Institutional Investors and Asset Prices

by

Morad Elsaify

Department of Business Administration
Duke University

Date: __________________________

Approved:

______________________________
Campbell R. Harvey, Advisor

______________________________
Simon Gervais

______________________________
Anna Cieślak

______________________________
Andrew Patton

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University

2021
ABSTRACT
Institutional Investors and Asset Prices
by
Morad Elsaify
Department of Business Administration
Duke University

Date: ______________________________________________________________________

Approved:

__________________________________________
Campbell R. Harvey, Advisor

__________________________________________
Simon Gervais

__________________________________________
Anna Cieślak

__________________________________________
Andrew Patton

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Business Administration in the Graduate School of Duke University

2021
Abstract

In this dissertation, I study the sources and consequences of heterogeneity in the behavior of different institutional investors in financial markets. In particular, I show that (i) institutional investors respond to the incentives generated by their organizational structures, (ii) the strategy a particular institutional investor pursues depends on their skill, and (iii) institutional investors have heterogeneous impacts on equilibrium market prices.

In the first chapter, I begin by documenting the substantial heterogeneity in portfolios across different types of investors. To explain this phenomenon, I build a model in which investors have different information processing capabilities. The model predicts that highly capable investors specialize in factor timing, hold more volatile and dispersed portfolios, and reduce average risk premia and volatility. Using novel empirical measures of investors’ capabilities and information choices, I find that hedge funds are the most capable investors, while insurance companies and pension funds are the least. Variation in factor timing ability is the primary driver of these differences. Investors’ portfolios exhibit properties consistent with the model’s predictions. Using a demand system approach, I show that hedge funds have the greatest per-dollar impact on expected returns, shrinking expected returns in the factor zoo by nearly 40% per $1 trillion of invested capital.

In the second chapter, I examine whether hedge fund managers respond to the incentives generated by their compensation contracts. To do this, I present a model in which hedge fund managers maximize their expected compensation subject to leverage constraints. This allows me to explore the impact of hedge funds’ prototypical contract structure on their dynamic risk allocations and on asset prices in general. One implication of the model is that risk taking varies as a function of a fund’s dis-
tance to its high-water mark. My empirical work is consistent with the implications of the model in that hedge funds at and furthest from their high-water marks take on significantly greater levels of risk. This increased risk is accomplished in part by investing in more volatile securities. Further, as more hedge funds approach their high-water marks, aggregate hedge fund risk taking increases, the security market line flattens, and betting against beta returns increase, consistent with evidence that these funds increase investment in high beta securities. This work highlights the importance of misaligned preferences between financial intermediaries and investors in explaining asset prices.
Acknowledgements

I am indebted to my committee chairs, Cam Harvey and Simon Gervais, for their constant encouragement, suggestions, and belief; I would not have the main ideas behind this dissertation were it not for them. I thank Anna Cieślak for pushing me to think about the economics behind all of my ideas and her (occasional) critical feedback. I also thank Andrew Patton for always providing a different perspective to my research. More than anything else, I am grateful to my entire committee for their willingness to volunteer their time to mentor me.

Beyond my committee, I especially thank Brian Weller for encouraging my interest in machine learning and computer science. I will fondly remember my regular walks through the quad with Sharique Hasan, which provided illuminating perspectives on work, life, and productivity, as well as much needed respites from research.

I am incredibly grateful to have had the support of so many of my family and friends. I’m particularly grateful to my brother, Amora, for his constant willingness to listen to ideas, challenges, and many mock presentations, as well as his encouragement throughout the job market. I also thank both of my parents for their support throughout my education. Finally, I am grateful to have had so many avenues to get away from work, including Criterion Thursdays, watching the Leafs lose with Michael Boutros and James Pinnington, regularly beating Arthur Tellis in chess, and trying nearly every restaurant in Durham with Usa Kerdnunvong.
Contents

Abstract iv
Acknowledgements vi
List of Figures x
List of Tables xi
1 Introduction 1
2 Which Investors Drive Factor Returns? 2
  2.1 Introduction ........................................... 2
  2.2 Model ................................................... 11
    2.2.1 Model Setup ..................................... 11
    2.2.2 Equilibrium ....................................... 16
    2.2.3 Relative Attention Allocation .................... 23
    2.2.4 Payoff Covariances ............................... 25
    2.2.5 Portfolio Characteristics .......................... 27
    2.2.6 Portfolio Composition ............................ 31
    2.2.7 Impact on Factors ................................ 32
  2.3 Mapping Theory to Data .............................. 34
    2.3.1 Proxies of Attention Allocation ................... 35
    2.3.2 Demand System Estimation ........................ 37
    2.3.3 Data ............................................. 40
  2.4 Empirical Evidence ................................. 44
    2.4.1 Identification of Attention Capacity .......... 44
    2.4.2 Attention Allocation ............................ 47
    2.4.3 Portfolio Characteristics ....................... 53
2.5 Investors’ Impact on Factors .............................................. 58
  2.5.1 Construction of Counterfactual Factors .............................. 59
  2.5.2 Impact on Risk Premia ................................................. 61
  2.5.3 Impact on Volatility .................................................... 69
2.6 Conclusion ........................................................................ 71

3 Hedge Fund Incentives, Risk Taking, and Asset Prices ............. 73
  3.1 Introduction .................................................................... 73
  3.2 Model ............................................................................. 79
    3.2.1 Setup ....................................................................... 79
    3.2.2 Mutual Funds ........................................................... 84
    3.2.3 Hedge Funds .............................................................. 84
    3.2.4 Equilibrium ............................................................... 92
  3.3 Data .............................................................................. 97
    3.3.1 Morningstar CISDM .................................................... 97
    3.3.2 Thomson Reuters S34 Institutional Holdings .................. 99
  3.4 Methodology ................................................................. 100
    3.4.1 Estimation of High-Water Mark Distance ....................... 100
    3.4.2 Measures of Risk ........................................................ 104
    3.4.3 Proxies of Hedge Fund Wealth Distribution and Market Quantities .................................................. 107
  3.5 Results ............................................................................ 112
    3.5.1 Fund-Specific Results .................................................. 112
    3.5.2 Equilibrium Results ..................................................... 122
  3.6 Conclusion ....................................................................... 129

viii
List of Figures

2.1 Portfolio Characteristics by Investor Type .............................. 3
2.2 Percent Ownership by Investor Type ............................... 41
2.3 Proxy for Attention Capacity .......................................... 46
2.4 Time Series Exposures by Investor Type ............................... 51
2.5 Portfolio Dispersion by Investor Type ............................... 54
2.6 Individual Portfolio Variance by Investor Type ....................... 55
2.7 Individual Autocorrelation by Investor Type ......................... 57
2.8 Scaled Repricing by Characteristic .................................... 63

3.1 Model Overview ............................................................ 81
3.2 Hedge Fund Manager Utility ........................................... 92
3.3 Comparative Statics ....................................................... 95
3.4 Proxies of Hedge Fund Wealth Distribution ........................... 109
3.5 Aggregate Hedge Fund Risk and SML Estimates ....................... 111
3.6 HWM Distance Through Time ........................................... 113

A.1 Repricing by Investor Type .............................................. 144
A.2 Scaled Repricing by Investor Type ...................................... 145
A.3 Cross Sectional Correlations Between Investors Over Time .......... 147
A.4 Inter-Asset Attention Allocation ....................................... 152
List of Tables

2.1 Estimated Attention Proxies, All Characteristics ........................................ 48
2.2 Estimated Correlation Attention Proxies ....................................................... 50
2.3 Factor Timing Decomposition ........................................................................ 52
2.4 Absolute Price Impact by Investor Type ......................................................... 64
2.5 Price Impact by Investor Type and Average Return ......................................... 65
2.6 Regression of Repricing on Returns ................................................................. 66
2.7 Scaled Repricing by Characteristic and Investor Type .................................... 68
2.8 Volatility Impact by Investor Type .................................................................. 70

3.1 Number of Funds Remaining After Each Filter ................................................. 98
3.2 Summary Statistics of Fund Characteristics ...................................................... 103
3.3 Summary Statistics of Company Characteristics ............................................... 106
3.4 Volatility by Distance Decile .......................................................................... 114
3.5 Fund Return Volatility Regressions ................................................................. 116
3.6 Average Long Volatility Regressions ............................................................... 119
3.7 Sources of Risk Regressions ............................................................................ 120
3.8 Hedge Fund AUM and Flows .......................................................................... 123
3.9 Average Hedge Fund Volatility Results ........................................................... 125
3.10 Security Market Line Slope Estimates ............................................................. 126
3.11 Betting Against Beta Regression Results ....................................................... 128

A.1 Estimated Attention Proxies Relative to Investment Universe ....................... 135
A.2 Estimated Attention Proxies Using Kalman Filter .......................................... 136
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.3 Estimated Attention Capacity Using Even and Odd Quarters</td>
<td>137</td>
</tr>
<tr>
<td>A.4 Effect of Short Positions, Aggregate Portfolios</td>
<td>138</td>
</tr>
<tr>
<td>A.5 Estimated Attention Capacity Using First and Second Half of Sample</td>
<td>139</td>
</tr>
<tr>
<td>A.6 Factor Timing Horizon</td>
<td>140</td>
</tr>
<tr>
<td>A.7 Estimated Investor Attention Capacity, $F_{Skill_i}$</td>
<td>141</td>
</tr>
<tr>
<td>A.8 Estimated Investor Timing Ability, $F_{Timing_i}$</td>
<td>142</td>
</tr>
<tr>
<td>A.9 Estimated Investor Selection Ability, $F_{Selection_i}$</td>
<td>142</td>
</tr>
<tr>
<td>A.10 Time Series Correlation/Covariance Matrix</td>
<td>146</td>
</tr>
<tr>
<td>A.11 Estimated Attention Proxies, All Characteristics</td>
<td>148</td>
</tr>
<tr>
<td>A.12 Model Parameters</td>
<td>152</td>
</tr>
<tr>
<td>A.13 Anomaly Characteristics</td>
<td>201</td>
</tr>
<tr>
<td>A.14 Anomaly Summary Statistics</td>
<td>208</td>
</tr>
<tr>
<td>A.15 Largest Investors</td>
<td>213</td>
</tr>
<tr>
<td>A.16 Model Parameters</td>
<td>231</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

While most investors have a common goal—to generate positive risk-adjusted returns—there is substantial heterogeneity in their organizational structures and portfolios they hold. This dissertation focuses on understanding both the roots and consequences of this heterogeneity across different classes of institutional investors. In particular, I focus on three questions.

First, how skilled are different institutional investors, and how does their skill affect the strategies they pursue and investment style? This is addressed in the first chapter, where I show that there is substantial heterogeneity in skill across different investor classes, and that factor timing ability drives almost all of this variation.

Second, how do different institutional investors impact equilibrium market prices, and how does this vary with their overall skill? This too is addressed in the first chapter, where I estimate demand functions for individual investors and compute counterfactual market prices as if a particular class of investors did not exist. I show that market impact varies drastically across these different investor classes, with some investors being heavily influential in setting equilibrium market prices.

Finally, do institutional investors optimize their portfolios in response to the incentives generated by their organizational structure? This is addressed in the second chapter, where I study differences in the compensation contracts between hedge funds and mutual funds. I show that hedge funds display time-varying risk preferences consistent with the incentives generated by their contract structures, and that these risk preferences impact equilibrium market prices.
Chapter 2

Which Investors Drive Factor Returns?

2.1 Introduction

Most investors—institutional and retail alike—have a common goal: to provide positive risk-adjusted returns for themselves or their clients. Despite this common goal, there is substantial heterogeneity in portfolio composition across investor classes. This heterogeneity is particularly noticeable in investors’ exposures to firm and stock characteristics in both the time series and the cross section. Figure 2.1 illustrates this heterogeneity by plotting the exposures of different investor types to a variety of characteristics in the factor zoo (Cochrane, 2011). Investors differ greatly in both the characteristics they tilt towards or against and the degree to which they deviate from the market portfolio (corresponding to a score of 0). In this paper, I propose an explanation for this phenomenon based on differences in information processing capabilities across investors and examine the implications for an investor’s strategy choice, portfolio allocation, and role in equilibrium price setting.

I start with a model of information choice and portfolio allocation. In the model, a single risk factor contains a persistent “fundamental” and a transitory “idiosyncratic” shock. Investors choose how to allocate their attention between these two components and then, using the information they acquire, make portfolio decisions. When allocating their portfolios, investors choose average exposures based on their information about the fundamental (factor selection), as well as changes in exposures based on their information about the idiosyncratic component (factor timing).
Differences in investors’ abilities to process information lead to variation in optimal information choices across investors, which carry a host of implications for portfolio choice and importance in the price formation process.

**Figure 2.1:** Portfolio Characteristics by Investor Type

![Figure 2.1](image)

*Note:* This figure plots the average characteristic decile (relative to the market portfolio) by investor type. Each quarter, firms are assigned into deciles for each of the 55 characteristics considered. An investor’s decile score is defined as the difference between the value-weighted average decile of her portfolio and the score of the market portfolio. A positive value indicates exposure to the long side of the characteristic relative to the market portfolio. Characteristics are sorted on the x-axis by the average hedge fund exposure. Details about the characteristics and their construction are given in Appendix A.1.4. The sample is from 1999Q1 to 2018Q4.

I demonstrate empirically that investors make information and portfolio choices that are consistent with the model. Using novel empirical proxies to measure investors’ information processing capabilities and choices, I document that hedge funds are the most capable investors and long-term investors (insurance companies and pension funds) are the least. Decomposing their ability into factor selection and tim-
ing strategies, I find that differences in investors’ capabilities are explained by their ability to time factors, rather than select factors. Using holdings-level data, I show that an investor makes portfolio decisions that are consistent with her information capabilities and choices.

Lastly, I quantify the importance of different types of investors in determining risk premia and volatility by re-estimating prices of risk under various counterfactual scenarios. Following the nascent literature on demand systems in asset pricing (Koijen and Yogo, 2019; Koijen et al., 2020), I find that hedge funds have the largest per-dollar impact on risk premia, as well as a sizable impact on volatility. In particular, hedge funds shrink expected returns in the factor zoo by nearly 40% and reduce volatility by 0.9 pp per annum per $1 trillion of invested capital. The effect on risk premia is concentrated among the factors with the highest expected returns, such as size and value. Within this group, expected returns fall by 3.3 pp per $1 trillion. For a reduction in hedge fund capital similar to that suffered in the Great Recession of 2008, this amounts to a 1.9 pp increase in risk premia, a rise of over 25%. The ability of other investors to impact factor returns is limited.

My model makes two key contributions to the literature studying attention allocation and portfolio choice. First, it considers the information choice problem as a function of investors’ total information processing capabilities. This approach helps explain observed differences in strategies and portfolio characteristics across investor classes. In doing so, I provide theoretically motivated empirical counterparts to investors’ attention capacities and information choices. Adding cross sectional heterogeneity in ability to the information choice problem complements existing work relying on time-varying market conditions to identify changes in attention allocations, as in Kacperczyk et al. (2016).

Second, my model considers an information choice problem in which there are
two types of uncertainty about an aggregate risk factor. Specifically, investors face uncertainty about the average payoff associated with the risk factor, as well as its idiosyncratic realization. Existing models focus on the information choice problem when there are only transitory shocks, either to specific assets (as in Van Nieuwerburgh and Veldkamp, 2010) or to an aggregate risk factor (Kacperczyk et al., 2016). My model allows for uncertainty about a persistent shock, which can arise due to time-varying risk premia that are difficult to forecast (as shown in Welch and Goyal, 2008) or the multitude of factors in the factor zoo.\(^1\) When there are hundreds of proposed risk factors, investors face uncertainty about not only each risk factor’s idiosyncratic realization, but also its average premium. This insight builds on attempts to account for the factor zoo that incorporates this fundamental uncertainty (Harvey et al., 2016; Chinco et al., 2019; Chen and Zimmerman, 2019).\(^2\)

The model generates several predictions for how information choices, portfolio characteristics, and price impact vary with attention capacity. An investor’s information choice depends on the uncertainty about the fundamental relative to uncertainty about the idiosyncratic component. When fundamental uncertainty is high, investors with more capacity allocate a greater fraction of their attention to learning about the idiosyncratic component. As investors use their information when making portfolio decisions, this implies that high capacity investors engage more in factor timing. Investors with less capacity, on the other hand, focus on learning about the fundamental, meaning they engage primarily in factor selection.

Differences in attention capacity and allocations imply differences in the dispersion, variance, and autocorrelation of investors’ risk factor exposures. The dispersion and variance of an investor’s portfolio are driven by two competing effects. An in-

\(^1\)My model can incorporate an arbitrary number of risk factors.

\(^2\)Gabaix (2014) and Martin and Nagel (2020) study learning using sparsity/shrinkage-based approaches when there are a potentially large number of covariates, as in the factor zoo.
vestor with greater attention capacity receives more precise signals, reducing the variance and dispersion of her portfolio. However, as she is aware of the quality of her information, she updates and acts on her beliefs more aggressively. The latter effect dominates, meaning that portfolio dispersion and variance in both the time series and the cross section are increasing in attention capacity.

The autocorrelation of an investor’s risk factor exposures depends on her information choices. An investor who does not learn about the idiosyncratic component will not react to the realization of its signal. As a result, she will have a much more stable portfolio over time. The relationship between attention capacity and portfolio autocorrelation therefore depends on the relationship between attention capacity and information choices. When fundamental uncertainty is high, an investor with higher capacity spends a greater fraction of her attention on the idiosyncratic component, resulting in a negative relationship between capacity and autocorrelation.

The final predictions from the model concern the role of attention capacity in determining equilibrium prices. In the model, prices are endogenously determined as an aggregation of all investors’ beliefs. Investors’ weights in this aggregation are proportional to the quality of their information. As a result, investors with superior information have a greater impact on prices. This is true in the pricing of both the fundamental and idiosyncratic components. Further, the model allows me to examine the direction of this price impact. Investors with superior information increase price informativeness by bringing prices closer to the actual factor realizations. Investors with less precise information, on the other hand, reduce price informativeness by reducing the average quality of investors’ information.

Investors’ impact on price informativeness has implications for both risk premia and volatility. Increasing price informativeness shrinks expected returns towards zero. This is because high capacity investors act as arbitrageurs by bringing prices closer
to their fundamental values. The same logic applies to the idiosyncratic component. An increase in price informativeness brings prices closer to a factor’s idiosyncratic realization at each point in time. As a result, high capacity investors also reduce the volatility of factor returns.

There are three key challenges to testing these predictions empirically. The first is obtaining a universe of factors to which investors plausibly allocate attention. I borrow from the large literature in asset pricing documenting the cross sectional relationship between characteristics and returns (Harvey et al., 2016; McLean and Pontiff, 2016). To that end, I gather a set of 55 accounting and price variables that have been documented to predict returns cross sectionally and form a long-short portfolio for each. The resulting factor universe generates a panel that is sufficiently large to conduct meaningful time series and cross sectional tests.

Second, investors’ attention capacities and information choices are fundamentally unobservable. However, my model provides a heuristic to identify both of these quantities using the covariance between an investor’s risk factor exposures and factor returns. Investors form portfolios that covary with their beliefs, an insight used in Kacperczyk et al. (2016). Those with more precise information have more accurate beliefs, meaning they are able to choose a portfolio that covaries highly with factor realizations. Thus, an investor’s attention capacity can be proxied by the covariance between her portfolio and factor returns. In the same vein, an investor’s information choice is revealed by the decomposition of this covariance into its cross sectional and time series dimensions. An investor who receives precise information about the fundamental is able to choose average exposures that covary positively with average returns. Thus, she will have a high cross sectional covariance between her portfolio and returns. Similarly, an investor who receives precise information about the idiosyncratic component is able to adjust her portfolio in anticipation of idiosyn-
cratic returns, meaning that she will have a high time series covariance between her portfolio and idiosyncratic returns.

The third challenge comes in quantifying the impact different investors have on prices. The relevant experiment considers how factors would differ in the absence of a particular investor class. Estimating such quantities requires a fully specified system of asset demand. Following Kojien and Yogo (2019) and Kojien et al. (2020), I estimate demand functions for each investor at each point in time as a function of observable stock characteristics. The resulting demand system allows me to estimate counterfactual factor returns under various scenarios. Using these counterfactuals, I can examine how risk premia and volatility change in the absence of a group of investors.

Empirically, I find evidence consistent with the model’s theoretical predictions. Conforming to conventional wisdom, I estimate that hedge funds have the highest attention capacity, while long-term investors and short-sellers have the least. The vast majority of this variation is due to differences in timing ability: hedge funds particularly excel at factor timing, whereas long-term investors and short-sellers do not. This supports the result that investors with high capacity spend relatively more attention on the idiosyncratic component, while those with low capacity spend the majority of their attention on the fundamental.

