix} \quad \text{not allowable}$$

$$\mathbf{P} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & -2/3 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1/4 \end{bmatrix} \quad \text{not allowable}$$
If we assume that the P matrix is allowable, then we can turn the ill defined system of equation (28) into an exactly identified system. Consider what happens when asset \( j \) is not involved in a view: then, the \( j^{th} \) column of \( P \) will be a zero vector. That means that the \( j^{th} \) row of \( P' \) will be a zero vector, which means that the \( j^{th} \) element of \( P'm \) is 0. Mathematically:

\[
\omega_{BL}^j = \frac{1}{1+\tau} \left( \omega_{eq}^j + 0 \right) = \frac{\omega_{eq}^j}{1+\tau}
\]  

Thus, if asset \( j \) is not involved in any views, then its BL weights should be exactly the same as its equilibrium weight, normalized by the confidence in the CAPM prior.

Now, consider the case where asset \( j \) is involved in a view \( k \) in a long (short) manner. Then, the \((k, j)\) element of \( P \) will be a +1 (-1); the rest of the entries in the \( j^{th} \) column will be zero. That means that the \( j^{th} \) row of \( P' \) will be a zero vector, except for the \( k^{th} \) column, which will be a +1 (-1). Thus, the \( j^{th} \) element of \( P'm \) will be \( +m_k \) \((-m_k)\), where \( m_k \) is the \( k^{th} \) element of the tilt vector \( m \). Mathematically:

\[
\omega_{BL}^j = \frac{1}{1+\tau} \left( \omega_{eq}^j \pm m_k \right)
\]  

All assets involved in the \( k^{th} \) view will be tilted by \( \pm m_k \): the sign depends on whether the view is long or short on the particular asset.

Thus, if there are \( l \geq k \) assets involved in \( k \) views, the system becomes an exactly identified \((k \times k)\) system if one restricts \( P \) to be allowable. Using the logic behind equations (29) and (30), each column of \( P' \) can be identified up to a sign: thus, there are \( 2^k \) allowable \( P \) matrices which we could infer from a given \( x \). Later on, we will prove that all \( 2^k \) lead to informationally equivalent solutions. After having identified an allowable \( P \), the system is exactly identified. This is because there are \((n - l)\) equations of the form \( 0=0 \) due to the fact that \((n - l)\) assets are not involved in any views. Furthermore, there are \((l - k)\) redundant equations as each asset in the same view has the same equation. Thus, there are \( k \) independent equations (they are independent since \( P \) is allowable and thus must have a rank of \( k \)) and \( k \) unknowns in \( m \) which means that there is a unique solution for \( m \).

By placing intuitive restrictions on \( P \), we have reduced the factorization in equation (28) to an exactly identified system of equations. Conditional on \( P \) being allowable, we can arrive at \( 2^k (P, m) \) pair solutions which are informationally equivalent. These restrictions on \( P \) are relatively loose and mesh well with how the BL model is used in practice. There are two justifications for the first condition of allowable \( P \) matrices:

1. Typically, the PM has views on a small number of assets compared to the size of his investment universe (i.e. \( k \ll n \)). Thus, an asset being involved in more than one view is not too common.

2. According to Theorem 1, the row reduced echelon form \((P_{ref}, \Omega_{ref}, v_{ref})\) triplet is always in the family of informationally equivalent solutions for a given \( \omega_{BL} \). Often times, a \( P \) matrix may not be allowable, yet its row reduced echelon form is allowable.
A good example of this is the situation in equation (5), where the first two \( P \) matrices are not allowable yet the row reduced echelon form is. Thus, the restriction of having an allowable \( P \) is really a restriction on at least one \( P \) matrix in a given family of informationally equivalent solutions being allowable. Clearly, the latter fact allows for a broader set of proprietary views by the PM.

As for justifying the second condition of allowable \( P \) matrices, recall that the proprietary views are statements on the random vector of expected asset returns \( \mu \). Thus, using different units of assets within the same view is less natural. For example, statements of the form "\( \mu_1 + 1/2\mu_2 = 3\% \)" are mathematically valid yet unintuitive given the setup. Note that as long as there exists a \( P \) matrix in the informationally equivalent set which satisfies:

\[
\text{mod}(P, c) = 0
\]  

for some scalar \( c \), then there exists a linear transformation of that matrix into a \( P \) matrix which satisfies the second condition of allowability. Note that the \( \text{mod}(.) \) operator above takes the modulo of each element in \( P \). According to the above, the view "\( 2\mu_1 + 2\mu_2 = 6\% \)" can be transformed to an informationally equivalent view "\( \mu_1 + \mu_2 = 3\% \)" which is consistent with condition 2 of the definition of allowability. If we solved the Inverse Problem, we would estimate the latter, which is informationally equivalent to the former.

Once we have identified \( m \), finding \( v \) is straightforward because there is a 1-1 transformation between the two: solving equation (27) for \( v \), we get:

\[
v = \Delta A m + \frac{P\pi}{1 + \tau}
\]  

The transformation is 1-1 since \( A \) is positive definite and \( P \) is full row rank. Note that assuming that \( P \) is full row rank is valid because of the first condition of allowability. Of course, up to now, we have not discussed how to get \( \Omega \); given the current setup, we will need to make a parametric assumption on it. This is discussed below in the key assumptions paragraph of the next section.

### 6.2 The Inference Algorithm for the No Noise Case

In this section, we present the inference algorithm we developed to solve the Inverse Problem and the assumptions needed for it to lead to a solution.

**Assumption 1** We use the same public information as the PM and define the “appropriate” investment universe.

**Assumption 2** The PM inputs the posterior returns distribution through a mean-variance optimizer in order to arrive at an allocation.
Assumption 3 For the \((P, \Omega, v)\) triplet that the PM inputs, there exists at least one informationally equivalent triplet whose \(P\) matrix is allowable.

Assumption 4 The elements in \(m\) are unique and nonzero: in other words, each view results in a distinct tilt magnitude from equilibrium.

Assumption 5 The view confidence matrix \(\Omega\) is given by:

\[
\Omega = P \tau \Sigma P'
\]

This form of \(\Omega\) is used heavily in the literature (see for example [50]) because \(\Omega\) is an unintuitive parameter matrix to set. This specification of \(\Omega\) gives more confidence to views which involve less volatile assets and vice versa. Intuitively, this makes sense: saying that Exxon-Mobil stock will return 5\% is a more confident view than saying that Dryships (a Nasdaq listed shipping company) will return 5\% as the latter is more volatile and thus a same magnitude view on it should give “less information”. In any case, this assumption is needed in order to reduce the number of unknowns. Nevertheless, it is intuitive, used by many in practice, and is still a parametric function of two unknowns (\(\tau\) and \(P\)). It should be noted that this assumption can be generalized to any \(\Omega\) which is a function of known parameters and (possibly) \(\tau\) and \(P\). However, we use the parametrization in equation (33) due to its intuition and widespread use.

Assumption 6 The risk aversion \(\Delta\) of the PM is the same as the market risk aversion \(\Delta_{mkt}\). Alternatively, assume that \(\Delta\) is known. Below, we proceed as if \(\Delta_{mkt} = \Delta\); the algorithm can easily accommodate having \(\Delta_{mkt} \neq \Delta\) and can be straightforwardly modified accordingly.

Given the above assumptions, we proceed to present the inference algorithm and then describe how each step is implemented: next to each step, we list what parts of the public information and/or what variables estimated in previous steps we need.

<table>
<thead>
<tr>
<th>Algorithm Steps</th>
<th>Data Needed at Each Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>Estimate the confidence in the prior (\tau) (\omega_{BL}, \omega_{eq})</td>
</tr>
<tr>
<td>Step 2</td>
<td>Estimate the number of views (k) (x \equiv (1 + \tau) \left(\frac{\omega_{BL} - 1}{1 + \tau} \omega_{eq}\right))</td>
</tr>
<tr>
<td>Step 3</td>
<td>Identify the (P) matrix (x, k)</td>
</tr>
<tr>
<td>Step 4</td>
<td>Identify the tilt vector (m) (x, k, P)</td>
</tr>
<tr>
<td>Step 5</td>
<td>solve for (v = \Delta Am + \frac{P\pi}{1 + \tau}) (P, \tau, \pi, \Sigma, \Delta)</td>
</tr>
</tbody>
</table>

Table 1: The 5 steps of the inference algorithm and the data needed for each step.
For now, assume that we have estimated \( \tau \). The next step is to estimate the number of views that the PM had. Recall that all assets which are not involved in a view have a 0 in the corresponding entry in \( x \). Furthermore, all assets in the same view must have the same absolute value in \( x \) because of the fact that \( P \) is allowable. Thus, estimating the number of views boils down to grouping the elements of \( |x| \) and counting the number of resulting groups with non-zero entries. This is represented graphically in Figure 2 below. After estimating the number of views \( k \), we must build the columns of \( P' \): each column corresponds to one view. For a given view \( k \), we find the indices of the elements of \( x \) which are in group \( k \). For every asset in that group, input a +1 in the corresponding entry of the column if its entry in \( x \) is positive and a -1 if the entry is negative. For all other entries of that \( P' \) column, input a zero. The process is illustrated in Figure 3. Having estimated all columns of \( P \), solving the resulting \((k \times k)\) system for \( m \) is straightforward: after obtaining \( m \), \( v \) can be found by using equation (32).

If the assumptions presented in the beginning of this section hold, then following the inference algorithm is guaranteed to result in at least one \((P, m)\) solution pair. The following theorem formalizes this and is proven in the appendix.

**Theorem 3** Assume that the assumptions in the beginning of the section hold. Then, there exists at least one solution pair \((P, m)\) to the factorization problem \( x = P'm \) if and only if:

1. Every element \( x_i \) of the \( x \) vector belongs to the set \( S \):

\[
x_i \in S, \quad \text{where} \quad S = \{0, \pm m_1, \pm m_2, \ldots, \pm m_k\}, \quad m_i \neq m_j \neq 0 \quad \forall i \neq j \quad (34)
\]

2. Each nonzero member of the set \( S' \), which contains the absolute values of the elements of \( S \), is present in \( |x| \).

Figure 2: The above figure gives a graphical depiction of the procedure of estimating \( k \). We look at the absolute value of the entries in \( x \) and count the number of distinct nonzero entries.

\[
|x| = \begin{bmatrix}
m_1 \\
m_1 \\
m_1 \\
m_1 \\
m_2 \\
m_2 \\
0
\end{bmatrix}
\]

Count number of distinct nonzero entries in \( |x| \)

Estimated number of views \( k=2 \)
If one follows the convention for generating the $P'$ columns above, the resulting $m$ vector will have only positive elements. In fact, one can follow any sign convention and arrive at informationally equivalent results, as long as it is consistent within a group (i.e. all elements in a group of $x$ of the same sign will get the same sign in the $P'$ column and all elements in a group of $x$ with different signs will get different signs in the $P'$ column). Furthermore, the ordering of the $P'$ columns does not matter. The following theorem formalizes this and is proved in the appendix.

**Theorem 4** All possible $(P, \Omega, v)$ solutions resulting from following the inference algorithm under different sign convention and/or different group ordering are informationally equivalent.

The inference algorithm is quite straightforward if we know what $\tau$ is. However, if we were to implement the inference algorithm in practice, we would not have access to the PM's $\tau$ and thus, we would have to get some sort of estimate of it. The issue of estimating $\tau$ is discussed in Section 6.4.

### 6.3 The Inference Algorithm when $x$ is Noisy

In the real world, we would expect the $x$ vector to be noisy because the PM may not be able to obtain exactly the BL proportions he desires due to market frictions such as trading costs and/or liquidity problems. Thus, the deviations from equilibrium for each asset may be corrupted. Another part where noise shows up is in the estimation of the market equilibrium weights $\omega_{eq}$. Overall, due to the noise, we would not expect to find duplicate elements in $|x|$. This makes it more challenging to estimate both the number of views and to partition $|x|$ into view groups. However, if the PM were a BL investor, we would expect to see clusters in $|x|$. Thus, in the noisy case, estimating $k$ and grouping the elements becomes a "clustering problem". Figure 4 depicts a real $|x|$ vector from the iShares Small Growth Fund (July 2009 holdings). This is one of the mutual funds in Section 8 whose holdings we apply the inference algorithm on. The x-axis is the deviation from equilibrium of each asset and the plot is a 1-dimensional plot. The inference algorithm requires us to group the assets into $k$ view groups, where $k$ is unknown. We must first decide which assets have 0 (or close to zero) deviation from equilibrium; in the figure, this is the left most cluster without a circle around it. Then, for the "non-zero view" elements, we must run an unsupervised algorithm to determine both the number of clusters and which elements belong to each of the clusters. This is harder than just running an unsupervised "clustering" algorithm because a view may only involve 1 or 2 assets whereas another one may involve many. Furthermore, the overall number of data-points can range from tens to thousands, yet clustering algorithms typically are designed for datasets which have millions of data points and each cluster has a comparable number of elements (biology applications). Nevertheless, so as to preserve generality, we proceed without making any structural assumptions on the noise except that
Figure 3: The above figure graphically depicts the procedure for generating the $P'$ matrix. We group the nonzero entries of $|x|$ into $k$ groups, where $k$ is estimated from step 2 of the algorithm. Then, for each group, a column of $P$ is generated in the following manner: for every asset in a given group, we input a +1 in the corresponding entry of the column if its entry in $x$ is positive and a -1 if its entry is negative. For all other entries of the column, we input a zero.

It is zero mean. Furthermore, we prove a “consistency” theorem which links Inverse Problem solutions when the number of clusters is misestimated.

The main approach we take to partitioning $|x|$ involves looping over a set $K$ of possible cluster numbers. For each $k \in \{2 \ldots k_{max}\}$, we cluster the data according to the K-means algorithm and compute a cluster validity index; we pick the $k$ which maximizes this index and keep the corresponding partition. After obtaining the partition, we find the center of each cluster by averaging the elements in that cluster; we then assign each element of $|x|$ the value of its cluster center; we call this new vector $|\mathbf{Y}|$. Finally, we create the vector $y$, which is defined as:

$$y = \text{sgn}(x) |\mathbf{Y}|$$  \hspace{1cm} (35)
Figure 4: Scatter plot of the elements in $|x|$ vector from the iShares Small Growth Fund July 2009 holdings. The x-axis is in (%) units.

where $\text{sgn}(x)$ is an $(n \times n)$ diagonal matrix whose $(i, i)$ element is the sign of $x_i$. Note that $y$, satisfies the two conditions in Theorem 3 and thus, steps 3-5 of the inference algorithm can be applied as in the no-noise case.

The cluster validity index which we used is a variant of the silhouette statistic.\textsuperscript{15} This validity index was chosen because it outperformed other validity indices in simulation, such as the Calinski-Harabasz index ([26]), the average consensus statistic ([53]), and the average fuzziness index ([14]).

For observation $|x_i|$, let $a(i)$ be the average distance to other points in its cluster, and $b(i)$ be the average distance to points in the nearest cluster besides its own, where nearest is defined by the cluster minimizing this average distance. Then the silhouette statistic for that element $s(i)$ is defined by:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Note that the silhouette statistic is between -1 and 1 by definition; element $i$ is well clustered if $s(i)$ is close to 1. We modify the above definition of the silhouette statistic slightly for any element in $|x|$ which is clustered alone in order to alleviate the bias introduced by the fact that elements clustered alone get a +1 when using the original silhouette statistic formula in equation (36). Clearly, this is undesirable as it will bias our estimate $\hat{k}$ upwards.

\textsuperscript{15}For a good discussion of K-means clustering and the silhouette statistic, see [52] and [43].
For an element \(|x_i|\) that is clustered alone, define as \(d_{\text{left}}\) as the distance to its closest neighbor on the left and \(d_{\text{right}}\) as the distance to its closest neighbor to the right. If \(|x_i|\) is not the smallest or largest element in \(|x|\), we define its silhouette statistic as:

\[
s(i) = \min \left\{ 1, \min \left\{ \frac{d_{\text{left}}, d_{\text{right}}}{\|x_i\| - \text{mean}(d_{\text{left}}, d_{\text{right}})} \right\} \right\}
\]  

(37)

If \(|x_i|\) is clustered alone and happens to be the leftmost element, we define its silhouette statistic as:

\[
s(i) = \min \left\{ 1, \frac{d_{\text{right}}}{\text{mean}(|x_j|)} \right\}
\]  

(38)

Analogously, if \(|x_i|\) is clustered alone and happens to be the rightmost element, we define its silhouette statistic as:

\[
s(i) = \min \left\{ 1, \frac{d_{\text{left}}}{\text{mean}(|x_j|)} \right\}
\]  

(39)

The optimal number of clusters \(k\) is that for which the resulting partition has the highest average \(s(i)\) over the data set:

\[
\hat{k} = \arg \max_{k \in K} \left( \frac{\sum_{i=1}^{n} s^k(i)}{n} \right)
\]

(40)

Note that \(s^k(i)\) is not defined for the \(k = 1\) cluster case. However, in practice, this is ok because one cluster would mean that either all the PM's assets are involved in one view or that he has no view on any of the assets; furthermore, a visual inspection of the data can easily pick up such cases.

Given the intricacies of deciding the number of clusters in \(|x|\), there is a chance of misestimating the number of clusters. The following theorem provides a consistency result between two \((P, m)\) factorizations which differ in the number of clusters.

**Theorem 5** Assume that the noisy vector \(\tilde{x}\) is observed. Consider two different factorizations \((P_k, m_k)\) and \((P_1, m_l)\) of \(\tilde{x}\) resulting from estimating two different view numbers \(k\) and \(l\) respectively, where \(l > k\). Define the resulting views vectors as \(v_k\) and \(v_l\) respectively. Define as \(L\) the set of all \((k \times l)\) matrices which satisfy the following 3 conditions:

1. All entries are 0 or +1.
2. Each column of \( A \in L \) has exactly one +1.

3. Each row must have at least one +1.

Then, there exists a matrix \( A \in L \) such that \( P_k = AP_1 \). Furthermore:

1. If \( \tilde{x} \) is not corrupted by noise, then \( v_k = Av_l \).

2. If \( \tilde{x} \) is corrupted by noise, then \( E(\Lambda v_l - v_k) = 0 \).

The above theorem basically says that if the number of views is overestimated, and if one knows the matrix \( \Lambda \), which "sums" the rows of \( P_1 \) in such a way so as to get \( P_k \) then one can reconstruct the correct views vector \( v_k \) by applying the same matrix \( \Lambda \) to \( v_l \). This means that overestimating the number of views does not cause a loss of information. There is a consistency between the various \((P, v)\) solutions of different dimensions. Note that such a \( \Lambda \) matrix is guaranteed to exist because the \( P \) matrices are allowable.

6.4 Estimating \( \tau \)

Recall that for the BL model to make economic and statistical sense, \( \tau \) must be smaller than 1. If there were no noise, one could loop over \( x(\tau) \) and keep all \( \tau \) which lead to a number of groups in \( x(\tau) \) (including the "no view" group) less than or equal to the number of assets \( n \). However, this is clearly not an optimal strategy as it may not lead to a unique \( \tau \). Furthermore, in practice, there is noise in \( x \) and thus the above approach doesn’t make sense. In theory, we can only pinpoint a unique \( \tau \) if we add an additional restriction to the type of views that the PM can have. To understand what type of restriction we would need, observe that:

\[
I'\omega_{BL} = \frac{1}{1 + \tau} (1 + I'P'm) = \frac{1 + \sum_{c=0}^{k} n_c m_c}{1 + \tau} \equiv \frac{1 + z}{1 + \tau} \tag{41}
\]

where the scalar \( z \) is the "net bullishness" of the overall views of the PM. If view \( k \) is a long (short) biased view, then \( I'P'_k = n_k \), where \( n_k \) is the sum of the elements in the \( k_{th} \) column of \( P' \), which is positive (negative). If view \( k \) is a portfolio view (i.e. same number of long positions as short positions), then \( I'P'_k = 0 \). Thus, intuitively, the quantity \( z \equiv \sum_{c=0}^{k} n_c m_c \) is the net bullishness of the PM. If \( z = 0 \), then we could solve equation (35) for \( \tau \):

\[
\tau = \frac{1}{I'\omega_{BL}} - 1 \tag{42}
\]
Of course, if the BL weights sum to more than 1, we know that the net bullishness cannot be 0. If they sum to less than one, then we can’t differentiate between a higher \( \tau \) and a more negative net bullishness given the information set that we have. Fortunately, in practice, we have a prior about a PM’s investment style focus (e.g. Large Cap stocks vs Small Cap Stocks or High Value stocks vs Low Value stocks): we can use this information to justify a practical way of estimating \( \tau \) and validating the result.

The idea for estimating \( \tau \) stems from the following piece of intuition: there is probably an aggregation of the PM’s portfolio into a 2x1 vector such that it is quite likely that he has only one long-short view. For example, take a portfolio and classify each element as a Large Cap or Small Cap stock. Do the same for the market portfolio and look at the difference. If for some \( \tau \in [0,1] \) the resulting difference looks like:

\[
x_2(\tau) = \begin{bmatrix} +a + \varepsilon_1 \\ -a + \varepsilon_2 \end{bmatrix}, \quad \text{where} \quad 1'x_2(\tau) < 0
\]  

(43)

where \( \varepsilon_1 \) and \( \varepsilon_2 \) are relatively small, then one can convincingly make the argument that:

\[
P_2' = \begin{bmatrix} +1 \\ -1 \end{bmatrix}, \quad m = a + \frac{\varepsilon_1 + \varepsilon_2}{2}, \quad \text{and} \quad 1'P_2'm = 0
\]  

(44)

meaning that it is valid to search for a \( \tau^* \) in the vicinity of \( \tau \) such that \( |1'x_2(\tau^*)| \) is minimized. Proceeding more formally, denote by \( x_n(\tau) \) the original \( x \) vector of portfolio holdings. One can create an arbitrary 2x1 aggregated portfolio by left-multiplying \( x_n(\tau) \) by a \((2xn)\) matrix \( \Lambda \in L \) where the family \( L \) is defined as in Theorem 5. Thus:

\[
x_2(\tau) = \Lambda x_n(\tau)
\]  

(45)

Essentially, we want to find a \((2xn)\) matrix \( \Lambda \in L \) and a \( \tau \) such that the entries in \( x_2(\tau) \) are of opposite sign yet of similar magnitude: then, we can make the argument in equations (43) and (44), and then find a \( \tau^* \) in the vicinity of \( \tau \) such that \( |1'x_2(\tau^*)| \) is minimized.

If there exists a region \( T \subseteq [0,1] \) of \( \tau \) value such that for all \( \tau \in T \), there exist at least two elements \( x_n^i(\tau) \) and \( x_n^j(\tau) \) in \( x_n(\tau) \) such that \( x_n^i(\tau)x_n^j(\tau) < 0 \), then there exists at least one \( \Lambda \in L \) such that \( x_2^i(\tau)x_2^j(\tau) < 0 \). One could then do the following minimization and get an estimate of \( \tau^* \):

\[
\min_{\Lambda \in L, \tau \in T} |1'\Lambda x_n(\tau)|
\]  

(46)

The closer the objective function gets to 0, the more confident we are about our estimate \( \tau^* \). Note that the above approach will yield a negative \( \tau^* \) if \( 1'x_2(\tau) > 0 \). Thus, a necessary condition for getting a non-negative \( \tau \) is to have a non-negative investment in the risk-free asset so that \( 1'x_2(\tau) \leq 0 \). If this is the case, then the minimization in equation (46) can be simplified to:

\[
\min_{\Lambda \in L, \tau \in T} -1'\Lambda x_n(\tau)
\]  

(47)
In any case, minimizing the argument in equation (47) is an NP-Hard problem. However, in practice, we typically have some idea on which $A$ matrices to try because as mentioned before, we typically have some idea of the PM’s investment strategy. Thus, the set of $A$ matrices we would need to search over is significantly reduced.


7 Simulations

In this section, we explore the robustness of our inference algorithm by running simulations. Given the problem setup, there are three main sources of noise which can corrupt $x$ in practice:

1. The PM may not be able to obtain exactly the BL weights he desires due to market frictions such as trading costs and/or liquidity problems.

2. The estimation of the market equilibrium weights may be noisy due to a misspecification of the PM’s investment universe.

3. The $\tau$ estimate may be noisy.

Our aim is to examine the robustness of steps 2-5 (see Table 1) of our inference algorithm: in other words, given an observed noisy $\tilde{x}$, how does each estimation step perform? Recall that from a single vector, we need to estimate the number of views $k$, the assets involved in each view, and the magnitude of each view. Since this is a sequential procedure, mistakes in a given stage are carried through. To build some intuition about the nature of what could go wrong, consider the following possible errors:

1. The number of views could be overestimated or underestimated: thus, the dimensionality of the problem is corrupted. In the “overestimation” case, some views will involve fewer assets and extra views are “created” that the PM does not have: for example, having the original view of “the portfolio that is long 1 unit of Apple, long 1 unit of IBM, short 1 unit of Exxon-Mobile, and short 1 unit of BP, will return 5%” is quite different from having the 2 views “the portfolio that is long 1 unit of Apple, long 1 unit of IBM, and short 1 unit of Exxon-Mobile will return 8%” and “the portfolio that short 1 unit of BP will return -3%”. Thus, if the PM has a pairwise sector bet (i.e. the original view), estimating two views which are directional and of which the second is asset specific is clearly not desirable from a practical point of view.

In the “underestimation” case, assets involved in different views are lumped into one view: for example, the PM may have two absolute views on two assets based on his analysis. Clearly, information about the PM’s input is lost if the two distinct views are lumped into one.

2. Even if the number of views is estimated correctly, the assets involved in a given view may be misspecified. For example, an estimated view may end up having either “additional” or “omitted” assets.
3. Even if both the number of views and the assets involved in each view are estimated correctly, the magnitude of each of the views is vulnerable to estimation error due to the noise in \( \tilde{x} \). Although undesirable, this is less of a problem because the dimensionality of the problem is not affected. Essentially, errors in \( \tilde{x} \) are filtered through the \( A \) matrix (see equation (32)). Since in this scenario, \( P \) is assumed to be estimated correctly, the magnitudes of the errors in \( v \) depend purely on the magnitudes of the errors in \( \tilde{x} \) vis a vis the errors in \( m \) resulting in step 4 of the algorithm.

In the simulations we present, we aim to understand the errors introduced in each of the three possible stages described above; we will present 3 simulation cases of varying complexity. In our simulations, we use a portfolio of 25 assets; the reason for using 25 assets is that in Section 8, we apply the inference algorithm to real mutual fund portfolios where we aggregate a fund’s holdings into 2x3 and 5x5 Fama-French portfolios. In each simulation, we start with a set of public information, a \( P \) matrix, and a \( \tau \). Then, we follow the steps below:

1. We vary the "strength" of the views vector input \( v \) from \((P\pi - 2\%)\) to \((P\pi + 2\%)\) in increments of 0.1\%, resulting in 41 different view vector inputs.

2. For each of these 41 view vectors \( v \) that we input, we create the corresponding \( x \) vector of equilibrium deviations.

3. For each of these 41 \( x \) vectors, we generate 2000 noisy observations \( \tilde{x} \) which are corrupted by adding a zero mean Gaussian noise random vector. We use a “low” noise level of standard deviation 1\% and a “high” noise level of standard deviation 2\%.

4. For each of the 2000 noisy \( \tilde{x} \) vectors resulting from each of the 41 \( x \) vectors, we run our inference algorithm. For each of the 41 different PM view vector inputs, we report the following:

   a) The average and median error in the estimation of the true number of views \( k \) over the 2000 trials.

   b) The average and median number of \( P \) rows that are identified correctly over the 2000 trials.

   c) For each of the 2000 trials where both the number of views \( k \) and the \( P \) matrix are estimated correctly, we calculate the distance between the estimated views magnitudes \( v^{\text{hat}} \) and the input views vector \( v^{\text{true}} \), where the distance measure is defined as:

\[
    d(v^{\text{hat}}, v^{\text{true}}) = \frac{\sum_{i=1}^{k} |v_i^{\text{hat}} - v_i^{\text{true}}|}{k}
\]  

(48)
We report the average and the median of this distance. It is worth noting that it would not make sense to report the performance of the \( v \) estimation statistics for the case where not all rows of \( P \) are estimated correctly: this is because \( v \) is a linear function of \( m \) (see equation (32)) and the matrix \( A \) is a function of \( P \Sigma P' \). Thus, if a set of rows of \( P \) are misestimated, this will corrupt not only the magnitude of the corresponding views but also the magnitude of the views corresponding to the rows of \( P \) that have been estimated correctly. The degree of this corruption will depend on the matrix \( \Sigma \). Therefore, it makes sense to quantify the distance measure proposed in equation (48) only if the estimated \( P \) is fully correct.

In the following 3 sections, we present and discuss the results for each of the three simulation cases in order of increasing complexity.

### 7.1 Case 1: One Portfolio View

In this section we present the simplest of the three simulation cases that we explored. It involves one portfolio view on 4 assets: more specifically, the view is of the form:

\[
a_1 - a_5 + a_{21} - a_{25} = v
\]

where \( a_i \) refers to asset \( i \). Figure 5 plots the mean and median estimated number of views for each of the 41 equilibrium deviations \( v - P \pi \) and under each of the two noise scenarios (low and high). Overall, the inference of \( k \) is very accurate: the larger the deviation of the views from equilibrium, the more accurate our estimate is. The number of views is estimated to be 1 for all 2000 trials for equilibrium view deviations above (in absolute value) 0.7% for the low noise case and 1.4% for the high noise case. This is very promising and underscores the robustness of Step 2 of our algorithm in this simple case. In practice, most of the PM’s views will deviate significantly from equilibrium: being able to pick up views that deviate so little from the equilibrium is a very significant result.

Figure 6 depicts the results for identifying the \( P \) matrix. More specifically, for each of the 41 equilibrium deviations \( v - P \pi \), we plot the mean and median number of the rows of \( P \) that are estimated correctly: we do the above for both the low and high noise cases. Again, the performance is very strong: the \( P \) matrix is estimated correctly 90% of the time for equilibrium view deviations above 0.3% (in absolute value) for the low noise case and 0.7% (in absolute value) for the high noise case. Notice that for some small equilibrium view deviations, although the number of views is estimated correctly, the \( P \) matrix may not be estimated correctly. This happens when clusters are close by and some elements may jump from one cluster to another without altering the estimate of the number of clusters.