Further, the model’s predictions regarding the characteristics of an investor’s portfolio are borne out in the data. Hedge funds have portfolios that are by far the most dispersed, most volatile, and least autocorrelated. This is consistent with the fact that hedge funds seem to have the most precise information and specialize in factor timing. Long-term investors, on the other hand, have portfolios that exhibit the least dispersion, lowest variance, and among the highest autocorrelation of all investors considered.
Finally, using the demand system to compute counterfactual risk prices, I estimate that hedge funds have the greatest per-dollar impact on asset prices. Per $1 trillion of invested capital, hedge funds move expected returns by an average of 3.97 pp per annum in absolute value. This effect is several times greater than all other investors, with the same reduction in the capital of long-term investors moving risk premia by only 0.4 pp in absolute value. Moreover, the direction of hedge funds’ impact on expected returns depends on the sign and magnitude of realized returns. Among the factors with the highest average realized returns (e.g., size and value), hedge funds decrease risk premia by 3.35 pp per $1 trillion of invested capital, compared with a 2.16 pp increase among factors with negative realized returns. Hedge funds also have a substantial impact on volatility; per $1 trillion of invested capital, hedge funds reduce volatility by 0.9 pp, a 2.9% decrease. By shrinking expected returns towards zero and reducing the volatility of factor returns, these results illustrate the important and unique role hedge funds play in the price formation process.

Related Literature

This paper contributes to three strands of literature. First, it builds on previous work studying information choice in the context of portfolio management. Among these, Van Nieuwerburgh and Veldkamp (2010) study how under-diversification can arise due to information acquisition. Kacperczyk et al. (2016) study how information choices varies with the business cycle. Specifically, they show that investors shift attention towards aggregate shocks during recessions and towards asset-specific shocks during booms. Abis (2020) builds on this to study how performance differs over time between quantitative investors, who can only learn about aggregate shocks, and discretionary investors. Sammon (2020) examines the the effect of the introduction of ETFs on information choice and price informativeness. I complement these papers
by studying how an investor’s information choice depends on her total attention. In
doing so, I not only document substantial heterogeneity in attention capacity across
investors, but also shed light on the stark differences in strategies pursued by the
cross section of investors and the implications this has for price informativeness.³

Second, I provide evidence on the roles different investors play in financial mar-
kets. There is a large body of work that examines the roles of “smart money” and
“dumb money” on asset prices. For example, Akbas et al. (2015) use flows to exam-
ine whether hedge funds and mutual funds attenuate or amplify mispricings. A cross
sectional predictor using a similar narrative is constructed in Frazzini and Lamont
(2008). Caglayan et al. (2020) use disagreement between hedge funds and other in-
stitutional investors to construct another cross sectional predictor. The performance
of different groups of investors is further explored in McLean et al. (2020). Rather
than rely on common narratives to decide who are the “smart” and “dumb” money
investors, I use theoretically motivated empirical measures of ability and information
choices to identify these investors in the data, and test whether they make optimal
information choices.

Lastly, I quantify the impact different investors have on both expected returns
and volatility using a demand system approach (Kojen and Yogo, 2019). Using
this demand system to compute counterfactual returns in the absence of a group of
investors, I provide evidence on the ability of investor classes to affect risk premia
and volatility. In a similar vein, Kojen et al. (2020) analyze heterogeneity in demand
curves across investor classes and determine which investors have the greatest impact
on stock prices. My analysis focuses on the systemic changes in prices along defined
stock characteristics to examine investors’ impacts on factor risk premia and volatility.

³Similarly, Dim et al. (2020) study how the rise of factor investing affects price informativeness
and Breugem and Buss (2019) examine how relative performance evaluation in the mutual fund
industry lowers information acquisition.
Other research in this growing literature studies the ability of investors to affect real change in firms’ environmental practices (Noh and Oh, 2020).

The rest of the chapter is organized as follows. Section 2.2 presents the model and derives the main results. The link between the theory and empirics, as well as the details of the demand estimation and data used, is discussed in Section 2.3. Section 2.4 presents estimates of investors’ attention capacities and allocations and empirically tests the model’s predictions. Section 2.5 presents and discusses the impact different investors have on risk premia and volatility. Section 2.6 concludes. The Appendix contains additional theoretical and empirical results, a detailed derivation of the model and all proofs, and a list of all characteristics used and their definitions.

2.2 Model

2.2.1 Model Setup

There are $T \geq 2$ periods beginning at $t = 0$. At $t = 0$, investors allocate attention across a set of shocks at all future dates. At each date $t \geq 1$, investors receive a set of signals based on their information choice at $t = 0$ and make portfolio allocation decisions. All payoffs are realized at the end of period $T$. Appendix A.2.1 contains a detailed derivation of the model and all proofs.

Assets

There is a single riskless asset and many risky assets. The price of the riskless asset is normalized to 1 and pays $r$ each period. The risky assets follow a 1-dimensional factor structure.\(^4\) That is, each asset $k$ has exposure $\beta_k$ to a single risk factor $f_t$, as

\(^4\)The assumption of a single risk factor is made for simplicity. The case with $N$ identical risk factors is equivalent. Appendix A.1.2 explores the problem with heterogeneous risk factors in
well as exposure to stock-specific noise. Each asset has period $t$ payoffs given by

$$v_{k,t} = \alpha_k + \beta_k f_t + \eta_{k,t};$$

(2.1)

where each asset has a final cumulative payoff of $v_k = \sum_t v_{k,t}$. The vector of stock-specific noise $\eta_t \sim N(0, \Sigma_\eta)$ is assumed to be sufficiently independent and the number of assets sufficiently large such that the law of large numbers applies. As a result, the asset-specific noise $\eta_{k,t}$ is diversified away by investors and it suffices to focus on the risk factor $f_t$ (Ross, 1976).\(^5\)

**Risk Factor**

The risk factor $f_t$ is driven by two components, both of which are unknown. The first—the “fundamental” component—is persistent across time, while the second—the “idiosyncratic” component—is independent of the fundamental and serially uncorrelated. Formally,

$$f_t = \mu + \epsilon_t,$$

(2.2)

where

$$\mu \sim N(0, \sigma_\mu) \text{ and } \epsilon_t \sim iid N(0, \sigma_\epsilon).$$

(2.3)

The final payoff of the risk factor is $f = \sum_t f_t = T\mu + \sum_t \epsilon_t$. The idiosyncratic component is the transitory deviation of a factor payoff from its fundamental value. Uncertainty about the fundamental stems from the factor zoo; with hundreds of characteristics potentially associated with returns, there is uncertainty surrounding

\(^5\)This reduces the economy to one in which the aggregate risk factor is the only asset. In this way, this model studies how investors optimally allocate attention to a given factor, rather than specific stocks, and abstracts from under-diversified or concentrated portfolios. Van Nieuwerburgh and Veldkamp (2010) study under-diversification as a product of optimal information choice.
each factor’s average payoff. This is similar to models of the factor zoo in Chinco et al. (2019) and Chen and Zimmerman (2019), where expected returns are drawn from a mean zero normal distribution.

Finally, the risk factor has stochastic supply to rule out fully revealing prices. Supply is given by \( \bar{s} + s_t \), where \( \bar{s} \geq 0 \) and \( s_t \sim N(0, \sigma_s) \).

**Investors and Portfolio Allocation Problem**

There is a unit mass of investors with mean-variance utility over per-period wealth.\(^6\) There are \( M \) types of investors indexed by \( m \), each of mass \( \psi_m \) (such that \( \sum_m \psi_m = 1 \)). Each investor \( i \), whose type is denoted \( m_i \), has risk aversion \( \gamma \) and initial wealth normalized to zero. Denote by \( \mathbb{E}_{i,t}[\cdot] \) and \( \mathbb{V}_{i,t}[\cdot] \) the expectation and variance conditional on investor \( i \)'s time \( t \) information set. Each investor chooses how many shares \( q_{i,t} \) of the risk factor to hold given her information set and prices \( p_t \) in order to maximize her expected utility \( u_{i,t} \) subject to the evolution of wealth \( W_{i,t} = q_{i,t} (f_t - r p_t) \).

The portfolio allocation problem is given by

\[
\max_{q_{i,t}} \quad u_{i,t} = \gamma \mathbb{E}_{i,t}[W_{i,t}] - \frac{\gamma^2}{2} \mathbb{V}_{i,t}[W_{i,t}]
\]

\[
s.t. \quad W_{i,t} = q_{i,t} (f_t - r p_t) .
\]

**Signals and Attention Allocation Problem**

At time \( t = 0 \), investors allocate attention across a set of signals they receive at all future dates.\(^7\) In this context, allocating “attention” means receiving a signal with greater precision. This can be thought of as a stand-in for skill, effort, technological

\(^6\)The lack of wealth effects makes this equivalent to having utility over terminal wealth.

\(^7\)Investors optimally do not update their attention allocations in response to the realization of their signals or investment decisions, so this restriction is without loss of generality.
resources, etc. Investors receive a sequence of noisy signals about the fundamental and idiosyncratic components at each time $t \geq 1$. Formally, the fundamental and idiosyncratic signals are given by

$$\theta_{\mu,i,t} = \mu + \xi_{\mu,i,t} \quad \text{and} \quad \theta_{\epsilon,i,t} = \epsilon_i + \xi_{\epsilon,i,t},$$

where $\xi_{\mu,i,t} \overset{iid}{\sim} N(0, K_{\mu,i,t}^{-1})$ and $\xi_{\epsilon,i,t} \overset{iid}{\sim} N(0, K_{\epsilon,i,t}^{-1})$ are independent of each other and independent across investors and time. The signals are therefore unbiased. The precisions of these signals, $K_{\mu,i,t}$ and $K_{\epsilon,i,t}$, are chosen by the investor subject to the constraints described below. In addition to these signals, investors can use some of their attention to learn about the fundamental and idiosyncratic components via prices.\(^8\) These signals are common across investors and unbiased with precisions $K_{p,\mu,t}$ and $K_{p,\epsilon,t}$, respectively.

The only ex-ante difference between investors of different types is their total attention capacity, which restricts the feasible choice of signal precisions. Investor $i$ is endowed with total attention capacity $K_{m,i} \geq 0$, which she can freely allocate between the fundamental and idiosyncratic signals across each of the $T$ trading dates subject to two constraints. The first is an information capacity constraint, which limits the sum of signal precisions across all dates to be below the investor’s total attention capacity, as in Kacperczyk et al. (2016):

$$K_{m,i} \geq \sum_{t=1}^{T} \left( K_{\mu,i,t} + K_{\epsilon,i,t} + \ell_{\mu,i,t} K_{p,\mu,t} + \ell_{\epsilon,i,t} K_{p,\epsilon,t} \right),$$

where $\ell_{\mu,i,t}$ and $\ell_{\epsilon,i,t}$ are indicators equal to 1 if investor $i$ decides to learn from

\(^8\)Standard rational expectations models typically assume it is costless to learn from prices (e.g., Admati, 1985). Theoretically, extracting a signal from prices involves perfect knowledge of the economy’s parameters and the attention allocation decisions of all investors in the economy. In practice, investors may face costs in obtaining such information and extracting such signals.
prices about the fundamental and timing components at time $t$, respectively. The formulation of the information capacity constraint (2.6) imposes that the costs of obtaining a signal from prices are equal to those of obtaining a signal of equal precision that is conditionally independent of prices.

This constraint can be thought of as a research budget. Investors only have a limited amount of resources to spend on learning about future payoffs, but some investors may be more efficient, exert more effort, or have larger endowments to spend. For example, some investors may have human capital or technological advantages or more aligned incentives that induce higher effort. These differences are reflected in an investor’s attention capacity.

The second constraint is a no-forgetting constraint, which is simply a non-negativity constraint on the chosen precisions:

$$K_{\mu,i,t}, \ K_{\epsilon,i,t} \geq 0 \ \forall t \in \{1, \ldots, T\}. \quad (2.7)$$

Each investor $i$ chooses the sequence of fundamental attention $\{K_{\mu,i,t}\}_{t=1}^{T}$, idiosyncratic attention $\{K_{\epsilon,i,t}\}_{t=1}^{T}$, and learning-from-price indicators $\{\iota_{\mu,i,t}, \iota_{\epsilon,i,t}\}_{t=1}^{T}$ in order to maximize the sum of her unconditional expected utility. Formally, the information choice problem is given by

$$\max_{\{K_{\mu,i,t}, K_{\epsilon,i,t}, \iota_{\mu,i,t}, \iota_{\epsilon,i,t}\}_{t=1}^{T}} \mathbb{E} \left[ U_i \right] = \sum_{t=1}^{T} \mathbb{E} \left[ u_{i,t} \right] \quad (2.8)$$

$$s.t. \quad K_{m_i} \geq \sum_{t=1}^{T} \left( K_{\mu,i,t} + K_{\epsilon,i,t} + \iota_{\mu,i,t} K_{p,\mu,t} + \iota_{\epsilon,i,t} K_{p,\epsilon,t} \right)$$

$$K_{\mu,i,t}, \ K_{\epsilon,i,t} \geq 0 \ \forall t \in \{1, \ldots, T\}.$$
Market Clearing

Prices are determined by the market clearing condition

\[ \int_0^1 q_{i,t} dt = \bar{s} + s_t, \quad (2.9) \]

where the right hand side is total (stochastic) supply and the left is integrated over each investor’s demand. Total demand can be alternatively expressed as

\[ \int_0^1 q_{i,t} dt = \sum_{m=1}^{M} \psi_m \bar{q}_{m,t}, \quad (2.10) \]

where \( \bar{q}_{m,t} \equiv \psi_{m}^{-1} \int_{\{i; \ m_i=m\}} q_{i,t} dt \) is the average demand of type \( m \) investors.

2.2.2 Equilibrium

This model makes an important deviation from previous work in rational expectations equilibrium models in that it generalizes the structure of risk in the economy. Previous work (e.g., Grossman and Stiglitz, 1980; Admati, 1985; Kacperczyk et al., 2016) assumes that each asset has a single aggregate shock that the investor learns about, making it possible to obtain closed form solutions when investors learn from prices. This is generally not possible when the number of shocks exceeds the number of assets, as in this case. This paper resolves this issue by assuming learning from prices is costly.\(^9\) Specifically, holding the precision fixed, the costs of obtaining a signal via prices are equivalent to those of obtaining a signal that is conditionally independent of prices. The result is that investors never choose to learn from prices when the supply is sufficiently volatile. The following result formalizes this.

\(^9\)Such costs reflect the difficulties of learning about all of the economy’s parameters, as well as the actions of all its participants, and are manifested in increased effort, labor costs, technological investment, etc.
Lemma 1. An investor who faces the attention allocation problem in (2.8) never learns from prices for sufficiently large supply variance $\sigma_s$.

The basic intuition for Lemma 1 is as follows. If an investor learns from prices about the fundamental, she will believe the realization is high when either (i) the idiosyncratic component has a high realization, or (ii) the supply shock is large. The former is beneficial from the investor’s perspective, since this increases the correlation between her beliefs and payoffs. The latter, however, will lead the investor to buy when the price is high and sell when the price is low. If the variance of the supply shock is sufficiently high, the latter effect dominates the former. A similar logic applies to learning about the idiosyncratic component.

In what follows, $\sigma_s$ will be assumed to be sufficiently high such that investors never learn from prices.\(^{10}\) It therefore suffices to focus on equilibria in which the investor obtains information that is conditionally independent of prices and ignores the information contained therein.\(^{11}\) The model is solved by first deriving each investor’s posterior beliefs, computing their optimal demand and equilibrium prices, and, finally, deriving the information choices of each investor. Each step is considered in turn.

Learning

Investors revise their beliefs about the fundamental and idiosyncratic components conditional on the realization of their signals via Bayesian updating. Investor $i$’s

\(^{10}\)There is some empirical evidence documenting that supply variance is large. For example, Cho (2020) estimates that noise trading is responsible for roughly 20% of daily stock market volatility.

\(^{11}\)Investors’ decisions to rationally ignore the information contained in prices can also be motivated by an anticipatory utility perspective, in which investors choose subjective beliefs about the informativeness of their own signals and that of prices. With sufficiently high risk aversion, Banerjee et al. (2019) shows that investors optimally choose to ignore the information contained in prices. Learning from prices is also absent in the characteristics-based demand model of Kojien et al. (2020).
posterior beliefs about the fundamental are normally distributed with mean $\hat{\mu}_{i,t}$ and precision $\hat{\sigma}_{\mu,i,t}^{-1}$, which are given by

$$\hat{\mu}_{i,t} = \hat{\sigma}_{\mu,i,t} \sum_{s=1}^{t} K_{\mu,i,s} \theta_{\mu,i,s} \quad \text{and} \quad \hat{\sigma}_{\mu,i,t}^{-1} = \sigma_{\mu}^{-1} + \sum_{s=1}^{t} K_{\mu,i,s}. \quad (2.11)$$

Similarly, beliefs about the idiosyncratic component are normally distributed with mean $\hat{\epsilon}_{i,t}$ and precision $\hat{\sigma}_{\epsilon,i,t}^{-1}$:

$$\hat{\epsilon}_{i,t} = \hat{\sigma}_{\epsilon,i,t} K_{\epsilon,i,t} \theta_{\epsilon,i,t} \quad \text{and} \quad \hat{\sigma}_{\epsilon,i,t}^{-1} = \sigma_{\epsilon}^{-1} + K_{\epsilon,i,t}. \quad (2.12)$$

The posteriors in Eqs. (2.11) and (2.12) reveal an important distinction between fundamental and idiosyncratic attention. Learning about the fundamental $\mu$ yields more precise beliefs in all future periods (hence the summation in Eq. (2.11)). Learning about the idiosyncratic component $\epsilon_t$, however, yields more precise beliefs only in the current period. In this way, fundamental attention can be thought of as a \textit{durable} good and idiosyncratic attention as a \textit{nondurable}.\textsuperscript{12}

Combining these yields the posterior beliefs of the risk factor payoff $f_t$:

$$\hat{f}_{i,t} = \hat{\sigma}_{f,i,t} \sum_{s=1}^{t} K_{\mu,i,s} \theta_{\mu,i,s} + \hat{\sigma}_{\epsilon,i,t} K_{\epsilon,i,t} \theta_{\epsilon,i,t}$$

$$\hat{\sigma}_{f,i,t}^{-1} = \left( \sigma_{\mu}^{-1} + \sum_{s=1}^{t} K_{\mu,i,s} \right)^{-1} + \left( \sigma_{\epsilon}^{-1} + K_{\epsilon,i,t} \right)^{-1}. \quad (2.13)$$

\textsuperscript{12}As the fundamental is assumed constant throughout the model, there is no depreciation to fundamental attention; the information obtained about the fundamental at time $t$ is equally useful at all later dates. In principle, one could allow the fundamental to evolve according to an AR(1) process, in which the persistence parameter represents a depreciation rate in fundamental attention.
Portfolio Allocation and Prices

Taking the first order condition of the portfolio problem (2.4) yields the usual optimal demand

\[ q_{i,t} = \gamma^{-1} \hat{\sigma}^{-1}_{f,i,t} \left( \hat{f}_{i,t} - r p_t \right) . \]  

(2.14)

Evaluating the market clearing condition (2.9) at the optimal demand (2.14) yields a function for prices that is linear in the fundamental, idiosyncratic component, and supply shocks:

\[ p_t = \frac{1}{r} \left[ a_t + b_t \mu + c_t \epsilon_t + d_t s_t \right] , \]  

(2.15)

where \( a_t, b_t, c_t \) and \( d_t \) are derived in Appendix A.2.1.

Attention Allocation

The optimal attention allocation can be solved by maximizing ex-ante expected utility given prices and an investor’s demand. Substituting optimal demand (2.14) into the per-period expected utility yields

\[ \mathbb{E} [ u_{i,t} ] = \frac{1}{2} \mathbb{E} \left[ \hat{\sigma}^{-1}_{f,i,t} \left( \hat{f}_{i,t} - r p_t \right)^2 \right] . \]  

(2.16)

To compute this expectation, note that \( \left( \hat{f}_{i,t} - r p_t \right) \) is normally distributed, as both posterior means \( \hat{f}_{i,t} \) and prices \( p_t \) are linear functions of normally distributed variables. Time \( t \) utility \( \hat{\sigma}^{-1}_{f,i,t} \left( \hat{f}_{i,t} - r p_t \right)^2 \) is therefore non-central \( \chi^2 \)-distributed. Evaluating its expectation,

\[ \mathbb{E} [ u_{i,t} ] = \frac{1}{2} \left\{ \hat{\sigma}^{-2}_{f,i,t} \mathbb{V}[\hat{f}_{i,t} - r p_t] + \hat{\sigma}^{-1}_{f,i,t} \mathbb{E}[\hat{f}_{i,t} - r p_t]^2 \right\} 
\]

(2.17)

\[ = \frac{1}{2} \hat{\sigma}^{-1}_{f,i,t} \left( 2b_t \hat{\sigma}_{\mu,i,t} + 2c_t \hat{\sigma}_{\epsilon,i,t} + v_t + a_t \right) - \frac{1}{2} , \]
where \( v_t \equiv \mathbb{V} [f_t - r_p t] \) is the variance of realized returns and derived in Appendix A.2.1.

In general, the marginal utility of allocating attention is not necessarily positive in Eq. (2.17). While increasing attention reduces the posterior variance, it also reduces the covariance between an investor’s forecast error \( \hat{f}_{i,t} - f_t \) and realized returns \( f_t - r_p t \). A high covariance increases utility, as this means an investor takes large positions when returns are high. When the price is very sensitive to realized returns (i.e., high \( b_t \) and \( c_t \)), marginal utility may become negative. This is similar to the “market breakdown” result in Battacharya and Spiegel (1991); if investors are too informed, trading can halt. To rule out such scenarios, Assumption 1 limits the maximum amount of attention with which each type of investor can be endowed. Technically, this imposes a bound on the maximum difference between price coefficients, which ensures marginal utility is always positive.