Figure 7 depicts the distance between the estimated and true \( v \) vectors for all the simulation trials where the number of views and the \( P \) matrix were estimated correctly. More
specifically, for each of the 41 equilibrium deviations \( v - P\pi \), for each of the trials where \( k \) and \( P \) are estimated correctly, we plot the mean and the median of the distance between the estimated and true \( v \): we do the above for both the low and high noise cases. Overall, the mean and median errors are mostly below 0.02% for the low noise case and below 0.06% for the high noise case. Note that there is a “gap” in the plot for the 0% and ±0.1% equilibrium view deviation points: this is because for such small deviations, either \( k \) and/or \( P \) were misestimated for all of the 2000 trials. Thus, we conclude that the last step of the algorithm performs very well if \( k \) and \( P \) are estimated correctly.

Overall, it seems that the noise in \( \tilde{x} \) has the potential to severely impact estimating \( k \); however, if \( k \) is estimated correctly, then \( P \) will be estimated correctly unless the noise corrupts some points so much that they actually switch clusters without affecting the number of clusters. If steps 2-3 of the algorithm go well, then essentially, the noise in \( \tilde{x} \) is transferred directly to \( m \). Recall that each element in \( m \) is a “tilt” from equilibrium due to a particular view; in our inference algorithm, within a group \( k \) of \( |\tilde{x}| \), we average its elements in order to get \( m_k \). If the noise is zero mean, then on average, \( m_k \) is unbiased; thus, the more elements a group has, the more likely it is that the estimated \( m_k \) is close to the true value. The vector \( m \) is then filtered by the matrix \( A \) (step 5 of the algorithm), which essentially is a covariance matrix. Given that the elements of \( A \) are much smaller than 1, small noises in \( m \) result in even smaller noises in \( v \). This is exactly why the errors in \( v \) in Figure 7 are on the order of hundredths of a percent compared to the order of the noise we added to \( x \) which is on the order of 1%.

![Estimated k per Equilibrium View Deviation](image)

Figure 5: Mean and median estimates of the number of views for each of the 41 equilibrium deviations \( v - P\pi \) and under each of the two noise scenarios (low and high) for Case 1.
Figure 6: Mean and median estimates of the number of the rows of the $P$ matrix that are estimated correctly for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 1.

Figure 7: Mean and median estimates of the distance between the estimated and true $v$ vectors for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 1.
7.2 Case 2: Three Views

In this section, we present a more complex example where the PM has three views which are not all portfolio views and which involve different numbers of assets:

\[ -a_{21} + a_{25} = v_1 \]
\[ +a_{16} + a_{17} - a_{20} = v_2 \]
\[ +a_1 = v_3 \]  

Figure 8 plots the mean and median estimated number of views for each of the 41 equilibrium deviations \( v - P\pi \) and under each of the two noise scenarios (low and high). As was the case in Case 1, the larger the deviation of the views from equilibrium, the more accurate our estimate of \( k \) is. The number of views is always estimated correctly for equilibrium deviations above 0.7% (in absolute value) for the low noise case and for deviations above 1.4% (in absolute value) for the high noise case. Although there are now three views instead of just one, the performance is as robust as that of Case 1. Clearly, the clustering algorithm we use in Step 2 is able to handle more complex cluster structures; this is a testament to our modified Silhouette Statistic criterion for selecting the number of clusters (see Section 6.3).

Figure 9 depicts the results for identifying the \( P \) matrix. More specifically, for each of the 41 equilibrium deviations \( v - P\pi \), we plot the mean and median number of the rows of \( P \) that are estimated correctly: we do the above for both the low and high noise cases. The \( P \) matrix is always estimated correctly for equilibrium deviations above 0.7% (in absolute value) for the low noise case and above 1.4% (in absolute value) for the high noise case.

Figure 10 depicts the distance between the estimated and true \( v \) vectors for all the simulation trials where the number of views and the \( P \) matrix were estimated correctly. More specifically, for each of the 41 equilibrium deviations \( v - P\pi \), for each of the trials where \( k \) and \( P \) are estimated correctly, we plot the mean and the median of the distance between the estimated and true \( v \): we do the above for both the low and high noise cases. Overall, the mean and median errors are mostly below 0.04% for the low noise case and below 0.08% for the high noise case. Note that the "gap" in the plots for points near an equilibrium deviation of 0% is due to the fact that either \( k \) and/or \( P \) were misestimated for all of the 2000 trials: compared to Case 1, the gap covers the same (small) band because our \( k \) and \( P \) estimates are as accurate as in Case 1. As was the case in Case 1, the last step of the algorithm performs very well if \( k \) and \( P \) are estimated correctly.

Overall, although this case has added complexity compared to Case 1, the performance of our algorithm is as robust.
Figure 8: Mean and median estimates of the number of views for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 2.

Figure 9: Mean and median estimates of the number of the rows of the $P$ matrix that are estimated correctly for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 2.
Figure 10: Mean and median estimates of the distance between the estimated and true $v$ vectors for each of the 41 equilibrium deviations $v - P\pi$ and under each of the two noise scenarios (low and high) for Case 2.
7.3 Case 3: Six Views

In this section, we present an even more complex example where the PM has six views which are not all portfolio views and which involve different numbers of assets:

\[ v_1 = a_4 + a_5 \]
\[ v_2 = a_6 - a_7 - a_8 - a_9 - a_{10} \]
\[ v_3 = a_{11} - a_{14} \]
\[ v_4 = a_{16} - a_{17} - a_{18} - a_{20} \]
\[ v_5 = a_{22} \]
\[ v_6 = a_{21} - a_{24} - a_{25} \] (51)

Figure 11 plots the mean and median estimated number of views for each of the 41 equilibrium deviations \( v - \Pi \pi \) and under each of the two noise scenarios (low and high). As expected, the larger the deviation of the views from equilibrium, the more accurate our estimate of \( k \) is. Although the median estimate of the number of views is equal to 6 for equilibrium deviations above 0.3% (in absolute value) for the low noise case and 0.6% (in absolute value) for the high noise case, the average remains elevated above 6 even for equilibrium view deviations of 2% in the high noise case. This indicates that there are some outlier cases where the number of views are misestimated: in any case, the trend shows that if deviations get larger towards 3%, the number of views will always be estimated correctly.

Figure 12 depicts the results for identifying the \( \Pi \) matrix. More specifically, for each of the 41 equilibrium deviations \( v - \Pi \pi \), we plot the mean and median number of the rows of \( \Pi \) that are estimated correctly: we do the above for both the low and high noise cases. As the equilibrium view deviation increases, the fidelity of our \( \Pi \) matrix estimate increases. Although the median number of correctly estimated \( \Pi \) rows settles at 6 for equilibrium deviations above 0.6% (in absolute value) for the low noise case and 1.1% (in absolute value) for the high noise case, the average remains slightly below 6 even for equilibrium view deviations of 2%. Part of this phenomenon is due to the fact that if for a certain trial, the number of views is misestimated, then not all 6 rows of \( \Pi \) can be estimated correctly. In any case, the trend shows that if deviations get larger towards 3%, the \( \Pi \) matrix is always be estimated correctly.

Figure 13 depicts the distance between the estimated and true \( v \) vectors for all the simulation trials where the number of views and the \( \Pi \) matrix were estimated correctly. More specifically, for each of the 41 equilibrium deviations \( v - \Pi \pi \), for each of the trials where \( k \) and \( \Pi \) are estimated correctly, we plot the mean and the median of the distance between the estimated and true \( v \): we do the above for both the low and high noise cases. Overall, the mean and median errors are mostly below 0.1% for the low noise case and below 0.2% for the high noise case. Note that the “gap” in the plots for points near an equilibrium deviation of 0% is larger than in the previous two cases (compare to Figures 7 and 10); this is merely a reflection of the fact that \( k \) and/or \( \Pi \) were misestimated for all of the 2000 trials in a larger band around 0% equilibrium deviation. However, as was the case in the previous 2 cases, the last step of the algorithm performs very well if \( k \) and \( \Pi \) are estimated correctly.
Figure 11: Mean and median estimates of the number of views for each of the 41 equilibrium deviations \( v - P\pi \) and under each of the two noise scenarios (low and high) for Case 3.

Figure 12: Mean and median estimates of the number of the rows of the \( P \) matrix that are estimated correctly for each of the 41 equilibrium deviations \( v - P\pi \) and under each of the two noise scenarios (low and high) for Case 3.
Figure 13: Mean and median estimates of the distance between the estimated and true \( \mathbf{v} \) vectors for each of the 41 equilibrium deviations \( \mathbf{v} - \mathbf{P}\pi \) and under each of the two noise scenarios (low and high) for Case 3.
8 Applying the Inference Algorithm to Mutual Fund Holdings

In order to test the performance of the inference algorithm on real data, we apply it to various mutual fund holding data obtained from Morningstar. However, we do not apply it at the individual stock level; rather, we apply it at a Fama-French (FF) portfolio level. Thus, our goal is to estimate the proprietary views of the PM on the FF portfolios: we focus on the 5x5 FF portfolios. There are two main reasons why we chose to focus on the FF portfolios:

1. Running the inference algorithm at an individual equity level would return views on individual stocks. However, it would be much more interesting to measure whether a PM has views on FF factors such as value or size because such funds typically self proclaim themselves as “Large Cap Growth”, “Mid Cap Value”, etc.

2. A typical equity portfolio manager holds between 100-500 stocks. However, the appropriate investment universe for a US equities PM has more than 7000 stocks. Thus, in practice, we would observe that the PM would have many assets with a 0 weight, which would be considered a deviation from the market equilibrium if interpreted strictly under the BL paradigm. However, in reality, there are numerous reasons why a PM may invest in much fewer stocks compared to the size of his investment universe. Some of the key reasons include limited time, attention, and liquidity.

8.1 Data Description and FF Portfolio Formation

We present the results for two mutual funds with different stated investment strategies. For each fund, we ran the inference algorithm on their portfolio holdings twice a year since 2005: thus, we were able to track each PM’s views over time. We chose funds that have a relatively large number of holdings and report those holdings on a regular basis. Furthermore, we chose funds which invest only in US equities and “risk-free” securities (such as money market funds) so that we can readily define the FF factors. Table 2 presents the two funds we analyzed along with their stated investing style and the average number of stock holdings over our sample period.

In order to organize the market and mutual fund stock level portfolios into FF portfolios, we used the market cap data from the CRSP database and the book value data from the Compustat database. We defined the FF factors exactly as in [34]. At a given time, for a given stock in a portfolio, we use the respective FF cutoffs for that time in order to classify the stock as belonging to one of the FF portfolios: the PM’s weight in a given FF portfolio is defined as the sum of the portfolio weights of the stocks in the mutual fund which fell into that particular FF portfolio category as defined by the FF cutoffs for that period. Thus, at a given time, for each fund, we obtained a 25x1 (6x1) vector for the 5x5 (2x3) FF portfolio case. We performed the analogous procedure to determine the market equilibrium weights at each time. In order to organize the market and mutual fund stock level portfolios into FF
portfolios, we used the market cap data from the CRSP database and the book value data from the Compustat database. We defined the FF factors exactly as in [34]. If the portfolio date was between July-December of year \( t \), we used the (market equity) ME at the end of June of year \( t \): furthermore, to get the book to market (BTM), we used the book value for the fiscal year ending in year \( t-1 \) and divided it by the ME at the end of December of year \( t-1 \). If the portfolio date was between January-June of year \( t \), we used the (market equity) ME at the end of June of year \( t-1 \): furthermore, to get the book to market (BTM), we used the book value for the fiscal year ending in year \( t-2 \) and divided it by the ME at the end of December of year \( t-2 \).

In order to estimate \( \Sigma \) and \( \Delta_{mkt} \) for the 2x3 and 5x5 FF portfolios, we used the reported returns series on Ken French’s website. We used monthly data for a 6 year rolling window starting 72 months before the portfolio date. The vector \( \pi \) was estimated by using equation (3). Each PM’s tau for a certain portfolio date was estimated according to the methodology presented in Section 6.4: the result is presented in Figure 14.

<table>
<thead>
<tr>
<th>Fund Name</th>
<th>Fund FF Strategy</th>
<th>Average Number of Assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fidelity Mid Cap Value Fund</td>
<td>Mid Cap Value Stocks</td>
<td>114</td>
</tr>
<tr>
<td>GMO US Core Equity Fund</td>
<td>Core Stocks</td>
<td>266</td>
</tr>
</tbody>
</table>

Table 2: Summary of the two mutual funds we analyzed.

In terms of annotating the results, we use letters to indicate the size and numbers to indicate the value of a portfolio: this is graphically depicted in the left panel of Figure 14. Thus, for example, in the 2x3 case, \( b_3 \) indicates a relatively small core stock (i.e. size is between 20\(^{th}\) and 40\(^{th}\) percentile, value is between 40\(^{th}\) and 60\(^{th}\) percentile).

For a given fund, the PM may have a different number of views involving different assets at different times. In our notation, this means that the P matrix has different dimensions.
and/or different rows at each time. Clearly, reporting a sequence of \((P, v)\) results is not a practical way of displaying results. Thus, for each fund, we separate the views by the number of assets that they involve and present those which best illustrate each fund’s overall views for clarity.

### 8.2 Mutual Fund Results

The *The Fidelity Mid Cap Value Fund* reports that its strategy is to invest in Mid Cap Value stocks. The *The GMO US Core Equity Fund* reports that its strategy is to invest in Core stocks but does not specify the size. Figure 15 plots all of the one asset equilibrium view deviations \(v - P\pi\) and Figure 16 plots all of the two asset equilibrium view deviations \(v - P\pi\) that were estimated over time for the *Fidelity Mid Cap Value Fund*. Figure 18 plots all of the one asset equilibrium view deviations \(v - P\pi\) and Figure 19 plots all of the two asset equilibrium view deviations \(v - P\pi\) that were estimated over time for the *GMO US Core Equity Fund*. In these plots, each point corresponds to a view at a given time: in other words, each point corresponds to a row of a \(P\) matrix. The y-axis depicts the estimated view deviations from equilibria in annualized \%: for a given view, this is simply the estimated view magnitude less the market implied expected return for that combination of assets.

Given the *Fidelity* fund strategy, we would expect it to be bullish for midsize value stocks and bearish for small growth stocks and/or large growth stocks compared to the market. In Figure 15, it is clear that compared to the market, the fund is consistently very bullish on the c5, d4, and d3 portfolios (i.e. Mid Cap Value and Core) and slightly bearish on the e1 and e2 portfolios (i.e. Large Cap Growth). In Figure 16, it is clear that the *Fidelity* fund is consistently bullish on mid cap and large value stocks in its 3 pairwise bets as well as for its d4+e4 and c4+e5 views. Thus, overall, although the fund advertises itself as a "Mid Cap Value" fund, it seems that it focuses on the larger mid cap value stocks.

Given the *GMO* fund strategy, we would expect it to be bullish for large core stocks and bearish for small stocks compared to the market. In Figure 18, it is clear that compared to the market, the fund is consistently very bullish on the e1 portfolio and is slightly bullish on the e2, e3, and e4 portfolios. Furthermore, it is bearish on the c1 and d1 portfolios. In Figure 19, it is clear that the fund is slightly bullish on combinations involving large core and growth stocks. Furthermore, it is slightly bullish for combinations of mid cap growth and core stocks. Thus, overall, although the fund advertises itself as a “Core” fund, it seems that it focuses on the larger mid cap core stocks and some large growth stocks.

Figure 17 depicts a scatter plot where each plotted point corresponds to a row of an estimated \(P\) matrix at a certain time for the *Fidelity* fund; in other words, each point represents a portfolio of FF portfolios that we have inferred that the PM had. The points are color coded depending on how many assets a particular view involves. The PM’s estimated view magnitude for each such view is plotted on the y-axis and the actual return of that portfolio over the subsequent month is on the x-axis: note that the returns are monthly and are not annualized. Figure 20 is the analogous plot for the *GMO* fund. For both funds, the betas are positive (0.03 and 0.05 respectively). However, the *GMO* fund does a slightly better job in forecasting future returns compared to the *Fidelity* fund.
Figure 15: The PM's estimated one asset equilibrium view deviations $v - P\pi$ over time. Note that the y-axis is in annualized %.

Figure 16: The PM's estimated two asset equilibrium view deviations $v - P\pi$ over time. Note that the y-axis is in annualized %.
Figure 17: Scatter plot of the PM’s estimated view magnitudes on the next month’s returns for the assets involved in his views (y-axis) versus the realized returns of those asset combinations over the next month (x-axis). The points are color coded to reflect the number of assets that each view involves. Note that the numbers are monthly percentages and are not annualized.

Figure 18: The PM’s estimated one asset equilibrium view deviations $v - P_\pi$ over time. Note that the y-axis is in annualized %.
Figure 19: The PM’s estimated two asset equilibrium view deviations \( v - P_r \) over time. Note that the y-axis is in annualized \%. 

Figure 20: Scatter plot of the PM’s estimated view magnitudes on the next month’s returns for the assets involved in his views (y-axis) versus the realized returns of those asset combinations over the next month (x-axis). The data from August 2008 is excluded. The points are color coded to reflect the number of assets that each view involves. Note that the numbers are monthly percentages and are not annualized.
8.3 Further Discussion of Results

In this chapter, we have presented the results of our inference algorithm at the 2x3 and 5x5 FF portfolio level. The funds we chose differed in investment style both in terms of stock size as well as in terms of stock "value". Overall, our estimated PM views are consistent with the stated objective of that fund. Short of being able to actually interview the portfolio managers in order to cross validate our results, examining their consistency with the funds' stated investment styles is the next best thing. Overall, our results are very encouraging in terms of picking out which assets belong in a given view and whether that view is bullish or bearish compared to the market equilibrium. Of the funds that we analyzed, the GMO US Equity Core Fund has the smallest average estimated view magnitude. This can be partly explained by our parametric assumption on $\Omega$, which is proportional to the estimated $\tau$ (see equation (33)). Thus, $\mathbf{A}$ is inversely proportional to $\tau$ and therefore, $v$ is as well (see equation (27)). All else being equal, a higher $\tau$ will result in lower view magnitudes.

Another phenomenon worth pointing out is that in the scatter plots depicting the PM’s estimated view magnitudes have regression betas are small in absolute value. The reason that the betas are small in absolute value comes from the fact that the y-axis variables are expected returns whereas the x-axis variables are realized returns. Thus, for the market implied expected returns scatter plots, the y-variable comes from historical data and is a "slow moving" average\textsuperscript{16} and therefore, it makes sense for the realized returns (x-axis) to exhibit a higher volatility and a wider range thus resulting in low betas (in absolute value) and small $R^2$ values.

\textsuperscript{16}Recall that both $\Delta_{mkt}$ and $\Sigma$ are estimated from 6 years of historical data
Part II
Investment Deviation from Fundamentals and Systemic Risk
9 Motivation for An Active Leveraged Sector

The recent economic crisis started as a financial crisis; as is typical with such crises, it quickly spread to the real economy. This crisis has highlighted the reality that over the past decades, the financial system has evolved into a market based system where leveraged financial intermediaries (called levered players from here on) play an active role instead of their traditional role of passively channeling capital from ultimate lenders to ultimate borrowers. This active behavior is driven by an imposed regulatory value at risk (VaR) constraint. However, adherence to the VaR constraint causes a series of unintended dynamics which are typically not captured by existing standard macroeconomic models; these dynamics tend to be nonlinear in nature and can be summarized as “invest more following good investment returns and divest more following bad investment returns”.

Over the past 25 years, levered players have grown in size significantly. A large part of this growth has come from the increasing willingness of the unleveraged sector to lend money to the leveraged sector: money market funds, which have grown rapidly over the past two decades, have been one of the primary channels from which the leveraged sector has been able to attract yield chasing investors. One of the key issues is that money market funds are generally viewed as riskless investments, even though they lend short term funds to leveraged players. Figure 21 plots the year end net funds invested in US money market accounts (blue line) and the percentage of these funds invested in the short term risky debt of leveraged financial intermediaries (red line), including repurchase agreements and commercial paper.\textsuperscript{17} The amount invested in US money market funds reached an astonishing 2.2 Trillion dollars in 2007; the percentage of these funds invested into risky levered debt varied between 70\% and 90\% before 2008, when it dropped to 60\%. Furthermore, this percentage seems to increase during boom periods (1986-1989, 1993-2000, 2003-2007) and drop during strenuous economic times. Figure 22 plots the difference between the the 30 day financial commercial paper rate and the 30 day risk free rate.\textsuperscript{18,19} The difference has been extremely low historically and even during the peak of the crisis, the differential was less than 0.25\%. Thus, it seems that investors are typically happy to invest in risky levered debt and receive a yield slightly higher than the prevailing risk free rate, especially during economic expansions. The relative ignorance of money market fund investors aided the growth of the leveraged sector, as it was able to borrow a large amount of funds through the money market channel at cheap interest rates due to the fact that investors viewed the money market as a very safe investment.

One of the biggest worries in the aftermath of the Lehman Brothers collapse was that money market funds would “break the buck”. Indeed, one of the primary factors which exacerbated the financial crisis was the inability of levered firms to roll over their short term borrowing as investors were spooked. This was so unprecedented by everyone, from investors to politicians, that the Federal Government was forced to provide temporary deposit insurance guarantees for money market funds in order to stop the panic.

Another important determinant of the ability to the leveraged sector’s ability to borrow is

\textsuperscript{17} The data is obtained from the Investment Company Institute (www.ici.org).

\textsuperscript{18} The latter is proxied by the one month Federal Funds Rate.

\textsuperscript{19} The rates are not annualized; they reflect the effective return of an investment over 30 days.
the short term risk free rate. This rate is essentially controlled by the Federal Reserve (Fed), which views it as a monetary policy tool used primarily to control inflation expectations. However, there is a large body of recent empirical evidence showing that the short term Fed Funds Rate (FFR) plays an important pricing role through the funding cost effects that it has on levered players; furthermore, levered players’ balance sheet aggregates have been shown to predict future real economic activity. This additional channel of monetary policy is termed the “risk taking” channel in [8] and is distinct from the “inflation expectations” channel. However, it is not the case that the FFR rate setting criteria formally take into account the workings of the leveraged sector: it is well documented that the Fed gradually adjusts the FFR in response to inflation and the output gap so as to achieve its mandate of sustaining “maximum employment” and low inflation; thus, there are large periods of time where the FFR is constant, and this has non-trivial effects on the leveraged sector.\(^{20}\)

In a nutshell, a constant risk free rate coupled with the positive feedback behavior of the leveraged sector aided by the willingness of the unleveraged sector to lend to the leveraged sector may have effects on real economy investment and systemic risk. In order to assess the effects that the levered players’ behavior has, we analyze an agent based model inspired by the recent empirical evidence of levered financial intermediary active balance sheet management. In our model, there are dynamic interactions with respect to real economy investment between three key players: the leveraged sector (e.g. hedge funds), the unleveraged sector (e.g. pension funds), and the Fed. The dynamic interaction between these players resulting from their localized optimization behavior dictates the amount of investment in the real economy at each point in time and the deviation of this investment amount from that which

\(^{20}\)The Fed sets the FFR only eight times a year and any rate adjustments are extremely gradual; if there is a change, it typically happens over a large time horizon at 0.25% increments.
is justified by the fundamentals.\textsuperscript{21}

In this paper, we find that under a constant risk free rate environment, the positive feedback effects resulting from the leveraged player’s VaR constraint adherence causes investment to deviate from fundamentals and an increase in the potential systemic cost of a default; the magnitude of these effects depends on the relative size of the leveraged sector to the unleveraged sector. In order to mitigate these undesired effects, we propose a FFR setting rule: among other parameters, the optimal FFR depends on the relative size of the leveraged sector to the unleveraged sector. Furthermore, the frequency of rate adjustment needs to be the same as the decision frequency of the leveraged and unleveraged player. However, by using only the risk free rate as a policy tool, it is not generally feasible to stabilize the potential systemic cost of a default and have investment reflect fundamentals at the same time.

The dependencies of the optimal FFR described above are interesting when thinking about the current monetary policy framework. In current monetary policy models, the relative size of the leveraged sector is not an input; this is in part due to the fact that leveraged financial intermediaries are viewed as passive players and did not used to have as much importance in the financial system as they do today. Furthermore, the current monetary policy norm is one of small gradual adjustments to the FFR; in contrast, the decision making of the leveraged and unleveraged players typically is of much higher frequency. Most investment banks, hedge funds, and pension funds make investment allocation decisions daily. Given the importance of the short term risk free rate in our model as a variable affecting real economy investment and potential systemic costs, it is the case that monetary policy and financial stability are linked.

In what follows, Section 10 presents the relevant literature, Section 11 presents the model setup, Section 12 presents the key results, Section 13 presents the optimal rate setting strategy for the FFR, Section 14 summarizes and concludes, and Section 16 is the Appendix.

\textsuperscript{21}The term fundamentals is intuitively related to the net present value (NPV) of the future cashflows; see Section 11.2 for a detailed description.
10 Literature Review

As the authors in [3] state, it is “difficult, if not impossible, to find a systemic risk measure that is at the same time practically relevant and completely justified by a general equilibrium model”; the authors also identify the need to bridge the gap between economic theory and actual regulations. Our approach is a step in this direction: the literature seems to lack a dynamic microfounded agent based model which generates phenomena commonly observed in the financial market and explicitly incorporates both an active leveraged sector and an active Fed which sets short term interest rates according to the financial system’s “state”. Our approach lies on the intersection of monetary policy, financial economics, and systemic risk, and we review the relevant literature from each field below.

The literature on systemic risk in the financial system is in its infancy as most of this literature has been written after the crisis of 2007; furthermore, most of it focuses on empirical methods in identifying and assessing various metrics of systemic risk. Our focus is on the effects arising from the localized decision making of the players, and our aim is to propose an optimal risk free rate setting rule; thus, we adopt a structural approach instead of a reduced form approach so as not to suffer from the Lucas Critique (see [46]), which states that reduced form models derived from observed relationships between various macroeconomic quantities produce relations which will differ depending on what macroeconomic policy regime is in place. Goodhart’s law (see [37]) nicely summarizes this point: “Any statistical relationship will break down when used for policy purposes”.

The importance of leveraged players, such as broker/dealers (B/D from here on) and hedge funds, in the supply of credit to the broader economy, has increased dramatically over the past 25 years due to the growth of securitization and the changing nature of the financial system toward one based on the financial markets (see [6]). Both [9] and [7] document that total financial assets of B/D’s have increased tremendously over the past 30 years, compared to both commercial bank assets and to US household assets. Thus, broker/dealers and other leveraged non-deposit taking institutions have taken a much larger role in intermediating between ultimate savers and ultimate borrowers. [6] finds that broker/dealers actively manage their leverage in a procyclical manner: when their asset values grow, they increase their leverage and vice versa; this behavior is the exact opposite of a passive player. The authors attribute this procyclical behavior to their VaR constraint adherence.

Risk management via VaR is the standard in the financial industry due to the relevant guidelines in the 1996 amendment on market risk regulation, where the Bank of International Settlement (BIS) chose VaR as the regulatory reporting tool for bank risk. There is a good deal of literature focused on measuring VaR (see for example [32]). In terms of studying the role of VaR in an asset pricing equilibrium context, [10] analyzes the dynamic portfolio selection effects of a VaR constrained representative investor in the familiar continuous time complete market Brownian motion setting. The authors find that stock market volatility and risk premium increase in down markets and decreases in up markets. Along a similar

\[\text{[65]}\]
line, [31] examines asset price dynamics when a large number of small myopic agents manage their risk via VaR constraints and update their beliefs of the returns distribution over time; the authors find that asset prices are lower, have time paths with deeper and longer troughs, and exhibit a greater degree of volatility.

Having discussed the relevant literature on VaR constraints, we now proceed to examine the significance of the leveraged sector with respect to the real economy. There is a large body of empirical evidence regarding the effect that levered players’ balance sheet quantities have on financial market prices and real economic variables. [6] finds that changes in short term borrowing by broker/dealers forecasts changes in financial market risk measures such as the CBOE VIX. [9] and [4] document that B/D asset growth forecasts durable consumption and housing investment. The empirical work in [8] and [5] builds on this theme by exploring the channel through which broker/dealer balance sheet quantities influence real economic variables. These authors hypothesize that the tightness of levered intermediaries balance sheet constraints determines their “risk appetite”. Risk appetite, in turn, determines the set of real projects that receive funding, and hence determines the supply of credit.

It is clear from the above empirical literature that levered financial intermediaries have become important players in the financial markets and that their procyclical balance sheet management has the capacity to affect both financial and real economic variables: this raises the need to explicitly incorporate an active levered sector into macroprudential policy prescriptions and models.

Before continuing, we present the key tenets of current monetary policy thinking: the term “thinking” is appropriate because the FFR decisions are not strictly rule based (see for example [54]). They are a result of the combination of model outputs and discretion and are typically influenced by the residing Fed Chairman’s philosophy. The Fed has a mandate from Congress to sustain “maximum employment” and low inflation; its primary tool in meeting its goals is the FFR. Under normal circumstances, the FOMC meets 8 times a year to decide the FFR. In models of monetary economics that are commonly used at central banks, the active role of financial intermediaries is ignored; banks and B/D’s are passive players that the central bank uses to implement monetary policy (see [7]). The “expectations channel” dominates current monetary thinking, where short-term rates matter only to the extent that they determine long term interest rates, which are seen as being risk-adjusted expectations of future short rates. Thus, an important theme in the expectations channel is managing the expectations of future short term rates by charting a path and communicating this path clearly to the market; in doing this, the thinking goes that the Fed can influence long rates, such as mortgage rates and corporate lending rates, which are important for the economy’s performance. According to Alan Blinder, a former Vice Chairman of the Fed’s Board of Governors, monetary policy has important macroeconomic effects only to the extent that “it moves financial market prices that really matter - like long-term interest rates, stock market values and exchange rates.” This type of thinking overlooks the effect that short term rates have levered players’ borrowing.

\[\text{footnote reference} 23\text{footnote text}\]

\[\text{footnote reference} 24\text{footnote text}\]
As one would expect, the Fed does not disclose the exact models it uses in setting the FFR. However, it has been shown that Taylor rules track the actual FFR quite well (see for example [29] or [30]). Thus, historically speaking, the Fed has been reacting to inflation and GDP growth data in its rate setting decisions, which is in accordance with its mandate. The Fed has a publication regarding a large-scale quarterly macroeconomic model that it uses to simulate the US economy called the “FRB/US” model (see [23]); this model contains 300 equations and identities although the number of stochastic “core” equations of the economic behavior of firms, households, and investors is around 50 equations. Although it is described by a large state-space, the empirical fits of the structural descriptions of macroeconomic behavior are comparable to those of reduced-form time series models, such as the Neo-Keynesian (NK) model. The publication does not explicitly present the FRB/US model but presents key characteristics about it. As is the case with the NK model, the FRB/US does not model financial intermediary balance sheet dynamics of the type described earlier in this section. Furthermore, it emphasizes the expectations role of the FFR as opposed to its pricing variable role.