**Assumption 1.** Attention capacity is uniformly bounded:

\[
\max_m K_m \leq \sqrt{\frac{\sigma_\epsilon (\sigma_\mu + \sigma_\epsilon)}{\sigma_\mu \sigma_\epsilon}}.
\] (2.18)

With this bound on attention capacity, the structure of the attention allocation problem (2.8) admits two simplifications that, together, yield a tractable solution. First, investors will never allocate attention to the fundamental after the first trading date. This is because fundamental attention is a perfectly durable good; a signal about the fundamental at time \( t \) affects beliefs identically across time. As Assumption 1 guarantees that marginal utility is always positive, investors prefer having information earlier rather than later.

**Lemma 2.** Given Assumption 1, investors (i) never allocate attention to the fundamental after time \( t = 1 \): \( K_{\mu,i,t} = 0 \ \forall t \geq 2 \) and (ii) allocate a constant amount of
attention to the idiosyncratic component: \( K_{e,i,s} = K_{e,i,t} \equiv K_{e,i} \forall s, t \in [1, \ldots, T] \).

This, in turn, converts what was a dynamic problem into \( T \) identical static problems. As investors receive information about the fundamental only at time \( t = 1 \), posterior beliefs about the fundamental are constant across time periods. The marginal utility of allocation attention to the idiosyncratic component is therefore symmetric across time, meaning that an investor’s optimal idiosyncratic attention is constant across time. Lemma 2 formalizes these two properties of the optimal information choice.

**Corollary 1.** The price coefficients are constant: \( a_t = a, b_t = b, c_t = c \) and \( d_t = d \) for all \( t \in [1, \ldots, T] \).

An immediate result of Lemma 2 is that the price coefficients are constant across time. This occurs because investors have the same quality of information at each point in time. Corollary 1 states this.

With these results, the resulting attention allocation can be derived. Denoting the Lagrange multipliers of the information capacity constraint \( \lambda_i \), the no-forgetting constraint for the fundamental \( \nu_{\mu,i,t} \), and the no-forgetting constraint for the idiosyncratic component \( \nu_{e,i,t} \), the first order conditions are given by

\[
\begin{align*}
[K_{\mu,i}] & : \quad \frac{1}{2} T \hat{\sigma}_{\mu,i}^2 \hat{\sigma}_{f,i}^{-2} \left\{ 2 (c - b) \hat{\sigma}_{e,i} + v + a^2 \right\} - \lambda_i + \nu_{\mu,i} = 0 \\
[K_{e,i}] & : \quad \frac{1}{2} \hat{\sigma}_{e,i}^2 \hat{\sigma}_{f,i}^{-2} \left\{ 2 (b - c) \hat{\sigma}_{\mu,i} + v + a^2 \right\} - \lambda_i + \nu_{e,i} = 0,
\end{align*}
\]

where the time subscripts have been dropped due to Corollary 1.

Focusing on the factors for which the no-forgetting constraints do not bind (i.e.,
\( \nu_{\mu,i} = \nu_{\epsilon,i} = 0 \), an investor’s attention allocation is

\[
K_{\mu,i} = \frac{-K_m - T\sigma_\mu^{-1} - T\sigma_\epsilon^{-1} + \sqrt{T \left( K_m + \sigma_\mu^{-1} + T\sigma_\epsilon^{-1} \right) \left( K_m + \sigma_\mu^{-1} + T\sigma_\epsilon^{-1} + \zeta \right)}}{T - 1}
\]

(2.20)

\[
K_{\epsilon,i} = \frac{K_m + \sigma_\mu^{-1} + \sigma_\epsilon^{-1} - \sqrt{T^{-1} \left( K_m + \sigma_\mu^{-1} + T\sigma_\epsilon^{-1} \right) \left( K_m + \sigma_\mu^{-1} + T\sigma_\epsilon^{-1} + \zeta \right)}}{T - 1}
\]

(2.21)

where \( \zeta \equiv \frac{2(T-1)(c-b)}{v+a^2} \) is a term that accounts for the impact prices have on an investor’s attention allocation decision. An investor wants to allocate attention to the component to which prices do not respond. As a result, if prices are more sensitive to the fundamental than the idiosyncratic component (i.e., \( b > c \)), an investor will shift her attention towards the idiosyncratic component and vice versa. When the prices are equally sensitive to both the fundamental and idiosyncratic components, prices do not affect an investor’s decision. The effect of prices is diminished when the variance of returns \( v \) is high, as this means the benefit of learning is large.

**Lemma 3.** Given Assumption 1, if \( \frac{\sigma_\mu^{-1} - \sqrt{T}\sigma_\epsilon^{-1}}{K_m} \in \left( -1, \frac{1}{\sqrt{T}} \right) \), the no-forgetting constraints do not bind for investor \( i \) for sufficiently large \( \sigma^2 + \sigma_s \).

In general, interior solutions are not guaranteed. If \( \sigma_\mu \) and \( \sigma_\epsilon \) are sufficiently different or if \( T \) is large, investors will be drawn to use their entire attention capacity to learn about the more volatile source of risk. However, ex-ante utility is concave in both fundamental and idiosyncratic attention, so a corner solution will be optimal only if the interior solution in Eqs. (2.20) and (2.21) is not feasible. Lemma 3 describes a sufficient condition to ensure the no-forgetting constraints do not bind. First, the range of parameters for which an interior solution exists is increasing with total attention capacity. This is due simply to the fact that attention has decreasing
returns to scale, so allocating large amounts of attention to one component makes
the other relatively more attractive. Second, when $s^2 + \sigma_s$ is large, the impact of
prices on investors’ attention allocation decisions is limited, which ensures a feasible
interior solution. Note that this is only a sufficient criterion; interior solutions may
well exist outside this interval.

**Proposition 1.** If $\sigma_i/\sigma_\mu$ is greater (less) than $\sqrt{T}$, the price coefficient on the funda-
mental is less (greater) than that on the idiosyncratic component: $b < c$ ($b > c$). The
price coefficients are equal iff $\sigma_i/\sigma_\mu = \sqrt{T}$.

Proposition 1 characterizes how the prices respond to fundamental and idiosyn-
cratic realizations. When the ratio of idiosyncratic to fundamental variance $\sigma_i/\sigma_\mu$
(the “variance ratio”) is high, prices are more responsive to the idiosyncratic com-
ponent $\epsilon_t$ than to the fundamental $\mu$. This is because, in this case, learning about
the idiosyncratic component is relatively more attractive. As the marginal utility
of allocating attention is increasing in variance, a higher value of $\sigma_i$ relative to $\sigma_\mu$
raises the marginal benefit of idiosyncratic attention, despite having to allocate that
attention each period. When the variance ratio is equal to $\sqrt{T}$, prices are equally
informative of both the fundamental and idiosyncratic terms.

This and many other results depend on whether $\sigma_i/\sigma_\mu$ is greater or less than $\sqrt{T}$.
Caution should be taken in interpreting this condition literally. An implicit assump-
tion in the model is that it is equally costly to learn about the fundamental and
idiosyncratic components (see Eq. (2.6)). In practice, learning about one component
may be easier than the other. For example, the fundamental component may be
easier to learn about because there is a longer time horizon over which investors can
acquire information. This can be incorporated by multiplying $K_{\mu,i,t}$ in the informa-
tion capacity constraint (2.6) by a constant. The next subsections examine how an
investor’s strategy choice, portfolio characteristics, and importance in price setting
depend on this variance ratio and her attention capacity.

2.2.3 Relative Attention Allocation

Fundamental and idiosyncratic attention reflect investors’ decisions to pursue two
distinct strategies: factor selection and factor timing. By paying attention to the
fundamental, investors use their information to make portfolio decisions based on
beliefs about long-term expected payoffs. In contrast, paying attention to the id-
iosyncratic component informs investors about payoffs only in the current period,
leading to portfolios that are strategically tilted each period. While investors with
greater attention capacities will pay more attention to both fundamental and idiosyn-
cratic components, the fraction of total attention that is paid to a component (i.e.,
relative attention) distinguishes between those who factor select and those who fac-
tor time. Thus, by looking at relative attention, the model can address the question:
which investors factor time?

Proposition 2. For attention capacity sufficiently uniformly bounded, if $\sigma_r/\sigma_n$ is
greater (less) than $\sqrt{T}$, relative attention to the fundamental $K_{r}/K_{n}$, is increasing
(decreasing) in total attention capacity $K_m$. Relative fundamental attention is con-
stant iff $\sigma_r/\sigma_n = \sqrt{T}$.

Proposition 2 provides the answer. When the variance ratio is low, the fraction
of total attention allocated to the idiosyncratic component is increasing in investors’
attention capacity. This means that factor timing is a luxury good, as it is dispro-
portionately pursued by investors with high attention capacity. This occurs because
investors allocate attention to a single fundamental shock but many idiosyncratic
shocks. As a result, investors learn a lot about the fundamental shock and relatively
little about each idiosyncratic shock. Because of this, high capacity investors essen-
tially exhaust most of the benefit from learning about the fundamental and instead 
focus on the idiosyncratic component.

This and most subsequent results require that investors’ attention capacities are 
uniformly bounded. This limits the distortionary effects of prices on investors’ at-
tention allocation decisions. Recall that $\zeta$ from Eqs. (2.20) and (2.21) is a term that 
adjusts investors’ information choices based on the price process. If this term is too 
extreme, it can “reverse” investors’ preferences. For example, if the variance ratio is 
less than $\sqrt{T}$, it is more productive to learn about the fundamental without consider-
ing the effect of prices. As a result, investors pay more attention to the fundamental, 
raising the sensitivity of the price to the fundamental. This makes learning about 
the fundamental less attractive. If the distortionary effect of prices is large enough, it 
can undo this mechanism. As such, this and later results require that the impact of 
prices be limited. This can be done by, as stated in the result, ensuring that attention 
capacity is uniformly bounded, or by assuming that $s^2 + \sigma_s$ is sufficiently large. This 
latter condition increases the denominator of $\zeta$, limiting the impact of prices.

2.2.4 Payoff Covariances

Attention allocation decisions affect portfolio allocation decisions. An investor’s in-
formation choice is reflected in the covariance between an investor’s portfolio and risk 
factor payoffs. An investor with greater capacity is able to choose a portfolio that 
 covaries more with realizations because her beliefs are more accurate. For the same 
reason, an investor’s cross sectional and time series covariances reflect her fundamen-
tal and idiosyncratic attention, respectively. An investor who receives a very precise 
signal about the fundamental is able to choose a portfolio that covaries cross section-
ally with average returns, while an investor who receives a precise signal about the idiosyncratic component chooses a portfolio that covaries with idiosyncratic returns. To that end, these quantities function as empirical proxies for investors’ attention capacity and information choices.

The covariance between an investor’s portfolio and payoffs—“total covariance”—is given by

\[ \mathbb{C} \left[ q_{i,t}, f_t \right] = \gamma^{-1} \hat{\sigma}_{f,t}^{-1} \left[ (\hat{\sigma}_{\mu,i} K_{\mu,i} - b) \sigma_{\mu} + (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) \sigma_{\epsilon} \right]. \] (2.22)

This can be decomposed into cross sectional and time series dimensions. Let superscripts $CS$ and $TS$ denote cross sectional and time series quantities, respectively. The cross sectional covariance between investors’ portfolios and payoffs—“fundamental covariance”—is

\[ \mathbb{C}^{CS} \left[ q_{i,t}, f_t \right] = \gamma^{-1} \hat{\sigma}_{f,t}^{-1} (\hat{\sigma}_{\mu,i} K_{\mu,i} - b) \sigma_{\mu}, \] (2.23)

while the time series covariance—“idiosyncratic covariance”—is given by

\[ \mathbb{C}^{TS} \left[ q_{i,t}, f_t \right] = \gamma^{-1} \hat{\sigma}_{f,t}^{-1} (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) \sigma_{\epsilon}. \] (2.24)

These two measures reflect investor $i$’s factor selection and timing skills relative to the average investor. The price coefficients $b$ and $c$ are weighted averages of $\hat{\sigma}_{\mu,i} K_{\mu,i}$ and $\hat{\sigma}_{\epsilon,i} K_{\epsilon,i}$, respectively, which are both increasing in attention capacity. Investors with above (below) average attention to a component will exhibit positive (negative) covariance. This is the byproduct of a general equilibrium model: the average investor must have a portfolio that does not covary with payoffs. Proposition 3 formalizes the basic intuition that investors with greater information choose portfolios that covary more with realized payoffs.
Proposition 3. Given Assumption 1, for attention capacity sufficiently uniformly bounded, total covariance is strictly increasing in attention capacity. Fundamental and idiosyncratic attention are increasing in attention capacity (at least one strictly so). Further, fundamental and idiosyncratic covariances are equal to zero for $K_m = \bar{K}_\mu$ and $K_m = \bar{K}_\epsilon$, respectively, where $\bar{K}_\mu$ and $\bar{K}_\epsilon$ are defined in Appendix A.2.1.

These covariances can also be used to examine the relative attention allocation decision from Proposition 2. The fraction of total covariance that comes from the cross section, $\frac{\sigma^S_{\mathcal{C}[\bar{n},\bar{f}]}}{\sigma^C_{\mathcal{C}[\bar{n},\bar{f}]}}$, reflects the fraction of attention allocated to the fundamental. If investors with greater attention capacity allocate relatively more attention to the idiosyncratic component (which occurs when $\sigma_i/\sigma_\mu < \sqrt{T}$), idiosyncratic covariance increases with attention capacity faster than fundamental covariance. This is because an extra unit of capacity is allocated in increasing proportions to the idiosyncratic component. As a result, for investors with above average attention to the fundamental, the fraction of total covariance due to the cross section is decreasing in attention capacity. For investors with below average attention capacity, the fraction is increasing because both covariances are negative. The same logic holds for $\sigma_i/\sigma_\mu > \sqrt{T}$, which is formalized in Proposition 4.

Proposition 4. Given Assumption 1, for attention capacity sufficiently uniformly bounded, if $\sigma_i/\sigma_\mu > \sqrt{T}$ and $K_{m_i} \geq \bar{K}_\epsilon$ ($K_{m_i} \leq \bar{K}_\mu$), the ratio of fundamental to total covariance $\frac{\sigma^S_{\mathcal{C}[\bar{n},\bar{f}]}}{\sigma^C_{\mathcal{C}[\bar{n},\bar{f}]}}$ is increasing (decreasing) in total attention capacity. Conversely, if $\sigma_i/\sigma_\mu < \sqrt{T}$ and $K_{m_i} \geq \bar{K}_\mu$ ($K_{m_i} \leq \bar{K}_\epsilon$), the ratio of fundamental to total covariance $\frac{\sigma^S_{\mathcal{C}[\bar{n},\bar{f}]}}{\sigma^C_{\mathcal{C}[\bar{n},\bar{f}]}}$ is decreasing (increasing) in total attention capacity. The covariance ratio is constant if $\sigma_i/\sigma_\mu = \sqrt{T}$.

The results in this section provide a framework to identify attention capacities and allocations among investors, allowing the theory to be tested empirically. High
capacity investors are those who have high overall covariance with realized payoffs, given by the sum of the time series and cross sectional covariances. Further, the breakdown of this covariance into its time series and cross sectional dimensions reveal investors’ information choices.

2.2.5 Portfolio Characteristics

The model generates several predictions regarding the characteristics of investors’ portfolios. These include the portfolio dispersion, variance, and autocorrelation. Each is considered in turn.

Dispersion

When updating beliefs, investors take into account not only the realization of their signals, but also the ex-ante quality of these signals. This affects portfolios in two ways. On the one hand, investors who receive more precise information have signals that are, in expectation, more concentrated. Thus, these investors deviate less from the average investor of that type. On the other hand, investors are aware of the quality of their signals, and hence, conditional on receiving the same signal realization, update their beliefs more aggressively if the signal is of higher precision.

The dispersion of portfolios among investors of the same type is dominated by the second of these two forces. To see this, recall that \( \bar{q}_{m,t} \equiv \psi_m^{-1} \int_{\{i: m_i = m\}} q_{i,t} di \) is the average demand of type \( m \) investors. The portfolio dispersion of investor \( i \) is the variance of her deviation from this portfolio, which is given by

\[
\mathbb{V}[q_{i,t} - \bar{q}_{m,t}] = \gamma^{-2} \hat{\sigma}^{-2}_{f,i} \left( \hat{\sigma}^2_{\mu,i} K_{\mu,i} + \hat{\sigma}^2_{e,i} K_{e,i} \right).
\]  

(2.25)

Portfolio dispersion is a weighted average of fundamental and idiosyncratic atten-
tion with weights $\hat{\sigma}^2_{\mu,i}$ and $\hat{\sigma}^2_{\epsilon,i}$, respectively. As both fundamental and idiosyncratic attention are increasing in attention capacity, it is clear that portfolio dispersion is increasing as well. This is stated in Proposition 5.

**Proposition 5.** Given Assumption 1, for attention capacity sufficiently uniformly bounded, portfolio dispersion $\nabla [q_{i,t} - \hat{q}_{m,t}]$ is increasing in attention capacity $K_{m,t}$.

**Variance**

The variance of investors’ portfolios—both in the time series and the cross section—again depends on their information choices and, more primitively, their attention capacities. As before, more precise information leads to more dispersed beliefs despite the signals being more concentrated. However, portfolio variance is also affected by the quality of investors’ signals relative to those of the average investor. To see this, the cross sectional variance of an investor’s portfolio is

$$\nabla^{CS} [q_{i,t}] = \gamma^{-2} \hat{\sigma}_{f,i}^{-2} \left\{ (\hat{\sigma}_{\mu,i} K_{\mu,i} - b)^2 \sigma_{\mu} + \hat{\sigma}_{\mu,i}^2 K_{\mu,i} \right\}$$  \hfill (2.26)

and the time series variance is

$$\nabla^{TS} [q_{i,t}] = \gamma^{-2} \hat{\sigma}_{f,i}^{-2} \left\{ (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c)^2 \sigma_{\epsilon} + \hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i} + d^2 \sigma_{s} \right\}.$$  \hfill (2.27)

These variances depend on investors’ attention capacity in three ways. First, an investor with greater capacity has a higher posterior precision $\hat{\sigma}_{f,i}^{-1}$, which increases the magnitude of her portfolio weights.$^{13}$ Second, more precise information increases the degree to which investors react to the noise in their signals (the second additive term in Eqs. (2.26) and (2.27)). These two effects unambiguously increase portfolio

$^{13}$To see this, recall from Eq. (2.14) that the optimal demand is linear in the posterior precision.
variance. The last effect is that greater information increases the degree to which investors react to the factor realization (the first additive term in Eqs. (2.26) and (2.27)). This results in an increase in variance for investors with above average capacity, but decreases variance for investors with below average capacity. Proposition 6 establishes the positive effects dominate; portfolio variance is increasing in attention capacity in both the time series and cross section.

**Proposition 6.** Given Assumption 1, for attention capacity sufficiently uniformly bounded, the variance of individual portfolios in the cross section ($\mathcal{V}^{CS}[q_{i,t}]$) and time series ($\mathcal{V}^{TS}[q_{i,t}]$) is increasing in attention capacity.

**Autocorrelation**

Portfolio autocorrelation reflects the degree to which an investor engages in factor timing. An investor who allocates more of her attention to the idiosyncratic component dynamically changes her portfolio more aggressively in response to the signals she receives, leading to lower autocorrelation. Investors who primarily allocate to the fundamental, on the other hand, will follow a more passive strategy. As a result, the relationship between attention capacity and portfolio autocorrelation depends on the relationship between attention capacity and the fraction of attention allocated to the idiosyncratic component.

**Proposition 7.** Given Assumption 1, for attention capacity sufficiently uniformly bounded, investor autocorrelation is increasing if $\sigma_/\sigma_\mu \geq \sqrt{T}$. Further, for sufficiently small $\sigma_s$, investor autocorrelation is decreasing in attention capacity if $\sigma_/\sigma_\mu < \sqrt{T}$.

Proposition 7 establishes the relationship between portfolio autocorrelation and attention capacity. If the fraction of attention allocated to the fundamental is increasing in attention capacity, so too does autocorrelation. However, if the opposite
is true, investor autocorrelation is decreasing if $\sigma_s$ is sufficiently small.

Notably, an investor who does not allocate any attention to the idiosyncratic component does not have a perfectly autocorrelated portfolio. This is due to two reasons. First, she is at an informational disadvantage. Due to market clearing, for every investor who takes positions that positively covary with idiosyncratic payoffs, there must be one who takes positions that negatively covary. Investors with the least knowledge about the idiosyncratic component are the only investors willing to take on these positions. As the idiosyncratic component is $iid$, portfolios will be time-varying. Second, the presence of stochastic supply necessitates some time-variation; investors will increase (decrease) their holdings when supply is high (low).