In terms of setting the FFR, it is well documented that there is a large degree of policy inertia (see for example [55] or [29]); in other words, the adjustment of the FFR towards the policymaker’s desired rate is typically “gradual” and persistent, which results in tightening or loosening FFR cycles. [30] shows that the source of the observed policy inertia in the U.S. is due to deliberate “interest-smoothing”. [30] also finds that credit and asset market conditions do not have a direct explanatory effect on the FFR. The authors also present suggestive narrative evidence from FOMC discussions and Fed governor speeches during which monetary policymakers explicitly framed their decisions in a policy smoothing context: in fact, disagreement among FOMC members is typically about the speed of interest-rate smoothing instead of whether it should be a practice.

Clearly, an important variable in determining the ability of levered players to actively manage their balance sheet is their cost of borrowing: they finance their investments in the real economy by borrowing short term funds. As explained in [8], the FFR determines other relevant short term interest rates, such as repo rates and interbank lending rates. Thus, the FFR is important in setting short-term interest rates more generally. [9] finds that an increase in the FFR decreases B/D short term borrowing and vice versa. Furthermore, [7] finds that an increase in the FFR decreases B/D assets and vice versa. [8] and [5] conclude from their empirical results that the FFR plays an explicit role in the transmission of monetary policy, which is independent of the expectations channel. The authors therefore argue that monetary and financial stability policy are related by their effects on levered players’ balance sheet management.

A key policy question which central bankers have pondered around the world is whether monetary policy should take account of asset prices and react to asset price bubbles. There are differing views on whether the Fed should explicitly respond to asset prices in its monetary policy: among others, [12] and [13] posit that an inflation-targeting approach is sufficient and that changes in asset prices should affect policy only through inflation and GDP expectation adjustment. This opinion is shared by Alan Greenspan, who’s famous speech in 2002 (see
states that it is not the role of the Fed to pop asset price bubbles. However, there also exist proponents of an activist Fed with respect to asset price bubbles, such as [19], [22], and [38]: the general logic is that asset prices may affect demand via direct and indirect wealth effects leading to altered consumption plans. In our model, we show that the optimal FFR policy needs to respond to the exogenous investment return due to the latter’s effect on the distribution of wealth in the financial system.

In terms of relevant theoretical models which incorporate active financial intermediaries, the two most relevant papers to our model are [25] and [8]. [25] presents a dynamic macroeconomic model in which financial experts borrow from less productive agents in order to invest in financial assets. They derive full equilibrium dynamics and argue that steady-state analysis misses important effects. Among other results, they find that volatility effects and precautionary hoarding motives lead to a destabilizing deleveraging phenomenon;25 prices play an important role in this process. The authors also find that endogenous volatility and correlation is generated in asset prices during deleveraging episodes.

[8] presents a one period simple model which involves a VaR constrained levered player borrowing from an unlevered player in order to invest in a risky asset; the unlevered player can also directly invest in the risky asset. In their model, there is no explicit risk free asset and there are no defaults. Furthermore, there is no coupling between the two players because the unlevered player will lend as much as the levered player wishes at a constant rate of interest which is deemed to be risk free (exogenously). Thus, leverage will be constant unless there is a change in the statistics of the returns distribution.

Our multiperiod model builds on the one period model in [8]; however, apart from being multiperiod, a key enrichment is that both the amount that L borrows and its cost is a result of market clearing. Furthermore, in our model, the return on the investment is exogenous, as opposed having the dividend amount be exogenous; we also do not assume a particular distribution on the return. Moreover, an additional enriching feature is that there is an explicit risk free asset in which both players can invest in. Our model’s added richness results in some interesting phenomena. For example, as we will see, L’s leverage can decrease even as L’s ownership of the investment increases. Thus, procyclical leverage is distinct from procyclical demand in our model.

Starting from our enriched multiperiod model, we define and study the evolution of two key metrics: the investment deviation from fundamentals and the potential systemic cost of a default by the leveraged player (which we sometimes abbreviate by the term “risk”). We find that as the relative size of the leveraged player increases, investment deviates more from fundamentals and risk “builds up”. We subsequently propose an optimal FFR setting rule which ensures that investment reflects fundamentals and show that this does not necessarily stabilize risk (although it may slow its growth).

---

25 Such a deleveraging phenomenon is similar to so called “margin spirals” (see for example [24]).
11 The Model Setup

In this section, we present the setup of the model. Overall, we seek to balance “model selection” tradeoffs on multiple levels; the driving philosophy of this work is that simple behavioral rules generate complex system dynamics. We seek to make the model simple enough such that useful policy prescriptions can be derived, yet sophisticated enough to capture the key flow of funds characteristics between the leveraged and unleveraged sectors. We also seek to focus on the financial sector and abstract the investment in the real economy as well as omit real economic variables such as labor income; this is done in order to isolate the key dynamics of high frequency financial variables (e.g. daily or weekly) induced by the leveraged sector’s behavior; the underlying philosophy here is that the endogenous risks generated by agent behavior in the financial system can build up and affect real economic variables and we present recent empirical findings on this later in this section.

Our model has three heterogeneous agents interacting; the leveraged sector (L), the unleveraged sector (U), and the Fed. At each point in time, the net investment depends on L’s ability to borrow from U as well as U’s direct investment. The uncertainty in the system is due to the real economy investment performance, whose value is driven by an exogenous random variable reflecting changes in sentiment about economic growth. The key coupling results from the market clearing condition, which is that the amount that L wants to borrow must equal the amount that U wants to lend: the interest rate that clears the market is called the risky debt rate. In the model, this ends up depending on the relative size of the leveraged player and on the risk free rate. Although in reality, there exist many heterogeneous levered players with different business lines, our aim is to understand the effect of macroprudential policies on aggregate variables, such as the amount of risky assets held by the levered sector as a whole and the flow of funds between the aggregated players.

In what follows, Section 11.1 presents the players and their preferences, Section 11.2 discusses the real economy investment, Section 11.3 presents the “game’s” solution, Section 11.4 presents the properties of the solution, Section 11.5 presents a simplification of the game solution, and Section 11.6 presents the two key metrics we focus on.

11.1 Players and Preferences

The financial system model is in discrete time and has 3 agents: a levered agent (L), an unlevered agent (U), and the Fed. Furthermore, there is a risky investment “asset” which represents the net investment in real economic projects which produces stochastic future returns. Moreover, there is a risk free asset, whose rate of return is controlled by the Fed.

- **The levered sector:** Think of L as consisting of investment banks, hedge funds, and other levered players of the financial system such as the SPVs of the shadow banking system. L is able to invest in the risky asset and the risk free asset. In addition to investing his equity at time $t$ (denoted by $w^L_t$), L can issue debt and use the proceeds (denoted by $b_t$) to further invest in his opportunity set. At each period, L invests so as to maximizes his expected equity in the next period, subject to a VaR constraint.
The VaR constraint limits the amount he can borrow \( b_t \) such that his probability of bankruptcy is less than \( c \), an externally imposed amount:

\[
\max_{b_t, q_t} E_t[w^L_{t+1}] = \max_{b_t, q_t} \left( w^L_t + b_t \right) (q_t E_t[V_{t+1}] + (1 - q_t)F_t) - b_t X_t
\]

\[
\text{st } P(w^L_{t+1} < 0) \leq c \Rightarrow P \left( V_{t+1} < \frac{b_t X_t - (b_t + w^L_t) (1 - q_t)F_t}{q_t (b_t + w^L_t)} \right) \leq c
\]

where \( V_{t+1} \) denotes the exogenous random asset return, \( X_t \) denotes the equilibrium interest rate that \( L \) pays on the amount he borrowed, \( E_t[\cdot] \) denotes the expectation given the information set at time \( t \), and \( q_t \) denotes the proportion of the money that \( L \) has at his disposal at time \( t \) which is directed towards the risky asset; \( q_t \) is between 0 and 1 by definition. It should be noted that \( L \)'s optimization results in a supply function \( b_t(X_t) \) of how much he wants to borrow. The actual amount he borrows at time \( t \) is determined by a market clearing condition where his supply of risky debt equals \( U \)'s demand of it.

The fact that \( L \) wants to maximize his expected equity is consistent with the institutional framework of how such institutions operate; the mission of the CEO is to maximize shareholder value, not to manage shareholder risk. This is one of the inherent frictions in corporate finance: if one believes that higher risk gives higher rewards, then the CEO must pursue these higher risk higher reward opportunities to satisfy shareholders. Of course, the debt-holders would prefer the low risk low reward strategy as this would decrease the probability of bankruptcy. The VaR constraint is a form of regulation put in place to limit the leveraged player’s riskiness. For a more in depth discussion of such issues, see [21].

- **The unlevered sector**: Think of \( U \) as consisting of pension funds, mutual funds, and other “real money” agents who invest their wealth (denoted by \( w^u_t \)) and are not allowed to borrow money and lever up.\(^{26}\) \( U \) has an additional investment option compared to \( L \), in which he can lend money to \( L \).\(^{27}\) However, \( U \) is risk averse and is a mean-variance optimizer in each period:

\[
\max_{a^u_t, a^d_t, a_t} E_t \left[ \frac{w^u_{t+1}}{w^u_t} \right] - \frac{1}{2\tau} \text{Var}_t \left[ \frac{w^u_{t+1}}{w^u_t} \right]
\]

\[
w^u_{t+1} = w^u_t + z_{t+1} = w^u_t \left( a^u_t V_{t+1} + a^d_t D_{t+1} + a_t F_t \right)
\]

\(^{26}\)Such agents typically invest household wealth invested in 401(k) and other saving accounts and are thus not allowed to lever up.

\(^{27}\)Sometimes, we will refer to this as “investing in \( L \)'s risky debt.”
where $\tau$ denotes the “risk willingness” of the agent, $z_{t+1}$ denotes $U$’s return, $a_t^v$ denotes $U$’s portfolio weight in the risky asset, $a_t^D$ denotes $U$’s portfolio weight in L’s risky debt, $a_t$ denotes $U$’s portfolio weight in the risk free asset, and $D_{t+1}$ denotes the return on investing in L’s risky debt; note that $D_{t+1} \neq X_t$ because although L has promised to return $X_t$ (a quantity agreed upon at time $t$) for every dollar borrowed, he may end up returning less than that if he goes bankrupt; in fact, it can be shown that:

$$D_{t+1} \equiv \min \left\{ \left( \frac{w_t^L}{w_t^v} \frac{1}{a_t^D} + 1 \right) V_{t+1}, X_t \right\}$$

because although $X_t$ is the return on the debt agreed upon at time $t$ through supply and demand, a bad enough $V$ will make L default, causing $U$ to seize any of L’s remaining assets.

For $U$ to be unlevered, the following must hold:

$$a_t^v + a_t^D + a_t = 1$$

$$a_t^v, a_t^D, a_t \geq 0$$

- **The Fed:** The Fed is the entity which sets the short term risk free rate $F_t$ (we may also refer to this as the short rate). As discussed in Section 3, the FFR determines other relevant short term interest rates. In this model, we assume that the Fed can exactly determine the short term risk free rate, although in practice, the short term risk free rate would be the yield on an extremely short term US government bond, which moves in tandem with the FFR, yet may theoretically deviate from this spread.

Figure 23 graphically depicts the model: the black arrows highlight what investment opportunities are available to each agent and the green arrows refer to the returns of each of the two assets.

The frequency of the agent’s decisions is relatively fast; one can imagine L making decisions at frequencies faster than weekly due to the fact that L is constantly pressured to return value to shareholders. Having $U$ make decisions at the same frequency as L is a simplification; however, although arguments can be made that households care about long term performance, an argument can also be made that real money managers, who manage household wealth, such as pension funds and mutual funds, compete intensely for customers and thus care about short term performance. Such fund managers do indeed face short term assessments of their performance. Given that portfolio manager bonus compensation practices are similar across firms such as investment banks, hedge funds, mutual funds, and pension funds, it is reasonable to assume that $U$ acts as fast as L.

---

28Note that the typical risk aversion parameter $\gamma \equiv \frac{1}{\tau}$. 
11.2 The Risky Investment

In the model, the risky asset can be viewed as an investment in the real economy. One can think of the investment funds channeled via the financial system as going to a set of "producers", who use the capital to undertake real economic activity, such as investing in houses, factories, etc. In our model, the producers accept whatever capital is given to them and go about investing in real economy projects. The fundamental value of these projects is an exogenous random variable and reflects short term changes in expectations and/or sentiment regarding the performance of the broader economy; we denote this exogenous random variable by $V$. As we will see, the agents' locally optimizing behavior in response to $V$ ends up affecting reinvestment in the real economy.

Overall, it is desirable that reinvestment in the economy is not affected by short term financial frictions. Real economy investments such as those mentioned above typically have a large horizon as such projects take time, real labor is involved, etc. Project plans should not have to be readjusted at every time due to the financial system's response to the exogenous process $V$, which captures short term changes in expectations/sentiments about the future cash flow potential of the real economy projects. Investment decisions should not be dictated by short term changes in sentiment; they should be made according to the long term potential that they may have.

In our model, due to the locally optimizing characteristic of the financial system, the producer is given additional "unplanned" capital to utilize following a "good" $V$ outcome, whereas an "unplanned" divestment of capital from the producer results from a "bad" $V$ outcome. Thus, an unplanned divestment resulting from a bad $V$ outcome will cause the
producer to have to liquidate/terminate parts of the project which have a good long term expected trend; an unplanned investment resulting from a good V outcome will give the producer additional capital to invest in additional long term projects with long term potential; however, since this excess capital may be unexpectedly “pulled” following a subsequent bad V, this long term positive project, which was initialized in response to a good V by hiring personnel, buying machinery, etc., may be discontinued due to another short term V outcome; thus, having the producer adjust long term investment undertaking in response to short term “noise” is not desirable.

We now proceed with the notation we will use. At time t, an amount of $P_t$ is invested in the project (risky asset); this amount is endogenously determined. However, the change in value per dollar of investment $V_{t+1}$ is an exogenous random variable. At time $t+1$, the fundamental value of the project is:

$$S_{t+1} = P_t V_{t+1}$$

In other words, $S_{t+1}$ reflects the time $t+1$ net present value (NPV) of all future profits of the investment. The NPV at time $t+1$ depends on the (endogenously determined) investment at time $t$ and the exogenous random variable $V_{t+1}$. Deviations of the amount reinvested at time $t+1$, $P_{t+1}$, from $S_{t+1}$ are a result of the interactions from the financial sector (i.e. L and U), not from the actual investment’s long term potential.

$V$ is modeled as an IID random variable, where the mean $m$ and standard deviation $s$ are known to L and U; furthermore, we assume that the $c^{th}$ percentile is a known value, where $c$ is the regulated maximum allowed probability of bankruptcy. However, we do not make an assumption on the type of distribution as this is not needed for our analysis.

11.3 The Solution

In this section, we present the model solution. We first present the demand functions of U and L, which are the result of their optimization problems; we then present the market clearing $(X, b)$ solution.

11.3.1 U’s Allocations

U’s optimization problem is presented in equation (53); we assume that the statistics $m$ and $s$ of the process for $V$ are known. Furthermore, in accordance with the behavioral assumption of U viewing L’s short term debt as quite safe,\textsuperscript{29} he assumes that the debt has an expected return of $X_t$, which is the endogenous market clearing risky debt rate, yet attaches a small standard deviation of $s_D$ to reflect its risky nature. In the same spirit, U assumes that $D_{t+1}$ and $V_{t+1}$ have a zero correlation. Thus:

$$\mathbb{E}_t [w^u_{t+1}] = \mathbb{E}_t \left[ a^u_t m + a^D_t X_t + a_t F_t \right]$$

\textsuperscript{29}This behavior was discussed in the introduction and comes from the observed reality of most people fearlessly using money market funds as if they were equivalent to risk free assets (i.e. government guaranteed bank accounts).
\[ V \rho[w_{t+1}^W] = (w_t^u)^2 \left( (a_t^v)^2 s^2 + (a_t^D)^2 s_D^2 \right) \]  

(57b)

Thus, at each time \( t \), \( U \) solves:

\[
\max_{a_t^v, a_t^D, a_t} E_t[z_{t+1}] - \frac{1}{2\tau} Var_t[z_{t+1}] \]

\[ \text{st } a_t^v + a_t^D + a_t = 1 \]

(58)

which leads to the following portfolio weights:

\[
a_t^v = c_v (m - F_t), \quad c_v = \frac{\tau}{s^2} \]

(59a)

\[
a_t^D = 2c_D (X_t - F_t), \quad c_D = \frac{\tau}{2s_D^2} \]

(59b)

\[
a_t = 1 - a_t^v - a_t^D \]

(59c)

Thus, \( U \) has a demand function of \( L \)'s risky debt; the proportion of his wealth that he will lend to \( L \) is a function of the interest rate \( X_t \) he will receive.

### 11.3.2 \( L \)'s Optimal Borrowing

We now turn to \( L \)'s problem, which is given in equation (52); as mentioned in Section 11.1, it can be shown that the optimal proportion of \( L \)'s investable funds invested in the risky asset is \( q_t^* = 1 \) (see the Appendix). Thus, \( L \)'s problem can be reduced to:

\[
\max_{b_t} E_t[w_{t+1}^L] = \max_{b_t} w_t^L m + b_t (m - X_t) \]

\[ \text{st } P(w_{t+1}^L < 0) \leq c = P \left( V_{t+1} < \frac{b_t X_t}{b_t + w_t^L} \right) \leq c \]

(60)

\( L \)'s optimization results in a supply function of debt, meaning the amount he wishes to borrow depends on the interest rate \( X_t \) that he will pay. By equating the demand and supply functions of \( L \)'s risky debt, \( X_t \) and \( b_t \) are jointly determined.

In fact, \( L \)'s objective function is concave since \( b_t \) and \( X_t \) move in the same direction: increasing the desired borrowing amount will actually increase the interest rate \( X_t \) demanded by \( U \) in order to incentivize \( U \) to allocate a larger proportion \( a_t^D \) of his wealth \( w_t^u \) to \( L \)'s risky debt. Thus, when optimizing, \( L \) will increase \( b_t \) until either:

1. He hits his VaR constraint. In this case, the optimal borrowing \( b_t \) as a function of \( X_t \), denoted by \( b_t^{VaR} \), can be found by manipulating the VaR constraint:
\[ P \left( V_{t+1} < \frac{b^V_{t+1}X_t}{b^V_{t+1} + w_t^L} \right) = c \Rightarrow \frac{b^V_{t+1}X_t}{b^V_{t+1} + w_t^L} = K \Rightarrow b^V_{t+1} = \frac{w_t^L K}{X_t - K} \]  

(61)

where \( K \equiv \Phi^{-1}(c) \) and \( \Phi(.) \) denotes the CDF of the random variable \( V \). Note that typically, \( K \) will be a number less than 1, signifying a loss from investment in the risky asset. For example, if \( V \) is a lognormal random variable whose logarithm is normally distributed with mean \( \mu = 8\% \) and \( \sigma = 15\% \), and \( c = 5\% \), then \( K = 0.846 \), meaning that \( L \) is solvent as long as his risky asset investment does not lose more than 15.4\% until the next period.

2. He hits his optimal borrowing \( b^{unc}_{t} \) before hitting his VaR constraint: thus, in this case, \( b^{unc}_{t} \leq b^V_{t+1} \). Note that under this case, the VaR constraint is not tight meaning that \( L \)’s probability of bankruptcy is smaller than \( c \).\(^{30}\)

### 11.3.3 The Equilibrium Solution

The exogenous process for \( V \) drives the system, yet the system’s investment response is determined endogenously and depends on the state variables of the system at a given time. There are two variables that are determined at each time \( t \): the amount \( L \) borrows at time \( t \), \( b_t \), and the interest rate that he has to pay at time \( t+1 \) on the amount he borrowed at time \( t \), \( X_t \).\(^{31}\) The \((X_t, b_t)\) solution determines the amount of investment in the risky asset \( P_t \).

It is clear that the time \( t \) investment \( P_t \) depends on the amount that \( L \) borrows from \( U \): in this model, since there is one supply and one demand for the risky debt, \( X_t \) and \( b_t \) are jointly determined by the market clearing condition. The interest rate \( X_t \) should be such that \( U \)’s demand for risky debt equals \( L \)’s supply of risky debt. Mathematically:

\[ a^P_t(X_t)w_t^L = b_t(X_t) \]  

(62)

where \( a^P_t \) is the fraction of \( U \)’s wealth invested in the risky debt. Note that for the amount that \( L \) borrows to be positive, the following must be true:

\[ X_t > F_t \]  

(63)

Armed with this relationship between the risky debt supply and demand, we can solve for the \((X_t, b_t)\) pairs in both of \( L \)’s modes of operation:

\(^{30}\)See the Appendix for a proof of this.

\(^{31}\)A note about the notation: if at time \( t \), I invest \$1 in the risk free asset, I will have \( F_t \) dollars at time \( t+1 \). If I had invested in the risky asset, I would have \( V_{t+1} \) dollars at time \( t+1 \). Although both are rates of return between times \( t \) and \( t+1 \), the subscript \( t \) on \( F_t \) refers to the fact that this quantity is known at time \( t \). The subscript \( t+1 \) on \( V_{t+1} \) reflects the fact that this quantity is a random variable observed at time \( t+1 \).
1. **VaR Constrained Mode:** In this case, we need to solve equations (61) and (62) simultaneously:

\[
\frac{w_t^L}{X_t - K} = 2c_D (X_t - F_t) w_t^u \Rightarrow X_t^+ = \frac{F_t + K \pm \sqrt{(F_t - K)^2 + \frac{2K w_t^L}{c_D w_t^u}}}{2}
\] (64)

However, the \(X_t^-\) solution is rejected since \(X_t^- \leq F_t\), which contradicts equation (63). Plugging in the expression for \(X_t\) into equation (62) gives \(b_t\):

\[
b_t = w_t^u c_D \left( K - F_t + \sqrt{(F_t - K)^2 + \frac{2K w_t^L}{c_D w_t^u}} \right)
\] (65)

Thus, both \(X_t\) and \(b_t\) are increasing functions of the relative size of \(L\) to \(U\), \(w_t^L / w_t^u\).

2. **Unconstrained Mode:** Plugging in equation (62) into L’s FOC and solving for \(b_t\) gives:

\[
b_t = c_D w_t^u (m - F_t)
\] (66)

Thus, according to equation (63):

\[
X_t = \frac{m + F_t}{2}
\] (67)

It is worth highlighting that in this mode of operation, \(b_t\) and \(X_t\) do not depend on the relative size of \(L\) to \(U\); furthermore, if \(F_t\) is constant, then \(b_t\) and \(X_t\) are constant.

From here on, we denote the relative size of \(L\) to \(U\), \(w_t^L / w_t^u\), by \(y_t\). To summarize, depending on L’s mode of operation, the \((X_t, b_t)\) solutions are:

\[
X_t = \min \left( \frac{m + F_t}{2}, \frac{F_t + K + \sqrt{(F_t - K)^2 + \frac{2K y_t}{c_D y_t}}}{2} \right)
\]

\[
b_t = \min \left( c_D w_t^u (m - F_t), w_t^u c_D \left( K - F_t + \sqrt{(F_t - K)^2 + \frac{2K y_t}{c_D y_t}} \right) \right)
\] (68)

It should be highlighted that \(a_t^D = \frac{b_t}{w_t^u}\). Thus, the solution \((X_t, b_t)\) implicitly gives an \(a_t^D\); as we will see, the change in \(a_t^D\) is key in determining investment deviation from fundamentals and risk.
Given the solution \((X_t, b_t, a_t^D)\), the prevailing time \(t\) investment can be found by aggregating the investment from both \(L\) and \(U\):

\[
P_t = w_t^L + b_t(X_t) + a_t^v(F_t)w_t^u = w_t^L + a_t^v(X_t)w_t^u + a_t^v(F_t)w_t^u
\]

(69)

11.4 Solution Properties

11.4.1 Levered Balance Sheet Dynamics

As we saw, \(L\) is maximizing his expected equity in the next period subject to a VaR constraint. The realized outcome of \(V\) will affect his constraint adherence and he will have to readjust his borrowing accordingly; this behavior is pro-cyclical as a “good” \(V\) outcome will cause his constraint to be loose; \(L\) will thus look to borrow more in order to invest. Analogously, a “bad” \(V\) outcome will cause him to be in violation of his constraint and thus, he will be forced to borrow less.

This is graphically illustrated in Figure 24. At time \(t\), \(L\) borrows \(b_t\). Thus, \(L\) owns \(w_t^L + b_t\) dollars worth of the risky asset, financed by \(w_t^L\) (his equity) and by \(b_t\) (the amount he borrows from \(U\)). At time \(t + 1\), before he rebalances his portfolio, \(L\) owns assets worth \((w_t^L + b_t)V_{t+1}\) and owes to \(U\) an amount of \(b_tX_t\). Thus, his time \(t + 1\) equity is:

\[
w_{t+1}^L = w_t^L d_{t+1}, \\
\]

\[
d_{t+1} = \max \left\{ \left( \frac{w_t^L + b_t}{w_t^L} \right) V_{t+1} - \frac{b_t}{w_t^L} X_t, 0 \right\}
\]

(70)

where \(d_{t+1}\) denotes the return on equity (ROE) and the max operator is there because equity cannot be negative due to its limited liability nature; thus, if \(V_{t+1}\) is small enough to cause the equity to go negative, \(U\) starts to lose money and the equity remains at 0. We denote this critical value of \(V_{t+1}\) as \(K\), which can be read off from equation (70) as:

\[
K \equiv \frac{b_t X_t}{w_t^L + b_t}
\]

(71)

It is worth observing that \(K\) is a constant because at each time \(t\), \(L\) rebalances his portfolio so that \(P(w_{t+1}^L < 0) = P(V_{t+1} < K) \leq c\).

In terms of balance sheet adjustments, if \(V_{t+1}\) is sufficiently “high”, then \(L\)’s increased equity at time \(t + 1\) will cause the VaR constraint to be loose; this means he can borrow more in order to buy more of the asset; he will do this until he hits his time \(t + 1\) VaR constraint since in his utility function, he is maximizing expected returns. This situation is graphically depicted in Figure 2, where a good \(V\) initially leads to an increase in \(L\)’s asset value and a decrease of the debt to equity ratio (middle panel); the debt to equity ratio decreases since the debt value does not increase as the asset value increases. Since in the middle panel, \(L\) has “slack”, he increases his position in the asset by borrowing more money and buying more of the asset until he hits his VaR constraint (right panel); this action makes \(L\) hold more of the asset after rebalancing than he would have held had he been passive. Analogously, if \(V_{t+1}\) is sufficiently “bad”, \(L\)’s decreased equity will cause him to be in violation of his time
$t+1$ VaR constraint; thus, he will borrow less at time $t+1$ and thus overall decrease his risky asset position.

This behavior is called "pro-cyclical", because a good fundamental outcome causes $L$ to further increase his position and a bad fundamental outcome causes $U$ to further decrease his position. This positive feedback response is the key dynamic in the model, which effectively causes funds to be sucked into the system from the risk free asset if $V$ is good and vice versa.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure24.png}
\caption{Graphical depiction of $L$'s balance sheet dynamics in response to a "good" $V_{t+1}$.}
\end{figure}

11.4.2 Solution Dependence on Relative Size

From the solution expressions $(X_t, b_t)$, it is clear that they are increasing functions of the relative size $y_t$ when $L$ is operating in the VaR constrained mode. More importantly, $a^D_t$ is a function of $y_t$; this means that a larger $y_t$ will attract a larger proportion of $U$'s wealth. This is one of the key properties of the model; as $y$ increases, more of $U$'s wealth is brought into the financial system from the risk free asset and channeled towards the risky investment.

If $L$ becomes large, then he ceases to be operating in the VaR constrained mode because his optimal borrowing is below the limit imposed by the VaR constraint. In this mode of operation, the solution $(X_t, b_t)$ is not dependent on $y_t$.

We now proceed to find what this "cutoff relative size" is, denoted by $\bar{y}_t$, where $b_{t}^{unc} = b_{t}^{VaR}$.

\begin{align*}
    b_{t}^{unc} &= b_{t}^{VaR} \Rightarrow \\
    m - F_t &= K - F_t + \sqrt{(F_t - K)^2 + \frac{2K}{cD} y_t} \Rightarrow \\
    \bar{y}_t &= \frac{cD}{2K} \left[ (m - K)^2 - (F_t - K)^2 \right]
\end{align*}
Thus, if $y_t$ is above this cutoff, then L’s optimal borrowing does not cause him to hit his VaR constraint and is given by equation (66). Therefore, his probability of bankruptcy is below $c$. Intuitively, this means that there is a natural leverage limit of the system; if L is much larger than U, then the interest rate on his debt will be quite high, meaning that his optimal borrowing amount will be lower than that dictated by his VaR constraint.