### 2.2.6 Portfolio Composition

An investor’s portfolio can be evaluated based on the realizations of risk factor payoffs. Specifically, ex-post, how do investors’ average positions in a risk factor depend their on attention allocations? Because of the private noise in an investor’s signals, an individual investor’s portfolio will not correspond exactly with ex-post fundamental payoffs. For this reason, the representative type $m$ investor’s time series average portfolio is examined. This is given by

$$\mathbb{E}^{TS} \left[ \tilde{\mu}_{m,t} \right] = \gamma^{-1} \hat{\sigma}_f^{-1} \left\{ \hat{\sigma}_{\mu,m} K_\mu - b \right\} \mu - a \right\}.$$

The usual components affecting portfolios are present. Average portfolio holdings are driven by the investor’s attention capacity relative to the market average and her posterior precision. Investors with above (below) average attention capacities take positions that are positively (negatively) correlated with the realization of the fundamental $\mu$. This is a key feature of a general equilibrium model; for every investor
who beats the market, there must be one who gets beaten. Proposition 8 states this formally.

**Proposition 8.** Given Assumption 1, average portfolio of type m investors $E^{TS}[q_{m,t}]$ is increasing in the realized fundamental $\mu$ if $K_{m_i} > \bar{K}_m$ and decreasing otherwise.

### 2.2.7 Impact on Factors

In this model, prices are an agglomeration of investors’ beliefs. As a result, to the extent an investor type impacts prices, they also mechanically affect returns. In this subsection, the relationship between an investor type’s attention capacity and their impact on returns—both average risk premia and volatility—is examined. To quantify this, the counterfactual of assuming the portfolios of type $m$ investors are replaced by the market portfolio is considered. An investor type’s impact is the difference between actual and counterfactual returns.

**Risk Premia**

Equilibrium risk premia are given by $E^{TS}[f_t - r p_t]$. Letting $p_{-m,t}$ denote the price under the counterfactual that all type $m$ investors instead held the market, consider the change in risk premia (normalizing for investor size):

$$
\Delta^m E^{TS} [f_t - r p_t] \equiv \psi_{m}^{-1} \left( E^{TS} [f_t - r p_t] - E^{TS} [f_t - r p_{-m,t}] \right) 
$$

$$
= \psi_{m}^{-1} r E^{TS} [p_{-m,t} - p_t]
$$

(2.29)

An investor type’s importance in price setting is driven solely by attention capacity. This affects prices in two ways. First, a group of investors with greater attention capacity holds portfolios that deviate more significantly from the market portfolio. As a result, switching to the market portfolio would have a drastic impact on prices.
Second, if those investors received very high quality signals, then, in their absence, the average information quality in the market falls, leading to a greater impact on prices.

**Proposition 9.** Suppose the risk factor is in zero net supply ($\bar{s} = 0$). The per-investor price impact of type $m$ investors $\Delta^m E^{TS} [f_t - rp_t]$ is given by $\kappa_m \mu$, where $\kappa_m$, defined in Appendix A.2.1, is negative and decreasing in $K_m$ for $K_m \geq \bar{K}_\mu$ and positive otherwise.

Proposition 9 characterizes how risk premia would change in the absence of an investor class. Investors with above average attention shrink risk premia towards zero. If the realized fundamental is positive (negative), then these investors reduce (increase) the average premium by taking positive (negative) positions in the factor. Further, for investors with above average attention capacities, the magnitude of their price impact is increasing in attention capacity. This is consistent with the notion that highly skilled investors act as arbitrageurs by identifying mispricings and trading against them. These investors have the greatest impact on prices.

**Volatility**

The time series volatility of factor returns is given by $\nabla^{TS} [f_t - rp_t]$. In a similar vein to the analysis of risk premia, let

$$\Delta^m \nabla^{TS} [f_t - rp_t] \equiv \psi^{-1}_m (\nabla^{TS} [f_t - rp_t] - \nabla^{TS} [f_t - rp_{-m,t}])$$

(2.30)

denote the difference between the observed volatility and the volatility that would be observed if all type $m$ investors held the market portfolio.

**Proposition 10.** The per-investor volatility impact of type $m$ investors $\Delta^m \nabla^{TS} [f_t - rp_t]$ is decreasing in $K_m$, negative if $K_m \geq \max \{\bar{K}_\mu, \bar{K}_\epsilon\}$, and positive if $K_m \leq \min \{\bar{K}_\mu, \bar{K}_\epsilon\}$
Much like the previous result, Proposition 10 characterizes how the volatility of factor returns would change if a class of investors instead held the market. Investors with above average attention to both the fundamental and idiosyncratic components reduce volatility, while those with below average attention in both components increase volatility. This occurs because investors with superior information are able to informatively trade. As a result, they increase their positions in anticipation of high idiosyncratic payoffs, and decrease their positions in anticipation of low idiosyncratic payoffs. This moves prices closer to the factor’s idiosyncratic payoff, offsetting the impact of the idiosyncratic component on returns, therefore reducing volatility in the time series.

2.3 Mapping Theory to Data

To test the theory, it would be ideal to know both (i) the attention capacities of each investor type and (ii) their attention allocations. However, both of these quantities are fundamentally unobservable. Attention capacity is a function of an investor’s incentives, employee skill, technological capabilities, etc., while allocations are the product of the internal division of an investor’s resources. However, these challenges can be overcome by observing that investors’ attention allocation decisions map to their downstream portfolio decisions.

The model itself generates a heuristic to map these portfolio decisions back to investors’ attention capacities and allocations. Proposition 3 states that investors with higher attention capacities have a higher covariance between exposures and payoffs overall, in the cross section, and in the time series. This means that the covariance between an investor’s portfolio and asset payoffs directly reveal her capacity and information choices. This allows for a simple rule to identify and compare attention
capacities across investors: those with higher overall payoff covariances have higher attention capacities. Thus, rather than rely on conventional wisdom, the model enables us to test it: do hedge funds truly exhibit greater ability? Do mutual funds add value to their clients? Are households “dumb money”?

In the same vein, a decomposition of the payoff covariance reveals investors’ information choices. The cross-sectional covariance between an investor’s portfolio and payoffs reveal her fundamental attention, while the time series payoff covariance reveals her idiosyncratic attention. Further, the ratio of cross-sectional to overall covariance is related to an investors’ relative attention allocations, as in Proposition 4. The model can therefore be tested by identifying investors’ attention capacities through the total payoff covariance, and then decomposing this into time series and cross-sectional components to determine both absolute and relative attention allocations.

Each of the subsequent theoretical results generates an empirical prediction regarding the relationship between attention capacity and either portfolio characteristics (Propositions 5, 6, 7) or investors’ importance in the price formation process (Propositions 9, 10). Testing the former is straightforward; however, estimating investors’ roles in determining asset prices requires specifying demand functions for each investor. To do this, I follow the recent literature on demand systems in asset pricing (Koijen and Yogo, 2019; Koijen et al., 2020). The next subsections provide further detail on the empirical proxies of attention, demand system estimation, and data used.
2.3.1 Proxies of Attention Allocation

In the model, an investor chooses quantities of risk factors to hold. This is akin to choosing the exposures of her portfolio to known risk factors. A high capacity investor is able to select a portfolio that ex-post covaries highly with factor returns. There are two components to this: (i) choosing systematic deviations from the market portfolio to adjust her average exposures (factor selection) and (ii) dynamically adjusting exposures in anticipation of idiosyncratic realizations (factor timing). A successful factor selection strategy is reflected in a high covariance between an investor’s average portfolio and average factor returns (cross sectional covariance), while a successful factor timing strategy is reflected in a high covariance between an investor’s deviation from her average portfolio and idiosyncratic factor returns (time series covariance). Letting $FSkill_i$ denote the covariance between investor $i$’s risk exposures and realized factor returns, these quantities can be recovered by a decomposition of this covariance into cross sectional and time series dimensions as follows:

$$FSkill_i \equiv \frac{1}{NT} \sum_{c=1}^{N} \sum_{t=1}^{T} q_{c,t}^i (f_c^t - \bar{f})$$

$$= \frac{1}{N} \sum_{c=1}^{N} q_c^i (\bar{f}_c - \bar{f}) + \frac{1}{N} \sum_{c=1}^{N} \left[ \frac{1}{T} \sum_{t=1}^{T} (q_{c,t}^i - \bar{q}_c^i) (f_c^t - \bar{f}_c) \right], \quad (2.31)$$

where $q_{c,t}^i$ is investor $i$’s exposure to characteristic $c$ at time $t$, $f_c^t$ is the time $t$ return associated with characteristic $c$, and bars denote averages. As investors with superior information choose portfolios that covary highly with realized returns, $FSkill_i$ proxies for an investor’s attention capacity.

The decomposition in the second line is similar to that in Daniel et al. (1997) and divides an investor’s capacity into a static component ($FSelection_i$) and a dynamic
one \( FTiming_{t,c} \). The static component measures an investor’s factor selecting ability. A high value of \( FSelection_i \) indicates that the investor holds an average portfolio that covaries highly with the (time series) average factor return \( \bar{f}^c \). This reflects investors’ attention to the fundamental component of factors; investors who allocate more attention to the fundamental exhibit high factor selecting ability.

The second term, \( FTiming_{t,c} \), measures an investor’s factor timing ability. This measures how an investor’s deviation from her average portfolio \( (q^c_{i,t} - \bar{q}^c_i) \) covaries with the idiosyncratic component of factor returns \( (f^c_i - \bar{f}^c) \). A high value indicates that an investor is able to anticipate these idiosyncratic returns and trade against them. As a result, this term is the product of an investor’s idiosyncratic attention.

Note that timing ability is defined per characteristic. In the model, an investor’s attention allocation is dependent on the idiosyncratic volatility of each characteristic. Hence, by defining this measure per characteristic, the relationship between timing ability and idiosyncratic volatility can be examined. In much of the empirical analysis, investors’ factor timing ability is averaged across characteristics, in which the subscript \( c \) is dropped. Finally, to ensure these measures are well estimated, the covariances are computed after aggregating the portfolios of all individual investors in each investor type.

### 2.3.2 Demand System Estimation

The demand system estimation closely follows Kojen et al. (2020). In this subsection, a brief overview of the methodology is presented. Kojen and Yogo (2019) and Kojen et al. (2020) provide a theoretical grounding of the model and further details of the estimation procedure.

Investors choose how to allocate their wealth between a set of stocks and an outside
asset based on observable characteristics. Let \( w_{i,s,t} \) denote investor \( i \)'s portfolio weight in stock \( s \) at time \( t \) and \( w_{i,0,t} \) her weight in the outside asset. Investor \( i \)'s demand for stock \( s \) (relative to the outside asset) at time \( t \) is given by

\[
\frac{w_{i,s,t}}{w_{i,0,t}} = \exp \left \{ \beta_{0,i,t} + \beta_{1,i,t} me_{s,t} + \beta_{i,t}^t x_{s,t} \right \} \epsilon_{i,s,t} \tag{2.32}
\]

where \( me_{s,t} \) is the log market capitalization of stock \( s \) at time \( t \). The vector of stock characteristics \( x_{s,t} \) includes log book equity, operating profitability, asset growth, dividend yield, and market beta.\(^{14}\) Note that the elasticity of demand with respect to stock characteristics are allowed to be time-varying. This is because, as in the model in Section 2.2, investors’ demand for characteristics changes with their belief about idiosyncratic payoffs.

Latent demand \( \epsilon_{i,s,t} \) is the component of demand that cannot be explained by the included characteristics, standardized to have mean 1. Further, latent demand is assumed to be exogenous to all characteristics other than market equity.\(^{15}\) Market equity is endogenous to latent demand. As a result, an instrument for market equity is necessary, which is discussed next.

**Instrument for Market Equity**

Plausibly exogenous variation in investors’ investment mandates is used to construct an instrument for market equity. An investment mandate is the set of securities that an investor considers when making portfolio choice decisions. As a result, variation

\(^{14}\)This is the same set of characteristics used in Kojen and Yogo (2019). An alternative set of characteristics based on principal components of the universe of characteristics yields similar results.

\(^{15}\)It is important to note that price variables, such as momentum, are dependent on market equity. As a result, including these variables in the vector of stock characteristics \( x_{s,t} \) may violate the identifying assumption.
in investment mandates *across investors* generates exogenous variation in market valuations.

Following Koijen and Yogo (2019), the investment mandate is computed as the union of stocks an investor has held in the previous 11 and current quarters, denoted by $S_{i,t}$.\(^{16}\) An instrument for market equity is constructed by computing market equities as if all other investors (excluding the household sector) held an equal-weighted portfolio within their investment mandate. That is,

$$
\widehat{m}e_{i,s,t} = \log \left( \sum_{i' \neq i, HH} A_{i',t} \frac{1[s \in S_{i',t}]}{1 + |S_{i',t}|} \right),
$$

where $A_{i,t}$ is the value of investor $i$’s portfolio, $1[s \in S_{i',t}]$ is an indicator equal to 1 if stock $s$ is in investor $i$’s investment mandate at time $t$, and $|S_{i,t}|$ is the number of stocks in investor $i$’s time $t$ investment mandate.

This instrument is exogenous to investor $i$’s portfolio decisions and depends only on the investment mandates of other investors. This construction allows portfolio decisions to be endogenous within an investor’s investment mandate. However, it rules out investment mandates being endogenous to latent demand.

The resulting moment condition is

$$
\mathbb{E}_t \left[ \left( \frac{w_{i,s,t}}{w_{i,0,t}} \exp \{ \bar{\beta}_{t,t} \bar{x}_{s,t} \} - 1 \right) z_{s,t} \right] = 0,
$$

where $\bar{\beta}_{t,t} = [\beta_{0,i,t}, \beta_{1,i,t}, \beta_{t,t}']'$, $\bar{x}_{s,t} = [1, me_{s,t}, x'_{s,t}]'$, and $z_{s,t} = [1, \widehat{m}e_{i,s,t}, x'_{s,t}]'$.

\(^{16}\)The investment universe is highly stable over time; roughly 95% of securities held in the current quarter had been held in at least one of the previous 11 quarters. Koijen and Yogo (2019) show that this instrument is strong.
Estimation Procedure and Shrinkage Penalty

Finally, to account for the fact that investors tend to hold very few securities, Kojien et al. (2020) add a ridge penalty to the moment condition (2.34) centered around a target set of coefficients $\beta^*_i$:

$$
\mathbb{E}_t \left[ \left( \frac{w_{i,s,t}}{w_{i,0,t}} \exp \left\{ \hat{\beta}_{i,t} x_{s,t} \right\} - 1 \right) z_{s,t} \right] + \text{diag} (\lambda_{i,t}) (\hat{\beta}_{i,t} - \beta^*_i) = 0. \tag{2.35}
$$

Eq. (2.35) can be estimated without any shrinkage ($\lambda_{i,t} = 0$) for investors with greater than 1,500 positive holdings in a quarter. For investors with less than 1,500 holdings, the shrinkage target is set to the average coefficients across all investors with greater than 1,500 holdings of that type. That is, there is a distinct shrinkage target for each investor type. The shrinkage penalty is set to $\lambda_{i,t} = 15N_{i,t}^{-0.8}$, where $N_{i,t}$ is the number of stocks held by investor $i$ in quarter $t$, for consistency with Kojien et al. (2020). The procedure to estimate Eq. (2.35) is given in Appendix C of Kojien et al. (2020).

Kojien and Yogo (2019) show that there is a unique equilibrium as long as $\beta_{1,i,t}$ is less than 1 for all investors and greater than $-1$ for at least one investor. As a result, the moment condition (2.35) is estimated subject to the restriction that $\beta_{1,i,t} < 1$. To do so, the unrestricted model is estimated first. If the estimated coefficient $\hat{\beta}_{1,i,t}$ is greater than 1, it is fixed at $\hat{\beta}_{1,i,t} = 0.99$ and the remaining coefficients are estimated under the same moment condition.

\[\text{Using 5-fold cross-validation to tune the shrinkage hyperparameters yields similar estimates.}\]
2.3.3 Data

To examine the model’s predictions empirically, both a comprehensive set of investors’ portfolios and a universe of risk factors are needed. The portfolio holdings come from Form-13F filings. These filings are required at a quarterly frequency for all institutional investment managers overseeing at least $100M in assets under management (AUM) and contain all long positions in U.S. equities greater than either $200,000 or 10,000 shares. These filings are obtained through FactSet over the period 1999Q1 through 2018Q4.\textsuperscript{18} FactSet also compiles holdings data from additional sources, including stakeholder holdings and fund-level reports. I follow the FactSet documentation in aggregating this data.\textsuperscript{19} To account for short positions, short interest data is gathered from Compustat. The universe of securities include all common shares (share codes 10, 11, 12, 18) listed on NYSE, AMEX, and NASDAQ (exchange codes 1, 2, and 3) that have non-missing values for return, price, and shares outstanding.

Investor are grouped into eight broad categories. FactSet provides a classification of investors, which are aggregated into six categories: Brokers, Hedge Funds, Investment Advisors, Long-Term (insurance companies and pension funds), Mutual Funds, and Private Banking. The remaining two categories are aggregate Short-Sellers and Households. Short-Sellers are the class of investors that aggregates all short interest holdings, while Households are constructed so that total holdings are equal to a security’s market capitalization plus short interest.\textsuperscript{20} Importantly, for these two classes, individual portfolios are not observed.

\textsuperscript{18}A separate sample, obtained directly through EDGAR beginning in 1999, parsed using string matching techniques, and verified using external price data, yields very similar holdings data and results. The data is constructed in a way that is similar to Backus et al. (2019), but updated to 2020Q4. The data and code can be accessed at https://elsafym.github.io/EDGAR-Parsing/.

\textsuperscript{19}See Appendix A.1.5 for details of this aggregation and further information on investor types.

\textsuperscript{20}In cases in which total holdings exceeds market cap plus short interest, institutional holdings are scaled back proportionally.
Figure 2.2: Percent Ownership by Investor Type

![Figure 2.2: Percent Ownership by Investor Type](image)

Note: This figure plots percent ownership by investor type over the sample period. Investor classifications come from FactSet. Investor types (excluding Households and Short-Sellers) are sorted by the final ownership share. Households are constructed from the residual of institutional ownership and market capitalizations. Short-Sellers are the aggregate class of short interest. Ownership shares sum to more than 1 due to the presence of Short-Sellers. The sample is from 1999Q1 to 2018Q4.

Figure 2.2 plots the share of ownership for each investor class. Changes in ownership shares have occurred gradually over time. Since 2000, the share of household ownership has fallen from roughly 50% to 25%. This has been absorbed mostly by investment advisors and mutual funds. Hedge fund ownership has risen markedly from near 0% to approximately 4% of total market capitalization.

A set of 55 risk factors are created based on characteristics that have been associated with cross-sectional return predictability (Harvey et al., 2016; McLean and Pontiff, 2016; Chen and Zimmerman, 2019). These contain a variety of accounting and price variables computed using data from Compustat and CRSP. The full set of characteristics, their definitions, and summary statistics are given in Appendix A.1.4. Characteristics that are negatively associated with returns (e.g., size) are multiplied by −1, so all characteristics considered here should be positively related.
to returns. Further, each characteristic is cross sectionally orthogonalized to market beta to ensure the resulting risk factors are ex-ante market neutral.\textsuperscript{21} Risk factors are created using value-weighted long-short portfolios. Specifically, each security \( s \) is cross sectionally assigned to deciles for each characteristic \( c \) at time \( t \), denoted \( d_{s,t}^c \).\textsuperscript{22} Factor portfolios are formed for each characteristic in the standard way by taking the difference between the value-weighted returns in decile 10 and the value-weighted returns in decile 1. The resulting factor return for characteristic \( c \) is denoted \( f_t^c \).

To score investors’ exposures to these risk factors, a characteristics-based demand approach is taken (Lettau et al., 2019; Kelly et al., 2019). An investor’s exposure to a characteristic is defined as the portfolio-weighted mean of the decile scores \( d_{s,t}^c \) less the decile score of the market portfolio. Letting \( w_{i,s,t} \) denote the weight of security \( s \) in investor \( i \)’s portfolio at time \( t \), her exposure to (i.e., holdings of) characteristic \( c \) is

\[
q_{i,t}^c = \sum_{s=1}^{S_t} w_{i,s,t} d_{s,t}^c - d_t^c,  \tag{2.36}
\]

where \( S_t \) is the number of securities in the sample at time \( t \) and

\[
\bar{d}_t^c = \frac{\sum_{s} M_{E,s,t} d_{s,t}^c}{\sum_{s} M_{E,s,t}} \tag{2.37}
\]

is the decile score from holding the market portfolio (i.e., using market capitalizations \( M_{E,s,t} \) as weights). Note that an investor’s demand for a characteristic is defined in terms of her deviation from the market portfolio.\textsuperscript{23}

\textsuperscript{21}This alleviates concerns that different investors may have different hedging concerns with respect to market risk.

\textsuperscript{22}Alternative methods, including equal-weighted deciles, rank scores, and z-scores, are also examined and yield similar results.

\textsuperscript{23}A potential concern with measuring exposures in this way is the fact that it omits the short side of investors’ portfolios. Appendix A.1.1 discusses robustness to this omission using short positions from mutual funds in CRSP holdings data. Omitting the short side of investors’ portfolios results
To qualify for inclusion in the factor portfolio or in determining investors’ exposures at a given date, a security must be common equity (share codes 10 and 11) and in the top 5% of market cap. The complement of securities (share codes 12 and 18 and bottom 5% market cap) is the outside asset. All investors with less than $1 million in the outside asset or fewer than 5 holdings in a quarter are moved to the household sector.

Finally, when estimating the demand system and computing counterfactuals, it is necessary to have a complete panel of covariates. As a result, missing characteristics are cross sectionally imputed using \( k \)-nearest neighbors.\textsuperscript{24} It is important to emphasize that this imputation is done only to compute the counterfactual valuations; securities with missing observations of a characteristic are not included in the factor construction or when computing investors’ exposures to that characteristic.

\subsection{2.4 Empirical Evidence}

The model examines how differences in attention capacities affect an investor’s penchant to select and time factors, as well as the downstream characteristics of her portfolio. To study this empirically, differences in attention capacities are explored at the investor class level. In principle, an investor’s attention capacity is an endogenous decision, which is a function of her investment mandate and compensation contract, as well as properties such as size, technological resources, number of researchers, etc. For example, hedge funds, whose contracts typically feature a performance fee to align incentives, may be expected to choose higher attention capacities than mutual funds, in estimates that are deflated by about 5-6%, and a correlation in excess of 0.9.

\textsuperscript{24}Cross-validation is used to tune the optimal number of neighbors. Setting an additional 5% of the data to missing completely at random, the number of neighbors that minimizes the mean-squared error is 16. Using softImpute (Mazumder et al., 2010) as an alternative imputation method yields similar results, but has a slightly higher MSE.
who are evaluated relative to a benchmark and thus have little incentive to deviate from this (Breugem and Buss, 2019). These properties tend to be relatively homogeneous across a class of investors, so it stands to reason that there should be meaningful heterogeneity in attention across investor classes. Appendix A.1.1 contains robustness tests for the main findings and additional results for attention capacities within an investor class.

2.4.1 Identification of Attention Capacity

The starting point for testing the model is to estimate attention capacities by investor type. Figure 2.3 presents these estimates as proxied by the covariance between an investor’s portfolio and factor returns. Hedge funds have far and away the greatest level of attention capacity, with an estimate that is statistically significant and an order of magnitude larger than other investors. On the other hand, brokers, long-term investors, and short-sellers have negative return covariances, indicating that these investors have low levels of attention capacity relative to the market. The remaining investors have estimates that are not statistically different from zero, indicating that these investors have attention capacities roughly equal to that of the average investor.

These estimates both agree with and diverge from conventional wisdom in several important ways. First, they affirm the notion that hedge funds are sophisticated investors in search of mispricings. The ability of hedge funds to select portfolios that covary highly with ex-post returns is unmatched by any other investor type. As such, this finding justifies the use of studying hedge fund’s behavior when constructing cross sectional return predictors or examining their role in determining prices (e.g., Akbas et al., 2015; Elsafi, 2020). Second, these results suggest that households are not “dumb money”; rather, they hold portfolios that perform about as well as the
average investor. Finally, although not statistically significant, investment advisors and mutual funds seem to provide some value to their clients, as the point estimates for their return covariance is 2-4 times that of households.

**Figure 2.3: Proxy for Attention Capacity**

![Figure showing the relationship between investor type and return covariance](image)

*Note:* This figure plots the estimate of investors’ attention capacities by investor type. This is proxied for by $F_{Skills}$, the covariance between an investor’s portfolio and factor returns. The estimated proxies are computed for the aggregate holdings by investor type over the entire sample. The covariance computed over the full set of characteristics is displayed in red. Characteristics are further split into two equal-sized groups based on their idiosyncratic volatility over the entire sample, computed as the standard deviation conditional on the sample mean. 95% confidence intervals are shown. Confidence intervals are computed by a stratified-block bootstrap. Blocks of 2 years (8 quarters) are resampled by characteristic 1,000 times. The sample is from 1999Q1 to 2018Q4.

Perhaps most interesting is the result that short-sellers hold portfolios that covary negatively with returns. While short-sellers are often thought of as sophisticated, there are three possible explanations for this. First, the short-selling category is an amalgamation of all short-selling activity, including those from households and those who may use short-selling to hedge their portfolios. Second, as short-selling is relatively costly, this is generally not an efficient way to bet on factor performance.
Third, while short interest is often thought of almost exclusively as the short side of hedge fund portfolios (Jiao et al., 2016), a back of the envelope calculation from Choi et al. (2020) estimate that less than 20% of hedge fund trading involves short-selling.\footnote{To examine whether the superior performance of hedge funds could be due to significant shorting, I examine a “hybrid” investor class that combines hedge funds with 20% of total short-selling positions in unreported results. All findings are consistent even when accounting for this. Further evidence against this hypothesis is found in Choi et al. (2020), which documents that hedge funds’ short sales are generally profitable.}

Figure 2.3 further divides the sample of characteristics into those with low and high idiosyncratic volatilities. In principle, investors may choose to allocate attention differently across a set of risk factors depending on their idiosyncratic volatilities. However, the point estimates across these two samples are not significantly different from one another. This justifies the modeling choice of only having a single risk factor or assuming investors have a fixed attention capacity for each risk factor and cannot substitute across factors.\footnote{In Appendix A.1.2, an extension of the model considering attention allocation across factors is presented.}

There are two caveats in interpreting these results. First, a crucial assumption in the model is that investors have a common objective function and the same hedging motives. While this assumption is reasonably well satisfied for most investor types, it is particularly poorly suited for brokers, who may be more concerned with generating volume than acquiring information about future payoffs. As a result, interpreting the results for brokers in the context of the model should be done with caution.

Second, one may wonder whether these results are driven by only observing the long side of investor’s portfolios. This is unlikely to be the case. As these are estimates of exposures with respect to long-short portfolios, an investor can easily have a negative exposure without shorting any individual stocks. Thus, for an investor who wants to manage exposures to these factors, there would be no reason to engage in any short-selling. Additionally, even if some investors were to manage these ex-
posures via short-selling, holding a long portfolio that covaries highly with ex-post returns nevertheless requires superior information. For these reasons, the absence of short-selling in estimating these exposures is unlikely to be driving differences across investor type.\textsuperscript{27}

\section{2.4.2 Attention Allocation}

Having identified investors’ attention capacities, the next step is to examine the allocation of this capacity to factor timing and selection strategies. As discussed in Section 2.2, the relationship between attention capacity and information choice depends on the variance of the fundamental relative to that of the idiosyncratic component.\textsuperscript{28} Decomposing an investor’s capacity into cross sectional and time series dimensions allows for a direct test of the relationship between attention capacity and allocation.

Table 2.1 presents the decomposition of investors’ total return covariance into time series and cross sectional dimensions. Across the board, the overwhelming majority of covariation between portfolios and returns is due to the time series. In fact, no investor displays statistically significant ability to select factors cross sectionally. The signs between $F_{Skill_i}$ and $F_{Timing_i}$, however, are all consistent; investors with above average overall capacity also have above average timing ability. Further, as with overall attention capacity, hedge funds are the only investors with a statistically significant ability to factor time relative to the average investor.

This provides strong evidence that an investor’s penchant to factor time is increasing in her attention capacity. To see this, note that $\frac{1.631}{1.712} = 95\%$ of hedge fund’s

\textsuperscript{27}In Appendix A.1.1, short positions from CRSP mutual fund data are used to show that exposures using long-only portfolios is nearly identical to exposures using both long and short positions.

\textsuperscript{28}This relationship also depends on the unmodeled difficulty of learning about the fundamental relative to the idiosyncratic component.
Table 2.1: Estimated Attention Proxies, All Characteristics

<table>
<thead>
<tr>
<th>Attention Proxy</th>
<th>$FSkill_i$ (Estimate, Percentile)</th>
<th>$FTiming_i$ (Estimate, Percentile)</th>
<th>$FSelection_i$ (Estimate, Percentile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hedge Fund</td>
<td>1.712** 99.1%</td>
<td>1.631*** 100.0%</td>
<td>0.081 54.2%</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>0.213 93.0%</td>
<td>0.081 77.7%</td>
<td>0.131 88.4%</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>0.145 83.0%</td>
<td>0.120 86.3%</td>
<td>0.025 58.0%</td>
</tr>
<tr>
<td>Household</td>
<td>0.061 60.6%</td>
<td>0.017 55.8%</td>
<td>0.044 58.6%</td>
</tr>
<tr>
<td>Private Banking</td>
<td>−0.166 27.0%</td>
<td>−0.222 12.5%</td>
<td>0.057 61.9%</td>
</tr>
<tr>
<td>Long-Term</td>
<td>−0.310** 0.6%</td>
<td>−0.412*** 0.0%</td>
<td>0.102 83.3%</td>
</tr>
<tr>
<td>Broker</td>
<td>−0.733*** 0.0%</td>
<td>−0.791*** 0.0%</td>
<td>0.058 74.4%</td>
</tr>
<tr>
<td>Short-Seller</td>
<td>−1.955** 1.5%</td>
<td>−1.349* 2.8%</td>
<td>−0.606 17.2%</td>
</tr>
</tbody>
</table>

*Note: This table presents estimates of attention capacity and allocations by investor type. $FSkill_i$ is the covariance between an investor’s portfolio and factor returns, while $FTiming_i$ and $FSelection_i$ decompose this covariance into time series and cross sectional covariance, respectively. The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with unpermuted factor returns. All covariances have been multiplied by 1,000 for readability. *p<0.1; **p<0.05; ***p<0.01.*

... covariance is due to the time series, compared with only $\frac{0.081}{0.213} = 38\%$ for investment advisors and $\frac{0.017}{0.061} = 28\%$ for households. This means that hedge funds—the investors with the greatest attention capacity—allocate the greatest proportion of their attention to factor timing. On the other hand, among investors with below average capacity, the vast majority of their attention seems to be allocated to the fundamental. This can be seen from the fact that, while brokers and long-term investors have below average overall attention capacity, they actually have above average ability to select factors. This means that their below average attention capacity is due entirely to their inability to time.

One may wonder whether differences in risk aversion or investment restrictions may be driving differences in these estimates across investor types. For example, a less risk averse (or less constrained) investor may choose exposures to characteristics that are larger in magnitude. Because the estimates in Table 2.1 are covariances, they scale with the magnitude of investors’ exposures. To account for this possibility, Table 2.2 presents the analogous quantities using the total, time series, and cross
sectional correlation, rather than covariance, between exposures and returns. While
this has the benefit of accounting for differences in risk appetites or constraints,
it masks differences in the confidence levels of different investors, a quantity that
endogenously depends on an investor’s attention capacity. Despite this, the estimates
of overall attention capacity, as well as factor timing and factor selection ability, are
qualitatively similar across the two measures.

Table 2.2: Estimated Correlation Attention Proxies

<table>
<thead>
<tr>
<th>Attention Proxy</th>
<th>FSkillCor&lt;sub&gt;i&lt;/sub&gt;</th>
<th>FTimingCor&lt;sub&gt;i&lt;/sub&gt;</th>
<th>FSelectionCor&lt;sub&gt;i&lt;/sub&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>0.035**</td>
<td>99.3%</td>
<td>0.052***</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>0.022</td>
<td>93.2%</td>
<td>0.012</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>0.014</td>
<td>82.6%</td>
<td>0.016</td>
</tr>
<tr>
<td>Household</td>
<td>0.003</td>
<td>58.6%</td>
<td>0.001</td>
</tr>
<tr>
<td>Private Banking</td>
<td>0.009</td>
<td>27.1%</td>
<td>0.017</td>
</tr>
<tr>
<td>Long-Term</td>
<td>-0.035**</td>
<td>0.7%</td>
<td>-0.063***</td>
</tr>
<tr>
<td>Broker</td>
<td>-0.050***</td>
<td>0.0%</td>
<td>-0.059***</td>
</tr>
<tr>
<td>Short-Seller</td>
<td>-0.033**</td>
<td>0.8%</td>
<td>-0.031**</td>
</tr>
</tbody>
</table>

Note: This table presents alternative estimates of attention capacity and allocations by investor type based on the correlation, rather than the covariance, between exposures and returns. FSkillCor<sub>i</sub> is the correlation between an investor’s portfolio and factor returns, while FTimingCor<sub>i</sub> and FSelectionCor<sub>i</sub> decompose this correlation into time series and cross sectional correlation, respectively. The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with unpermuted factor returns. *p<0.1; **p<0.05; ***p<0.01.

Investors’ timing ability can be more closely visualized in Figure 2.4, which plots innovations in investors exposures (q<sub>t</sub><sup>c, t</sup> − q<sub>t</sub><sup>c</sup>) as a function of idiosyncratic factor returns (f<sub>t</sub><sup>c</sup> − ¯f<sup>c</sup>). This reinforces the measures of timing ability in Table 2.1, but also shows how this ability—or disability—is accomplished. The striking feature is the lack of variation across time in the portfolios of long-term investors, mutual funds, and investment advisors, all of whom exhibit either negative or no timing ability. Across all characteristics considered, these investors simply do not deviate much from their average exposures. For example, it is rare for long-term investors to deviate from their average portfolios by more than 0.2 deciles. This pattern is similar
for both brokers and households, although to a lesser extent.

The picture is very different for hedge funds. There is substantial variation across time relative to their average exposures. The magnitudes are quite large; a shift in excess of 0.5 deciles occurs fairly frequently for hedge funds. This is consistent with the model’s predictions. Because hedge funds receive more precise signals about idiosyncratic factor returns, they react more aggressively. This generates a positive relationship between their portfolio and idiosyncratic returns, as well as substantial variation in their portfolios across time. The opposite is true for long-term investors, who receive less precise information.

Interestingly, while short-sellers have significantly negative timing ability, they seem to exhibit substantial variation in their portfolios (even more so than hedge funds). As mentioned before, interpreting short-sellers in the context of the model may be problematic. While the other investor classes are well-defined, short-sellers are an amalgamation of all short-selling across the seven other investor types. As such, short-sellers do not represent a separate entity, but rather an investor class that combines the activities of all other investors.

Factor timing ability can arise due to reaction to either previous realized returns or contemporaneous return surprises. While this deviates slightly from the setup of the model, which abstracts from autocorrelation in the idiosyncratic component of factor payoffs, it is nevertheless instructive to consider which component of factor returns investors are learning about and reacting to. To address this, the factor returns are assumed to follow an AR(1) process with a factor-specific autocorrelation \( \rho_c \):\(^{29}\)

\[
f_t^c = (1 - \rho_c) \bar{f}_t^c + \rho_c f_{t-1}^c + \epsilon_t^c. \tag{2.38}
\]

\(^{29}\)The results are not sensitive to additional autoregressive lags or moving average terms.
Figure 2.4: Time Series Exposures by Investor Type

![Graphs showing time series exposures by investor type](image)

Note: This figure plots innovations investors’ exposures as a function of the realized idiosyncratic factor return for each investor type. Innovations in exposures are measured as the deviation from an investor type’s average exposure in a characteristic at a given date. The idiosyncratic factor return is the difference between the return at a given date and the average return over the full sample. The line of best fit is plotted through each investor’s cloud of points. The sample is from 1999Q1 to 2018Q4. *p<0.1; **p<0.05; ***p<0.01.
Table 2.3: Factor Timing Decomposition

<table>
<thead>
<tr>
<th></th>
<th>Return $f_i^c$</th>
<th>Expected Return $\rho_c f_{i-1}^c$</th>
<th>Return Surprise $\epsilon_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>1.631***</td>
<td>99.9%</td>
<td>0.070</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>0.081</td>
<td>77.7%</td>
<td>-0.008</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>0.120</td>
<td>85.0%</td>
<td>-0.009</td>
</tr>
<tr>
<td>Household</td>
<td>0.017</td>
<td>54.4%</td>
<td>-0.029</td>
</tr>
<tr>
<td>Private Banking</td>
<td>-0.222</td>
<td>13.0%</td>
<td>-0.051**</td>
</tr>
<tr>
<td>Long-Term</td>
<td>-0.412***</td>
<td>0.0%</td>
<td>-0.045***</td>
</tr>
<tr>
<td>Broker</td>
<td>-0.791***</td>
<td>0.1%</td>
<td>-0.060**</td>
</tr>
<tr>
<td>Short-Seller</td>
<td>-1.349*</td>
<td>3.1%</td>
<td>0.056</td>
</tr>
</tbody>
</table>

Note: This table decomposes factor timing ability $FT_{iming_i}$ into expected and surprise components. Returns are modeled as an AR(1) process: $f_i^c = (1 - \rho_c) f_i^c + \rho_c f_{i-1}^c + \epsilon_i^c$. The first column presents the time series covariance between exposures and contemporaneous returns ($f_i^c$). The second column presents the time series covariance between exposures and the expected component of returns ($\rho_c f_{i-1}^c$). The third column presents the time series covariance between exposures and return surprises ($\epsilon_i^c$). The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with unpermuted factor returns. All covariances have been multiplied by 1,000 for readability. *p<0.1; **p<0.05; ***p<0.01.

Estimating this model for each factor, timing ability can arise from covariation with either expected ($\rho_c f_{i-1}^c$) or surprise ($\epsilon_i^c$) components. Table 2.3 presents this decomposition. No investor has a statistically significant positive reaction to the expected component of factor returns, while private banking, long-term investors, and brokers react negatively. The lack of positive timing ability arising from the expected component seems to suggest that investors are not reacting optimally to the information contained in previous returns. Interestingly, statistically significant negative timing ability to the expected component indicates that certain investors react in the opposite direction of expected future returns. On the other hand, the reaction to the return surprise more closely mirrors the overall timing ability in Table 2.1 (and repeated in the first column of Table 2.3).
2.4.3 Portfolio Characteristics

Given the above results, the remaining predictions concerning the properties of investors’ portfolios—dispersion, volatility, and autocorrelation—can be tested. As these tests relate to the portfolios of individual investors, the two aggregate investor classes (households and short-sellers) are omitted.

Dispersion

Proposition 5 states that investors with greater attention capacity exhibit higher within-type dispersion in their portfolios. The intuition is as follows. Investors with greater capacity receive more precise information, so their signals are more concentrated. However, they update their beliefs more aggressively in response to their signals. The second effect dominates, leading to an increasing relationship between capacity and dispersion. Based on the estimates of investors’ attention capacities in Figure 2.3, the model would predict that hedge funds have the highest dispersion, while brokers and long-term investors have the lowest.

Figure 2.5 presents dispersion by investor type. The estimates presented are pooled across all, high volatility, and low volatility characteristics; however, the resulting estimates are similar. The portfolios of hedge funds are far and away the most dispersed. In contrast, the portfolios of long-term investors are the least dispersed. Investment advisors, mutual funds, and private banking all have levels of dispersion that are roughly similar, which reflects their status as having roughly average capacity. These results are all consistent with the model’s predicted relationship.

Diverging from the theoretical predictions are brokers. Despite having below average attention capacity, brokers have the second highest dispersion. This likely stems from the fact that brokers’ particular holdings are often dictated by client
Figure 2.5: Portfolio Dispersion by Investor Type

![Portfolio Dispersion by Investor Type](image)

Note: This figure plots dispersion in individual investors' portfolios by investor type. Portfolio dispersion is defined as the variance of an investor’s portfolio relative to the aggregate portfolio of that type (i.e., $\nabla [q_{t,i} - \bar{q}_{m,t}]$). Dispersion is computed by characteristic and the resulting estimates are pooled across characteristics. Characteristics are further split into two equal-sized groups based on their idiosyncratic volatility. 95% confidence intervals are shown.

demand. As a result, dispersion among brokers likely reflects heterogeneous demand of different clients, rather than brokers beliefs about their payoffs.

Variance

The variance of investors' portfolios is examined next. In a similar vein to the predictions concerning dispersion, the theory states that both the time series and cross sectional variance of an investor’s portfolio should be positively related to attention capacity. As before, investors with greater attention capacity update their beliefs more aggressively, and this effect dominates the fact that the variance of their signals is lower. This generates the positive relationship.
Figure 2.6: Individual Portfolio Variance by Investor Type

![Figure 2.6: Individual Portfolio Variance by Investor Type](image)

**Note:** This figure plots the time series and cross sectional variance in investors' portfolios by investor type. The sum of the cross sectional and time series variance is an investor’s total portfolio variance. The estimates presented are the (equal-weighted) average of the variances of all investors of a given type. 95% confidence intervals are shown.

The variance of an investor’s portfolio is decomposed into a cross sectional and time series dimensions as follows:

\[
\mathbb{V} [q_{i,t}^c] = \frac{1}{NT} \sum_{c=1}^{N} \sum_{t=1}^{T} (q_{i,t}^c - \bar{q}_i)^2 \\
= \underbrace{\frac{1}{NT} \sum_{c=1}^{N} (\bar{q}_i^c - \bar{q}_i)^2}_{= \mathbb{V} \text{CS} [q_{i,t}^c]} + \underbrace{\frac{1}{NT} \sum_{c=1}^{N} \sum_{t=1}^{T} (q_{i,t}^c - \bar{q}_i^c)^2}_{= \mathbb{V} \text{TS} [q_{i,t}^c]}. \tag{2.39}
\]

Using this decomposition, Figure 2.6 plots the average cross sectional and time series variance of investors’ portfolios by type. Hedge funds exhibit the greatest variance in both the cross section and the time series. In fact, their overall variance
is nearly twice as high as the next highest group of investors. This reflects their status as having far and away the largest attention capacity. Interestingly, the variance is roughly evenly split between the cross sectional and time series dimensions despite the focus of hedge funds on factor timing. Further evidence in support of the theory is that long-term investors have the lowest variance along both dimensions. This means that long term investors hold portfolios that are (i) closest to the market portfolio on average and (ii) deviate the least from their average portfolios.