11.4.3 Relative Size Growth Barrier

The relative size of L to U will grow if $V_{t+1} > \bar{V}_{t+1}$. Solving analytically for this value $\bar{V}_{t+1}$:

$$d_{t+1}(V_{t+1}) = z_{t+1}(V_{t+1}) \Rightarrow$$

$$\frac{w^L_t + b_t}{w^L_t} V_{t+1} - \frac{b_t}{w^L_t} X_t = (a^P_t V_{t+1} + a^D_t X_t + a_t F_t)$$

(73)

Since the ratio of L’s assets to his equity $\frac{w^L_t + b_t}{w^L_t}$ is his leverage $l_t$, we can rewrite the above as:

$$V_{t+1}(l_t - a^P_t) = (l_t - 1 + a^D_t) X_t + a_t F_t \Rightarrow$$

$$\bar{V}_{t+1} = \frac{l_t - 1 + a^D_t}{l_t - a^P_t} X_t + \frac{a_t}{l_t - a^P_t} F_t$$

(74)

Notice that $\bar{V}_{t+1}$ is a convex combination of $X_t$ and $F_t$ since $l_t - 1 + a^D_t + a_t = l_t + a^P_t$ and $l_t \geq 1$. Note that in deriving the above, we assumed that $\bar{V}_{t+1} > K$ and used the corresponding “non-default” $d_{t+1}$ and $z_{t+1}$ expressions. If $\bar{V}_{t+1}$ had been less than $K$, then $d_{t+1} = 0 < z_{t+1}$; the only case where $d_{t+1} = z_{t+1} = 0$ is when $V_{t+1} = 0$ and $a_t F_t = 0$, leading to $\bar{V}_{t+1} = 0$.

Overall, the take home message of this section is that the “hurdle” $\bar{V}_{t+1}$ depends on $y_t$.

11.4.4 Leverage Versus Size

We now turn to examine how L’s leverage depends on $y_t$ for the case where he is operating in the VaR constrained mode.

Recall that L’s leverage is defined as the value of his assets over his equity. At each rebalancing time, L adjusts his portfolio such that his VaR constraint is tight. This however does not mean that his leverage stays constant. Starting from equation (71), the following expression for L’s borrowing $b_t$ can be derived:

$$d_{t+1}(V_{t+1}) = z_{t+1}(V_{t+1}) \Rightarrow$$

$$\frac{w^L_t + b_t}{w^L_t} V_{t+1} - \frac{b_t}{w^L_t} X_t = (a^P_t V_{t+1} + a^D_t X_t + a_t F_t)$$

(73)

This is a trivial and unrealistic case which happens with 0 probability and thus we ignore the case where $\bar{V}_{t+1} = 0$. 

11.4.4 Leverage Versus Size

We now turn to examine how L’s leverage depends on $y_t$ for the case where he is operating in the VaR constrained mode.

Recall that L’s leverage is defined as the value of his assets over his equity. At each rebalancing time, L adjusts his portfolio such that his VaR constraint is tight. This however does not mean that his leverage stays constant. Starting from equation (71), the following expression for L’s borrowing $b_t$ can be derived:
\[ K = \frac{b_t X_t}{w^L_t + b_t} \Rightarrow b_t = w^L_t \frac{K}{X_t - K} \]  

L’s leverage \( l_t \) is defined as the ratio of his assets to his equity. Using equation (75), it is easy to express \( l_t \) as:

\[ l_t \equiv \frac{w^L_t + b_t}{w^L_t} = \frac{X_t}{X_t - K} \]  

Since \( X_t \) is an increasing function of \( y_t \), L’s leverage decreases as his relative size increases.\(^{33}\)

Since the relative size \( y_t \) increases if \( d_{t+1} > z_{t+1} \) and vice versa, L outperforming U in a period will lead to L having lower leverage in the next period and vice versa.

### 11.4.5 Solution Dependence on Outcome

In this section, we show that \( y_{t+1} \) and thus \((X_{t+1}, a^D_{t+1}, b_{t+1})\) are increasing in \( V_{t+1} \), and \( l_{t+1} \) is decreasing in \( V_{t+1} \). Note that we examine this for the case where \( V_{t+1} > K \); for \( V_{t+1} < K \), \( y_{t+1} = 0 \), and thus \( y_{t+1} \) will not vary with \( V_{t+1} \).

Starting from the fact that \( y_{t+1} = y_{t+1} x_{t+1} \):

\[ \frac{d y_{t+1}}{d V_{t+1}} = y_t \frac{d_t z_{t+1} - z_{t+1} d_{t+1}}{z_{t+1}^2} \]

\[ = \frac{y_t}{z_{t+1}^2} \left( l_t z_{t+1} - (l_t V_{t+1} - (l_t - 1) X_t) \right) \]

\[ = \frac{y_t}{z_{t+1}^2} \left( l_t (a^D_t X_t + a_t F_t) + (l_t - 1) X_t a_t^\gamma \right) \]

\[ > 0 \]  

where we have used the fact that \( l_t \geq 1 \) since L will always invest all of his equity plus any additional amount he borrows from U.

Thus, \( y_{t+1} \) is increasing in the \( V_{t+1} \), meaning that \((X_{t+1}, a^D_{t+1}, b_{t+1})\) are also increasing in \( V_{t+1} \) whereas \( l_{t+1} \) is decreasing in \( V_{t+1} \).

### 11.4.6 Dynamics of \( P_t \)

The dynamics of the investment process \( P_t \) are driven by the dynamics of the state variable \( y_t \): the time \( t + 1 \) investment \( P_{t+1} \) will depend on \( y_t \) as well as on the exogenous return \( V_{t+1} \). Assuming that at time \( t + 1 \), \( V_{t+1} \geq K \) and thus L is solvent, \( P_{t+1} \) can be expressed with the help of equation (69) as:

\[ P_{t+1} = w^L_{t+1} + b_{t+1} + a^v_{t+1} w^u_{t+1} \]

\(^{33}\)Note that \( X_t \geq K \) is guaranteed in the model if \( F \geq K \) is assumed, which is realistic since \( K \) is less than 1 by definition and \( F \) is greater than 1 in practice.
\[ \begin{align*}
&= (w_t^u + b_t)V_{t+1} - b_t X_t + w_{t+1}^u (1 - a_{t+1}) \\
&= (w_t^u + b_t)V_{t+1} - b_t X_t + (w_t^u a_t^u V_{t+1} + w_t^u a_t^D X_t + w_t^u a_t F_t) (1 - a_{t+1}) \\
&= (w_t^u + b_t)V_{t+1} - b_t X_t + w_t^u a_t^u V_{t+1} + w_t^u a_t^D X_t + w_t^u a_t F_t - w_{t+1}^u a_{t+1} \\
&= P_t V_{t+1} + (w_t^u a_t F_t - w_{t+1}^u a_{t+1}) \\
&= P_t V_{t+1} + w_t^u (a_t F_t - z_{t+1} a_{t+1})
\end{align*} \]

where we use the debt supply/demand relationship in equation (62) throughout the proof. Furthermore, in line 2 of the proof above, we use the fact that U’s weights sum to 1.

The above dynamics for \( P_{t+1} \) state that the investment tomorrow is equal to the fundamental value of the investment plus an additional term, which is equal to the inflow of funds into the system from the proceeds of U’s time \( t \) risk free asset investment minus the outflow of funds arising from U’s time \( t+1 \) investment in the risk free asset. Figure 25 graphically depicts these one period investment “flow” dynamics. At time \( t+1 \), there is a net inflow of money into the U/L system from both their risky asset investment proceeds and from their risk free asset investment proceeds (green arrows); these inflows determine \( y_{t+1} \), and U and L rebalance accordingly; this rebalancing determines the time \( t+1 \) risky asset and risk free asset investments (red arrows); the time \( t+1 \) deviation of the investment from its fundamental value depends on the endogenous interactions of U and L at time \( t+1 \).

---

Figure 25: Graphical depiction of the system’s monetary inflows and outflows

In order to gain some intuition on the system behavior, we consider two limiting cases:

\[ \text{Note that in the figure, the relative size } y_{t+1} \text{ is different from the relative size } y_t. \]
1. **No Risk Free Asset**: If the system did not have a risk free asset investment option, then reinvestment would always reflect fundamentals. Mathematically, this case is described by setting $a_t = a_{t+1} = 0$ in equation (78); this results in $P_{t+1} = P_t V_{t+1} \equiv S_{t+1}$. Intuitively, all net inflows from the proceeds of the risky asset would ultimately be reinvested in it: $L$ and $U$ may trade, yet ultimately, $U$’s investment in $L$’s debt is channeled to the risky asset.

2. **Only U**: If the system did not have $L$, $U$ could only invest in the risky asset and the risk free asset. In such a case, there would be underinvestment in the risky asset during “good times” and overinvestment during “bad times”. Intuitively, this is a result of $U$’s risk aversion: assuming a constant risk free rate, in the case that the risky asset outperforms the risk free asset, $U$ will divert some of the risky asset profits to the risk free asset. If the risky asset underperforms the risk free asset, he will divert some of his risk free asset profits to the risk free asset. Mathematically:

\[
P_{t+1} = a_{t+1}^r w_{t+1}^u
= a_t^r w_t^u (a_t^r V_{t+1} + (1 - a_t^r) F_t)
= P_t (a_t^r V_{t+1} + (1 - a_t^r) F_t)
= P_t z_{t+1}
\]

where we have used the fact that if the risk free rate is constant, then $a_t^r = a_{t+1}^r$. Thus, there are two cases:

**Outperformance**: $V_{t+1} > F_t \Rightarrow z_{t+1} < V_{t+1} \Rightarrow P_{t+1} < P_t V_{t+1}$

**Underperformance**: $V_{t+1} < F_t \Rightarrow z_{t+1} > V_{t+1} \Rightarrow P_{t+1} > P_t V_{t+1}$

\[\text{(79)}\]

11.5 **Solution Simplification**

The expressions for the solution $(X_t, b_t)$ have a square root term which makes their analysis in closed form difficult. We thus present a linearized approximation to the $(X_t, b_t)$ solution; this simplified solution is accurate when $y_t << 1$.\(^{35}\) This assumption is realistic in practice and the Appendix provides a rough estimate of $y_t$ using worldwide data.

\(^{35}\)We sometimes refer to the simplified solution case as the “simplified model”. Furthermore, we sometimes refer to the non-simplified solution as the “full model”.  

82
11.5.1 The Linearization

A Taylor series approximation of the square root term for $X_t$ gives:

$$
X_t = \frac{F_t + K + \sqrt{(F_t - K)^2 + \frac{2K}{cd}y_t}}{2}
$$

$$
= \frac{1}{2} \left[ F_t + K + (F_t - K) \sqrt{1 + \frac{1}{(F_t - K)^2} \frac{2K}{cd}y_t} \right]
$$

$$
\approx \frac{1}{2} \left[ F_t + K + (F_t - K) \left( 1 + \frac{1}{2} \frac{2K}{(F_t - K)^2} \frac{2K}{cd}y_t \right) \right]
$$

$$
= F_t + \frac{K}{2cd} \frac{y_t}{F_t - K}
$$

(81)

Note that $X_t$ is still increasing in $y_t$, yet the dependence is now linear; note also that it is still greater than or equal to $F_t$, as desired.

In order for the above approximation to be valid, we assume that:

$$
\frac{1}{(F_t - K)^2} \frac{2K}{cd}y_t \leq \epsilon_1 << 1 \tag{82}
$$

Given the simplified expression for $X_t$, we present the resulting expressions for other key variables:

$$
\begin{align*}
 b_t &\equiv w_t^u 2cd(X_t - F_t) = w_t^u \frac{K}{F_t - K} y_t = w_t^u (l_t - 1) y_t \\
 l_t &\equiv \frac{w_t^l + b_t}{w_t^l} = 1 + \frac{w_t^u}{w_t^l} \frac{K}{F_t - K} y_t = 1 + \frac{K}{F_t - K} = \frac{F_t}{F_t - K} \\
 a_t^D &\equiv \frac{b_t}{w_t^u} = \frac{K}{F_t - K} y_t = (l_t - 1) y_t \\
 a_t^v &\equiv c_v (m - F_t) \\
 a_t &\equiv 1 - c_v (m - F_t) - \frac{K}{F_t - K} y_t = 1 - c_v (m - F_t) - (l_t - 1) y_t
\end{align*}
$$

(83)

11.5.2 Simplified Model Properties

11.5.2.1 The Dynamics of $y_t$ and $z_t$

\footnote{In the Appendix, we give an in depth discussion about the effect that the $X_t$ approximation has on various model properties.}
We now proceed to derive the dynamics of $z_t$ and $y_t$ for the simplified model. If $V_{t+1} \geq K$, the dynamics of U’s ROE are:

$$z_{t+1} = a_i^V V_{t+1} + a_i^D X_t + a_i F_t$$

$$= a_i^V V_{t+1} + \frac{b_t}{w_t^u} X_t + a_i F_t$$

$$= a_i^V V_{t+1} + \frac{(w_t^L + b_t) K}{w_t^u} + (1 - a_i^D - a_i^V) F_t$$

$$= a_i^V V_{t+1} + (1 - a_i^V) F_t + y_t K + a_i^D K - a_i^D F_t$$

$$= a_i^V V_{t+1} + (1 - a_i^V) F_t + y_t K + (l_t - 1) y_t K - (l_t - 1) y_t F_t$$

$$= a_i^V V_{t+1} + (1 - a_i^V) F_t + \frac{F_t}{F_t - K} y_t K - \frac{K}{F_t - K} y_t F_t$$

$$= a_i^V V_{t+1} + (1 - a_i^V) F_t$$

(84)

where in the third line we have used the fact that L’s VaR constraint is tight, i.e. $b_t X_t = K (w_t^L + b_t)$. Thus, U’s ROE is a convex combination of the exogenous return $V_{t+1}$ and of $F_t$. We now derive the dynamics of $y_t$, by starting from the observation that $y_{t+1} z_{t+1} = \frac{w_{t+1}^L}{w_t^u}$. Thus:

$$y_{t+1} z_{t+1} = \frac{w_{t+1}^L}{w_t^u}$$

$$= \frac{w_t^L + b_t}{w_t^u} V_{t+1} - \frac{b_t X_t}{w_t^u}$$

$$= y_t V_{t+1} + a_i^D V_{t+1} - \frac{b_t X_t}{w_t^u}$$

$$= y_t V_{t+1} + (l_t - 1) y_t V_{t+1} - \frac{(w_t^L + b_t) K}{w_t^u}$$

$$= y_t V_{t+1} + (l_t - 1) y_t V_{t+1} - y_t K - a_i^D K$$

$$= l_t y_t V_{t+1} - y_t K - (l_t - 1) y_t K$$

$$= l_t y_t V_{t+1} - l_t y_t K$$

$$= y_t l_t (V_{t+1} - K)$$

(85)

where again, we have made use of the VaR constraint tightness in the fourth line. Thus, $y_{t+1}$ is given by:

$$y_{t+1} = \frac{l_t (V_{t+1} - K)}{z_{t+1}} = \frac{l_t (V_{t+1} - K)}{a_i^V V_{t+1} + (1 - a_i^V) F_t}$$

(86)

84
The relative size of $L$ to $U$ in the next period increases (decreases) when $L$'s ROE of $d_{t+1} \equiv l_t(V_{t+1} - K)$ is higher (lower) than $U$'s ROE of $z_{t+1} \equiv a_t^V V_{t+1} + (1 - a_t^V)F_t$.

The above derivations were for the case where $V_{t+1} \geq K$. If at some time $V_{t+1} < K$, then $L$ defaults and $y_{t+1} = 0$.

The nice feature of the simplified model is that if $F_t$ is constant, then $L$'s and $U$'s ROEs are IID since they are only a function of the contemporaneous $V$ outcome; this is not the case in the full model due to the fact that leverage varies with $y_t$.

### 11.5.2.2 Relative Size Growth Barrier

In this section, we elaborate on the process $y_t$ in the simplified model. Firstly, it should be noted that when $d_{t+1} = z_{t+1}$, the relative size stays constant. This happens if $V_{t+1} = F_t$.

To see why:

\[
\begin{align*}
  d_{t+1} = z_{t+1} \Rightarrow l_t(V_{t+1} - K) &= a_t^V V_{t+1} + (1 - a_t^V)F_t \\
  V_{t+1} &= \frac{l_t K + (1 - a_t^V)F_t}{l_t - a_t^V} \\
  V_{t+1} &= \frac{F_t K}{F_t - K} + (1 - a_t^V)F_t \\
  V_{t+1} &= F_t
\end{align*}
\]

(87)

Thus, it is straightforward to see that:

\[
y_{t+1} \begin{cases} > y_t & \text{if } V_{t+1} > F_t \\ = y_t & \text{if } V_{t+1} = F_t \\ < y_t & \text{if } V_{t+1} < F_t \end{cases}
\]

(88)

Thus, the hurdle for $y_t$ to grow is for $V$ to be larger than the risk free rate; in contrast to the full model, this hurdle does not depend on $y_t$ and thus makes the analysis more simple.

We can also see that $y_{t+1}$ is an increasing function of $V_{t+1}$ since:

\[
\frac{dy_{t+1}}{dV_{t+1}} = \frac{l_t z_{t+1} - l_t (V_{t+1} - K)a_t^V}{z_{t+1}^2} = \frac{l_t (1 - a_t^V)F_t + a_t^V K}{z_{t+1}^2} > 0
\]

(89)

### 11.5.2.3 Solution Dependence on Relative Size

From the relationships in (83), it is clear that $a^D_t$, $b_t$, and $X_t$, are increasing in the relative size $y_t$. This is exactly the same as for the full model case.

However, an interesting aspect of the simplified solution is that leverage does not depend on $y_t$; if $F_t$ is constant, so is $l_t$. Clearly, if leverage is not dependent on $y_t$, the effects of
L’s interaction with U will be more pronounced than in the full model case, where leverage is (slightly) decreasing in $y_t$. Thus, the result that demand is pro-cyclical in the simplified model is expected since we have already seen that this is the case in the full model.

11.5.2.4 Solution Dependence on Outcome

Since $y_{t+1}$ is an increasing function of the outcome $V_{t+1}$, so are $a_{t+1}^D$, $X_{t+1}$, and $b_{t+1}$ since they are increasing in $y_{t+1}$. This is the same as in the full model case. Leverage however depends only on the risk free rate and is thus not a function of $V_{t+1}$.

11.5.2.5 Simplified dynamics of $P_t$

The simplified solution variables in (83) allow us to simplify the expression for $P_t$ in (78) when the risk free rate $F_t = F$ is constant. Expanding the second term on the RHS of (78):

$$a_tF - a_{t+1}z_{t+1} = (1 - a^v)F - a_t^D F - (1 - a^v)z_{t+1} + a_{t+1}^D z_{t+1}$$

$$= (1 - a^v)(F - z_{t+1}) + (l - 1)d_{t+1}y_t - (l - 1)Fy_t$$

$$= (1 - a^v)a^v(F - V_{t+1}) + (l - 1)y_t(d_{t+1} - F)$$

$$= (1 - a^v)a^v(F - V_{t+1}) + (l - 1)y_t(V_{t+1} - F)$$

$$= (V_{t+1} - F)[(l - 1)y_t - (1 - a^v)a^v]$$

Thus:

$$P_{t+1} = P_tV_{t+1} + w_t^a(V_{t+1} - F)[(l - 1)y_t - (1 - a^v)a^v]$$

11.6 Metrics of Interest

We now proceed to define the two metrics of interest for the system which we will focus on.

11.6.1 Investment Deviation From Fundamentals

As discussed in Section 11.2, it is desirable for (re)investment to reflect fundamentals. Mathematically, the deviation from fundamentals at time $t$, $\Delta_t$, is defined as:

$$\Delta_t = P_t - S_t$$

In other words, there is a deviation from fundamentals when the amount reinvested in the economy either includes a net inflow from the risk free asset (positive deviation from fundamentals) or a net outflow to the risk free asset (negative deviation from fundamentals). The only way that investment will reflect fundamentals is if at each time, there are no “additional” financial flows between the financial system and the producers. To relate to this
intuitively, think of this situation as buying and holding a stock; everyday, its value changes based on exogenous (with respect to you) factors. On a given day, the stock has a value; on that day, you could sell some of your ownership and put the money in the bank or you could take some money from your bank account and buy some more. Both these situations involve “additional” flows; having investment reflect fundamentals in the economy is intuitively like you not doing anything on that day and thus implicitly “reinvesting” an amount equal to the value of the stock.

With the help equation (78), it is clear that the time $t+1$ deviation from fundamentals, $\Delta_{t+1}$, is increasing in $V_{t+1}$ since $-a_{t+1}$ and $z_{t+1}$ are increasing in $V_{t+1}$.

### 11.6.2 Risk

By the term risk, we refer to the potential systemic cost of a default by L. If L is very small and defaults, the repercussions intuitively should be less severe than if L is very large. In our model, if L defaults, then U seizes control of L’s assets, whose value is smaller than the amount owed to U by L. In the subsequent period, L does not exist, and the system is reduced to one involving only U (see Section 11.4.6). The post default period investment will drop due to the fact the proceeds that U is able to recover from L will be shifted to the risk free asset. The potential systemic cost is not something that is realized unless L defaults; nevertheless, it is a quantity that can be observed at each time $t$, and its meaning would be: “what would happen to investment tomorrow if there was an exogenous return of $V_b$ causing L to default”? The system is more “risky” when this potential systemic cost is high, as future risky asset investment will be low.

Assume we are at time $t$ and that a bad outcome $V_b < K$ occurs, causing L to go bankrupt. The time $t+1$ post default investment amount is:

$$P_{t+1}^b = a_{t+1}^v w_{t+1}^u = a_{t+1}^v w_{t+1}^u ( (a_t^v + a_t^D + y_t) V_b + a_t F_t)$$

where we have used equation (54) to determine the return of U’s risky debt holding. We can also rewrite $P_t$ as:

$$P_t = w^L_t + (a_t^v + a_t^D) w^u_t = w^u_t (y_t + a_t^v + a_t^D)$$

Thus, assuming that the risk free rate is constant, the post default investment $P_{t+1}^b$ as a percentage of the time $t$ investment is:

$$\frac{P_{t+1}^b}{P_t} = a^v \left( V_b + \frac{w^u_t a_t F_t}{P_t} \right) = a^v \left( V_b + \frac{a_t F_t}{y_t + a_t^v + a_t^D} \right) = a^v \left( V_b + \frac{(1 - a_t^v - a_t^D) F_t}{y_t + a_t^v + a_t^D} \right)$$

The larger the second term in the parenthesis in equation (94) is, the smaller the impact of a default by L is. We define the time $t$ risk $R_t$ as:

---

We use the terms “risk” and “potential systemic cost” interchangeably. Think of “risk” as a term reflecting post default investment as opposed to a term reflecting the probability of a default.
Now, let's examine how the potential systemic cost could build up under a constant risk free rate environment. If \( y_t \) grows, so does \( X_t \), meaning that \( a^D_t \) increases and \( a_t \) decreases. Thus, the numerator of \( R_t \) grows and its denominator decreases, thus making the overall potential systemic cost increase. The extreme case occurs as \( a_t \) goes to 0, sending \( R_t \) to infinity. This means that the post default investment as a percent of \( P_t \) will be \( a^V V_b \). To put this into perspective, let's use some reasonable annualized numbers for a back of the envelope calculation. Assume we are in a low rate environment (\( F_t = 1.02 \)) and a relative size of \( y_t = 1\% \) is large enough to attract a weight of \( a^D_t = 20\% \); furthermore, assume U's direct investment \( a_v = 30\% \), and the bad outcome \( V_b \) is a modest 0.9, which is below \( K \) (i.e. a 10\% drop from \( t \) to \( t + 1 \)). Then, \( \frac{P_{t+1}}{P_t} \approx 83\% \); in other words, the post default investment drops by 17\%!

Figure 26 illustrates our concept of risk. In both panels, U has the same \( a^v \). However, the left panel has a low potential systemic cost since \( y_t \) is small and thus U invests only a modest amount of his portfolio into L's debt. The right panel has high potential systemic cost since \( y_t \) is large and thus U invests a large amount of his portfolio into L's debt.

Figure 26: Graphical illustration of risk. The left panel system, where \( y_t \) is small, has a lower potential systemic cost than the right panel system, where \( y_t \) is large and L has attracted a larger proportion of U's portfolio into his risky debt.

Clearly, the important determinant of risk in our model is the relative size of L, \( y_t \). It should be noted that the risk at \( t + 1 \) is increasing in \( V_{t+1} \) since both \( y_{t+1} \) and \( a^D_{t+1} \) are increasing in \( V_{t+1} \).

\[
R_t = \frac{P_t}{w^a_t a_t F_t} = \frac{y_t + a^v_t + a^D_t}{a_t F_t} = \frac{y_t + a^v_t + a^D_t}{(1 - a^v_t - a^D_t)F_t}
\]

38 We use annualized numbers for simplicity; in reality, when looking at daily or weekly frequencies, we would use realistic numbers for those frequencies. The simulation results presented later on are done assuming a weekly frequency.
12 Key Results

In this section, we present the key results regarding the evolution of investment deviation from fundamentals and risk in our model. We focus on the case where the risk free rate is constant so as to understand the natural evolution of our metrics. As discussed in Section 3, the FFR is constant over long periods of time compared to the assumed frequency the decision making by L and U. A constant risk free rate implies that \( F_t = F_{t+1} = F \) and thus \( a_t^v = a_{t+1}^v = a^v \). For the relationships we derive that hold true in general, we do not drop the time subscript. It should be noted that for many cases, closed form solutions can only be derived for the simplified model case. Thus, also present simulations which compare the full model to the simplified model.

In what follows, Section 12.1 presents the expected growths of U and L, Section 12.2 presents the expected evolution of investment and its deviation from fundamentals, Section 12.3 presents the expected evolution of risk, and Section 12.4 presents simulation results comparing the full model results with the simplified model results.

12.1 The Expected Growth of U and L

Because we are interested in examining the system evolution due to L’s behavior, we focus on expectations conditional on no default. Overall in this section, to simplify notation, we do not explicitly annotate the expectation operator as conditional; all expectations should be assumed to be conditioned on no default occurring unless explicitly stated otherwise.

12.1.1 One Step Case

For both the full and simplified model, the one step expected growth of L is higher than U. Denote the non-default conditional mean of \( V \) as \( m' \); clearly, this is (slightly) larger than the unconditional mean of \( V \), \( m \).

If \( F_t < m \), then conditional on no default occurring, L’s expected growth is higher than U’s since:

\[
E_t \left[ d_{t+1} \right] = E_t \left[ \frac{w_t^L V_{t+1} + b_t (V_{t+1} - X_t)}{w_t^L} \right] > m'
\]

\[
E_t \left[ z_{t+1} \right] = E_t \left[ a_t^v V_{t+1} + a_t^D X_t + a_t F_t \right] < m'
\]

since \( X_t \leq \frac{m + F_t}{2} < m < m' \).

We have already seen that the relative size of L to U, \( y_t \), is very important for investment and risk. Since we have not assumed a particular distribution on \( V \), evaluating the conditional on no default \( E_t [y_{t+1}] = y_t E_t \left[ \frac{d_{t+1}}{z_{t+1}} \right] \) is not feasible. However, a first order approximation of \( E_t \left[ \frac{d_{t+1}}{z_{t+1}} \right] \) via the “Delta method” is in fact \( E_t \left[ \frac{d_{t+1}}{z_{t+1}} \right] \), which is greater than 1.
This is an increasingly accurate approximation as the mass of the V distribution becomes more concentrated around m.

12.1.2 n-Step Case

For the simplified model, we now proceed to derive the n step forward looking expectation of relative size when the risk free rate is constant at $F_t = F$ over the interval of interest.

$$E_t[y_{t+n}] = y_t E_t \left[ \prod_{i=1}^{n} \frac{d_{t+i}}{z_{t+i}} \right] = y_t \left( E_t \left[ \frac{d_{t+i}}{z_{t+i}} \right] \right)^n > y_t$$

since $E_t \left[ \frac{d_{t+i}}{z_{t+i}} \right] > 1$. Note that the expectation of the product becomes the product of expectations because in the simplified model, the term $\frac{d_{t+i}}{z_{t+i}}$ is IID since it’s just a function of $V_{t+i}$, which is IID.

For the full model, deriving a closed form solution for $E_t[y_{t+n}]$ is further complicated by the fact that the ratios $\frac{d_{t+i}}{z_{t+i}}$ and $\frac{d_{t+j}}{z_{t+j}}$ (where $1 \leq i, j \leq n$) are correlated. However, later in this section, we present simulation results and compare them to the simplified model.

12.2 Expected Future Investment and Deviation from Fundamentals

In this section, we explore the investment deviation from fundamentals and its expected evolution for the one step and n–step cases for the simplified model.

12.2.1 Deviation From Fundamentals Magnitude

From the $P_t$ dynamics in (91), it is clear that there are two cases. If the outcome is “good”, meaning $V_{t+1} > F_t$, then $\Delta_{t+1}$ is increasing in $y_t$; if the outcome is “bad”, meaning $V_{t+1} < F_t$, then $\Delta_{t+1}$ is decreasing in $y_t$. Define $y^c$ as:

$$y^c = \frac{(1 - a^v)a^v}{(l - 1)l} \quad (97)$$

Note that $y^c$ is constant since $F$ is constant. If $y_t > y^c$, then a good outcome will cause a positive deviation from fundamentals, which is increasing in $y_t$ and $V_{t+1}$; analogously, a bad outcome will cause a negative deviation from fundamentals, which becomes more negative as $y_t$ increases or as $V_{t+1}$ becomes worse.

Overall, if the relative size is above a critical value, the magnitude of the deviation from fundamentals will be increasing in $y_t$. A good outcome will cause $a_{t+1}$ to decrease such that there is a net inflow of funds from the risk free asset; a bad outcome will cause $a_{t+1}$ to increase such that there is a net outflow of funds to the risk free asset. These cases are depicted in the top panel of Figure 27.
The situation where $y_t < y^c$ is more subtle; here, $y_t$ is very small, which means that $L$’s behavior does not have an important effect on the system and the system resembles one with $U$ only in the limit. When $V_{t+1} > F_t$, $L$’s growth is not enough to prevent a net outflow of funds to the risk free asset; $a_{t+1}$ does not drop enough so as to prevent this. The analogous statement can be made when $V_{t+1} < F_t$; $a_{t+1}$ does not increase enough so as to prevent a net inflow of funds from the risk free asset. These cases are depicted in the bottom panel of Figure 27.