It is important to note that the variance in Figure 2.6 is the average variance of individual portfolios. This contrasts with the variation in aggregate portfolios shown in Figure 2.4. On both an individual and aggregate basis, hedge funds deviate from the market and from their average portfolios the most. That is, the variation does not “average out” when aggregating across an investor type; rather, there is comovement in hedge fund portfolios.

**Autocorrelation**

The last prediction concerning investors’ portfolio characteristics relates autocorrelation to attention capacity. In particular, an investor’s portfolio will exhibit a low autocorrelation if she spends relatively more of her attention on factor timing. Given that investors with higher capacities tend to allocate relatively more of their attention to factor timing (Table 2.1), autocorrelation should be decreasing in capacity.

Figure 2.7 shows the average autocorrelation by investor type. There are two main points. First, brokers and hedge funds have the lowest autocorrelation. As before, brokers may have autocorrelations that are lower due to differences in their objective function. Hedge funds’ low autocorrelation reflects their status as successful factor timers.

Second, long-term investors have significantly higher autocorrelations than hedge
Figure 2.7: Individual Autocorrelation by Investor Type

Note: This figure plots the quarterly autocorrelation of individual investors’ portfolios by investor type. The estimates presented are the (equal-weighted) average of the autocorrelations of all investors of a given type. 95% confidence intervals are shown.

funds, but the average autocorrelation is lower than that of investment advisors and mutual funds. This reflects the fact that these investors have far lower attention capacities than hedge funds and allocate a greater fraction of their capacity to the fundamental. The fact that investment advisors and mutual funds have higher autocorrelation is likely due to the presence of passive funds, which are generally static across time.

Overall, the model provides a lens through which investors’ attention capacities, information choices, and portfolio characteristics can be examined. Hedge funds are identified as the highest capacity investors, while long-term investors and short-sellers have the least capacity. Further, almost all of the variation in this capacity is due to variation in the time series, providing evidence that factor timing is only pursued by
high capacity investors.

These results map to downstream portfolio decisions. Hedge funds hold the most dispersed portfolios and their portfolios exhibit the highest variance in both the time series and cross section. Long-term investors have the least dispersion and variance. This is consistent with the model in that these two quantities should be positively related to attention capacity. Furthermore, hedge funds exhibit the lowest autocorrelation and long-term investors have among the highest, again consistent with the model.

2.5 Investors’ Impact on Factors

Having provided evidence that the model accurately captures several key features of investors’ portfolios, two predictions remain to be tested. These concern the impact each investor group has on risk premia and volatility (Propositions 9 and 10). According to the model, investors with above average attention capacities increase the information contained in prices. As a result, prices become more responsive to both the fundamental and idiosyncratic components, shrinking risk premia towards zero and reducing volatility. Investors with below average capacity have the opposite effect. The next subsection presents the methodology to test these predictions using the demand system detailed in Section 2.3.2.

2.5.1 Construction of Counterfactual Factors

The starting point is to properly define the counterfactual with which to examine the impact investors have on prices. The counterfactual considered proportionally redistributes the total value of all investors’ portfolios within a group to all other investors, similar to that in Section 2.2.7. Letting \( A_{i,t} \) denote investor \( i \)'s AUM, the
counterfactual from redistributing type \( m \) assets assigns each investor a new AUM defined by\(^{30}\)

\[
A^{-m}_{i,t} = \begin{cases} 
0 & \text{if } m_i = m \\
A_{i,t} + \sum_{m_i \neq m} A_{i,t} \sum_{m_i=m} A_{i,t} & \text{if } m_i \neq m
\end{cases}
\] (2.40)

This ensures the aggregate AUM is unchanged. However, as investors are free to substitute into or out of the outside asset, this does not mean that the sum of market values in the investment universe will remain constant. Counterfactuals are estimated for all investor types bar the household sector.\(^{31}\)

Under these counterfactual scenarios, equilibrium market values are re-estimated with the altered assets under management.\(^{32}\) Denoting these by \( ME_{s,t}^{-m} \), counterfactual stock returns are computed as the percent change in these market equities, accounting for dividends:

\[
r_{s,t}^{-m} = \frac{ME_{s,t}^{-m} - ME_{s,t-1}^{-m} + D_{s,t}}{ME_{s,t-1}^{-m}},
\] (2.41)

where \( D_{s,t} \) is the total payout of stock \( s \) at time \( t \).\(^{33}\) In rare cases, returns computed in this way (that is, via Eq. (2.41) under no counterfactual) deviates from the return data in CRSP. To ensure the results are not driven by these differences in the estimation, all return data that differs by greater than 5 basis points or 0.5% from the CRSP returns in absolute value are dropped. This drops 5% of the security-date observations.

---

\(^{30}\)For short-sellers, their AUM is the value of their short positions multiplied by \(-1\), so assets are taken away from other investor classes.

\(^{31}\)The household sector is omitted because it makes up a very large portion (\(0.5\%\)) of ownership in the early part of the sample. This approach has poor accuracy for such a large redistribution.

\(^{32}\)Appendix C of Kojien and Yogo (2019) provide a method to compute equilibrium market values, which are unique by the restrictions imposed in the estimation.

\(^{33}\)It is assumed that the dividend policy is unchanged in each of the counterfactuals considered. In principle, companies may adjust their dividend policies in light of the change in their market values. However, given the high correlation between the counterfactual market values and actual market values (median of 0.999), such changes are likely to be minimal.
Finally, factors are “reconstructed” under these counterfactual returns and market values. In doing so, it is assumed that decile assignments are consistent across both the baseline and all counterfactual scenarios. While this is unlikely to hold for the price-based characteristics in particular, it is a reasonable assumption for two reasons. First, the panel of market values and returns under the counterfactual scenarios are very highly correlated with the baseline. The median correlation between baseline and all the counterfactuals considered is 0.999 for market value and 0.973 for returns. As such, these differences are unlikely to have a substantial impact on the characteristics. Second, as the factors are formed using long-short portfolios, this is only an issue for stocks that would shift into or out of the first or tenth deciles under the counterfactual. Given the similarity of market values and returns between the baseline and counterfactuals, such shifts are unlikely. Further, as returns are averaged over deciles, which contain hundreds of stocks, small differences in decile composition are unlikely to generate substantial differences in factor returns.

Taking the decile assignments as given, factor returns under the various counterfactuals are rebalanced quarterly and computed in the usual way. Within deciles 1 and 10, returns are value-weighted by (counterfactual) market value, and the factor return is given by the difference between returns to decile 10 and returns to decile 1. The resulting factor return for characteristic $c$ at time $t$ under the counterfactual of omitting type $m$ investors is denoted $f_{t}^{c,-m}$.

2.5.2 Impact on Risk Premia

The impact a group of investors has on risk premia is reflected in the degree to which factor returns change under the counterfactual of redistributing that group’s AUM.
This yields the first measure of price impact:

$$Repricing_m^c = \mathbb{E} \left[ f_t^c - f_t^{c-m} \right], \quad (2.42)$$

In words, $Repricing_m^c$ is the difference between the observed expected return of factor $c$ and the expected return that would be observed if the AUM of type $m$ investors were redistributed as in Eq. (2.40). There are two elements affecting this: (i) the difference between the portfolios of type $m$ investors and that of others, and (ii) the share of the market owned by type $m$ investors. The model implies that, while differences in size are exogenous, heterogeneous portfolios are driven by differences in attention capacities. An investor with greater information will hold portfolios that deviate more from the market and are more correlated with realized returns. To isolate the role of information processing capabilities in impacting prices, the next measure scales differences in factor returns by the relative size of the investor group:

$$Scaled \ Repricing_m^c = \mathbb{E} \left[ A_{m,t}^{-1} \left( f_t^c - f_t^{c-m} \right) \right], \quad (2.43)$$

where $A_{m,t}$ is the total AUM of investor group $m$ at time $t$ in trillions of dollars. Note that $A_{m,t}$ is allowed to be time-varying to reflect changes in ownership shares over time. This allows the resulting estimates to be compared across investor classes of different sizes, as they are expressed per $1$ trillion.\textsuperscript{34}

Figure 2.8 presents estimates of scaled repricing by investor type across characteristic (the order is the same as in Figure 2.1). As with their exposures, hedge funds and short-sellers exhibit the greatest variation in repricing. This is a product of the fact that they hold portfolios that differ most from the market. Further, the estimates are negatively correlated with investors’ deviations from the market. For

\textsuperscript{34}The results are similar when normalizing by share of ownership, rather than trillions of dollars.
Figure 2.8: Scaled Repricing by Characteristic

![Scaled Repricing by Characteristic](image)

Note: This figure plots Scaled Repricing by investor type for each characteristic considered. Scaled Repricing is defined as the difference between the observed return and the counterfactual return that would have been observed if a group of investors’ AUM were proportionally distributed to other investors, expressed in percentage points per $1T. A positive value indicates that an investor type reduces expected returns. Characteristics are sorted on the x-axis in the same way as Figure 2.1 for comparison. Details about the characteristics and their construction are given in Appendix A.1.4. The sample is from 1999Q1 to 2018Q4.

example, hedge funds and short-sellers have correlations between scaled repricing and average exposures of -0.39 and -0.17, respectively. This negative relationship is to be expected: an investor that holds the long side of a factor puts upward price pressure on that factor, which results in a lower expected return. Hence, their repricing is negative.

Some of the estimates have sizable magnitudes. For example, per $1 trillion of invested capital, hedge funds increase the idiosyncratic volatility premium by over a 3 pp per quarter. This estimate may seem high; however, it is important to note two
things. First, idiosyncratic volatility is a factor where hedge funds hold one of the most differentiated positions from other market participants. As such, they should have a large impact on prices. Second, repricing is not determined solely by the degree to which hedge funds deviate, but also the degree to which other investors are willing to take on the positions vacated by hedge funds in the counterfactual. A very positive repricing indicates that other investors are unwilling to accept these positions.

<table>
<thead>
<tr>
<th>Table 2.4: Absolute Price Impact by Investor Type</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image.png" alt="Table" /></td>
</tr>
</tbody>
</table>

Note: This table presents estimates of price impact magnitudes by investor type. **Repricing** is defined as the (time series) average difference between an observed return and the counterfactual return that would be observed if a particular class of investors had their AUM proportionally redistributed to other investors. **Scaled Repricing** adjusts Repricing by differences in assets under management by normalizing by AUM (in trillions). The absolute value of these quantities, averaged across the set of factors considered, is presented. All measures are expressed in percentage points. Standard errors are in parentheses. All estimates are statistically significant at the 1% level.

An investor group’s overall impact on risk premia is measured by the *magnitude* by which expected returns change. Table 2.4 presents the average absolute value of the raw and scaled repricing measures. Overall, investment advisors have the largest absolute repricing. Adjusting for size, however, hedge funds have far and away the largest absolute impact. Per $1 trillion in invested capital, hedge funds change the average risk premium by 1 pp per quarter in absolute value. This is an order of magnitude larger than mutual funds, investment advisors, and long-term investors. Brokers and short-sellers also have sizable impacts on risk premia. This is consistent
with recent empirical evidence highlighting the important role of brokers in asset pricing (Adrian et al., 2014; He et al., 2017).

Proposition 9 contains another prediction. Investors with the greatest attention capacities not only have the largest impact on prices, but also shrink premia towards zero. To test this, Table 2.5 splits the 55 characteristics into three groups: those with negative, low, and high average returns.\(^{35}\) Hedge funds’ ability to affect expected returns is concentrated among risk factors with the most extreme average returns. Among factors with negative realized returns, a $1 trillion reduction in hedge fund capital increases the average risk premium by 0.54 pp per quarter, compared with a decrease of 0.84 pp among factors with the highest realized average returns. Investment advisors and mutual funds exhibit repricing that is decreasing in the average factor return as well. This is consistent with their having above average attention capacity, though not statistically significant.\(^{36}\)

The ability of different types of investors to “shrink” expected returns can be explicitly quantified. Recall from Proposition 9 that repricing is linear in the realized fundamental and that the slope is negative and decreasing for investors with above average attention and positive otherwise. That is, in the regression

\[
Scaled \text{ Repricing}^c_m = \alpha_m + \beta_m \bar{f}^c + \epsilon_m, \tag{2.44}
\]

the intercept \(\alpha_m\) should be zero for all investors and \(\beta_m\) should be negative and decreasing for above average capacity investors and positive otherwise. The coefficient \(\beta_m\) is interpreted as the percent increase in magnitudes of expected returns caused by a specific investor type.

\(^{35}\) Characteristics are split into “low” and “high” groups based on the median expected return among characteristics with positive expected returns.

\(^{36}\) The attention capacity of investment advisors is marginally significant, with a \(p\)-value of 0.14.
Table 2.5: Price Impact by Investor Type and Average Return

<table>
<thead>
<tr>
<th></th>
<th>Average Factor Return</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scaled Repricing</td>
<td>Scaled Repricing</td>
<td>Scaled Repricing</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td><strong>0.540</strong>*</td>
<td>-0.013</td>
<td><strong>-0.837</strong>*</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td><strong>0.088</strong>*</td>
<td>0.027</td>
<td><strong>-0.081</strong></td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>0.003</td>
<td>-0.051</td>
<td><strong>-0.065</strong></td>
</tr>
<tr>
<td>Private Banking</td>
<td>-0.053</td>
<td>0.009</td>
<td>-0.057</td>
</tr>
<tr>
<td>Long-Term</td>
<td><strong>-0.037</strong>*</td>
<td>0.022</td>
<td>-0.046</td>
</tr>
<tr>
<td>Broker</td>
<td>0.175</td>
<td>0.099</td>
<td>-0.068</td>
</tr>
<tr>
<td>Short-Seller</td>
<td>0.304</td>
<td>-0.024</td>
<td><strong>-0.268</strong></td>
</tr>
</tbody>
</table>

Note: This table presents estimates of price impact by investor type, split on the realized factor returns. *Scaled Repricing* is the average difference between an observed return and the counterfactual return that would be observed if a particular class of investors had their AUM proportionally redistributed to other investors, expressed in percentage points per $1T of redistribution. Factors are split into three categories based on the realized average factor return. Factors are assigned to Negative if they have a negative average return. The remaining factors are assigned to Low if they have a below median positive return and High otherwise. The average across the set of factors within a group is presented. Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 2.6 presents estimates of Eq. (2.44) by investor type. Notably, the three investor types with the greatest attention capacities—hedge funds, investment advisors, and mutual funds—have a negative and significant slope. Further, hedge funds are able to shrink expected returns by 39.8% per $1 trillion of invested capital, an effect that is an order of magnitude larger than other investor types. For example, investment advisors (mutual funds) shrink expected returns by only 5.5% (3.6%), consistent with hedge funds having much higher attention capacities. No other investor type is able to impact expected returns in a systematic way.

Interestingly, none of the coefficients $\beta_m$ are positive for the investor types presented. Across all investors, the estimated coefficients should average to zero; if all investors had impacts that were increasing in the average risk premium, then the observed risk premium would be too high and markets would not clear. This indicates that the household sector’s repricing is negatively related to expected returns.

Table 2.7 displays the top and bottom 5 characteristics in terms of scaled repricing.
Table 2.6: Regression of Repricing on Returns

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$R^2$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hedge Fund</td>
<td>0.001</td>
<td>-0.398***</td>
<td>0.164</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.124)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>0.000*</td>
<td>-0.055***</td>
<td>0.184</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.016)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>0.000</td>
<td>-0.036*</td>
<td>0.053</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.021)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private Banking</td>
<td>0.000</td>
<td>-0.025</td>
<td>0.019</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.025)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long-Term</td>
<td>0.000</td>
<td>-0.009</td>
<td>0.009</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broker</td>
<td>0.001</td>
<td>-0.047</td>
<td>0.015</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.053)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short-Seller</td>
<td>0.001</td>
<td>-0.155</td>
<td>0.036</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.111)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This table presents estimates of a regression of Scaled Repricing on realized average factor returns by investor type. Scaled Repricing is the average difference between an observed return and the counterfactual return that would be observed if a particular class of investors had their AUM proportionally redistributed to other investors, expressed in percentage points per $1T of redistribution. The short factor is omitted for Short-Sellers. Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.
Table 2.7: Scaled Repricing by Characteristic and Investor Type

<table>
<thead>
<tr>
<th>Type</th>
<th>Top 5 Characteristics (Scaled)</th>
<th>Bottom 5 Characteristics (Scaled)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Factor</td>
<td>Repricing</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>ivol</td>
<td>3.46</td>
</tr>
<tr>
<td></td>
<td>rome</td>
<td>2.47</td>
</tr>
<tr>
<td></td>
<td>mom12m</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td>rnoa</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>rooa</td>
<td>1.98</td>
</tr>
<tr>
<td>Inv. Advisor</td>
<td>oprof</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>roa</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>rnoa</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>mom6m</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>shvol</td>
<td>0.25</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>rnoa</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>oprof</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>roaa</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>shvol</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>repurch</td>
<td>0.27</td>
</tr>
<tr>
<td>Private Banking</td>
<td>momrev</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>oprof</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>roaa</td>
<td>0.37</td>
</tr>
<tr>
<td></td>
<td>shvol</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>repurch</td>
<td>0.27</td>
</tr>
<tr>
<td>Long-Term</td>
<td>oprof</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>eqissa</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>growth</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>ltrev</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>roea</td>
<td>0.14</td>
</tr>
<tr>
<td>Broker</td>
<td>strev</td>
<td>1.63</td>
</tr>
<tr>
<td></td>
<td>valuem</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>indrevlv</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>indrev</td>
<td>1.02</td>
</tr>
<tr>
<td></td>
<td>shvpl</td>
<td>0.86</td>
</tr>
<tr>
<td>Short-Seller</td>
<td>indrev</td>
<td>3.17</td>
</tr>
<tr>
<td></td>
<td>strev</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td>age</td>
<td>2.20</td>
</tr>
<tr>
<td></td>
<td>indmomrev</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td>indrevlvc</td>
<td>1.75</td>
</tr>
</tbody>
</table>

Note: This table presents the top and bottom 5 characteristics by Scaled Repricing for each type of investor. Scaled Repricing is the average difference between an observed return and the counterfactual return that would be observed if a particular class of investors had their AUM proportionally redistributed to other investors, expressed per $1T. Vol Impact is the difference between the observed factor volatility and the volatility that would be observed if a particular class of investors had their AUM proportionally redistributed to other investors, expressed per $1T. All measures are expressed in percentage points. Appendix A.1.4 contains a detailed description of the characteristics considered.
for each investor type. It is interesting to note the substantial differences in magnitudes and the factors affected. For example, per $1 trillion, long-term investors hardly move prices at all; they most impact the value factor and the magnitude is only 32 basis points per quarter. On the other hand, hedge funds decrease the value premium by 1.9 pp per $1 trillion of invested capital. Brokers and short-sellers have repricing on the order of 1-2 pp per quarter in absolute value for several momentum- and reversal-based characteristics, while mutual funds and investment advisors have impacts closer to that of long-term investors.

2.5.3 Impact on Volatility

Turning attention to the impact investors have on volatility, note again that this can be measured as simply the difference between observed and counterfactual volatility. Denoting the volatility of characteristic $c$ as $\sigma_c$, the measure of type $m$’s impact on volatility is

$$Vol \ Impact^c_m = \sigma_c - \sigma_c^{-m}, \quad (2.45)$$

where the $-m$ superscript represents the counterfactual of redistributing the AUM of type $m$ investors as usual. This measure suffers from the same issue as the raw repricing measure in that it scales with the size of an investor group, which is taken to be exogenous. Normalizing by the average AUM of that investor type makes the resulting scaled volatility impact size-invariant:

$$Scaled \ Vol \ Impact^c_m = \bar{A}_m^{-1} (\sigma_c - \sigma_c^{-m}), \quad (2.46)$$

where $\bar{A}_m$ is the average total AUM of investor type $m$.

Table 2.8 presents estimates of both measures. The first column, which contains
the raw volatility impact, shows that investment advisors have the greatest impact on volatility. In the absence of investment advisors, the average quarterly volatility would rise by 86 basis points. However, when accounting for the size of each investor type, short-sellers, brokers and hedge funds emerge as the most impactful, with a 55, 41, and 22 basis point increase in quarterly volatility per $1T reduction in invested capital, respectively. Finally, the last column expresses the scaled volatility impact as a percent of each characteristic’s baseline volatility. The resulting estimates are relatively small, with hedge funds (short-sellers) reducing volatility by an average of 2.9% (5.7%).