Table 3 summarizes the possible scenarios:

<table>
<thead>
<tr>
<th>Outcome/State</th>
<th>$y_t &gt; y^c$</th>
<th>$y_t &lt; y^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{t+1} &gt; F_t$</td>
<td>A</td>
<td>C</td>
</tr>
<tr>
<td>$\Delta_t &gt; 0$</td>
<td>$</td>
<td>\Delta_t</td>
</tr>
<tr>
<td>$</td>
<td>\Delta_t</td>
<td>$ increasing in $y_t$</td>
</tr>
<tr>
<td>$V_{t+1} &lt; F_t$</td>
<td>B</td>
<td>D</td>
</tr>
<tr>
<td>$\Delta_t &lt; 0$</td>
<td>$</td>
<td>\Delta_t</td>
</tr>
<tr>
<td>$</td>
<td>\Delta_t</td>
<td>$ decreasing in $y_t$</td>
</tr>
</tbody>
</table>

Table 3: This table depicts the four possible deviation from fundamentals scenarios.
Figure 27: Graphical depiction of the mechanism whereby investment deviates from fundamentals for the case where $y_t > y^*$ (top panel) and $y_t < y^*$ (bottom panel).
12.2.2 One Step Expected Investment

Assume that we are currently at time \( t \), and that the risk free rate \( F_t = F \) is constant. Starting from the fact that total investment at a given time is equal to the sum of L and U’s investment at that time:

\[
P_{t+1} = w_t^L + w_t^u(a^v + a_t^D) \\
= w_t^L d_{t+1} + w_t^u z_{t+1} a^v + w_t^u z_{t+1} a_t^D
\]  \( (98) \)

Since \( a_t^D = (l - 1)y_{t+1} \), \( P_{t+1} \) can be expressed as:

\[
P_{t+1} = w_t^L d_{t+1} + w_t^u z_{t+1} a^v + w_t^u z_{t+1}(l - 1)y_{t+1} \\
= w_t^L d_{t+1} + w_t^u z_{t+1} a^v + w_t^u z_{t+1}(l - 1)y_{t+1} d_{t+1} \\
= w_t^L d_{t+1} + w_t^u z_{t+1} a^v + w_t^L (l - 1)d_{t+1} \\
= w_t^L l d_{t+1} + a^v w_t^u z_{t+1}
\]  \( (99) \)

In a similar manner, \( P_t \) can be expressed as:

\[
P_t = w_t^L + w_t^u a^v + w_t^u a_t^D \\
= w_t^L + w_t^u a^v + w_t^u (l - 1)y_t \\
= w_t^L + w_t^u a^v + w_t^L (l - 1) \\
= w_t^L l + w_t^u a^v
\]  \( (100) \)

Thus, since \( E_t[d_{t+1}] > m' > E_t[z_{t+1}] \):

\[
E_t[z_{t+1}] \leq E_t \left[ \frac{P_{t+1}}{P_t} \right] \leq E_t[d_{t+1}]
\]  \( (101) \)

Thus, investment in the next period will grow faster than U’s expected ROE and slower than L’s expected ROE.

To isolate the marginal effect of L’s growth sucking in a higher percentage of U’s funds, consider an identical economy at time \( t \) where U does not adjust \( a^D \) in response to L’s relative size; mathematically, this would mean \( a_t^D = a_t^D \). In this alternative setup, the expected next period investment would be below expected fundamentals and lower than the expected investment discussed above. Denote the investment under this alternative setup as \( P_t^a \). Then:
\[ P_{t+1} - P_{t+1}^a = w_t^L d_{t+1} + w_t^u z_{t+1} a^v + w_{t+1}^u z_{t+1+1} a^D - (w_t^L d_{t+1} + w_{t+1}^u z_{t+1} a^v + w_{t+1}^u z_{t+1} a^D) \]
\[ = w_t^u z_{t+1} (a_{t+1}^D - a_t^D) \]  
(102)

Since all time \( t \) quantities are the same; the only difference is \( U \)'s time \( t+1 \) response. Thus:

\[ E_t [P_{t+1} - P_{t+1}^a] = w_t^u E_t (z_{t+1} a_{t+1}^D - z_{t+1} a_t^D) \]
\[ = w_t^u E_t ((l_t - 1) z_{t+1} y_{t+1} - (l_t - 1) z_{t+1} y_t) \]
\[ = w_t^u E_t ((l_t - 1) d_{t+1} y_t - (l_t - 1) z_{t+1} y_t) \]
\[ = w_t^u (l_t - 1) y_t E_t (d_{t+1} - z_{t+1}) \]
\[ > 0 \]  
(103)

Where the last inequality follows from (96).\(^{39}\)

Thus, in the alternative setup, the expected investment is lower. This makes sense intuitively if one considers that since the risky investment fundamental value is expected to grow at a rate higher than the risk free rate, then the expected reinvestment should be higher when \( a^D \) varies pro-cyclically.

### 12.2.3 One Step Expected Deviation from Fundamentals

Using the investment dynamics expression in (91):

\[ E_t [P_{t+1} - S_{t+1}] = w_t^u (m^F - F) [(l - 1) y_t - (1 - a^v) a^v] \]  
(104)

Thus, the expected deviation from fundamentals next period is an increasing function of the relative size \( y_t \). Furthermore, if \( y_t > y^F \), it is expected to be positive.

In the alternative setup, the investment in the next period is below fundamentals. Using the expression in (78) and setting \( a_{t+1} = a_t \):

\[ P_{t+1}^a = P_t V_{t+1} + a_t w_t^u (F_t - z_{t+1}) \Rightarrow \]
\[ E_t [P_{t+1}^a] = P_t m^F + a_t w_t^u (F_t - E_t z_{t+1}) \Rightarrow \]
\[ E_t [P_{t+1}^a] < P_t m^F \]  
(105)

Where the last inequality follows from (96).

\(^{39}\)Note that (103) holds in the full model when \( E[a_{t+1}^D] > E[a_{t+1}^D] \); the benefit of the simplified model is that the change in \( a^D \) is linearly related to the change in \( y \), allowing for a closed form expression.
Thus, the absence of U’s reallocation in response to L’s size causes expected investment in the next period to be lower than it is in our model; furthermore, it is below \( m' \), which is the expected fundamental return conditional on no default. Thus, in this alternative setup, on average, funds are divested from the risky investment.

12.2.4 \( n \)-Step Expected Investment

Analogously to the one step derivations for the simplified model, we can derive the \( n \) step forward looking expectation when the risk free rate is constant at \( F_t = F \) over the interval of interest.

\[
P_{t+n} = w_t^L + w_t^u (1 - a_{t+n}) = w_t^L \prod_{i=1}^{n} d_{t+i} + w_t^u \prod_{i=1}^{n} z_{t+i} (1 - a_{t+n})
\]

(106)

The interesting behavior is how investment is expected to evolve when L is constrained by the VaR. Under this case, we can expand \( a_{t+n} \):

\[
a_{t+n} = 1 - a^v - (l - 1)y_{t+n}
\]

\[
= 1 - a^v - (l - 1)\frac{d_{t+n}}{z_{t+n}}y_{t+n-1}
\]

\[
= 1 - a^v - (l - 1)\sum_{i=1}^{n} d_{t+i} y_i
\]

(107)

Thus:

\[
\prod_{i=1}^{n} z_{t+i} a_{t+n} = (1 - a^v) \prod_{i=1}^{n} z_{t+i} - y_l (l - 1) \prod_{i=1}^{n} d_{t+i}
\]

(108)

With this, we can expand the RHS of equation (106):

\[
P_{t+n} = w_t^L \prod_{i=1}^{n} d_{t+i} + w_t^u \prod_{i=1}^{n} z_{t+i} (1 - a_{t+n})
\]

\[
= w_t^L \prod_{i=1}^{n} d_{t+i} + w_t^u \prod_{i=1}^{n} z_{t+i} - w_t^u \prod_{i=1}^{n} z_{t+i} a_{t+n}
\]

\[
= w_t^L \prod_{i=1}^{n} d_{t+i} + w_t^u \prod_{i=1}^{n} z_{t+i} - w_t^u \left[ (1 - a^v) \prod_{i=1}^{n} z_{t+i} - y_l (l - 1) \prod_{i=1}^{n} d_{t+i} \right]
\]

95
Taking expectations on both sides gives:

\[
E_t[P_{t+n}] = w_t^L l \prod_{i=1}^{n} d_{t+i} + a^v w_t^U \prod_{i=1}^{n} z_{t+i}
\]

where \( \bar{d} \) and \( \bar{z} \) are just shorthand notations of the expectations of L’s and U’s ROEs respectively; as discussed in Section 12.1, the expectation of the product becomes the product of the expectations because \( d_{t+i} \) and \( z_{t+i} \) are IID variables in the simplified model.

Thus, the expected investment at horizon \( n \), \( E_t[P_{t+n}] \), is expected to grow faster than U’s expected ROE \( \bar{z} \) but less than L’s expected ROE \( \bar{d} \):

\[
\bar{z} E_t[P_{t+n-1}] \leq E_t[P_{t+n}] \leq \bar{d} E_t[P_{t+n-1}]
\]

Note that since \( \bar{z} > \bar{d} \), this means that the expected future one step investment increases faster than the risk free rate. Even more interesting is the fact that since \( \bar{d} > m \), the expected investment may grow at a rate greater than that implied by its expected fundamental return.

12.2.5 \( n \)-Step Expected Deviation From Fundamentals

The \( n \)-step expected deviation from fundamentals depends on the growth rate of expected investment. Define as \( g_{t+n} \) the growth rate of the expected investment between times \( (t + n - 1) \) and \((t + n)\). It is given by:

\[
g_{t+n} \equiv \frac{E_t[P_{t+n}]}{E_t[P_{t+n-1}]} = \frac{w_t^L \bar{d}^n + a^v w_t^U \bar{z}^n}{w_t^L \bar{d}^{n-1} + a^v w_t^U \bar{z}^{n-1}}
\]

It is straightforward to show that \( \frac{dg_{t+n}}{dn} > 0 \) by using the fact that \( \bar{z} < \bar{d} \). This means that the growth rate of expected investment grows over time. Thus, in the simplified model, as time progresses without any defaults occurring, the investment is expected to grow faster than U’s ROE and this growth rate is itself growing.

Recall that the deviation from fundamentals at time \( t + n \), which we denote as \( \Delta_{t+n} \), is defined as:

\[40\] Again, this stems from the assumption \( F_t < m \).
\[
E_t[D_{t+n}] = E_t[P_{t+n} - P_{t+n-1}V_{t+n}] = E_t[P_{t+n}] - E_t[P_{t+n-1}m']
\]  
(113)

where we have utilized the fact that the time \( t + n - 1 \) is independent of the time \( t + n \) fundamental outcome \( V_{t+n} \). Since \( g_{t+n} \) is increasing in \( n \), so is \( E_t[D_{t+n}] \); thus, although the deviation from fundamentals may start out negative, it will become positive after some time.

At time \( t \), we can find the time horizon \( n \) where this deviation is expected to go positive:

\[
E_t[D_{t+n}] > 0 \Rightarrow \\
E_t[P_{t+n}] > E_t[P_{t+n-1}m'] \Rightarrow \\
w_t l \bar{d}^n + a^v w_t \bar{z}^n > \left( w_t l \bar{d}^{n-1} + a^v w_t \bar{z}^{n-1} \right) m' \Rightarrow \\
w_t l \bar{d}^{n-1} (\bar{d} - m') > a^v w_t \bar{z}^{n-1} (m' - \bar{z}) \Rightarrow \\
w_t l \bar{d}^{n-1} \frac{K}{F - K} (m - F) > a^v w_t \bar{z}^{n-1} (1 - a^v)(m - F) \Rightarrow \\
w_t l (l - 1) \bar{d}^{n-1} > a^v (1 - a^v) w_t \bar{z}^{n-1} \Rightarrow \\
y_t > \frac{a^v (1 - a^v) \bar{d}^{n-1}}{l(l - 1) \bar{z}^{n-1}}
\]  
(114)

Thus, the time \( n \) where the deviation is expected to first swing into positive territory is the smallest \( n \) for which the above inequality holds. Note that since the RHS of the above is decreasing in \( n \), the condition will eventually be true.

Before continuing, one should keep in mind that the simplified solution is reflective of the full model solution when \( y \) is small. As \( y \) stops being small, the simplified solution overstates the amount that \( L \) borrows and thus expected future investment; in the full model, a point would be reached where \( L \) can’t borrow as much as is needed to sustain the dynamics of \( E_t[P_{t+n}] \) described above; of course, that happens after \( L \) has grown relatively large and thus has attracted a significant portion of \( U \)'s wealth. In any case, while the above relationships hold for the simplified solution, their accuracy with respect to the full model decreases on average as \( n \) increases.

### 12.3 Expected Risk

In this section, we show that the expected potential systemic cost is expected to grow over time in the simplified model. We first present the one step case and then proceed to the \( n \)-step case.

#### 12.3.1 One Step Expected Risk

Assume that we are currently at time \( t \), and that the risk free rate \( F_t = F \) is constant. Starting from the expression for potential systemic cost in equation (95) and using the fact
that $a_t^D = (l - 1) y_t$ in the simplified model:

$$R_t = \frac{y_t + a^v + a_t^D}{(1 - a^v - a_t^D)F} = \frac{ly_t + a^v}{(1 - a^v)F - (l - 1)y_t F}$$  \hspace{1cm} (115)$$

We already have established in Section 11.6 that risk is an increasing function of $y_t$. Although we cannot evaluate $E_t[R_{t+1}]$ in closed form, we can show that it is higher than $R_t$.

$$E_t[R_{t+1}] = E_t \left[ \frac{ly_{t+1} + a^v}{(1 - a^v)F - (l - 1)y_{t+1} F} \right]$$

$$= \frac{1}{E_t[(1 - a^v)F - (l - 1)y_{t+1} F]} \left[ \frac{ly_{t+1} + a^v}{(1 - a^v)F - (l - 1)y_{t+1} F} \right] - \frac{\text{cov}_t \left( \frac{ly_{t+1} + a^v}{(1 - a^v)F - (l - 1)y_{t+1} F}, (1 - a^v)F - (l - 1)y_{t+1} F \right)}{E_t[(1 - a^v)F - (l - 1)y_{t+1} F]}$$

$$= \frac{E_t[(1 - a^v)F - (l - 1)y_{t+1} F]}{E_t[1 - a^v] + a^v}$$

$$> \frac{1}{(1 - a^v)F - (l - 1)E_t[y_{t+1}] F}$$

$$= R_t$$  \hspace{1cm} (116)$$

where we have used the well known identity that $E \left[ \frac{A}{B} \right] = \frac{E[A]}{E[B]} - \text{cov} \left( B, \frac{A}{B} \right)$ in the second line. Furthermore, we have used the fact that:

$$\text{cov}_t \left( \frac{ly_{t+1} + a^v}{(1 - a^v)F - (l - 1)y_{t+1} F}, (l - 1)y_{t+1} F \right) > 0$$  \hspace{1cm} (117)$$

since both functions are increasing in $V_{t+1}$. Furthermore, we used that $E_t[y_{t+1}] > y_t$.

Thus, the one step expected risk higher than the current risk. This result is intuitive since $y_t$ is expected to grow in the next period.

12.3.2 $n$-Step Expected Risk

The $n$ step result comes naturally from the one step result in the previous section. Assuming that the risk free rate is constant between times $t$ and $t+n$, the process $R_t$ is a submartingale and thus $E_t[R_{t+n}] > R_t$. Intuitively, this is clear: if $y_t$ is expected to increase in the future, then the ratio of total system funds invested in the risky asset over those invested in the risk free asset should decrease as $L$, who is growing faster than $U$, attracts more of $U$’s wealth to his risky debt.
12.4 Simulation Results for Full and Simplified Model

In this section, we present simulation results in order to compare the behavior of the full and simplified models. We first give an overview of the simulations and then present some simulation results.

12.4.1 Simulation Overview

As the goal is to compare the two solutions under the assumptions presented in this section, we initialize the models with a “small” $y_0$ and observe the average evolution of the key variables over time for a constant risk free rate. For each simulation, we use 2000 trials and 150 time periods, where each time period represents 1 week. We use parameter values which reflect this frequency of decision making; all rates of return presented are not annualized unless stated.

For each simulation, we discard the trials for which a default occurs as our aim is to examine the behavior of the system when $L$ exists and is adhering to his VaR constraint. However, as part of our risk results, we depict the expected cost of a default at each time were a default to occur at that time. After discarding the default trials, we average the results of the remaining trials in the cross section in order to estimate the expectation of each variable. It should be noted that in each simulation, very few trials are discarded due to default since the imposed probability of bankruptcy is on the order of $[0.0001, 0.00001]$.

In terms of the process $V$, we use a log-normal random variable due to its ubiquitous use in asset modeling in finance and economics (see for example [18]). In our model description, we have not assumed a particular distribution and we have seen that the evolution depends primarily on the relationships between key parameters such as $m$, $s$, $F$, and $K$. Thus, as long as the parameters of the distribution satisfy the constraints discussed in Sections 11 and 12, our results should not differ qualitatively. We ran simulations for two other reasonable distribution choices, triangular and uniform, and found that the results did not differ qualitatively.

With respect to variable initialization, we always start U’s wealth at 1; the initial $y$ values are on the order of .001%; we choose these values so that we can observe whether the two models are close under the assumption of small $y$ and because they are consistent with where the relative size of the leveraged sector would have been during the late 1980s, when the shadow banking system started to take off (see for example [6]). Of course, depending on the the simulation environment, the relative size will grow with different speed and have different implications for risk and investment.

12.4.2 Additional Variables Definitions

In the simulations that follow, there are some additional variables that we depict either explicitly or to compare to a given result. We present definitions of these here.

- **Cost:** We present a measure called “cost”, which is what percent change in investment \( \frac{P_{t+1}}{P_t} - 1 \) would occur at time $t+1$ if there were a bad outcome $V_{t+1}^{b} = 0.9$ causing a
Simulation Parameters | m | s | F | y₀ | c
--- | --- | --- | --- | --- | ---
Simulation 1 | 8% | 4% | 2% | 0.001% | 0.001%
Simulation 2 | 8% | 4% | 3% | 0.001% | 0.001%
Simulation 3 | 8% | 7% | 2% | 0.001% | 0.001%

Table 4: This table depicts the parameter values used for the 3 simulations we present. All quoted rates are annualized.

default. This is given by equation (94) and is related to risk buildup.

- **Inverse Risk:** Instead of presenting risk as defined in (95), we present its inverse (i.e. \(\frac{a\alpha F}{y_1 + a^2 + a^3 y}\)), which is directly related to the cost metric discussed above. The maximum value that the inverse risk can attain is that where \(y_t = 0\), and we plot this for comparison; it is called “Safest” on the respective plots.

- **Crossover \(y\):** This is simply \(\bar{y}\) as presented in equation (72).

### 12.4.3 Simulation Results

We now proceed to present some simulation results. Clearly, there are many parameters one could play around with and hundreds of pages could be filled with plots; however, we pick the ones we find most illustrative and compare them. Each simulation will depict 8 plots organized in 4 figures: the variables depicted (in order of presentation) will be investment, cost, \(a^D\), inverse risk, \(y\), leverage, \(X\); we also depict the pair \((\epsilon_1, \epsilon_2)\), which are variables that determine the accuracy of our linearized solution approximation and described in the Appendix. For our linearized solution approximation to be accurate, these epsilons should be much smaller than 1. Table 4 depicts the key parameters used for each simulation. Each of the following sections will present and discuss a simulation.

#### 12.4.3.1 Simulation 1: Baseline

We use this simulation as our base case and will change some parameters in subsequent simulations so as to compare and contrast. Figures 28, 29, 30, and 31, present the simulation results. The simplified model is very close to the full model, especially when \(y\) tends to be small. However, towards the end of the period, the simplified model \(y\) tends to grow more than the full model \(y\). This is due to the fact that in the full model, as time goes on and \(y\) tends to increase, leverage tends to decrease (see Figure 30); this is in contrast to the simple model, where leverage does not depend on \(y\) and thus is constant over time. Thus, in the simplified model, there are more extreme outcomes of \(y\) at longer horizons. The \(\epsilon_1\) plot (Figure 31, bottom panel) is a testament to the fact that the approximation on average is getting worse with time; recall that \(\epsilon_1 << 1\) for the simplified model to track the full model well. However, \(\epsilon_2\) stays small throughout, meaning that the differential between \(X\) and \(F\) remains small, and this can be seen in the top panel of Figure 31.
In terms of average evolution, although investment is below its fundamental path in the beginning,\textsuperscript{41} as $y$ tends to grow, investment starts to increase above fundamentals and ultimately surpasses its fundamentals path. Furthermore, as $y$ tends to grow, L sucks in a higher percentage of U’s funds (Figure 29, top panel); this causes inverse risk to decrease over time on average and increases the average system cost of a 10\% drop in the fundamentals. As expected, for small $y$, the drop is less than fundamentals, but this drop becomes more severe as $y$ tends to increase: it reaches as low as $-17\%$ for the full model, which is far larger of a drop than the fundamental’s 10\% drop.

\textsuperscript{41}The fundamentals path is given by $P_t = P_0 * m_t$; if the average investment in the system has a higher slope than this path at a certain time, then average investment was above its fundamentals at that time and vice versa.
Figure 28: Simulation 1: Plot of the simulated average investment (top panel) and default cost (bottom panel) for the simplified and full models.
Figure 29: Simulation 1: Plot of the simulated average $a^D$ (top panel) and inverse risk (bottom panel) for the simplified and full models.
Figure 30: Simulation 1: Plot of the simulated average $y$ (top panel) and leverage (bottom panel).
Figure 31: Simulation 1: Plot of the simulated average $X$ (top panel) for the simplified and full models as well as the simplified model epsilon approximations (bottom panel).
12.4.3.2 Simulation 2: Lower Risk Free Rate

In this simulation, we lower the risk free rate to 1% (annualized), from 3% (annualized) which was the case in Simulation 1. In this case, since \( K \) is the same as in Simulation 1, we know that the simplified model leverage will be higher. Thus, we expect the simplified model to track the full model less well on average as time progresses. Furthermore, for both models, we expect faster \( y \) increases, higher \( a^D \), and higher default costs, compared to Simulation 1. Figures 32, 33, 34, and 35, present the Simulation 2 results.

The simplified model is close to the full model, especially when \( y \) tends to be small. However, towards the end of the period, the simplified model \( y \) tends to grow more than the full model \( y \) on average. As was the case in Simulation 1, there are more extreme outcomes of \( y \) at longer horizons due to the fact that the constant leverage assumption does not hold so well as \( y \) grows; the \( \epsilon_1 \) (Figure 35, bottom panel) shows that the approximation on average is worse than in Simulation 1.

For both models, compared to Simulation 1, the average investment rebounds from below the fundamentals path sooner; this is due to the fact that average \( y \) increases faster in this simulation. Also, \( a^D \) tends to grow faster (Figure 33, top panel) and inverse risk decreases over time (Figure 33, bottom panel). In this simulation, inverse risk starts out lower because a lower risk free rate cases \( a^v \) to be higher. Also, note that inverse risk drops more over time compared to Simulation 1 since every increase in \( a^D \) “counts” more now that \( a^v \) is higher. In terms of the cost, it is more severe on average than in Simulation 1 for all times; furthermore, it reaches as low as \(-21\%\) for the full model, which is far larger of a drop than the fundamental’s 10% drop.
Figure 32: Simulation 2: Plot of the simulated average investment (top panel) and default cost (bottom panel) for the simplified and full models.
Figure 33: Simulation 2: Plot of the simulated average $a^D$ (top panel) and inverse risk (bottom panel) for the simplified and full models.
Figure 34: Simulation 2: Plot of the simulated average $y$ (top panel) and leverage (bottom panel).
Figure 35: Simulation 2: Plot of the simulated average $X$ (top panel) for the simplified and full models as well as the simplified model epsilon approximations (bottom panel).
12.4.3.3 Simulation 3: Higher Variance

In this simulation everything is the same as in Simulation 1 except for the fact that we have raised the standard deviation of $V$. Intuitively, this will cause $L$ to be more risk averse because $K$ will decrease due to the increased standard deviation. Note that although we have assumed that the standard deviation is higher, we have kept $c_v$ constant because we want to isolate the effect of $L$ being more “risk averse”. Figures 36, 37, 38, and 39 present the Simulation 3 results.

The simplified model is close to the full model overall as $y$ does not tend to grow very fast; indeed, the $\epsilon_1$ (Figure 39, bottom panel) shows that the approximation on average is very good as it remains much smaller than 1.

For both models, compared to Simulation 1, average investment remains below its fundamentals path. Only towards the end does the slope of the average investment curve increase above the slope of the fundamental’s path, meaning that investment is growing above fundamentals on average towards the end. This is due to the fact that the average $y$ is much lower than in Simulation 1; a lower average $y$ also results in lower $a^D$ (Figure 37, top panel) and a smaller decrease of inverse risk (Figure 37, bottom panel). In terms of the cost, it is less severe on average than in Simulation 1 for all times; in fact, it is smaller on average than the fundamental’s 10% drop.

Overall, in this simulation, the fact that $L$ is effectively more “risk averse” due to $K$ being lower causes him to be less leveraged; note that this is similar to lowering the regulated probability of bankruptcy. Since $L$ is more risk averse, the expected growth rate of $y$ is smaller, meaning that a smaller percentage of U’s money is sucked into L’s debt. The result that increasing the risky investment fundamental variance actually causes a slower buildup of systemic risk is actually quite interesting and demonstrates that a regulation which causes $L$ to calculate his VaR risk using a very conservative $s$ slows the increase of systemic risk; however, the cost is that investment is tends to be below fundamentals.
Figure 36: Simulation 3: Plot of the simulated average investment (top panel) and default cost (bottom panel) for the simplified and full models.
Figure 37: Simulation 3: Plot of the simulated average $a^D$ (top panel) and inverse risk (bottom panel) for the simplified and full models.
Figure 38: Simulation 3: Plot of the simulated average $y$ (top panel) and leverage (bottom panel).
Figure 39: Simulation 3: Plot of the simulated average $X$ (top panel) for the simplified and full models as well as the simplified model epsilon approximations (bottom panel).
13 The Optimal Risk Free Rate Rule

Having examined the system under a constant risk free rate environment, we now turn to examine what the Fed could do in theory in order to mitigate investment deviation from fundamentals. The important theme of this section is that the optimal level of the risk free rate depends on the relative size of L to U. The same V outcome can have different response implications in economies with different $y_t$; for example following a good outcome, the optimal response is to raise the risk free rate in economies where $y_t$ is above a threshold; if $y_t$ is below this threshold, then the optimal response is to lower rates! The finding that relative size matters is an important result, given the literature on the FFR discussed in Section 3, where the risk free rate adjustment is gradual and follows a Taylor type rule which does not involve the relative size of the financial sector.

In what follows, the risk free rate setting logic is presented in Section 13.1. We then present the optimal rule for the full model case in Section 13.2 and for the simplified model in Section 13.3. For the simplified model, we are able to analyze the dependence of the optimal risk free rate on $y_t$ in closed form in Section 13.4, and show in Section 13.5 that at a given time, it is generally not possible to stabilize risk and have investment reflect fundamentals. Section 13.6 reflects on the overall findings.

13.1 The Logic Behind the Rule

As shown in Section 12, the frictions arising from the interaction between L and U cause investment to deviate from fundamentals. Following a good fundamental outcome, L borrows more from U, which may cause net investment to increase above fundamentals and vice versa. As discussed in Section 11.2, it is desirable that reinvestment in the economy is not affected by financial frictions. Once the producer makes an initial project decision in the real economy, he does not want to have to readjust his project plans due to short term financial system responses to V. The only way this can happen is if at each time, the amount reinvested in the project is exactly that justified by fundamentals; this is equivalent to there being no additional financial flows between the financial system and the producer at each time. Thus, the adjustment of $F$ should be such that it counterbalances the financial system frictions resulting from the interaction of U and L.

Intuitively, we are faced with a flow management problem; the goal is to get the situation depicted in Figure 40, where real economy investment is recycled; achieving this is equivalent to having the risk free investment amount recycled. The control variable $F$ affects all 3 of U’s investment allocations; however, $y$ affects only the red arrows, corresponding to U’s risky debt and risk free asset investment. The goal is to find the rule for $F$ which cancels out the effect of $y$ such that at each period, there is no net flow between the real economy and the financial system as well as between the Fed and the financial system.

According to equation (78), the reinvestment at a given time is the sum of the fundamental term and a correction factor which reflects the net inflow of funds from the Fed to the financial system. At a given time $t$, after having observed $y_t$, the Fed can set the rate $F_t$ (which is the risk free rate earned at time $t+1$ on a risk free investment made at time $t$) so
as to set this net inflow of funds to 0.\(^{42}\)

![Diagram](image)

**Figure 40:** Graphical depiction of the situation where financial frictions do not affect real economy investment.

At time \(t\), if there is no default, the investment in the risky asset is:

\[
P_t = P_{t-1} + \left[ w_{t-1}a_{t-1}(F_{t-1}, y_{t-1})F_{t-1} - w_{t-1}z_t(V_t, F_{t-1})a_t(F_t, y_t) \right]
\]

where we have highlighted the dependencies of each variable. U’s ROE between time \(t-1\) and \(t\) is \(z_t\), which is a function of the risk free rate \(F_{t-1}\) and the random variable \(V_t\). Furthermore, the time \(t\) proportion of U’s wealth in the risk free asset depends on \(F_t\) and \(y_t\). Thus, varying \(F_t\) will affect only \(a_t(F_t, y_t)\). To set the second term to 0, the optimal risk free rate \(F^*_t\) should be set such that:

\[
a^*_t(F^*_t, y_t) = \frac{a_{t-1}(F_{t-1}, y_{t-1})F_{t-1}}{z_t(V_t, F_{t-1})} = \frac{a_{t-1}(F_{t-1}, y_{t-1})F_{t-1}}{a^*_t V_t + a^*_t X_t + a_{t-1} F_{t-1}}
\]

Notice that the optimal \(a^*_t \geq 0\) since \(a_{t-1}\) cannot be lower than 0. Furthermore, \(a^*_t \leq 1\); it will be 1 only if \(a_{t-1} = 1\). This situation is the trivial case where U invests all of his money in the risk free asset and L invests his own money in the risky asset; then of course, all of U’s risky asset profits will be reinvested as \(a_t\) will be 1 for all \(t\) and all of L’s profits will be reinvested in the risky asset. Although investment will reflect fundamentals in this case, it is a trivial situation and we will not focus on it.

An interesting observation is that if U’s ROE is higher than the risk free rate, the optimal \(a^*_t\) decreases compared to \(a_{t-1}\). This might be counterintuitive at first, yet we will see that

\(^{42}\)A note about the notation: if at time \(t\), I invest 1$ in the risk free asset, I will have \(F_t\) dollars at time \(t+1\). If I had invested in the risky asset, I would have \(V_{t+1}\) dollars at time \(t+1\). Although both are rates of return between times \(t\) and \(t+1\), the subscript \(t\) on \(F_t\) refers to the fact that this quantity is known at time \(t\). The subscript \(t+1\) on \(V_{t+1}\) reflects the fact that this quantity is a random variable observed at time \(t+1\).
this decrease may be smaller than the decrease that would have happened if the risk free rate were held constant. Intuitively, if there is a very good \( V \), \( U \)'s wealth grows; if \( a_t \) did not decrease, then some of the proceeds of the investment would actually be divested and sent to the risk free asset. If the system were uncontrolled, then since \( y_t \) would grow by a large amount, \( a_t^D \) would rise significantly at the cost of \( a_t \), which would decrease significantly; this would cause reinvestment to be above fundamentals as funds will be sucked out from the Fed and into the risky asset. Thus, the controlled \( a_t^* \) should be somewhere between these two extreme cases.

The Fed needs to set \( F_t \) such \( U \)'s risk free weight becomes \( a_t^* \). Since \( U \)'s weights must sum to 1, we can solve for \( F_t^* \):

\[
1 - a_t^*(F_t^*) = a_t^D(y_t, F_t^*)
\]

Note that \( F_t^* < m \) if \( a_t^* < 1 \). This is due to the fact that \( a_t^D \geq 0 \) (since \( X_t(F_t) \geq F_t \)). If \( a_t^* = 1 \), then the LHS of (120) can be 1 only if \( F_t = m \) since then, \( L \) will be in the non-VaR constrained mode since \( \bar{y}_t = 0 \), making \( X_t = m, a_t^D = 0 \); also, \( a_t^* \) will be 0. Thus, if \( a_t^* < 1 \), then \( F_t < m \) for the LHS to be less than 1.

Let's think about the "balance" in equation (120). As \( V_t \) increases, the RHS will decrease. If the risk free rate is not changed, the LHS will decrease as well since \( y_t \) will increase, causing \( a_t^D \) to increase. The whole game in how rates need to be adjusted is whether the RHS or LHS changes more for a certain \( V_t \) outcome. Roughly speaking, the LHS is less sensitive to \( V_t \) if \( y_{t-1} \) is smaller; in this case, the RHS move will be more extreme and \( F_t \) will need to be adjusted to make the LHS follow the RHS. If \( y_{t-1} \) is relatively large, then the LHS side move is more extreme than the RHS side move, meaning that \( F_t \) will need to be adjusted to compensate the large LHS move and bring it in line with the RHS. We further analyze this for both the full model and the simplified model in the next sections.

### 13.2 Risk Free Rate Rule: Full Model Case

For the full model case, equation (120) can be rewritten as:

\[
X_t(F_t, y_t) = F_t \left( 1 + \frac{c_v}{2c_D} \right) + \frac{1 - c_v m - a_t^*}{2c_D}
\]

where \( X_t(F_t, y_t) \) is given by equation (68). The \( F_t^* \) which solves the above equation is the optimal risk free rate. One can solve it in closed form and there is a unique solution; however, the resulting expression does not provide much intuition about how \( F_t^* \) compares to \( F_{t-1} \) and \( y_{t-1} \). Thus, we proceed to graphically illustrate these points. Figure 41 shows the graphical solution of (121). The red line is the RHS (which we denote \( f(a_t^*) \) as it is a function of \( a_t^* \)) and the blue line is the LHS. The green line shows the intersection point, whose x-axis value is the optimal \( F^* \). The green arrows illustrate how each curve would move if \( V_t \) increased. The RHS has a y-axis crossing of \( \frac{1 - c_v m - a_t^*}{2c_D} \); thus, as \( V_t \) increases, the red line would move up since \( a_t^* \) would decrease. The LHS, which is represented by the blue curve, would adjust as follows; its belly portion would move towards the upper left since \( y_t \) increases as \( V_t \) increases;
its linear portions would expand as the kinks would move towards eachother (the kinks are depicted by the 2 dotted black lines). In the limit, the blue curve would tend to the line \( \frac{m + F_t}{2} \). 43

It is straightforward to see that there is a unique intersection point between the red line and blue curve. The red line has a slope which is greater than 1; the blue curve has a slope of 0.5 in its linear regions; in its belly region, the slope is:

\[
\frac{dX_t}{dF_t} = \frac{1}{2} \left( 1 + \frac{F_t - K}{\sqrt{(F_t - K)^2 + \frac{2K}{c_D} y_t}} \right) \leq 1
\]

where the inequality is strict if \( y_t > 0 \). Thus, there will be exactly one intersection point and therefore a unique \( F_t^* \).

Table 5 illustrates the possible cases for \( V_t \) and the corresponding \( a_t^* \) and \( y_t \) movements compared to \( a_{t-1}^* \) and \( y_{t-1} \) respectively. \( V_{t-1}^F \) is the value which \( V_t \) must surpass in order for U’s ROE \( z_t \) to be greater than \( F_{t-1} \). \( \bar{V}_{t-1} \) is the value which \( V_t \) must surpass in order for \( y_t \) to go above \( y_{t-1} \). The expressions are simple to derive and are given below:

\[
z_t = F_{t-1} \equiv V_{t-1}^{zF} = F_{t-1} - \frac{a_{t-1}^D}{a_{t-1}^z} (X_{t-1} - F_{t-1}) \leq F_{t-1}
\]

\[
\bar{z}_t = d_t \equiv \bar{V}_{t-1} = \frac{a_{t-1}^D + l_{t-1} - \frac{1}{a_{t-1}^z}}{l_{t-1} - a_{t-1}^z} X_{t-1} + \frac{a_{t-1}}{l_{t-1} - a_{t-1}^z} F_{t-1} \geq F_{t-1}
\]

where we have used the fact that \( X_{t-1} \geq F_{t-1} \) (the inequalities are strict if \( y_{t-1} > 0 \)).

For \( \bar{V}_{t-1} \), note that it is a weighted average of \( X_{t-1} \) and \( F_{t-1} \) where the weights are positive (since leverage is always greater or equal to 1) and sum to 1.

<table>
<thead>
<tr>
<th>( V_t )</th>
<th>( V_{t-1}^{zF} )</th>
<th>( V_{t-1}^{zF} )</th>
<th>( V_t )</th>
<th>( V_{t-1}^{zF} )</th>
<th>( V_{t-1}^{zF} )</th>
<th>( V_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( z_t &lt; F_{t-1} )</td>
<td>( z_t &gt; F_{t-1} )</td>
<td>( z_t &gt; F_{t-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( a_t^* &gt; a_{t-1} )</td>
<td>( a_t^* &lt; a_{t-1} )</td>
<td>( a_t^* &lt; a_{t-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( y_t &lt; y_{t-1} )</td>
<td>( y_t &lt; y_{t-1} )</td>
<td>( y_t &gt; y_{t-1} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: This table depicts the possible \( a_t^* \) and \( y_t \) movements compared to \( a_{t-1}^* \) and \( y_{t-1} \) respectively.

43The linear portions of the blue curve occurs when \( F_t \) is such that \( \frac{m + F_t}{2} \) is smaller than the square root expression for \( X_t \) in (68). Thus, the kink locations occur at:

\[
\bar{F}_t^k = K \pm \sqrt{(m - K)^2 - \frac{2K}{c_D} y_t}
\]

Thus, these kinks move towards the center as \( V_t \) increases because \( y_t \) increases with \( V_t \).
For the “middle case” where \(V_{t-1} < V_t < \overline{V}_{t-1}\), it is easy to see that \(F_t^* < F_{t-1}\) by looking at the balance equation (120). If \(F_t = F_{t-1}\), then the LHS would increase and the RHS would decrease. Thus, the optimal response must be to raise rates so that \(F_t > F_{t-1}\), making the LHS decrease.

However, the direction of rates is not obvious for the left and right cases in Table 5 as in both cases, which side of the balance equation drops (if \(V_t > \overline{V}_{t-1}\)) or increases (if \(V_t < V_{t-1}^{zF}\)) more depends on the relative size at time \(t-1\), \(y_{t-1}\). Figure 42 illustrates this point for the case where \(V_t > \overline{V}_{t-1}\). In the top panel, the optimal rate \(F_t^*\) increases from \(F_{t-1}\) due to the fact that \(y_{t-1}\) is relatively large; thus, in response to a good outcome which increases L’s relative size, the optimal response is to raise rates. However, in the bottom panel, when \(y_{t-1}\) is relatively small, the optimal response is to lower rates in response to a good outcome which increases L’s relative size; this is an interesting result that may seem counterintuitive, yet makes sense in the context of the balance equation; when \(y_{t-1}\) is small, even though a positive outcome occurs, it is not enough to suck in enough funds from the risk free asset in order for investment to reflect fundamentals. Essentially, the system resembles the case where only U exists (see Section 12.2 for a brief description) and thus, rates must be lowered so that all investment proceeds are reinvested.

The analogous behavior exists for the case where \(V_t < \overline{V}_{t-1}\); when \(y_{t-1}\) is large, the optimal response is to lower rates in response to a bad \(V_t\). However, if \(y_{t-1}\) is small, the optimal response is to raise rates following a bad outcome.

From the graphical analysis of the optimal rate rule, it is clear that the relationship between \(F_t^*\), \(V_t\), and \(y_{t-1}\) is fairly complex. The critical size of \(y_{t-1}\) for which the optimal response to a good \(V_t\) involves a rate increase as opposed to a rate decrease depends on \(V_t\) itself. Essentially, the question is whether the uncontrolled \(a_t\), denoted by \(a_t^{unc}\), would be less than or greater than \(a_t^*\); in the former case, the optimal response is to raise the risk free rate, whereas in the latter case, the optimal response is to lower it. Which case we are in depends on \(y_{t-1}\); finding this critical \(y_{t-1}^c\) boils down to evaluating whether the LHS or RHS decreases more if \(F_t = F_{t-1}\) for a fixed \(V_t\). Solving for this \(y_{t-1}^c\) can be achieved by fixing a \(V_t\), evaluating the LHS of the balance equation with \(F_t = F_{t-1}\), and comparing it to the RHS of the balance equation. It turns out that \(y_{t-1}^c\) depends on the actual outcome \(V_t\); this makes assessing the sensitivity of \(F_t^*\) to \(y_{t-1}\) and \(V_t\) in closed form a hairy task. We thus turn to analyze the optimal risk free rate rule in the simplified solution setting, which approximates the full solution well for small \(y_t\) as discussed in Section 12.
Figure 41: Graphical depiction of the optimal $F_t^*$ (top panel) and a zoomed in version (bottom panel).
Figure 42: Graphical depiction of the optimal $F_t^*$ response to $V_t > \bar{V} t - 1$ for the case where $y_{t-1}$ is larger (top panel) and for the case where $y_{t-1}$ is smaller (bottom panel). Note that for both cases, all parameters are the same except $y_{t-1}$. 
13.3 Risk Free Rate Rule: Simplified Model Case

For the simplified model case, equation (120) can be rewritten as:

\[ 1 - a_t^v - (l_t - 1)y_t = a_t^* \]  

(125)

One can easily solve the above for \( F_t^* \) and the solution will be unique since the LHS is a monotonically increasing function of \( F_t \). However, our aim in this section and the next is to understand when the optimal response involves a rate cut or a rate rise; thus, we aim to understand the relationship between \( F_t^* \) and \( F_{t-1} \).

The advantage of the simplified solution in the risk free rate analysis is that leverage depends only on the risk free rate. Furthermore, using the notation in Section 13.2, \( \bar{V}_t = V_t^{zF} \) since the following set of inequalities holds:

\[
V_t > F_{t-1} \Rightarrow d_t > V_t > z_t > F_{t-1} \\
V_t < F_{t-1} \Rightarrow d_t < V_t < z_t < F_{t-1}
\]

(126)

Thus, it is the case that one of the following two situations is true: \( (a_t^* < a_{t-1}, y_t > y_{t-1}) \) or \( (a_t^* > a_{t-1}, y_t < y_{t-1}) \). Thus, the middle case of Table 5 does not exist in the simplified model.

Table 6 summarizes the four possible scenarios: we refer to cases as A through D as depicted in the table for brevity.

<table>
<thead>
<tr>
<th>Outcome/State</th>
<th>( y_{t-1} &gt; y_{t-1}^c )</th>
<th>( y_{t-1} &lt; y_{t-1}^c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( V_t &gt; F_{t-1} )</td>
<td>( a_t^{unc} &lt; a_t^* &lt; a_{t-1} )</td>
<td>( a_t^* &lt; a_t^{unc} &lt; a_{t-1} )</td>
</tr>
<tr>
<td>( V_t &lt; F_{t-1} )</td>
<td>( a_t^{unc} &gt; a_t^* &gt; a_{t-1} )</td>
<td>( F_t^* &lt; F_{t-1} &lt; V_t )</td>
</tr>
</tbody>
</table>

Table 6: This table depicts the four possible risk free rate adjustment scenarios of the simplified model. We refer to these cases as A through D as denoted in the table.

From Table 6, we can see that the optimal risk free rate movement depends on both the relative size in the previous period and on the outcome. If \( y_{t-1} > y_{t-1}^c \), then the optimal response is to raise rates following a good outcome and vice versa; however, if \( y_{t-1} < y_{t-1}^c \), then the optimal response is the opposite. While this latter situation may seem counterintuitive, it is not; this latter situation occurs when \( y_{t-1} \) is very small and thus, a good outcome does not increase the relative size appreciably, meaning that U's weights will not change much, thus causing some of the investment proceeds to be funneled out; this can only be mitigated via a rate cut. The analogous argument can be made for the bad outcome case, as relative
size does not decrease appreciably and thus, U’s weights will not change much, thus causing some funds from the risk free asset to be brought into the system; this can only be mitigated via a rate rise.

We now proceed to derive \( y_t^c \) as well as to prove the inequalities between \( F_t^* \) and \( V_t \) presented in Table 6.

13.3.1 Derivation of \( y_t^c \)

Recall that \( y_t^c \) is the critical value of \( y_{t-1} \) which differentiates whether there should be a rate increase or decrease in response to a certain \( V_t \). There should be a rate increase if the uncontrolled \( \alpha_t^{unc} \) is smaller than \( \alpha_t^* \) and vice versa:

\[
\begin{align*}
\alpha_t^{unc} < \alpha_t^* \Rightarrow \alpha_t^{unc} < \frac{\alpha_{t-1} F_{t-1}}{z_t} \\
\Rightarrow \alpha_t^{unc} z_t < \alpha_{t-1} F_{t-1} \\
\Rightarrow (1 - \alpha_{t-1}^w) z_t - (l_{t-1} - 1) y_t z_t < (1 - a_{t-1}^v) F_{t-1} - (l_{t-1} - 1) y_{t-1} F_{t-1} \\
\Rightarrow (1 - a_{t-1}^v) z_t - (l_{t-1} - 1) y_{t-1} d_t < (1 - a_{t-1}^v) F_{t-1} - (l_{t-1} - 1) y_{t-1} F_{t-1} \\
\Rightarrow (1 - a_{t-1}^v) (z_t - F_{t-1}) < (l_{t-1} - 1) y_{t-1} (d_t - F_{t-1}) \\
\Rightarrow (1 - a_{t-1}^v) a_{t-1}^v (V_t - F_{t-1}) < (l_{t-1} - 1) y_{t-1} \left( \frac{F_{t-1}}{F_{t-1} - K} (V_t - K) - F_{t-1} \right) \\
\Rightarrow (1 - a_{t-1}^v) a_{t-1}^v (V_t - F_{t-1}) < (l_{t-1} - 1) y_{t-1} \mu_{t-1} (V_t - F_{t-1}) \\
\end{align*}
\]

(127)

If \( V_t > F_{t-1} \), then there should be a rate increase if:

\[
y_{t-1} > \frac{(1 - a_{t-1}^v) a_{t-1}^v}{l_{t-1} (l_{t-1} - 1)} \equiv y_t^c \]

(128)

Else, there should be a rate decrease. If \( V_t < F_{t-1} \), then there should be a rate increase if \( y_{t-1} < y_t^c \); else, there should be a rate decrease.

13.3.2 Derivation of Relationship Between \( F_t^* \) and \( V_t \)

The inequalities between \( F_t^* \) and \( V_t \) presented in Table 6 will be useful in assessing the sensitivity of \( F_t^* \) to \( y_{t-1} \) as well as in assessing the effect that the optimal \( F_t^* \) has on systemic risk.

For cases C and D, proving these bounds is trivial. By examining Table 6, it is clear that \( F_t^* < F_{t-1} < V_t \) for case C and \( F_t^* > F_{t-1} > V_t \) for case D.

Let’s consider case A now: showing that \( F_t^* < V_t \) is equivalent to showing that \( a_t(V_t) > a_t(F_t^*) \).

\[
a_t(V_t) > a_t(F_t^*)
\]

124
\[ 1 - a_t^v(V_t) - (l_t(V_t) - 1)y_t > \frac{a_{t-1}F_{t-1}}{z_t} \]
\[ = (1 - a_t^v(V_t))z_t - \frac{K}{V_t - K}y_{t-1}d_t > (1 - a_{t-1}^v)F_{t-1} - (l_{t-1} - 1)y_{t-1}F_{t-1} \]
\[ = (1 - a_t^v(V_t))z_t - \frac{K}{V_t - K}F_{t-1} - K(V_t - K)y_{t-1} > (1 - a_{t-1}^v)F_{t-1} - \frac{K}{F_{t-1} - K}F_{t-1}y_{t-1} \]
\[ = (1 - a_t^v(V_t))z_t > F_{t-1}(1 - a_{t-1}^v) \] (129)

The last relationship in (129) is true in case A since \( z_t > F_{t-1} \) and \( a_t^v(V_t) < a_{t-1}^v \).

Now, let’s consider case B: showing that \( F_t^* > V_t \) is equivalent to showing that \( a_t(V_t) < a_t(F_t^*) \). Following along the lines of the steps in (129):

\[ a_t(V_t) < a_t(F_t^*) \Leftrightarrow (1 - a_t^v(V_t))z_t < F_{t-1}(1 - a_{t-1}^v) \] (130)

The relationship in (130) is true in case B since \( z_t < F_{t-1} \) and \( a_t^v(V_t) > a_{t-1}^v \).

Having established the directionality of the optimal rate change, we explore the dependency of the rate adjustment magnitude on \( y_{t-1} \) in the next section.

### 13.4 Sensitivity of Optimal Rate in Simplified Model

Having bounded the optimal risk free rate in each of the 4 cases of the simplified model, we now proceed to examine its sensitivity to the relative size of \( L \) more explicitly. Before continuing, we touch on the reasons why this is important. Firstly, as discussed in Section 3, the relative size of the leveraged sector is currently not a factor which monetary policy considers; yet, even within our simple setup, the active role of the leveraged sector affects the channeling of funds to the real economy. Furthermore, we have seen that in order for investment to reflect fundamentals at each time, the Fed needs to act in every decision period; currently, this is not the case. However, as we have seen in Section 12, when the risk free rate is held constant, \( y_t \) tends to increase. Thus, when the Fed does act, the interesting question is whether the actual risk free rate magnitude change should depend on \( y_t \). Intuitively, it should, and the longer the Fed doesn’t act and \( y_t \) builds up, the more extreme this rate change should be when the Fed does act.

Table 7 summarizes the results we find for the magnitude of rate adjustment. For cases A and B, the magnitude of the rate adjustment is increasing in \( y_{t-1} \); thus, for economies with a larger relative size of \( L \) to \( U \), for the same outcome, the optimal risk free rate adjustment is bigger. The opposite holds for cases C and D.

### 13.4.1 Derivation

Assume that we are at time \( t \) and the Fed will set \( F_t^* \). We know that the balance equation for the simplified model in (125) must hold. We proceed to examine the sensitivity of \( F_t^* \) on \( y_{t-1} \). We fix \( V_t \) and take the respective derivative on both sides:
### Outcome/State

<table>
<thead>
<tr>
<th>$V_t &gt; F_{t-1}$</th>
<th>$y_{t-1} &gt; y_{t-1}^c$</th>
<th>$y_{t-1} &lt; y_{t-1}^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_t^* &gt; F_{t-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$</td>
<td>F_t^* - F_{t-1}</td>
<td>$ increasing in $y_{t-1}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$V_t &lt; F_{t-1}$</th>
<th>$F_t^* &lt; F_{t-1}$</th>
<th>$F_t^* &gt; F_{t-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>F_t^* - F_{t-1}</td>
<td>$ increasing in $y_{t-1}$</td>
</tr>
</tbody>
</table>

Table 7: This table depicts the rate adjustment magnitude results for cases as A through D.

$$
 c_v \frac{dF_t^*}{dy_{t-1}} + \frac{K}{(F_t^* - K)^2} y_t \frac{dF_t^*}{dy_{t-1}} - (l_t - 1) \frac{dy_t}{dy_{t-1}} \frac{F_{t-1}}{z_t} \frac{da_{t-1}}{dy_{t-1}} = 0
$$

(131)

Note that on the RHS, the fact that only $a_{t-1}$ depends on $y_{t-1}$ is a result of the fact that we are interested in assessing the impact of relative size at $t - 1$ on the optimal rate at $t$; this is interesting for the case where the Fed is not acting in every period, as is the case in practice. Thus, $F_{t-1}$ is not the time $t - 1$ optimal rate; rather, it is the rate which was set at some prior time $t - T$; in other words, the risk free rate has been constant over the interval $[t - T, t - 1]$.

For (131) to hold, one of the following must be true:

$$
 \frac{dF_t^*}{dy_{t-1}} > 0 \iff -(l_t - 1) \frac{dy_t}{dy_{t-1}} < \frac{F_{t-1}}{z_t} \frac{da_{t-1}}{dy_{t-1}}
$$

(132a)

$$
 \frac{dF_t^*}{dy_{t-1}} < 0 \iff -(l_t - 1) \frac{dy_t}{dy_{t-1}} > \frac{F_{t-1}}{z_t} \frac{da_{t-1}}{dy_{t-1}}
$$

(132b)

We proceed to examine the condition under which (132a) holds:

$$
 -(l_t - 1) \frac{dy_t}{dy_{t-1}} < \frac{F_{t-1}}{z_t} \frac{da_{t-1}}{dy_{t-1}} \iff

- \frac{K}{F_t^* - K} (F_t^* - F_{t-1}) > (l_t - 1) F_{t-1} \iff

- \frac{K}{F_t^* - K} (V_t - K) > (l_t - 1) F_{t-1} \iff

\frac{K}{F_t^* - K} (V_t - K) > \frac{K}{F_{t-1} - K} F_{t-1} \iff

\frac{V_t - K}{F_t^* - K} > 1
$$

(133)

When we are in cases A and C, $V_t > F_t^*$ (see Table 6), and thus, (132a) holds, meaning that $\frac{dF_t^*}{dy_{t-1}} > 0$. In other words, if $V_t > F_{t-1}$, then $F_t^*$ is an increasing function of $y_{t-1}$.
In case A, this means that the risk free rate increase $F_t^* - F_{t-1}$ is increasing in $y_{t-1}$: said another way, the risk free rate rise is larger if L’s relative size at the previous time is larger and the outcome $V_t$ is “good” (i.e. above $F_{t-1}$).

In case C, this means that the magnitude of the risk free rate decrease, $|F_t^* - F_{t-1}|$, goes towards 0 as $y_{t-1}$ increases: said another way, the risk free rate cut is smaller if L’s relative size at the previous time is larger and the outcome $V_t$ is “good”.

In an analogous fashion to (133), one can show that when we are in cases B and D, (132b) holds since $F_t^* > V_t$. This means that $\frac{dF_t^*}{dy_{t-1}} < 0$. In other words, if $V_t < F_{t-1}$, then $F_t^*$ is an increasing function of $y_{t-1}$.

In case B, this means that the magnitude of the risk free rate decrease, $|F_t^* - F_{t-1}|$, is increasing in $y_{t-1}$: said another way, the risk free rate cut is larger if L’s relative size at the previous time is larger and the outcome $V_t$ is “bad” (i.e. below $F_{t-1}$).

In case D, this means that the risk free rate increase, $F_t^* - F_{t-1}$, is decreasing in $y_{t-1}$: said another way, the risk free rate increase is smaller if L’s relative size at the previous time is larger and the outcome $V_t$ is “bad”.

To summarize, the magnitude of the risk free rate change is increasing in $y_{t-1}$ if $y_{t-1} > y_{t-1}^c$. However, if $y_{t-1} < y_{t-1}^c$, then the magnitude of the risk free rate change is decreasing in $y_{t-1}$.

### 13.5 Risk Under The Optimal Risk Free Rate Rule

We now proceed to examine how the potential systemic cost (i.e. risk) changes in the simplified model if $F_t^*$ is set according to the optimal risk free rate rule. We summarize the main findings in the two bullet points below:

- The potential systemic cost increases for the controlled system whenever $V_t > F_{t-1}$. For case A, this increase is smaller than the increase in the uncontrolled system; for case C, this increase is larger than in the uncontrolled system.

- The potential systemic cost decreases for the controlled system whenever $V_t < F_{t-1}$. For case B, this decrease is less than the decrease in the uncontrolled system; for case D, this decrease is larger than in the uncontrolled system.

#### 13.5.1 Derivation

Recall from equation (95) that risk or potential systemic cost is given by:

$$R_t = \frac{P_t}{w_t a_t F_t}$$

We denote the time change in risk from time $t-1$ to $t$ as $\gamma_t \equiv \frac{R_t}{R_{t-1}}$. By using (134), we can express $\gamma_t$ as:

$$\gamma_t \equiv \frac{R_t}{R_{t-1}} = \frac{P_t}{w_t a_t F_t} = \frac{P_t}{P_{t-1}} \frac{a_t F_{t-1}}{a_{t-1} F_t} \frac{1}{F_t}$$

127
For each of the 4 cases in Table 6, we proceed to evaluate $\gamma_t$ for both the uncontrolled (denoted $\gamma_t^{unc}$) and controlled (denoted $\gamma_t^*$) cases. Before continuing, it is worth pointing out that $R_t$ is a decreasing function of $F_t$ since increasing $F_t$ decreases the numerator (through $P_t$) and increases the denominator (through both $a_t$ and $F_t$).

Proceeding, $\gamma_t^*$ can be simplified to:

$$
\gamma_t^* = \frac{P_t}{P_{t-1}} \frac{a_{t-1}F_{t-1}}{a_tz_t} \frac{1}{F_t} = \frac{V_t}{F_t^*}
$$

since under the optimal control, $a_t^*z_t = a_{t-1}F_{t-1}$ and $P_t = P_{t-1}V_t$.

Below, we prove our claims for all four cases:

1. **Case A:** We know from Table 6 that $a_t^{unc} < a_t^*$. This means that $a_t^{unc}z_t < a_{t-1}F_{t-1}$ and $P_t > P_{t-1}V_t$ (see equation (78)); furthermore, since this is the uncontrolled case, $F_t = F_{t-1}$. Using these facts along with the expression in (135) gives:

$$
\gamma_t^{unc} > \frac{V_t}{F_{t-1}}
$$

Furthermore, since we are in case A, $F_{t-1} < F_t^* < V_t$, and therefore:

$$
\gamma_t^* = \frac{V_t}{F_t^*} > 1
$$

Thus, under the optimal risk free rate setting rule, the potential systemic cost increases. However, it increases less than the uncontrolled case since:

$$
\gamma_t^{unc} > \frac{V_t}{F_{t-1}} > \frac{V_t}{F_t^*} = \gamma_t^*
$$

In this case, stabilizing risk would mean that a rate $F_t^*$ would have to be applied, where $F_t^* > F_t^*$; however, the cost of this would be to push investment below fundamentals since $a_t^* > a_t^*$ and thus $P_t < P_{t-1}V_t$.

2. **Case B:** We know from Table 6 that $a_t^{unc} > a_t^*$. This means that $a_t^{unc}z_t > a_{t-1}F_{t-1}$ and $P_t < P_{t-1}V_t$ (see equation (78)); furthermore, since this is the uncontrolled case, $F_t = F_{t-1}$. Using these facts along with the expression in (135) gives:

$$
\gamma_t^{unc} < \frac{V_t}{F_{t-1}}
$$

Furthermore, since we are in case B, $V_t < F_t^* < F_{t-1}$, and therefore:

$$
\gamma_t^* = \frac{V_t}{F_t^*} < 1
$$
Thus, under the optimal risk free rate setting rule, the potential systemic cost decreases. However, it decreases less than the uncontrolled case since:

$$\gamma_t^{unc} < \frac{V_t}{F_{t-1}} < \frac{V_t}{F_t} = \gamma_t^* \quad (142)$$

In this case, stabilizing risk would mean that a rate $F_t^r$ would have to be applied, where $F_t^r < F_t^*$; however, the cost of this would be to push investment above fundamentals since $a_t^r < a_t^*$ and thus $P_t > P_{t-1}V_t$.

3. **Case C:** We know from Table 6 that $a_t^{unc} > a_t^*$. This means that $a_t^{unc}z_t > a_{t-1}F_{t-1}$ and $P_t < P_{t-1}V_t$ (see equation (78)); furthermore, since this is the uncontrolled case, $F_t = F_{t-1}$. Using these facts along with the expression in (135) gives:

$$\gamma_t^{unc} < \frac{V_t}{F_{t-1}} \quad (143)$$

Note that from the above formula, it is unclear whether risk increases or decreases in the uncontrolled case since $\frac{V_t}{F_{t-1}} > 1$. However, we know from (95) that $y_t > y_{t-1}$ means that $a_t^P > a_{t-1}^P$ and $a_t < a_{t-1}$; thus, risk does in fact increase in the uncontrolled case. Since we are in case C, $F_t^* < F_{t-1} < V_t$, and therefore:

$$\gamma_t^* = \frac{V_t}{F_t} > 1 \quad (144)$$

Thus, under the optimal risk free rate setting rule, the potential systemic cost increases. Furthermore, it increases more than in the uncontrolled case since:

$$\gamma_t^{unc} < \frac{V_t}{F_{t-1}} < \frac{V_t}{F_t} = \gamma_t^* \quad (145)$$

In this case, stabilizing risk would mean that a rate $F_t^r$ would have to be applied, where $F_t^r > F_t^*$; however, the cost of this would be to push investment below fundamentals since $a_t^r > a_t^*$ and thus $P_t < P_{t-1}V_t$.