Table 2.8: Volatility Impact by Investor Type

<table>
<thead>
<tr>
<th></th>
<th>Vol Impact</th>
<th>Scaled Vol Impact</th>
<th>% Scaled Vol Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hedge Fund</td>
<td>-0.158*** (0.031)</td>
<td>-0.225*** (0.045)</td>
<td>-2.902*** (0.482)</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>-0.857*** (0.146)</td>
<td>-0.132*** (0.022)</td>
<td>-1.814*** (0.314)</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>-0.173** (0.078)</td>
<td>-0.055** (0.025)</td>
<td>-0.908*** (0.292)</td>
</tr>
<tr>
<td>Private Banking</td>
<td>-0.015 (0.020)</td>
<td>-0.029 (0.040)</td>
<td>-0.614 (0.467)</td>
</tr>
<tr>
<td>Long-Term</td>
<td>-0.052*** (0.014)</td>
<td>-0.074*** (0.020)</td>
<td>-0.667*** (0.195)</td>
</tr>
<tr>
<td>Broker</td>
<td>-0.076*** (0.012)</td>
<td>-0.412*** (0.064)</td>
<td>-4.518*** (0.688)</td>
</tr>
<tr>
<td>Short-Seller</td>
<td>-0.257*** (0.068)</td>
<td>-0.554*** (0.145)</td>
<td>-5.782*** (1.467)</td>
</tr>
</tbody>
</table>

*Note: This table presents estimates of the impact each investor type has on the volatility of factor returns. Vol Impact is the difference between the observed factor volatility and the volatility that would be observed if a particular class of investors had their AUM proportionally redistributed to other investors. Scaled Vol Impact scales expresses volatility impact in terms of impact per $1T in AUM, while % Scaled Vol Impact expresses this impact as a percentage of the observed volatility. Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01."

While the signs are all negative, the relative magnitudes are mostly consistent with the theory and attention capacities estimated in Section 2.4. With the exception of brokers and short-sellers, hedge funds reduce volatility the most, while private banking and long term investors reduce it the least. This is consistent with Proposition 10, which states that the volatility impact is decreasing with attention capacity.

The fact that all of the investors tabulated reduce volatility diverges from the
theory. This indicates that households must increase volatility. This may be the byproduct of the construction of the household sector. As households are constructed as the “residual” to ensure markets clear, it is natural to think that their portfolios will exhibit large variation that comes purely from taking the opposite side of all institutional investors. As a result, while the magnitudes of these estimates are informative, their sign should be interpreted with caution.

Overall, these results highlight the roles different investors play in price setting. Hedge funds act as prototypical arbitrageurs in that they are responsible for the greatest repricing, shrink expected returns towards zero, and the reduce volatility of factor returns. Long-term investors, investment advisors, and mutual funds exhibit very little ability to impact prices on a per-dollar basis. Further, the evidence suggests a relatively important role for brokers in affecting premia and volatility, consistent with recent theories on intermediary asset pricing (He and Krishnamurthy, 2013).

2.6 Conclusion

Portfolios are not homogenous across investor classes. As a result, some investors must over-perform and others must under-perform. This raises the questions of which investors have skill and how they allocate their skill across strategies. I first develop a theoretical model to study the allocation problem and generate predictions that depend on this allocation. Using the model as a guide, I provide new empirical measures to both estimate investor ability and examine how they allocate their attention. The model is further able to match several key characteristics about investors’ portfolios and how they relate to investors’ abilities.

Using a demand system, I examine the impact different investors have on risk premia and volatility. This allows me to estimate each investor’s importance in the
price formation process both in aggregate and across risk factors. In particular, I show that the investors with the greatest ability—hedge funds—have the greatest impact on asset prices. In particular, their impact shrinks expected returns associated with these characteristics towards zero. In this way, hedge funds affirm their status as the prototypical arbitrageurs.

This work presents several avenues for future research. Having documented substantial differences in ability across investors, micro-founding the source of heterogeneity is particularly interesting. For example, this could reflect frictions between investor and client, such as benchmark evaluation in mutual funds (Breugem and Buss, 2019), flow-performance relationships (Chevalier and Ellison, 1997), or compensation contracts (Elsaify, 2020). It would also be interesting to use differences in investors’ portfolios to recover beliefs about whether a risk factor is under- or over-priced using a revealed preference approach à la Berk and van Binsbergen (2016) and Barber et al. (2016).
Chapter 3

Hedge Fund Incentives, Risk Taking, and Asset Prices

3.1 Introduction

Traditional asset pricing models rely on the assumption that financial intermediaries act as a “veil” for household preferences. Implicit in this assumption is that intermediaries not only face identical constraints, but also have the same objectives as households. While previous work has documented theoretical and empirical violations of the former, evidence on the latter’s impact on aggregate asset prices is scarce.\(^1\) In practice, the preferences of financial intermediaries and consumers can be misaligned for a variety of reasons.\(^2\) Perhaps most saliently, discrepancies in objectives between an asset manager and an investor arise due to asymmetric or convex compensation schemes, in which the manager is rewarded for any upside but does not face symmetric losses following poor performance.

This paper uses hedge funds as a laboratory to study the impact of preference

\(^1\)See e.g., Gromb and Vayanos (2002, 2018); Gårleanu and Pedersen (2011); He and Krishnamurthy (2013); Adrian et al. (2014); He et al. (2017).

\(^2\)Other studies of endogenous hedge fund risk taking include Aragon and Nanda (2012); Buraschi et al. (2014). Aragon and Nanda (2012) study determinants of hedge fund risk shifting around high-water marks, while Buraschi et al. (2014) show that endogenous risk taking can affect measures of hedge fund alphas. Lou (2012); Vayanos and Woolley (2013) explore how flow-based incentives can explain return predictability. Cuoco and Kaniel (2011); Basak and Pavlova (2013); Albuquerque et al. (2019) study benchmarking and relative performance evaluation and their impact on indexed vs. non-indexed assets. Koijen (2014); Christoffersen and Simutin (2017); Heater (2017) study risk shifting in mutual funds; in particular, Christoffersen and Simutin (2017) examine how the presence of a benchmark induces investment in higher beta securities.
misalignment between manager and investor on aggregate asset prices and allocations. In particular, I examine to what extent hedge fund managers alter portfolio risk and composition in response to the incentives generated by their contract structures as well as assess the impact that these actions have on asset prices. I show both theoretically and empirically that hedge fund risk taking varies as a function of a fund’s distance to its high-water mark. Specifically, funds close to their high-water marks take on significantly greater levels of risk by, in part, investing in more volatile securities. Further, this changing demand for risky securities affects the market risk premium and asset allocations across different types of investors.

Hedge funds serve as an ideal context to examine this for two reasons. First, hedge fund managers face an explicit and time-varying asymmetry in their compensation schemes. The oft-cited “2/20” compensation structure features a 2% flat fee of total assets under management and a 20% performance fee on all returns exceeding a threshold, known as the high-water mark (HWM).\(^3\) In addition to this structure, investor flows act as a disciplining device on fund managers (Chevalier and Ellison, 1997; Berk and Green, 2004; Feng et al., 2011). Persistent poor performance leads to continued outflows, which can eventually result in liquidation that is costly to the manager. These two induced nonlinearities—one of which is an explicit feature of managers’ contract structure—lead to managerial preferences that are misaligned with those of investors (Goetzmann et al., 2003; Panageas and Westerfield, 2009; Lan et al., 2013; Drechsler, 2014). Further, the importance of these asymmetries depends on the previous performance of the fund, generating time-variation in the degree of misalignment.

Second, hedge funds are lightly regulated. They are afforded freedoms that are largely unique in the asset management industry, such as the ability to sell short and

\(^3\)Typically, the high-water mark is the running maximum value of the fund and is reset annually.
trade on leverage. In addition, they often have lockup and restriction periods that limit the withdrawal of capital and can erect walls to block withdrawals entirely, which protect them against deviations from fundamentals (e.g., Shleifer and Vishny, 1997). The effect of this is twofold. The lack of regulation allows hedge fund managers to potentially pursue strategies in line with their own interests rather than those of their investors. Additionally, these extra flexibilities serve to reduce adjustment costs; as hedge funds have more tools at their disposal, changing portfolio composition should be less costly. Thus, to the extent that the objective and/or constraints of the hedge fund’s problem are dynamic, it stands to reason that the solution—a set of portfolio weights—should be dynamic as well.

Exploiting these features of the hedge fund industry, I build a heterogeneous agent static equilibrium model and empirically test its predictions. A continuum of risk-neutral hedge fund managers maximize their expected compensation from a prototypical contract subject to liquidation risk, while a single mutual fund manager has risk averse preferences aligned with those of her investors. Managers invest in either a risky or less risky asset subject to a binding leverage constraint and expected returns are determined in equilibrium.

The nonlinearities in the hedge fund’s objective function lead to either relatively risk-loving or risk-averse behavior depending on the distance to a fund’s high-water mark. Hedge fund managers overseeing funds with AUM at or near their HWMs choose the asset to maximize their expected performance fee. Due to the convexity in this compensation, these managers will prefer the risky asset, holding expected returns constant. Similarly, funds far below their HWMs want to minimize the probability of liquidation. For funds below the liquidation threshold, this entails “gambling for resurrection”—taking excess risk in order to clear the threshold—while, for funds above the threshold, this entails reducing risk. All other managers face linear
compensation schemes and, as a result, are risk-neutral.

Two parameters are critical to determining the expected returns and holdings of the two assets. The first is the mass of hedge funds relative to that of mutual funds. In equilibrium, expected returns are determined by the combination of hedge funds and mutual funds. If the ratio of hedge funds to mutual funds is small, the degree to which nonlinearities in the hedge fund contract structure matter for asset prices is limited: expected returns will be determined primarily by mutual funds, who will be the marginal agents. As a result, prices will not fully reflect the preferences of hedge fund managers, who are relatively less risk averse. This makes the risky security underpriced from the hedge fund managers’ perspective, leading to large investment in this asset. Conversely, if hedge funds make up a large portion of the market, their impact on asset prices will be large. As they are relatively less risk averse, this will reduce the risk premium on the risky asset. This lower premium reduces hedge fund investment in that asset.

The second is the distribution of hedge fund AUM relative to their HWMs. Absent hedge funds, the risky asset trades at a premium due to the risk aversion of mutual fund managers. The model shows that the aggregate risk aversion of the hedge fund industry and thus the risk premium depends on the shape of this distribution. If it is highly skewed towards the tails (many funds below their liquidation threshold or close to their HWMs), hedge funds will be relatively less risk averse, generating high demand for the risky asset. This will dominate the effect of mutual fund risk aversion, causing the risky asset to trade at a discount relative to the less risky asset. As hedge fund managers are relatively less risk averse, they will predominately hold

---

4While hedge fund assets under management make up only a small percentage of total managed money, the general impact on asset prices is dependent on trading size and volume. As hedge funds have the ability to take on (significant amounts of) leverage, they are likely to matter more than what their AUM would suggest.
the risky asset. If, on the other hand, the hedge fund wealth distribution has thinner
tails, there will be less demand for the risky asset from hedge funds, which will raise
its expected return above the less risky asset.

In line with this intuition, I derive simple conditions characterizing the cases in
which expected excess returns will be equal and in which the risky asset will trade
at a premium or discount relative to the less risky asset. Demand for the risky asset
can be met by a combination of three portions of the hedge fund wealth distribution:
those close to their HWMs, those below the liquidation threshold, and those facing
linear compensation schemes. The first two sets of hedge funds strictly prefer the
risky asset, while the latter is indifferent. If the total mass of managers in these three
regions is sufficiently small, the risky asset will trade at a premium relative to the
less risky asset. As a result of the hedge fund wealth distribution, the hedge fund
sector is more risk averse in aggregate. On the other hand, if the mass of the first
two regions is sufficiently large, the risky asset will trade at a discount, as the hedge
fund sector is less risk averse in aggregate. Otherwise, the expected returns of these
assets will be equal in equilibrium.

Empirically, I find evidence consistent with the model’s predictions. Combining
self-reported hedge fund returns with snapshots of company holdings reported via
Form 13-F, I am able to examine both fund return volatility and the risk composition
of the underlying portfolio. Sorting hedge funds into bins based on the distance to
their HWMs, funds in the top (closest to the threshold) and bottom (furthest from
the threshold) deciles have volatilities that are 0.7 and 3.8 percentage points higher
than that of the median fund, respectively. Further, management companies in the
top and bottom quintiles invest in securities that are on average 2.1 and 2.7pp more
volatile than those of the median fund, select into securities that are 3.3 and 4.0pp
more volatile, and, crucially, chase volatility. That is, they select into securities that
are expected to be 2.9 and 2.2pp more volatile over the following year. Funds closest
to their HWMs increase the volatility of their portfolio through investment in stocks
with higher market betas, while those furthest from the threshold do so through
investment in stocks with greater idiosyncratic volatilities.

The distribution of fund wealth exhibits substantial time series and cross sec-
tional variation. Under 14.3% of funds are below their HWMs as of 2003, compared
with 59.4% in 2008. Exploiting this variation, I compute two nonparametric proxies
capturing relevant measures of risk seeking in the distribution of hedge fund wealth.
On aggregate, hedge funds invest in riskier securities when many funds are near their
HWMs. A one standard deviation increase in the fraction of funds at their HWMs
increases the volatility of the mean hedge fund investment by 1.6pp relative to the
average security. Further, the same shock reduces the slope of the security market
line by 6pp per annum and increases the spread between returns of low and high
beta securities (betting against beta) by 8.4pp per annum. This is in line with the
finding that funds close to their HWMs increase the average beta of their portfolios,
which pushes down the expected return of high beta stocks, causing a flattening of
the security market line.

The opposite pattern emerges when hedge funds fall far below their HWMs. A one
standard deviation increase in the fraction of funds in the bottom decile of the HWM
distance distribution steepens the security market line and mitigates betting against
beta. This occurs primarily due to exit and outflows; following the same shock,
net outflows from the hedge fund industry exceed 4%. Further, this is concentrated
among poorly performing hedge funds; firms in the bottom decile are associated with
net outflows of over 17% relative to the median fund. As a result, their preferences
are reflected less in subsequent asset prices.

Overall, this work highlights the importance of differences in preferences—both
among intermediaries and between intermediaries and investors—in understanding
the impact of intermediaries on asset prices. Time-variation in preferences arising
from managerial compensation affects risk premia, as well as allocations and risk ex-
posures across different types of agents. This complements existing research focusing
on occasionally binding intermediary constraints. While these constraints generally
amplify risk premia in crises (i.e., when the constraint binds), the impact of these
constraints in non-crises is less clear. In contrast, hedge fund managers’ preferences
are highly dynamic and exhibit substantial cross-sectional and time series variation.
As a result, they provide a lens into the impact of intermediaries on asset prices in
all states of the economy.

The chapter proceeds as follows. Section 3.2 introduces the model and derives its
solution. The data used in the empirical tests of the model are discussed in Section
3.3. Section 3.4 details the construction of main variables and methodology employed.
In Section 3.5, the main results are presented. Section 3.6 concludes. All proofs are
in the Appendix.

3.2 Model

I present a simple model to illustrate how asymmetric compensation structures in
the hedge fund industry can affect portfolio allocations across agents and asset prices
in general.

3.2.1 Setup

The economy has two assets $j = 1, 2$ traded at time $t = 0$. At $t = 1$, all uncertainty
is resolved and payoffs are realized. Net returns are independent across assets and
given by
\[ r_j = \alpha_j + f_j, \] (3.1)
where \( \alpha_j \) is the expected return of asset \( j \) and \( f_j \) is a mean zero random variable. To keep the model simple, I assume \( f_1 \) and \( f_2 \) are bounded random variables with the same support, and \( f_1 \) is more prone to extreme outcomes than \( f_2 \). Specifically, \( f_1 \) is distributed uniformly over \([-\sigma, \sigma]\), while \( f_2 \) is given by the sum of two independent uniforms, each distributed over \([-\sigma/2, \sigma/2]\).\(^5\) That is,
\[ f_1 \sim U[-\sigma, \sigma] \]
\[ f_2 = \xi_1 + \xi_2, \quad \xi_i \overset{iid}{\sim} U[-\sigma/2, \sigma/2] \]
Thus, the pdfs of \( f_1 \) and \( f_2 \) are given by
\[ g_1(x) = \begin{cases} \frac{1}{2\sigma} & x \in [-\sigma, \sigma] \\ 0 & \text{otherwise} \end{cases} \] (3.3)
and
\[ g_2(x) = \begin{cases} \frac{1}{\sigma} + \frac{x}{\sigma^2} & -\sigma \leq x \leq 0 \\ \frac{1}{\sigma} - \frac{x}{\sigma^2} & 0 < x \leq \sigma \\ 0 & \text{otherwise} \end{cases} \] (3.4)

The model is depicted in Figure 3.1. There are two investment vehicles in the economy, hedge funds \((H)\) and mutual funds \((M)\). There is a mass \( \theta \) of hedge funds that are each run by their own manager and mass \( 2 - \theta \) of mutual funds overseen by a single manager.\(^6\) Managers assign funds to invest in one of assets 1 or 2, but

\(^5\)To see that \( f_2 \) is less prone to extreme outcomes, note that, for \( 0 < c < \sigma \), \( P(f_1 > c) = \frac{\sigma - c}{2\sigma} > \frac{\sigma - c}{2\sigma} \left( \frac{\sigma - c}{\sigma} \right) = P(f_2 > c) \). Similarly, for \(-\sigma < c < 0\), \( P(f_1 < c) > P(f_2 < c) \).

\(^6\)Mutual funds are run by a single manager to reflect the high degree of concentration within this
cannot mix between the two.

**Figure 3.1: Model Overview**

Managers

\[ i, j, k, \ldots \]

Funds

\[ H_1, H_2, H_3, \ldots \]

Mutual Funds

\[ M_1, M_2, M_3, \ldots \]

Assets

\[ \text{mass } \theta, 1 \]

\[ \text{mass } 2 - \theta, 2 \]

This figure describes the modeling framework. Black arrows indicate manager relationships (i.e., manager \( i \) manages \( H_1 \), while manager \( z \) manages \( M_1, M_2, M_3, \ldots \)). Red (blue) lines indicate investment of that fund in asset 1 (2). There is a mass \( \theta \) of hedge funds and a mass \( 2 - \theta \) of mutual funds.

Restricting these assets to be mutually exclusive allows for the interpretation of these assets as fully-levered trading strategies. As hedge fund managers are risk-neutral (discussed below), funds subject to a leverage constraint will optimally choose to have the maximum exposure to the strategy they select. The composition of the assets has two interpretations. First, these assets can be thought of as strategies resulting from investing in high-risk securities (asset 1) or in low-risk securities (asset 2). Alternatively, \( f_1 \) can be expressed as \( f_1 = 2 \xi_1 \), and \( \xi_i \) can be thought of as risk factors. With this interpretation, asset 1 has a loading of two on a single risk factor, while asset 2 has unit loading on two independent risk factors. In this way, asset 1 can be thought of as an undiversified strategy, while asset 2 represents a more diversified strategy.

Each hedge fund manager is risk-neutral, but faces payoffs that are non-linear in outcomes as a result of their compensation contracts. Managers receive compensation via three sources. First, they receive a management fee equal to a fraction \( m \) of the total wealth \( W \) managed by the fund. Second, managers have a fraction \( i \) of the industry. For modeling purposes, as mutual funds are assumed homogenous, this allows for mixing between assets that would not occur without this assumption.
their own capital invested in the fund and thus receive this fraction of fund returns. Third, managers receive a fraction \( p \) of returns on investor funds (i.e., not their own investment) only if the fund clears its high-water mark, which grows at an exogenous rate of \( g \leq \min \{ \alpha_1, \alpha_2 \} \). Hence, fund value must clear \( (1 + g) W_{\text{HM}} \) to earn a performance fee. Finally, in the case of poor performance, hedge funds are subject to liquidation risk. If fund wealth falls below a fraction \( \gamma \) of its HWM, the fund is liquidated with probability \( \delta \), with an associated cost \( A \) to the manager.

In practice, the HWM is reset every year to the running maximum of the fund. Motivated by this, let \( \tilde{W} \equiv \frac{W}{W_{\text{HM}}} \in [0, 1] \) be the normalized wealth-HWM ratio, hereafter simply referred to as “wealth.”\(^7\) The mass \( \theta \) of hedge funds have their wealth \( \tilde{W} \) distributed along the unit interval with mass function \( f(\cdot) \) and cdf \( F(\cdot) \). The payoff to a hedge fund manager from choosing asset \( j \in \{1, 2\} \) is given by

\[
u_j^{H}(\tilde{W}) = W \mathbb{E} \left[ m + i r_j + p (1 - i) H_j \left( 1 + r_j - \frac{1 + g}{\tilde{W}} \right) - A \delta L_j \right], \tag{3.5}
\]

where \( H_j = 1 \{ f_j > \bar{c}_j \} \) and \( L_j = 1 \{ f_j \leq c_j \} \),

\[
\bar{c}_j \equiv \max \left\{ \frac{1 + g - \tilde{W} (1 + \alpha_j)}{\tilde{W}}, -\sigma \right\}, \tag{3.6}
\]
\[
c_j \equiv \min \left\{ \frac{\gamma - \tilde{W} (1 + \alpha_j)}{\tilde{W}}, \sigma \right\},
\]

and \( \bar{c}_j (c_j) \) is the minimum (maximum) realization of \( f_j \) above (below) which a fund clears its HWM (can be liquidated).\(^8\) If \( H_j = 1 \), then a performance fee is collected on all investor returns exceeding \( \frac{1 + g}{\tilde{W}} \). Similarly, if \( L_j = 1 \), then the fund is liquidated with probability \( \delta \). Finally, since \( \tilde{W} \) is simply a scalar and is common to both assets,

\( ^7\)Strictly speaking, before the HWM is reset, a fund can have wealth exceeding its HWM (i.e., \( \tilde{W} > 1 \)). The general results are unchanged when accounting for this possibility.