4. **Case D:** We know from Table 6 that $a_t^{unc} < a_t^*$. This means that $a_t^{unc}z_t < a_{t-1}F_{t-1}$ and $P_t > P_{t-1}V_t$ (see equation (78)); furthermore, since this is the uncontrolled case, $F_t = F_{t-1}$. Using these facts along with the expression in (135) gives:

$$\gamma_t^{unc} > \frac{V_t}{F_{t-1}} \quad (146)$$

Note that from the above formula, it is unclear whether risk increases or decreases in the uncontrolled case since $\frac{V_t}{F_{t-1}} < 1$. However, we know from (95) that $y_t < y_{t-1}$ means that $a_t^P < a_{t-1}^P$ and $a_t > a_{t-1}$; thus, risk does in fact decrease in the uncontrolled case.
Since we are in case D, \( F_t^* > F_{t-1} > V_t \), and therefore:

\[ \gamma_t^* = \frac{V_t}{F_t^*} < 1 \quad (147) \]

Thus, under the optimal risk free rate setting rule, the potential systemic cost decreases. Furthermore, it decreases more than in the uncontrolled case since:

\[ \gamma_{t,\text{unc}} > \frac{V_t}{F_{t-1}} > \frac{V_t}{F_t^*} = \gamma_t^* \quad (148) \]

In this case, stabilizing risk would mean that a rate \( F_t^* \) would have to be applied, where \( F_t^* < F_t^* \); however, the cost of this would be to push investment above fundamentals since \( a_t^* < a_t^* \) and thus \( P_t > P_{t-1}V_t \).

Thus, if \( y_{t-1} \) is above \( y_{t-1}^* \), following the optimal control rule “tames” the risk change in the system both on the upside and on the downside. However, the opposite behavior occurs when \( y_{t-1} < y_{t-1}^* \), where we see that the risk change is amplified both on the upside and on the downside for the controlled case, compared to the uncontrolled case.

### 13.6 Optimal Risk Free Rate Discussion

In this section, we have presented a rule which the Fed could follow in every period to set the risk free rate so as to have investment reflect fundamentals. Furthermore, we showed that at a given time, one cannot generally have investment reflect fundamentals and risk not change. Essentially, the problem can be viewed as managing the net inflow/outflow of funds from the financial sector. For both the full and simplified model, the logic of setting \( a_t \) so as to balance these flows is the same. The dependence of \( a_t \) on the rate \( F_t \) is more complex in the full model; however, when \( y_t \) is small, the resulting simplified and full model risk free rates will be very similar.

As we saw, the most interesting case is when \( y_{t-1} > y_{t-1}^* \); if the Fed does not act in every period, we know that the relative size tends to increase over time. We saw that the magnitude of the rate adjustment \( |F_t^* - F_{t-1}| \) is larger for larger \( y_{t-1} \) if \( y_{t-1} > y_{t-1}^* \) (cases A and B). In these cases, this adjustment magnitude would be even larger if the aim were to stabilize the potential systemic cost. Thus, if the Fed does not act in every period, then the rate adjustment it will have to make when it does in fact act will tend to be higher.

The behavior is also interesting when \( y_{t-1} < y_{t-1}^* \). In this case, following a good outcome, the uncontrolled potential systemic cost is lower than the controlled potential systemic cost. Thus, the optimal response goes hand in hand with increasing potential systemic cost; this is because a good \( V \) when \( L \) is small does not affect \( a^D \) appreciably and rates need to be lowered to incentivize \( U \) to lend more to \( L \) so investment can reflect fundamentals. The analogous situation holds when \( y_{t-1} < y_{t-1}^* \) and there is a bad outcome. Again, in this case, \( a^D \) does not drop appreciably and thus rates need to be increased so that reinvestment does not end up being above fundamentals.
Overall, our results in this section highlight that considering the relative size of the leveraged sector should be incorporated into macroeconomic and financial stability models. The risk free rate has important implications for leverage and potential systemic cost through its effect on L’s behavior.

14 Summary

In this part of the thesis, we presented an aggregate model of the financial system which incorporates an active levered financial intermediary who acts in a procyclical manner. We then analyzed the evolution of two important metrics, and proposed an optimal risk free rate setting rule to mitigate some of the systemic effects. Focusing on leveraged financial intermediary behavior was motivated by the recent empirical literature which finds that leveraged financial intermediaries actively manage their balance sheet and expand them during economic expansions and vice versa. The main model ingredients/motivations are listed below:

- The allocation of household wealth to money market funds over the past 30 years has increased dramatically, fueling the levered sector’s impressive growth. Such money market funds typically pay little more than the risk free rate, implying that most investors are agnostic about the type of market exposure they have in bad outcomes.

- Financial regulators have imposed VaR as a measure of risk for levered financial intermediaries. Although institutions such as investment banks aim to maximize their return on equity, they are bound by this VaR constraint on how much risk they can take. Adhering to this type of constraint induces procyclical behavior, whereby the levered intermediary will increase his borrowing in good times and vice versa. Adhering to a VaR constraint actually makes good sense for institutions such as banks, who report to their shareholdes; clearly, they don’t want to default, but if they do, they don’t care what the possible losses are as the shareholders are wiped out irrespectively and any additional losses are born by the debt holders.

- It has been documented that financial intermediary asset growth forecasts real economic variables such as investment. This is not unexpected since financial intermediaries channel investment funds to the real economy. The fact that their balance sheet adjustments impact the real economy is an important motivation of why we choose to focus on aggregate behavior and understand how this channeling of funds works.

- Monetary economists view the short term risk free rate from an “inflation expectation management” view point. However, insofar as the level of this rate impacts the ability of the levered sector to borrow in order to lend, it has real implications for the channeling of investment and systemic risk.

Although the model has a relatively simple setup, the resulting behaviors are complex and nonlinear. The most important finding is that the relative size of the levered sector is
an important determinant of investment and risk. In “good times”, the levered sector will increase its borrowing from the unlevered sector, sucking in funds from the risk free asset to the real economy; as the levered sector gets larger, this behavior causes larger investment deviation from fundamentals. Furthermore, the potential systemic cost of a default by L gets larger as the relative size of the levered sector gets larger. In “bad times”, the levered sector will decrease its borrowing, pushing funds out of the economy into the risk free asset; a larger levered sector causes investment to fall more than that which is dictated by fundamentals. The key effect of L’s behavior is how it changes the indirect risky investment exposure of the unlevered sector.

Perhaps the most interesting aspect of our model is that economic overinvestment goes hand in hand with the buildup of risk. One of the factors that made the recent financial crisis quite memorable was the fact that very few people had a sense of the buildup of the potential systemic cost in the years leading up to the crisis. Au contraire, during the boom years of 2003-2007, the economy was perceived to be strong, investment was high (too high in hindsight), growth was strong, and the financial sector was raking in huge profits. However, what most people missed is that risk was building up quietly; in retrospect, the large growth of the levered sector’s size which was aided by its ability to borrow at low rates should have rang alarm bells about the possible systemic cost of a default. Our model is able to capture the tradeoff between investment and risk and shows that the underlying factor between both is the relative size of the leveraged sector.

In our model, the risk free rate is an important variable as it affects the cost of borrowing of the leveraged sector. We proposed a risk free rate setting rule which counterbalances the frictions in the supply of funds caused by the interactions of the leveraged and unleveraged sectors. We found that the optimal rate depends on the relative size of the leveraged sector; furthermore, we found that it is generally not possible to set the risk free rate such that time t investment reflects fundamentals and systemic risk does not change. The dependence of the risk free rate on relative size is not always intuitive at first sight. Roughly speaking, when the relative size of the leveraged sector is very small, the optimal response to a “good” fundamental outcome, which increases the relative size of the leveraged sector, is to actually lower the risk free rate; in this case, systemic risk will be higher than if there were no rate adjustment. In contrast, when the relative size of the leveraged sector is larger, the optimal response to the same good outcome is to raise the risk free rate; in this case, the potential systemic cost will increase but be lower than if there were no rate adjustment. Furthermore, the magnitude of this rate rise is increasing in the size of the leveraged sector. The analogous result holds for the case of a bad outcome; a small leveraged sector actually calls for a rate increase whereas a larger leveraged sector calls for a rate decrease.

Current macroeconomic models mostly view the short term risk free rate as a tool for setting inflation expectations. However, the importance of the risk free rate in our model is that of a contemporaneous regulating force on the investment channeling process; its level affects the risk taking capability of the leveraged sector and thus, the relative size of the leveraged sector needs to be a factor in determining its optimal level. Indeed, the fact that the risk free rate affects both investment and risk suggests that monetary policy and financial
stability are linked and thus, macroeconomic models need to be adjusted to incorporate an active levered financial intermediary sector.

Of course, it is typically the case that “you can’t manage what you can’t measure”. We have seen that investment, the buildup of risk, and the “right” level of the risk free rate depends on the relative size of the levered sector. While data for leveraged players such as broker/dealers is available, this is not the case for many other leveraged players, such as hedge funds and shadow banking system entities. Thus, from a macroprudential regulatory perspective, it is imperative to be able to accurately measure the whole size of the levered sector. Such data will allow policy makers to quantitatively incorporate financial sector frictions in their optimal policies.
15 Appendix for Part I of Thesis

In this appendix, we provide all of the proofs promised in Part I of this thesis.

15.1 Proof of Equations (13) and (14)

The derivation of equation (14) from equation (12) is a straightforward application of the matrix identity:

$$[A + BCB']^{-1} = A^{-1} - A^{-1}B[C^{-1} + B'A^{-1}B]^{-1}B'A^{-1}$$

where $B$ is a rectangular matrix. Starting from equation (10), we can derive equation (13) as follows:

$$\mu_{BL} = \Sigma_{BL}[(\tau\Sigma)^{-1}(\pi + P'\Omega^{-1}v)]$$

$$= [\tau\Sigma - \tau^2\Sigma'\tau\Sigma\pi + \Omega^{-1}\Pi\Sigma] [(\tau\Sigma)^{-1}(\pi + P'\Omega^{-1}v)]$$

$$= \pi - \tau\Sigma\pi'(\tau\Sigma\pi' + \Omega)^{-1}P\pi + \tau\Sigma\pi'\Omega^{-1}v - \tau^2\Sigma\pi'(\tau\Sigma\pi' + \Omega)^{-1}\Pi\Sigma\pi'\Omega^{-1}v$$

$$= \pi - \tau\Sigma\pi'(\tau\Sigma\pi' + \Omega)^{-1}P\pi + \tau\Sigma\pi'[I - (\Pi\Sigma\pi' + \Omega)^{-1}\Pi\Sigma\pi']\Omega^{-1}v$$

$$= \pi - \tau\Sigma\pi'(\tau\Sigma\pi' + \Omega)^{-1}P\pi$$

$$+ \tau\Sigma\pi'[\left((\Pi\Sigma\pi' + \Omega)^{-1}\Pi\Sigma\pi' + \Omega\right) - \left((\Pi\Sigma\pi' + \Omega)^{-1}\Pi\Sigma\pi'\Omega^{-1}v$$

$$= \pi - \tau\Sigma\pi'(\tau\Sigma\pi' + \Omega)^{-1}P\pi + \tau\Sigma\pi'\left(\Pi\Sigma\pi' + \Omega\right)^{-1}\left((\Pi\Sigma\pi' + \Omega) - \Pi\Sigma\pi'\right)\Omega^{-1}v$$

$$= \pi - \tau\Sigma\pi'(\tau\Sigma\pi' + \Omega)^{-1}P\pi + \tau\Sigma\pi'(\tau\Sigma\pi' + \Omega)^{-1}v$$

$$= \pi + \tau\Sigma\pi'(\tau\Sigma\pi' + \Omega)^{-1}(v - P\pi)$$

15.2 Proof of Theorem 1

We call $\omega_{BL}^*$ the BL weights vector resulting from $(P_*, \Omega_*, v_*)$. To prove this theorem, we use the result of Theorem 2, which is proved in Section 15.3. With the help of the factorization in Theorem 2, we can express $\omega_{BL}^*$ as:

$$\omega_{BL}^* = \frac{\omega_{eq} + P'*m_*}{1 + \tau}$$

Recall that $m_*$ can be written as:

$$m_* = \frac{1}{\Delta}A_*^{-1}\left(v_* - \frac{P_*\pi}{1 + \tau}\right) = \frac{1}{\Delta}\left(\Omega_* + \frac{P_*\Sigma P_*'}{1 + \tau}\right)^{-1}\left(v_* - \frac{P_*\pi}{1 + \tau}\right)$$
Using the fact that \((P_*, \Omega_*, v_*) = (BP, B\Omega B', Bv)\):

\[
m^* = \frac{1}{\Delta} \left( \frac{\Omega_*}{\tau} + \frac{P_*\Sigma P_*'}{1+\tau} \right)^{-1} \left( v_* - \frac{P_*\pi}{1+\tau} \right)
\]

\[
eq \frac{1}{\Delta} \left( \frac{B\Omega B'}{\tau} + \frac{BP\Sigma P'B'}{1+\tau} \right)^{-1} B \left( v - \frac{P\pi}{1+\tau} \right)
\]

\[
eq \frac{1}{\Delta} (B')^{-1} \left( \frac{\Omega}{\tau} + \frac{P\Sigma P'}{1+\tau} \right)^{-1} \left( v - \frac{P\pi}{1+\tau} \right)
\]

\[
eq (B')^{-1} m
\]

Thus:

\[
\omega_{BL}^* = \frac{\omega_{eq} + P'_* m_*}{1+\tau} = \frac{\omega_{eq} + P'_* (B')^{-1} m}{1+\tau} = \frac{\omega_{eq} + P'B'(B')^{-1} m}{1+\tau} = \omega_{BL}
\]

(153)

It is easy to show that the row reduced echelon form triplet is a special case. We get \(P_{\text{rref}}\) by first performing an “LDU decomposition” of \(P\) and then performing more row operations so that there are zeros both above and below the resulting pivots.\(^{44}\)

\[
MP = \text{LDU}
\]

(154)

\[
U = ZP_{\text{rref}}
\]

(155)

where \(L\) is a \((k \times k)\) lower triangular matrix, \(M\) is a \((k \times k)\) permutation matrix which puts the rows of \(P\) in the right order in order to factor it, \(U\) is an upper triangular matrix with \(1^*\) on its diagonal resulting from the LDU factorization, \(D\) is a \((k \times k)\) diagonal matrix which is set so that \(U\) has \(1^*\) on its diagonal, and \(Z^{-1}\) is a \((k \times k)\) upper triangular matrix which performs row operations on \(U\) so that there are zeros above the pivots. Thus:

\[
P_{\text{rref}} = (\text{LDZ})^{-1}MP
\]

(156)

Note that since \(M\) is a permutation matrix, it is invertible; moreover, \(M^{-1} = M'\). Also, note that the product \(\text{LDZ}\) describes the row operations needed to go from \(P_{\text{rref}}\) to \(MP\); since the rows of \(P_{\text{rref}}\) contain a basis of the row space of \(MP\), \(\text{LDZ}\) is an invertible transformation. Thus, the product \((\text{LDZ})^{-1}M\) is also invertible. Thus, the appropriate \(B\) matrix is \(B_{\text{rref}} \equiv (\text{LDZ}^{-1})M\). Therefore, the triplet \((P_{\text{rref}}, \Omega_{\text{rref}}, v_{\text{rref}}) \equiv (B_{\text{rref}}P, B_{\text{rref}}\Omega B'_{\text{rref}}, B_{\text{rref}}v)\) results in the same allocation \(\omega_{BL}\).

\(^{44}\) A good presentation of LDU decompositions is presented in [58].
15.3 Proof of Theorem 2

The proof of Theorem 2 is a bit tedious as it involves many algebraic manipulations. We start by rewriting equation (16) with the help of equation (12):

$$\omega_{BL} = \frac{1}{\Delta} (\Sigma + \Sigma_{BL})^{-1} \mu_{BL}$$

$$= \frac{1}{\Delta} (\Sigma + \Sigma_{BL})^{-1} \Sigma_{BL} \left[ (\tau \Sigma)^{-1} \pi + P'\Omega^{-1} \nu \right]$$

(157)

Our first objective is to simplify the \((\Sigma + \Sigma_{BL})^{-1} \Sigma_{BL}\) term. With the help of the identity in equation (149):

$$\Sigma_{BL}^{-1} = \Sigma_{BL}^{-1} - \Sigma_{BL}^{-1} (\Sigma_{BL}^{-1} + \Sigma^{-1})^{-1} \Sigma_{BL}^{-1}$$

and thus:

$$\Sigma_{BL}^{-1} = \Sigma_{BL}^{-1} - \Sigma_{BL}^{-1} (\Sigma_{BL}^{-1} + \Sigma^{-1})^{-1}$$

(158)

(159)

By plugging in the expression for \(\Sigma_{BL}\) of equation (11):

$$\Sigma_{BL}^{-1} (\Sigma_{BL}^{-1} + \Sigma^{-1})^{-1} = [(\tau \Sigma)^{-1} + P'\Omega^{-1} P] \left[ (\tau \Sigma)^{-1} + P'\Omega^{-1} P + \Sigma^{-1} \right]^{-1}$$

(160)

$$= [(\tau \Sigma)^{-1} + P'\Omega^{-1} P] \left[ \frac{1 + \tau}{\tau} \Sigma^{-1} + P'\Omega^{-1} P \right]^{-1}$$

$$= (\tau \Sigma)^{-1} \left[ \frac{1 + \tau}{\tau} \Sigma^{-1} + P'\Omega^{-1} P \right]^{-1} + P'\Omega^{-1} P \left[ \frac{1 + \tau}{\tau} \Sigma^{-1} + P'\Omega^{-1} P \right]^{-1}$$

$$= \left( \frac{1 + \tau}{\tau} \Sigma^{-1} + P'\Omega^{-1} P \right)^{-1} + P'\Omega^{-1} P \left[ \frac{1 + \tau}{\tau} \Sigma^{-1} + P'\Omega^{-1} P \right]^{-1}$$

$$= \left( (1 + \tau) I_n + \tau P'\Omega^{-1} P \Sigma \right)^{-1} + P'\Omega^{-1} P \left( \frac{1 + \tau}{\tau} \Sigma^{-1} + P'\Omega^{-1} P \right) \left( (1 + \tau) I_n + \tau P'\Omega^{-1} P \Sigma \right)^{-1}$$

$$= \left( (1 + \tau) I_n + \tau P'\Omega^{-1} P \Sigma \right)^{-1} + P'\Omega^{-1} P \left( (1 + \tau) I_n + \tau P'\Omega^{-1} P \Sigma \right)^{-1}$$

Thus, by plugging in the result of equation (160) into equation (159), we get:

$$\Sigma_{BL}^{-1} = I_n - \Sigma_{BL}^{-1} \left( \Sigma_{BL}^{-1} + \Sigma^{-1} \right)^{-1}$$

$$= I_n - [I_n + \tau P'\Omega^{-1} P \Sigma] \left( (1 + \tau) I_n + \tau P'\Omega^{-1} P \Sigma \right)^{-1}$$

136
\[ I_n = (1 + \tau) I_n + \tau P' \Omega^{-1} P \Sigma \] 

\[ (1 + \tau) I_n + \tau P' \Omega^{-1} P \Sigma \] 

\[ = I_n - \left[ (1 + \tau) I_n + \tau P' \Omega^{-1} P \Sigma \right]^{-1} \]

\[ = I_n - \left[ (1 + \tau) I_n + \tau P' \Omega^{-1} P \Sigma \right]^{-1} \]

\[ = \frac{\tau}{1 + \tau} \left[ I_n + \frac{\tau}{1 + \tau} P' \Omega^{-1} P \Sigma \right]^{-1} \]

Thus:

\[ \omega_{BL} = \frac{1}{\Delta} (\Sigma + \Sigma_{BL})^{-1} \mu_{BL} \]

\[ = \frac{1}{\Delta} (\Sigma + \Sigma_{BL})^{-1} \Sigma_{BL} \left[ (\tau \Sigma)^{-1} \pi + P' \Omega^{-1} v \right] \]

\[ = \frac{1}{\Delta} \frac{\tau}{1 + \tau} \left[ I_n + \frac{\tau}{1 + \tau} P' \Omega^{-1} P \Sigma \right]^{-1} \left[ (\tau \Sigma)^{-1} \pi + P' \Omega^{-1} v \right] \]

The inverse term \( I_n + \frac{\tau}{1 + \tau} P' \Omega^{-1} P \Sigma \) in equation (162) is complicated; in order to obtain the simple and intuitive result of Theorem 2, we must find a way to simplify the inverse term: this is shown in Theorem 6 below:

**Theorem 6** The following equality holds:

\[ \left[ I_n + \frac{\tau}{1 + \tau} P' \Omega^{-1} P \Sigma \right]^{-1} = I_n - \frac{1}{1 + \tau} P' A^{-1} P \Sigma \]

where \( A = \frac{\Omega}{\tau} + \frac{P \Sigma P'}{1 + \tau} \).

**Proof** With the help of the identity in equation (149), we can rewrite \( A \) as:

\[ A^{-1} = \tau \Omega^{-1} - \tau \Omega^{-1} P \left[ \left( \frac{\Sigma}{1 + \tau} \right)^{-1} + \tau P' \Omega^{-1} P \right]^{-1} \tau P' \Omega^{-1} \]

\[ \frac{1}{1 + \tau} P' A^{-1} P \Sigma = \frac{\tau P' \Omega^{-1} P \Sigma}{1 + \tau} - \tau P' \Omega^{-1} P \left[ \left( \frac{\Sigma}{1 + \tau} \right)^{-1} + \tau P' \Omega^{-1} P \right]^{-1} \frac{\tau P' \Omega^{-1} P \Sigma}{1 + \tau} \]

\[ = \frac{\tau P' \Omega^{-1} P \Sigma}{1 + \tau} - \tau P' \Omega^{-1} P \left( I_n + \tau P' \Omega^{-1} P \frac{\Sigma}{1 + \tau} \right) \left( \frac{\Sigma}{1 + \tau} \right)^{-1} \frac{\tau P' \Omega^{-1} P \Sigma}{1 + \tau} \]

137
\[
\frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau} - \tau P'\Omega^{-1}P \left( \frac{\Sigma}{1 + \tau} \right) \left[ I_n + \tau P'\Omega^{-1}P \frac{\Sigma}{1 + \tau} \right]^{-1} \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau}
\]

With the help of the result in equation (165), we can expand the RHS of equation (163):

\[
I_n - \frac{1}{1 + \tau} P' A^{-1} P \Sigma = I_n - \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau} + \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau} \left[ I_n + \tau P'\Omega^{-1}P \frac{\Sigma}{1 + \tau} \right]^{-1} \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau}
+ \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau} \left[ I_n + \tau P'\Omega^{-1}P \frac{\Sigma}{1 + \tau} \right]^{-1} \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau}
\]

\[
= \left[ I_n + \tau P'\Omega^{-1}P \frac{\Sigma}{1 + \tau} \right]^{-1} + \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau} \left[ I_n + \tau P'\Omega^{-1}P \frac{\Sigma}{1 + \tau} \right]^{-1} \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau}
\]

\[
\text{(166)}
\]

\[
= \left[ I_n + \tau P'\Omega^{-1}P \frac{\Sigma}{1 + \tau} \right]^{-1} - \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau}
+ \tau P'\Omega^{-1}P\Sigma \left[ I_n + \tau P'\Omega^{-1}P \frac{\Sigma}{1 + \tau} \right]^{-1} \left[ I_n + \tau P'\Omega^{-1}P \frac{\Sigma}{1 + \tau} \right]^{-1} \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau}
\]

\[
= \left[ I_n + \tau P'\Omega^{-1}P \frac{\Sigma}{1 + \tau} \right]^{-1} - \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau} + \frac{\tau P'\Omega^{-1}P\Sigma}{1 + \tau}
\]

\[
= \left[ I_n + \tau P'\Omega^{-1}P \frac{\Sigma}{1 + \tau} \right]^{-1}
\]

Q.E.D.

With the help of the above theorem, we can express the BL weights of equation (162) as follows:

\[
\omega_{BL} = \frac{1}{\Delta} \frac{\tau}{1 + \tau} \left[ I_n + \frac{\tau}{1 + \tau} P'\Omega^{-1}P\Sigma \right]^{-1} \left[ (\tau \Sigma)^{-1} \pi + P'\Omega^{-1}v \right] \quad (167)
\]

\[
= \frac{1}{\Delta} \frac{\tau}{1 + \tau} \left[ I_n - \frac{1}{1 + \tau} P' A^{-1} P \Sigma \right] \left[ (\tau \Sigma)^{-1} \pi + P'\Omega^{-1}v \right]
\]

With a little bit of algebra:

138
\[ \omega_{BL} = \frac{1}{\Delta} \frac{\tau}{1 + \tau} \left[ I_n - \frac{1}{1 + \tau} P'A^{-1} P\Sigma \right] [(\tau \Sigma)^{-1} \pi + P'\Omega^{-1} v] \]  

\[ = \frac{1}{1 + \tau} \left[ \frac{\tau}{\Delta} (\tau \Sigma)^{-1} \pi + \frac{\tau}{\Delta} P'\Omega^{-1} v - \frac{\tau}{\Delta (1 + \tau)} P'A^{-1} P\Sigma (\tau \Sigma)^{-1} \pi - \frac{\tau}{\Delta (1 + \tau)} P'A^{-1} P\Sigma P'\Omega^{-1} v \right] \]  

\[ = \frac{1}{1 + \tau} \left[ \frac{1}{\Delta} \Sigma^{-1} \pi + \frac{\tau}{\Delta} P'\Omega^{-1} v - \frac{1}{\Delta (1 + \tau)} P'A^{-1} P\pi - \frac{\tau}{\Delta (1 + \tau)} P'A^{-1} P\Sigma P'\Omega^{-1} v \right] \]  

Recall that by the standard CAPM assumption, the equilibrium weights vector is given by:

\[ \omega_{eq} = \frac{1}{\Delta} \Sigma^{-1} \pi \]  

Thus:

\[ \omega_{BL} = \frac{1}{1 + \tau} \left[ \omega_{eq} + \frac{\tau}{\Delta} P'\Omega^{-1} v - \frac{1}{\Delta (1 + \tau)} P'A^{-1} P\pi - \frac{\tau}{\Delta (1 + \tau)} P'A^{-1} P\Sigma P'\Omega^{-1} v \right] \]  

The last term inside the brackets in equation (170) can further be simplified by using the fact that \( A \equiv \frac{\Omega}{\tau} + \frac{P\Sigma P'}{1 + \tau} \) and thus:

\[ \frac{\tau}{\Delta (1 + \tau)} P'A^{-1} P\Sigma P'\Omega^{-1} v = \frac{\tau}{\Delta} P'A^{-1} \left( A - \frac{\Omega}{\tau} \right) \Omega^{-1} v \]

\[ = \frac{\tau}{\Delta} P'\Omega^{-1} v - \frac{\tau}{\Delta} P'A^{-1} \frac{\Omega}{\tau} \Omega^{-1} v \]  

\[ = \frac{\tau}{\Delta} P'\Omega^{-1} v - \frac{1}{\Delta} P'A^{-1} v \]  

Plugging the result of equation (171) into equation (170), we have:

\[ \omega_{BL} = \frac{1}{1 + \tau} \left[ \omega_{eq} + \frac{\tau}{\Delta} P'\Omega^{-1} v - \frac{1}{\Delta (1 + \tau)} P'A^{-1} P\pi + \frac{1}{\Delta} P'A^{-1} v \right] \]  

\[ = \frac{1}{1 + \tau} \left[ \omega_{eq} + P\left( \frac{1}{\Delta} A^{-1} v - \frac{1}{\Delta (1 + \tau)} A^{-1} P\pi \right) \right] \]  

\[ = \frac{1}{1 + \tau} \left( \omega_{eq} + P'm \right) \]  

where:

\[ m = \frac{1}{\Delta} A^{-1} v - \frac{1}{\Delta (1 + \tau)} A^{-1} P\pi = \frac{1}{\Delta} A^{-1} \left( v - \frac{P\pi}{1 + \tau} \right) \]
15.4 Proof of Theorem 3

Backward Statement Proof: Pick a \( P' \) according to the method outlined in steps 2 & 3 of the inference algorithm. If each element of \( x_i \) belongs to the set \( S \) and \( |x| \) contains each nonzero member of the set \( S' \) at least once, then the vector \( x \) lies on a \( k \)-dimensional subspace of \( R^n \). By construction, the \( (nxk) \) matrix \( P' \) satisfies the restriction in assumption 3 and thus, the \( k \) columns of \( P' \) form an orthogonal basis for a \( k \)-dimensional subspace of \( R^n \). Thus, \( x \) is in the column space of \( P' \) and the equation \( x = P'm \) will have exactly one solution vector \( m \). Furthermore, each entry in \( m \) will be distinct since \( |x| \) contains each nonzero element of the set \( S' \) at least once. Now, recall that for an arbitrary \( x \), there are many \( P' \) matrices that can be constructed via the steps in the inference algorithm which still satisfy the assumptions; the columns of \( P' \) can be generated in any order and/or the column generating rule in step 3 of the inference algorithm can be changed to: "for every asset in a given group, input a +1 in the corresponding entry of the column if the deviation (i.e. its entry in \( x \)) is negative and a -1 if the deviation is positive (i.e. its entry in \( x \)): for all other entries of the column, input a zero". Since every such \( P' \) is consistent with the allowability conditions, then the equation \( x = P'm \) will have exactly one solution for each such \( P' \).

Forward Statement Proof: if there exists a pair \((P, m)\) for which the \((nxk)\) matrix \( P' \) satisfies assumption 3 and the \((kxl)\) vector \( m \) satisfies assumption 4, then it is trivial to see that each entry in \( x = P'm \) will be in the set \( S \) and that every entry of the set \( S' \) will be present in \( |x| \); this is because each row of \( P' \) will either be zero (in which case the corresponding element of \( x \) will be zero), or it will have a single nonzero entry, say in column \( j \), in which case the corresponding element of \( x \) will be \( \pm m_j \).

15.5 Proof of Theorem 4

Assume we use two different consistent conventions in factoring the vector \( x \): under convention 1, we get \((P_1, m_1)\) and under convention 2, we get \((P_2, m_2)\). The difference between \( P_1 \) and \( P_2 \) is that a set of rows differ in sign. Define by \( M \) a \((kxk)\) permutation matrix such that \( P_2 = MP_1 \). Then:

\[
\begin{align*}
P_2'm_2 &= P_1'm_1 \\
P_1'M'm_2 &= P_1'm_1 \quad (174)
\end{align*}
\]

where the last equality follows from the fact that since \( P \) is allowable, it must have full row rank and thus the nullspace of \( P' \) is empty. Now, we seek the relationship between \( v_2 \) and \( v_1 \):

\[
v_2 = \Delta \left( \frac{P_2 \Sigma P_2'}{\tau} + \frac{P_2 \Sigma P_2'}{1+\tau} \right) m_2 + \frac{P_{2\pi}}{1+\tau}
\]

140
\[ = \Delta M \left( \frac{P_1 \Sigma P_1'}{\tau} + \frac{P_1 \Sigma P_1'}{1 + \tau} \right) m_2 + \frac{M P_1 \pi}{1 + \tau} \]  
\[ = \Delta M \left[ \left( \frac{P_1 \Sigma P_1'}{\tau} + \frac{P_1 \Sigma P_1'}{1 + \tau} \right) m_1 + \frac{P_1 \pi}{1 + \tau} \right] \]
\[ = M v_1 \]

Note also that:
\[ \Omega_2 = \frac{P_2 \Sigma P_2'}{\tau} = M P_1 \Sigma P_1' M' = M \Omega_1 M' \]  
(176)

Thus, the triplet \((P_2, \Omega_2, v_2) = (M P_1, M \Omega_1 M', M v_1)\) is in the informationally equivalent set.

### 15.6 Proof of Theorem 5

**Part 1 Proof:** Of course, if there is no noise added, then by following the procedure for estimating the number of views presented in Section 6.2, we would never misestimate the number of views. However, the theorem’s result in the “no noise” case is nice in a theoretical sense and serves as a foreground for understanding part 2 of the theorem. Since there is no noise, \(\bar{x} \equiv x\). Recall that \(v_l\) can be expressed as:

\[ v_l = \Delta A_l m_l + \frac{A P_1 \pi}{1 + \tau} \]
\[ = \Delta \left( \frac{P_1 \Sigma P_1'}{\tau} + \frac{P_1 \Sigma P_1'}{1 + \tau} \right) m_l + \frac{P_1 \pi}{1 + \tau} \]  
(177)

Since it was the observation of \(x\) that led to the factorization \(x = P_1 m_l\), we have:

\[ v_l = \Delta P_1 \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1 + \tau} \right) x + \frac{P_1 \pi}{1 + \tau} \]  
(178)

Multiplying both sides by the \((k \times l)\) matrix \(A\) we have:

\[ \Lambda v_l = \Delta \Lambda P_1 \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1 + \tau} \right) x + \frac{\Delta P_1 \pi}{1 + \tau} \]  
(179)

Moreover, since \(x = P_k m_k\), we have that:

\[ \Lambda v_l = \Delta P_k \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1 + \tau} \right) P_k m_k + \frac{P_k \pi}{1 + \tau} = v_k \]  
(180)
This concludes the proof of the first part of the theorem.

**Part 2 Proof:** The proof of part 2 of the theorem is similar to the proof of part 1. However, due to the fact that $\tilde{x}$ is now noisy, any factorization of it using the inference algorithm will be approximate because in Step 4 of the inference algorithm, in order to apply theorem 3, we create the vector $y$ from $\tilde{x}$ as described in equation (35); consequently, we estimate the vector $\tilde{m}$ (we use the tilde notation to indicate that this is the estimate resulting from a noisy $\tilde{x}$). Of course, the size of $\tilde{m}$ depends on how many views we estimate that the manager has in Part 2 of the inference algorithm. Analogously to the proof of part 1 of the theorem above, we denote the case where we think he has $l$ views by $\tilde{m}_l$ and the case where we think he has $k$ views by $\tilde{m}_k$, where $l > k$. In order to proceed with the proof, we must describe mathematically how to arrive to $\tilde{m}_l$ given our estimate $\tilde{P}_1$ and our observation $\tilde{x}$. Without loss of generality, assume that the vector $\tilde{x}$ is ordered such that assets in the same view are grouped next to each other. For the case where we estimated $l$ views, $\tilde{x}$ contains $l + 1$ blocks (the first $l$ for the views and the last one for the assets on which there is no view and thus have a 0 entry); we denote the set of elements in block $j$ by $C_j$. Denote each of the entries of $\tilde{x}$ as $\tilde{x}_i^j$, where $i = 1, \ldots, n$ denotes the entry of $\tilde{x}$ and $j = 1, \ldots, l + 1$ denotes which view-group we classified the element as being in; note that we are not assuming that this is the correct classification as our goal is to show consistency between two estimated $(P, m)$ pairs where each pair was estimated with a different estimate of the number of views. Given this setup, the $j^{th}$ entry of $m_l$ takes on the value $m_l^j$, where $m_l^j > 0$ and is given by:

$$\tilde{m}_l^j = \frac{1}{|C_j|} \sum_{d \in C_j} |\tilde{x}_d^j| \quad j = 1, \ldots, l$$  \hfill (181)

where $|C_j|$ is the cardinality of the set $C_j$. Note that $\tilde{m}_l$ has $l$ elements because there is not an entry in it for the group of assets on which there is no view.

In order to generalize equation (181) to a linear matrix notation, we define the matrices $S$ and $B_1$ as follows:

$$S = \begin{bmatrix}
  \text{sgn}(x_1) & 0 & \ldots & 0 \\
  0 & \text{sgn}(x_2) & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & \ldots & 0 & \text{sgn}(x_n)
\end{bmatrix} \quad (182)$$
Note that the size of the diagonal matrix $S$ is $(n \times n)$ and the size of $B_1$ is $(l \times (|C_1| + \cdots + |C_{l+1}|))$. The matrix $B_1$ is an "averaging" matrix; when applied to $\tilde{x}$, row $j$ averages the entries of block $|C_j|$. Note that the zero columns of $B_1$ correspond exactly to the $l + 1$ group of $\tilde{x}$. With the above definitions in mind, we generalize equation (181) to a linear matrix notation:

$$\tilde{m}_i = B_1 S \tilde{x}$$

(184)

since:

$$|\tilde{x}| = S \tilde{x}$$

(185)

The important observation is that if Step 3 of the inference algorithm is followed for constructing $\hat{P}_1$, then:

$$B_1 S = D_1 \hat{P}_1$$

(186)

where $D_1$ is an $(l \times l)$ diagonal matrix of the form:

$$D_1 = \begin{bmatrix}
\frac{1}{|C_1|} & 0 & \cdots & 0 \\
0 & \frac{1}{|C_2|} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \frac{1}{|C_l|}
\end{bmatrix}$$

(187)

With this in mind, we can proceed from equation (177):

$$\tilde{v}_l = \Delta \hat{P}_1 \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1 + \tau} \right) \hat{P}_1 \tilde{m}_l + \frac{\hat{P}_1 \pi}{1 + \tau} \Rightarrow$$

$$\Lambda \tilde{v}_l = \Delta \Lambda \hat{P}_1 \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1 + \tau} \right) \hat{P}_1 \tilde{m}_l + \frac{\Lambda \hat{P}_1 \pi}{1 + \tau} \Rightarrow$$

(188)
\[ \Lambda \tilde{w}_t = \Delta \hat{P}_k \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1+\tau} \right) \hat{P}'_1 \hat{P}_1 \tilde{x} + \frac{\hat{P}_k \pi}{1+\tau} \]

For the case where we estimate the number of views to be \( k \), we can analogously write:

\[ \tilde{v}_k = \Delta \hat{P}_k \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1+\tau} \right) \hat{P}'_k \tilde{m}_k + \frac{\hat{P}_k \pi}{1+\tau} = \Delta \hat{P}_k \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1+\tau} \right) \hat{P}'_1 \hat{D}_1 \hat{P}_1 \tilde{x} + \frac{\hat{P}_k \pi}{1+\tau} \]  
(189)

Subtracting equation (189) from equation (188) and using the fact that \( \tilde{x} = x + \eta_x \), where \( \eta_x \) is the zero mean noise:

\[ \Lambda \tilde{w}_t - \tilde{v}_k = \Delta \hat{P}_k \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1+\tau} \right) \left( \hat{P}'_1 \hat{D}_1 \hat{P}_1 - \hat{P}'_k \hat{D}_k \hat{P}_k \right) (x + \eta_x) \]  
(190)

Now, since \( x \) is the vector of deviations which is not corrupted by noise:

\[ \hat{D}_1 \hat{P}_1 x = B_1 \hat{s} x \equiv m_i \]  
(191)

Notice that the above is not "m tilde" because we are referring to the \( m \) vector which is identified when \( x \) has no noise.

We are thus in a position to say:

\[ \hat{P}'_1 \hat{D}_1 \hat{P}_1 x = \hat{P}'_1 m_i = x \]  
(192)

Analogously:

\[ \hat{P}'_k \hat{D}_k \hat{P}_k x = \hat{P}'_k m_k = x \]  
(193)

Plugging in equations (192) & (193) back into equation (190):

\[ \Lambda \tilde{w}_t - \tilde{v}_k = \Delta \hat{P}_k \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1+\tau} \right) \left( \hat{P}'_1 \hat{D}_1 \hat{P}_1 - \hat{P}'_k \hat{D}_k \hat{P}_k \right) \eta_x \]  
(194)

Since \( \eta_x \) is a zero mean random vector:

\[ E [\Lambda \tilde{w}_t - \tilde{v}_k] = E \left[ \Delta \hat{P}_k \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1+\tau} \right) \left( \hat{P}'_1 \hat{D}_1 \hat{P}_1 - \hat{P}'_k \hat{D}_k \hat{P}_k \right) \eta_x \right] \]
\[ = \Delta \hat{P}_k \left( \frac{\Sigma}{\tau} + \frac{\Sigma}{1+\tau} \right) \left( \hat{P}'_1 \hat{D}_1 \hat{P}_1 - \hat{P}'_k \hat{D}_k \hat{P}_k \right) E (\eta_x) \]  
(195)

This concludes the proof of the second part of the theorem.
16 Appendix for Part 2 of Thesis

16.1 Proof of $q_l = 1$

In this section, we prove that $L$ will never invest in the risk free asset if $K < F < m$; one could argue that putting some money in the risk free asset would allow him to borrow more and thus possibly make him better off; however, this is not the case.

We drop all time indicators for brevity as this is a static problem. Denote by $q$ the proportion of $L$’s net assets (his equity and the amount he borrows) which he invests in the risky asset and by $1 - q$ that which he invests in the risk free asset. Our goal is to show that if $F < m$, he will always invest a proportion of $q = 1$ in the risky asset. $L$’s objective function is:

$$C(q) = q(w_L + b(q))m + (1 - q)(w_L + b(q))F - b(q)X(q)$$

$$= q[w_L(m - F) + b(q)(m - F)] + b(q)(F - X(q)) + w_L F$$

(196)

$L$ satisfying his VaR constraint means:

$$c = P\{q(w_L + b(q))V + (1 - q)(w_L + b(q))F - b(q)X(q) < 0\} \Rightarrow$$

$$K = \frac{b(q)X(q) - (1 - q)(w_L + b(q))F}{q(w_L + b(q))} \Rightarrow$$

$$b(q) = w_L \frac{qK + (1 - q)F}{X - (qK + (1 - q)F)} \Rightarrow$$

$$b(q) = w_L \frac{f(q)}{X - f(q)}$$

(197)

where we have defined $f(q)$ as a shorthand for $qK + (1 - q)F$. Note that since we assumed $F > K$, then $f(q) \leq F$. Furthermore, note that $X(q)$ is given by:

$$X(q) = \frac{F + f(q) + \sqrt{(F - f(q))^2 + \frac{2K}{cD}}}{2}$$

(198)

Deriving the above is exactly analogous to the derivation of $X_t$ in equation (64). It is worth noting that $X(q) \geq F$.

Proceeding, we expand (196) with the help of (197) as:

$$C(q) = q[w_L(m - F) + b(q)(m - F)] + b(q)(F - X(q))$$

We drop the $w_L F$ term as it is not dependent on $q$. 

145
The goal is to examine whether $C(q = 1) > C(q < 1)$.

For brevity, we will use $X$ and $f$ instead of $X(q)$ and $f(q)$, keeping in mind that they are functions of $q$. Taking the derivative with respect to $q$:

$$C'(q) = \frac{w^L}{(X - f)^2} \left[ ((m - F)X + q(m - F)X' + f'(F - X) - X'f)(X - f) - (q(m - F)X + f(F - X))(X' - f') \right]$$

Manipulating the term in brackets gives:

$$\begin{align*}
[X'(q(m - F) - f) + X((m - F) - f') + f'F][X - f] - [X(q(m - F) - f) + fF][X' - f'] = \\
[q(m - F) - f][X'(X - f)X' - f(X' - f')] + X[f'(X - f) - f(X' - f')] + X(X - f)[(m - F) - f'] = \\
[q(m - F) - f + F][Xf' - X'f] + X(X - f)[(m - F) - f'] = \\
[q(m - F) - kq - (1 - q)F + F][Xf' - X'f] + X(X - f)[(m - F) - k + F] = \\
q(m - K)[Xf' - X'f] + (m - K)X(X - f) = \\
(m - K)[q(Xf' - X'f) + X^2 - Xf] = \\
(m - K)[X(qf' + X - f) - qfX']
\end{align*}$$

If (201) is positive, then $C(q)$ is increasing in $q$, meaning that $L$ will choose $q = 1$. We have assumed that $m > K$. We now show that the term in the brackets is positive:

$$qf' + X - f = q(K - F) + X - qK - (1 - q)F = X - F \geq 0$$

where the inequality is strict if $y > 0$. Furthermore, we know that $X'(q) < 0$ since:

$$\begin{align*}
X'(q) &= \frac{1}{2}f'(q) \left[ 1 + \frac{F - f(q)}{\sqrt{(F - f(q))^2 + \frac{2K}{c_D}y}} \right] \\
&= \frac{1}{2}(K - F) \left[ 1 + \frac{F - f(q)}{\sqrt{(F - f(q))^2 + \frac{2K}{c_D}y}} \right] < 0
\end{align*}$$

Thus, $C'(q) > 0$, meaning that $L$ will always choose to allocate all of his borrowing to the risky investment, as long as $F < m$. 

146
16.2 Proof that $P(V_{t+1} \leq K) \leq c$ in non-VaR Constrained Mode

In Section 11.3.3, we saw that in the non-VaR constrained case, $b_{t}^{inc} \leq b_{t}^{VaR}$ and that the probability of default is less than or equal to $c$; in this section, we show why. The key is that $\frac{b_{t}X_{t}}{w_{t}^{L} + b_{t}}$ is an increasing function of $b_{t}$. From the coupling condition:

$$a_{t}^{D} \equiv \frac{b_{t}}{w_{t}^{U}} = 2\sigma_{D}(X_{t} - F_{t}) \Rightarrow X_{t} = \frac{b_{t}}{2\sigma_{D}w_{t}^{U}} + F_{t}$$  \hspace{2cm} (204)

From here on, we drop the $t$ subscripts for brevity. From the above, $\frac{b_{t}X_{t}}{w_{t}^{L} + b}$ can be rewritten as:

$$\frac{b_{t}^{2}}{2\sigma_{D}w_{t}^{U}} + bF \frac{w_{t}^{L} + b}{w_{t}^{L} + b}$$ \hspace{2cm} (205)

The above is an increasing function of $b$ since:

$$\left(\frac{b_{t}^{2}}{2\sigma_{D}w_{t}^{U}} + bF\right)' = \frac{b_{t}^{2} + bw_{t}^{L} + Fw_{t}^{L}}{(w_{t}^{L} + b)^{2}} > 0$$ \hspace{2cm} (206)

Thus, if L’s optimal unconstrained borrowing, $b_{t}^{inc}$, does not violate the VaR constraint, then $\frac{b_{t}^{inc}X_{t}(b_{t}^{inc})}{w_{t}^{L} + b_{t}^{inc}} \leq \frac{b_{t}^{VaR}X_{t}(b_{t}^{VaR})}{w_{t}^{L} + b_{t}^{VaR}}$ and $b_{t}^{VaR} \leq b_{t}^{VaR}$. Since the CDF function $\phi(.)$ is a monotonically increasing function, then the probability of default is less than or equal to $c$ in the non-VaR constrained case.

16.3 The Simplified Model Approximation

As discussed in Section 12, in order to arrive at closed form expressions, we linearized the solution of the full model presented in Section 11. The approximation is accurate when $y_{t}$ is small. In this section, we present the approximation we made and analyze it.

16.3.1 An Empirical Justification

The assumption that $y_{t} \ll 1$ in practice is actually quite valid. The total equity of all levered financial intermediaries primarily includes that of hedge funds and other “alternative investment” institutions as well as that of investment banks. The total equity of unlevered players is by definition equal to their assets: unlevered players primarily consist of mutual funds and pension funds. Table 8 shows estimates of the total world equity of the key levered and unlevered players obtained at the end of 2009. \textsuperscript{46} The data for all institutions except banks is obtained from “TheCityUK” (see [2]), which compiles the data from various sources including governmental agencies, supranationals (OECD, IMF, and BIS), and investment managers. The data for banks is obtained from “The Banker” (see [1]).

\textsuperscript{46}Note that for levered players such as hedge funds and private equity funds, the term “assets under management” (AUM) is used instead of the term equity.
Clearly, the classification is very crude as institutions within a given category may be very heterogeneous. However, the aim is to use the data as a rough first order estimate of the relative size of the levered sector to the unlevered sector.

The vast majority of hedge funds and private equity funds are indeed levered as leverage is part of their business model. As for pension funds and mutual funds, they are typically not leveraged: in many countries, laws are in place to prevent such institutions from levering up due to the societal role that they play. Sovereign wealth funds are also typically unlevered as their job is to manage their country’s fiscal surplus. As for private wealth funds, the $26 Trillion number referenced in Table 8 refers to the wealth of high net worth individuals which was not invested through conventional investment management routes (pension funds, mutual funds, etc.) or alternative investment management routes (hedge funds, private equity funds, etc.). The vast majority of such HNWI’s outsource their investment decisions to their private wealth manager and do not typically employ leverage.

The two categories which are hard to classify are banks and insurance companies. Clearly, banks are levered institutions, yet the leverage amount and “behavior” of an investment bank such as Goldman Sachs is markedly different from that of a regional “savings and loan” bank. Furthermore, it is difficult to separate out the “investment banking” activities from the retail banking activities of mega banks such as Citigroup, due to the size and complexity of such modern “one stop shops”. Given the high level of data aggregation for banks, it is impossible to make a definitive statement on whether they are leveraged or not in the context of our model: a portion of the bank industry is definitely leveraged and acts to maximize its ROE subject to a bankruptcy constraint.

As for insurance companies, their traditional role has been to invest the premiums they receive into assets so as to be able to match their future liabilities (to policy holders). Thus, with respect to our model setup, they would probably fit into the unlevered category as they do not typically borrow in order to invest. However, they are tricky to classify as purely unlevered players within the context of our model, given their business model and all regulations imposed on them.

If we assume that one third of the banks fit into our levered category, the ratio \( y_t \) is approximately 5.9% (6.9% if insurance companies are excluded from the unlevered category). Even if we assume that all banks fit into our levered category, the ratio \( y_t \) only increases to approximately 9.6% (11.2% if insurance companies are excluded from the unlevered category).

Furthermore, one must bear in mind that the Table 8 numbers are for 2009. As mentioned in Section 3, there are numerous empirical studies which document the rapid increase in size over the past 30 years of levered players such as broker/dealers and hedge funds. This fact enhances our assumption of \( y_t \) being significantly smaller than 1 over the past 30 years.

The discussion in this section highlights the lack of appropriate data and the difficulty in estimating \( y_t \). Given its critical importance, an important regulatory objective is to be

---

47 A high net worth individual (HNWI) is defined as someone who has more than $1 Million in investable assets.

48 One third is the estimate of the size of US broker/dealers as a percentage of US commercial banks given in [7].
able to appropriately taxonomize various financial system players and obtain the appropriate data.

<table>
<thead>
<tr>
<th>Institution Type</th>
<th>Total Worldwide AUM in 2009 (US$ Trillions)</th>
<th>Levered or Unlevered?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pension Funds</td>
<td>28.0</td>
<td>Unlevered</td>
</tr>
<tr>
<td>Mutual Funds</td>
<td>22.9</td>
<td>Unlevered</td>
</tr>
<tr>
<td>Private Wealth (ex conventional investment management)</td>
<td>26</td>
<td>Unlevered</td>
</tr>
<tr>
<td>Sovereign Wealth Funds</td>
<td>3.8</td>
<td>Unlevered</td>
</tr>
<tr>
<td>Private Equity Funds</td>
<td>2.5</td>
<td>Levered</td>
</tr>
<tr>
<td>Hedge Funds</td>
<td>1.7</td>
<td>Levered</td>
</tr>
<tr>
<td>Top 1000 Banks (Tier 1 Capital)</td>
<td>4.9</td>
<td>Unclear</td>
</tr>
<tr>
<td>Insurance Company Funds</td>
<td>13.6</td>
<td>Unclear</td>
</tr>
</tbody>
</table>

Table 8: This table presents the total worldwide assets under management (AUM) of the major levered and unlevered institution types.
16.3.2 The Approximation

Recall that in order for the Taylor series approximation of $X_t$ in (81) to be valid, we assumed that:

$$\frac{1}{(F_t - K)^2} \frac{2K}{c_D} y_t \leq \epsilon_1 \ll 1 \tag{207}$$

We now discuss the effect that the approximation in $X_t$ has on two key identities. Recall that the resulting $(X_t, b_t)$ pair must satisfy equations (71) and (62): the former comes from the VaR restriction being tight and the latter comes from the supply/demand equation for L’s risky debt. If we impose that L satisfies his VaR constraint and plug in the simplified expression for $X_t$ in (81):

$$K = \frac{b_t X_t}{w_t^L + b_t} \Rightarrow$$

$$b_t = \frac{w_t^L K}{X_t - K} \Rightarrow$$

$$b_t = \frac{w_t^L K}{F_t + \frac{K}{F_t - K} \frac{y_t}{2c_D} - K} \Rightarrow$$

$$b_t = \frac{w_t^u y_t}{F_t + \frac{K}{F_t - K} \frac{y_t}{2c_D} - K} \Rightarrow$$

$$b_t = \frac{w_t^u y_t}{\frac{F_t - K}{K} + \frac{1}{F_t - K} \frac{y_t}{2c_D}} \Rightarrow$$

$$b_t = \frac{w_t^u y_t (l_t - 1)}{1 + \frac{K}{2c_D (F_t - K)^2} \frac{y_t}{2c_D}} \Rightarrow$$

$$b_t = \frac{w_t^u y_t (l_t - 1)}{1 + \frac{K}{2c_D (F_t - K)^2}}$$

Thus, according to the VaR constraint in (208), the amount invested in the risky debt should be a bit smaller than that presented in equation (83). However, note that the assumption in (82) is sufficient in justifying that $b_t \approx w_t^u y_t (l_t - 1)$ since the second term in the denominator is “small”. To see why:

$$\frac{1}{(F_t - K)^2} \frac{2K}{c_D} y_t \leq \epsilon_1 \Rightarrow \frac{1}{(F_t - K)^2} \frac{K}{2c_D} y_t \leq \frac{\epsilon_1}{4} \ll 1 \Rightarrow \frac{K y_t}{2c_D (F_t - K)^2} \approx 1 \tag{209}$$

Thus, the assumption in (82) allows us to assume that the identities in equations (71) and (62) still hold. As the assumption in (82) becomes less accurate, then $w_t^u y_t (l_t - 1)$ will increasingly overstate the amount that L borrows. If in a given trial, $y_t$ grows enough such
that the assumption in (82) becomes invalid, then we expect the simplified model behavior to deviate from the full model behavior.

An interesting aspect of the simplified model is that assuming that the assumption in (82) holds means that:

\[ X_t - F_t = \frac{K}{2cD} \frac{y_t}{F_t - K} \leq \epsilon_1 \frac{F_t - K}{4} < \epsilon_1 << 1 \]  

(210)

where we have used that \( F_t - K \leq 1 \).\(^{49}\) We denote \( \epsilon_1 \frac{F_t - K}{4} \) by \( \epsilon_2 \).

What the relationship in (210) means is that the differential between \( X_t \) and \( F_t \) will be small; this is consistent with the fact that short term lending rates to leveraged financial intermediaries have been very close to the risk free rate historically (see Figure 22 and the discussion in Section 2). In our model, in order for \( U \) to invest a non-negligible portion of his wealth into \( L \)’s debt, he must have a high risk tolerance for it: mathematically, this means that \( s_D \) must be small so that \( 2cD \) is such that while \( \frac{K}{2cD} \frac{y_t}{F_t - K} \ll 1 \), \( a^D_t = y_t \frac{K}{F_t - K} \) is not negligible.

Assuming that \( s_D \) is small is consistent with the historical fact that only 2 money market funds have “broken the buck” (i.e. lost money) over the past 40 years. Furthermore, it is consistent with the behavioral argument presented in [42], where the authors argue that in the years prior to the financial crisis, investors viewed short term financial intermediary debt as virtually default free and thus had no information gathering incentive; this resulted in them taking portfolio positions that resembled the market more than they thought or desired. This observation is related to systemic risk within the context of our model: \( U \)’s excess exposure to the investment asset if \( L \) defaults increases as \( a^D_t \) increases.

If \( U \) is proxied by the US mutual fund industry,\(^{50}\) then a very rough estimate of \( a^D_t \) over time is presented in Figure 43. Since 1990, it has varied roughly between 20% and 40%, and its ebbs and flows have coincided with economic expansions and contractions respectively. The drop in the early 1990s coincides with a recession and the aftermath of the Savings and Loan crisis. The increase in the late 1990s followed by the drop in the early 2000s mirrors the tech fueled expansion, which was then followed by the recession of the early 2000s. It then steadily grows during the credit boom years of 2003-2007 and then drops from 2008 onwards. Of course, when looking at Figure 43, one must keep in mind that the data is annual and that the short term risk free rate varied over the period. Through Figure 43, we are not trying to give a precise historical estimate of \( a^D_t \) as we lack the data to do so and clearly, there are many different types of financial intermediary debt with different horizons, covenants, etc.; rather, our aim is to give a ballpark figure of \( a^D_t \) and highlight that it is a non-negligible part of \( U \)’s portfolio.

\(^{49}\)Since \( F_t < m \), unless \( m \) is a wildly high number and/or \( K \) is close to 0, this is the case. For this not to hold for all \( F_t < m \), the “distance” between the mean and the \( \epsilon^{th} \) percentile would have to be over 100% in return terms. For realistic returns distributions at decision frequencies such as daily, weekly, and annually, \( m - K \) will be less than one.

\(^{50}\)We use this example because of data availability.
US Mutual Funds % Allocation to Financial Intermediary Debt

Figure 43: Estimated percentage allocation of US mutual fund assets into levered financial intermediary debt.

16.4 Variance of Next Period Investment

In this section, we examine how the variance of next period’s investment depends on $y_t$ when the risk free rate is constant. Given the expression for $P_{t+1}$ in (91) in the simplified model, it is clear that the current relative size of $y_t$ plays a role in the dispersion of the investment change in the next period. Conditional on no default at time $t+1$, the variance is:

$$Var_t \left( \frac{P_{t+1}}{P_t} \right) = \left( 1 + \frac{w_t^u}{P_t} ((l-1)ly_t - (1-a^v)a^v) \right)^2 \left(s'\right)^2$$

(211)

where $s'$ is the standard deviation of $V_{t+1}$ conditional on no default. Note that the term in parentheses above is always non-negative since:

$$a^v(1-a^v)w_t^u \leq a^v w_t^u \leq P_t$$

(212)

From (211), it is clear that the variance is an increasing function of $y_t$, and that when $y_t > y^c$, this variance is larger than $(s')^2$.

To isolate the effect of L’s procyclical behavior, consider an identical economy at time $t$ where $U$ does not adjust $a^D$ in response to L’s relative size; mathematically, this would mean $a_t = a_{t+1}$. In this alternative setup, the variance of the investment change would be lower. Denote the next period investment under this alternative setup as $P_{t+1}^a$. Then:
\[
\frac{P^a_{t+1}}{P_t} = V_{t+1} + \frac{w^u_t}{P_t} a_t (F - z_{t+1}) \\
= V_{t+1} - \frac{w^u_t}{P_t} a_t a^v (F - V_{t+1})
\]

Thus, the variance is:

\[
Var_t \left( \frac{P^a_{t+1}}{P_t} \right) = \left( 1 - \frac{w^u_t}{P_t} a_t a^v \right)^2 (s')^2 < (s)^2
\]

The inequality follows from:

\[
a^v a_t w^u_t < a^v w^u_t \leq P_t
\]

Thus, we have shown that the variance of next period's investment is an increasing function of \(y_t\) and will be above the fundamental variance if \(y_t > y^f\). This "extra" variance, compared to the alternative setup, is a result of the fact that \(a^D\) will increase following a good outcome and decrease following a bad outcome; the magnitude of these changes in \(a^D\) depend on the relative size of \(L\) to \(U\). In the alternative setup, variance is below the fundamental value due to the fact that a constant \(a^D\) "attenuates" the reinvestment response following both good and bad outcomes: a good outcome means that \(L\) grows more than \(U\); if \(a^D\) is constant, then the amount that \(L\) borrows grows only as fast as \(U\)'s wealth, meaning that some of the overall investment value will be shifted to the risk free asset. Analogously, a bad outcome means that \(L\) shrinks more than \(U\); if \(a^D\) is constant, then the amount that \(L\) borrows decreases less than \(L\)'s size, meaning that some funds are shifted from the risk free asset to \(L\)'s debt and thus the risky investment.
References


[54] W. Poole. The fed’s monetary policy rule. 2006.