\( ^8\)\( \bar{c}_j \) and \( c_j \) are technically functions of \( \tilde{W} \), but this is suppressed for simplicity.
it is omitted without loss of generality. Hedge funds simply choose \( j \in \{1, 2\} \) that maximizes utility given in Eq. (3.5).

A single mutual fund manager oversees the entire \( 2 - \theta \) mass of mutual funds and has mean-variance utility with risk aversion parameter \( \lambda \). The manager chooses a fraction \( \eta^M \in [0, 1] \) of funds indiscriminately to invest in asset 1, while the remaining \( 1 - \eta^M \) funds invest in asset 2. Unlike hedge funds, the mutual fund manager is only concerned with the return of the aggregate mutual fund industry.\(^9\) That is, her utility is given by

\[
u^M (\eta^M) = \mathbb{E} \left[ 1 + \eta^M r_1 + (1 - \eta^M) r_2 \right] - \frac{\lambda}{2} \mathbb{V} \left[ \eta^M r_1 + (1 - \eta^M) r_2 \right]. \tag{3.7}\]

**Equilibrium.** Let the set of hedge funds that strictly prefer asset 1 and are indifferent between both assets be denoted \( \Omega^S \) and \( \Omega^I \), respectively. Formally, they are defined as

\[
\Omega^S \equiv \{ \bar{W} : u_1^H (\bar{W}) > u_2^H (\bar{W}) \} \tag{3.8}
\]
\[
\Omega^I \equiv \{ \bar{W} : u_1^H (\bar{W}) = u_2^H (\bar{W}) \}. \tag{3.9}
\]

The demand for asset 1 by hedge funds, \( \eta^H \), is then restricted to

\[
\eta^H \in \left[ \int_{\Omega^S} f(x) \, dx, \int_{\Omega^S \cup \Omega^I} f(x) \, dx \right]. \tag{3.10}
\]

An equilibrium consists of expected returns, \( \alpha_1 \) and \( \alpha_2 \), and a set of demands \( \eta^H \) and \( \eta^M \) such that:

1. Each hedge fund chooses asset \( j \in \{1, 2\} \) to maximize utility given in Eq. (3.5).

\(^9\)This is motivated by the concentration in the mutual fund industry. The five (ten) largest fund complexes managed 51\% (61\%) of all mutual fund and ETF assets in the United States as of 2018 (https://www.ici.org/pdf/2019_factbook.pdf).
2. $\eta^M \in [0, 1]$ maximizes the utility of the mutual fund manager given in Eq. (3.7).

3. Markets clear: $\theta \eta^H + (2 - \theta) \eta^M = 1$.

### 3.2.2 Mutual Funds

The first order condition of the mutual fund’s problem is

$$ [\eta^M] : \alpha_1 - \alpha_2 + \frac{\lambda \sigma^2}{6} - \lambda \eta^M \frac{\sigma^2}{2} = \nu_0 - \nu_1, \quad (3.11) $$

where $\nu_0$ and $\nu_1$ are the Lagrange multipliers on the constraint $\eta^M \in [0, 1]$. The solution is then given by

$$ \eta^M = \begin{cases} 
0 & \alpha_1 - \alpha_2 \leq -\frac{\lambda \sigma^2}{6} \\
\frac{6(\alpha_1 - \alpha_2) + \lambda \sigma^2}{3\lambda \sigma^2} & \alpha_1 - \alpha_2 \in ( - \frac{\lambda \sigma^2}{6}, \frac{\lambda \sigma^2}{3}] \\
1 & \alpha_1 - \alpha_2 > \frac{\lambda \sigma^2}{3}. 
\end{cases} \quad (3.12) $$

This expression is a natural result from standard asset allocation models. The manager shifts weights towards the asset with higher returns, but balances this against the standard deviation of the assets. If expected returns are sufficiently far apart, the boundary constraints bind. Asset 1 has twice the variance of asset 2, so, if the expected returns are the same, the manager will allocate $\frac{1}{3}$ of the funds to asset 1 and $\frac{2}{3}$ to asset 2.
3.2.3 Hedge Funds

Before analyzing the hedge fund problem, it is necessary to impose an assumption on the support of the distribution $\sigma$. If $\sigma$ is too large, it may be possible for a fund to be either liquidated or clear its HWM depending on the realization of $f_j$. For tractability, Assumption 2 rules this out.

**Assumption 2.** Let $\alpha = \min \{\alpha_1, \alpha_2\}$ and $\bar{\alpha} = \max \{\alpha_1, \alpha_2\}$. $\sigma$ is sufficiently small:

$$\sigma < \frac{(1+g-\gamma)(1+\bar{\alpha})}{1+g+\gamma}.$$

With Assumption 2, the analysis is simplified to three regions (possibly with no mass) of the wealth distribution. In region $\mathcal{L} \equiv \left( \frac{\gamma}{1+\bar{\alpha}+\sigma}, \frac{\gamma}{1+\bar{\alpha}-\sigma} \right)$, there exist realizations of both $f_1$ and $f_2$ such that fund value falls below the liquidation threshold $\gamma$. In region $\mathcal{H} \equiv \left( \frac{1+g}{1+\bar{\alpha}+\sigma}, 1 \right)$, there exist realizations of both $f_1$ and $f_2$ such that fund value clears $(1+g)$ HWM. The remaining portion of the distribution, $\mathcal{R} \equiv [0,1] \setminus (\mathcal{L} \cup \mathcal{H})$, contains funds where (i) liquidation is certain with either asset, (ii) liquidation is only possible with the lower expected return asset, (iii) neither liquidation nor clearing the HWM is possible, or (iv) clearing the HWM is possible only with the higher expected return asset. In any case, the behavior of region $\mathcal{R}$ funds is straightforward to characterize and given in Proposition 11.

**Proposition 11.** Given Assumption 2, any fund with wealth $\tilde{W} \in \mathcal{R}$ will always prefer the higher expected return asset. If both assets yield the same expected return, then funds will be indifferent between assets.

The intuition behind Proposition 11 is as follows. Clearing the HWM is beneficial to managers, while falling below the liquidation threshold is costly. Suppose asset $j$ has a lower expected return than asset $k$. If a fund in region $\mathcal{R}$ can clear its HWM, it is only possible with asset $k$. As a result, the manager will strictly prefer asset
Similarly, if fund wealth can fall below $\gamma$, it is only possible with asset $j$. To insure against this possibility, the manager will select asset $k$. If both assets offer the same return, then it is impossible to either clear the HWM or fall below the liquidation threshold. In this case, the manager faces no asymmetric payoffs, so, by risk neutrality, is indifferent between the two assets.

**Region $\mathcal{L}$.** By Assumption 2, funds in region $\mathcal{L}$ can fall below the threshold $\gamma$ with either asset, but cannot earn a performance fee. Thus, their utility can be simplified to\(^{10}\)

$$u^H_j(\bar{W}) = m + i\alpha_j - A\delta \Pr(f_j \leq c_j). \tag{3.13}$$

Evaluating this probability for both assets,

$$u^H_1(\bar{W}) = m + i\alpha_1 - \frac{A\delta(\gamma - c_1)}{2\sigma}$$

$$u^H_2(\bar{W}) = \begin{cases} m + i\alpha_2 - \frac{A\delta(\gamma - c_2)}{2\sigma} & c_2 \leq 0 \\ m + i\alpha_2 - \frac{A\delta(\gamma - c_2)}{2\sigma^2} & c_2 > 0. \end{cases} \tag{3.14}$$

**Lemma 4.** Given Assumption 2, if $\bar{W} \in \mathcal{L}$ and $\alpha \equiv \alpha_1 = \alpha_2$, then for $\bar{W} < \frac{\gamma}{1+\alpha}$, asset 1 is strictly preferred: $u^H_1(\bar{W}) > u^H_2(\bar{W})$. For $\bar{W} > \frac{\gamma}{1+\alpha}$, asset 2 is strictly preferred: $u^H_1(\bar{W}) < u^H_2(\bar{W})$. Managers are indifferent for $\bar{W} = \frac{\gamma}{1+\alpha}$: $u^H_1\left(\frac{\gamma}{1+\alpha}\right) = u^H_2\left(\frac{\gamma}{1+\alpha}\right)$.

Lemma 4 characterizes the optimal behavior of region $\mathcal{L}$ managers when the expected returns are the same. There is a cutoff, $\frac{\gamma}{1+\alpha}$, above which managers prefer the less risky asset (asset 2) and below which they prefer the riskier asset (asset 1). This cutoff is the threshold below which the probability of falling into the liquidation region exceeds 50%. That is, funds below this cutoff need to get “lucky” to clear $\gamma$,\(^{10}\)See the left panel of Figure 3.2 for a plot of this utility for the two assets.
while funds above the threshold will most likely clear $\gamma$. As a result, funds below this cutoff are more willing to take risk, and prefer the higher risk asset, while those above this threshold want to avoid risk and prefer the lower risk asset.

**Proposition 12.** Let $d(\cdot) \equiv u_1^H(\cdot) - u_2^H(\cdot)$. Given Assumption 2 and $\tilde{W} \in \mathcal{L}$. If $\alpha_1 > \alpha_2$,

1. if $d(W^*_L) \geq 0$, then $u_1^H(\tilde{W}) \geq u_2^H(\tilde{W})$ for all $\tilde{W}$

2. if $d(W^*_L) < 0$, then there exists $\frac{\gamma}{1+\alpha_2} < W^L < W^*_L < \tilde{W}^L < \frac{\gamma}{1+\alpha_1-\sigma}$ such that $u_1^H(\tilde{W}) > u_2^H(\tilde{W})$ for all $\tilde{W}$ except for $\tilde{W} \in [W^L, \tilde{W}^L]$, in which case $u_1^H(\tilde{W}) \leq u_2^H(\tilde{W})$,

where $W^*_L \equiv \frac{2\gamma}{2(1+\alpha_2)-\sigma}$. Further, if $i = 0$, the condition $d(W^*_L) \geq 0$ is equivalent to $\alpha_1 - \alpha_2 \geq \frac{\sigma}{4}$.

If, instead, $\alpha_1 < \alpha_2$,

3. if $d(W^{**}_L) \leq 0$, then $u_1^H(\tilde{W}) \leq u_2^H(\tilde{W})$ for all $\tilde{W}$

4. if $d(W^{**}_L) > 0$, then there exists $\frac{\gamma}{1+\alpha_2+\sigma} < W^L < W^{**}_L < \tilde{W}^L < \frac{\gamma}{1+\alpha_2}$ such that $u_1^H(\tilde{W}) < u_2^H(\tilde{W})$ for all $\tilde{W}$ except for $\tilde{W} \in [W^L, \tilde{W}^L]$, in which case $u_1^H(\tilde{W}) \geq u_2^H(\tilde{W})$,

where $W^{**}_L \equiv \frac{2\gamma}{2(1+\alpha_2)+\sigma}$. Further, if $i = 0$, the condition $d(W^{**}_L) \leq 0$ is equivalent to $\alpha_1 - \alpha_2 \leq -\frac{\sigma}{4}$.

In conjunction with Lemma 4, Proposition 12 completely characterizes the behavior of managers overseeing funds with $\tilde{W} \in \mathcal{L}$. The intuition can most clearly be seen in the case in which $i = 0$. When $\alpha_1 \neq \alpha_2$, two results can occur. If $\alpha_1$ and $\alpha_2$ are sufficiently far apart, then all funds in region $\mathcal{L}$ will prefer the higher expected return asset. However, if the expected returns are sufficiently close, there will be a double-thresholding strategy. Suppose $\alpha_1 > \alpha_2$. As in Lemma 4, all funds with
wealth below $\frac{\gamma}{1+\alpha_2}$ will prefer the riskier asset, as they are relatively less risk-averse. If $\alpha_1 \gg \alpha_2$, then all funds with wealth above $\frac{\gamma}{1+\alpha_2}$ will also prefer asset 1. This is because the benefit of asset 1’s higher expected return outweighs the extra risk that high wealth funds want to avoid. However, for $\alpha_1$ and $\alpha_2$ sufficiently close, there will be a subset of funds with $\bar{W} > \frac{\gamma}{1+\alpha_2}$ for whom it is optimal to choose asset 2 despite its lower expected return.

The question then becomes: what wealth levels benefit most from this risk? Importantly, it is not the very highest wealth levels in this range. Rather, it is a subset of wealth levels strictly between $\frac{\gamma}{1+\alpha_2}$ and $\frac{\gamma}{1+\alpha_1}$. For funds very close to $\frac{\gamma}{1+\alpha_2}$, the added risk from asset 1 is not very valuable: they are likely to fall below the threshold $\gamma$, but the shift in the distribution (higher $\alpha$) outweighs the different shape of the distributions. For funds very close to $\frac{\gamma}{1+\alpha_1}$, falling below $\gamma$ is a rare event, and thus insuring against it via higher expected returns is more beneficial than altering the distribution shape. A similar logic holds for $\alpha_1 < \alpha_2$.

The left panel of Figure 3.2 illustrates the mechanism.\textsuperscript{11} When $\alpha_1 = \alpha_2$ (solid black, solid red), for $\bar{W} > \frac{\gamma}{1+\alpha}$, the utility from asset 2 is greater than that of asset 1. These are the funds for which falling below $\gamma$ is unlikely and thus are relatively risk averse. However, for $\bar{W} < \frac{\gamma}{1+\alpha}$, asset 1 has a greater utility than asset 2, as these funds are likely to fall below $\gamma$. As $\alpha_2$ rises (dashed red), funds with $\bar{W} > \frac{\gamma}{1+\alpha}$ still prefer asset 2; however, there is a subset of funds with $\bar{W} < \frac{\gamma}{1+\alpha}$ that switch from asset 1 to asset 2. These funds are at the extremes of this region. A similar phenomenon occurs as $\alpha_2$ falls (dotted red). As $\alpha_2$ continues to rise or fall (not pictured), all funds in region $\mathcal{L}$ will prefer asset 2 or asset 1, respectively.

**Corollary 2.** Let $D^L_1(\alpha_1, \alpha_2)$ denote the demand for asset 1 by funds with wealth

\textsuperscript{11}See Appendix A.2.2 for the parameter values.
\( W \in \mathcal{L} \). Given Assumption 2, \( D_1^\mathcal{L} (\alpha_1, \alpha_2) \) is given by

\[
D_1^\mathcal{L} (\alpha_1, \alpha_2) = \begin{cases} 
F \left( \frac{\gamma}{1+\alpha} \right) - F \left( \frac{\gamma}{1+\alpha+\sigma} \right) & \alpha_1 = \alpha_2 \\
F \left( \frac{\gamma}{1+\alpha_1-\sigma} \right) - F \left( \frac{\gamma}{1+\alpha_1+\sigma} \right) & \alpha_1 > \alpha_2 \cap d(W^*_\mathcal{L}) \geq 0 \\
F \left( \frac{\gamma}{1+\alpha_1} \right) - F \left( \frac{\gamma}{1+\alpha_1+\sigma} \right) + F(W^\mathcal{L}) - F(W^\mathcal{L}) & \alpha_1 > \alpha_2 \cap d(W^*_\mathcal{L}) < 0 \\
0 & \alpha_1 < \alpha_2 \cap d(W^*_\mathcal{L}) \geq 0 \\
F(W^\mathcal{L}) - F(W^\mathcal{L}) & \alpha_1 < \alpha_2 \cap d(W^*_\mathcal{L}) < 0,
\end{cases}
\]

where \( W^*_\mathcal{L} \) and \( W^*_\mathcal{L}^* \) are defined in Proposition 12. The demand for asset 2 in region \( \mathcal{L} \) is implicitly defined by

\[
D_2^\mathcal{L} (\alpha_1, \alpha_2) \equiv F \left( \frac{\gamma}{1+\bar{\alpha} - \sigma} \right) - F \left( \frac{\gamma}{1+\bar{\alpha} + \sigma} \right) - D_1^\mathcal{L} (\alpha_1, \alpha_2).
\]

Using Proposition 12, it is straightforward to characterize the demand for assets 1 and 2 from funds in region \( \mathcal{L} \). Corollary 2 formalizes the demand. If \( \alpha_1 \) and \( \alpha_2 \) are far apart, all funds will demand the higher expected return asset. However, if \( \alpha_1 \) and \( \alpha_2 \) are close, only the most and least wealthy funds in this region will demand the higher expected return asset. For \( \alpha_1 = \alpha_2 \), the region is split, with wealthy funds preferring asset 2 and less wealthy funds preferring asset 1.

**Region \( \mathcal{H} \).** Funds in region \( \mathcal{H} \) cannot fall below \( \gamma \), but can earn a performance fee (by Assumption 2). As a result, their utility can be simplified to\(^ {12} \)

\[
u_j^H (\bar{W}) = m + i \alpha_j + p (1 - i) P (f_j > \bar{c}_j) \mathbb{E} \left[ 1 + r_j - \frac{1 + g}{\bar{W}} | f_j > \bar{c}_j \right] = m + i \alpha_j + p (1 - i) P (f_j > \bar{c}_j) (\mathbb{E} [f_j | f_j > \bar{c}_j] - \bar{c}_j).
\]

\(^ {12}\)See the right panel of Figure 3.2 for a plot of this utility for the two assets.
Evaluating these for assets 1 and 2 gives

\[ u_H^1(\bar{W}) = m + i\alpha_1 + p(1 - i)\frac{(\sigma - \bar{c}_1)^2}{4\sigma} \]
\[ u_H^2(\bar{W}) = \begin{cases} 
  m + i\alpha_2 + p(1 - i)\left(\frac{\sigma^3 - 3\sigma^2\bar{c}_2 + 3\sigma\bar{c}_2^2 + \bar{c}_2^3}{6\sigma^2}\right) & \bar{c}_2 \leq 0 \\
  m + i\alpha_2 + p(1 - i)\frac{(\sigma - \bar{c}_2)^3}{6\sigma^2} & \bar{c}_2 > 0.
\end{cases} \]  

(3.18)

**Lemma 5.** Given Assumption 2, if \( \bar{W} \in \mathcal{H} \) and \( \alpha \equiv \alpha_1 = \alpha_2 \), then asset 1 is strictly preferred: \( u_H^1(\bar{W}) > u_H^2(\bar{W}) \). Further, for \( \alpha_1 > \alpha_2 \), asset 1 is strictly preferred: \( u_H^1(\bar{W}) > u_H^2(\bar{W}) \).

Lemma 5 establishes an unsurprising result: funds that can earn a performance fee prefer the riskier asset, as this maximizes the expected value of a performance fee. This holds for cases in which both the expected returns are the same (\( \alpha_1 = \alpha_2 \)), and when the riskier asset has a higher expected return (as a result of preference monotonicity). Lemma 5 applies for all funds in region \( \mathcal{H} \), as these are funds for which clearing the HWM occurs with positive probability with either asset.

**Proposition 13.** Given Assumption 2, for \( \bar{W} \in \mathcal{H} \), if \( \alpha_1 < \alpha_2 \),

1. if \( d(W_{\mathcal{H}}^*) \leq 0 \), then \( u_H^1(\bar{W}) \leq u_H^2(\bar{W}) \) for all \( \bar{W} \)

2. if \( d(W_{\mathcal{H}}^*) > 0 \), then there exists \( \frac{1 + g}{1 + \alpha_1 + \sigma} < W^H < W_{\mathcal{H}}^* < W^H < \frac{1 + g}{1 + \alpha_2 - \sigma} \) such that \( u_H^1(\bar{W}) < u_H^2(\bar{W}) \) for all \( \bar{W} \) except for \( \bar{W} \in [W^H, W^H_{\mathcal{H}}] \), in which case \( u_H^1(\bar{W}) \geq u_H^2(\bar{W}) \).

where \( W_{\mathcal{H}}^* \equiv \frac{2(1 + g)}{2(1 + \alpha_2) + \sigma - \sqrt{4(\alpha_1 - \alpha_2) + \sigma}} \). Further, if \( i = 0 \), the condition \( d(W_{\mathcal{H}}^*) \leq 0 \) is equivalent to \( \alpha_1 - \alpha_2 \leq -\frac{2\sigma}{g} \).

The intuition for Proposition 13 can be seen from the case in which \( i = 0 \). For \( \alpha_1 < \alpha_2 \) sufficiently close, there is a subset of funds in the middle of region \( \mathcal{H} \) that will
prefer asset 1 despite its lower returns. This echoes the result from Proposition 12: the funds in the middle of the region benefit the most from the added risk, as funds in the tails of region $\mathcal{H}$ either earn a performance fee with very low probability or with a virtual guarantee, so the added risk is not very beneficial. With this, Corollary 3 characterizes the demand from funds in region $\mathcal{H}$.

**Corollary 3.** Let $D^\mathcal{H}_1(\alpha_1, \alpha_2)$ denote the demand for asset 1 by funds with wealth $\bar{W} \in \mathcal{H}$. Given Assumption 2, $D^\mathcal{H}_1(\alpha_1, \alpha_2)$ is given by

$$D^\mathcal{H}_1(\alpha_1, \alpha_2) = \begin{cases} 
1 - F \left( \frac{1+g}{1+\alpha_2+\sigma} \right) & \alpha_1 \geq \alpha_2 \\
0 & \alpha_1 < \alpha_2 \cap d(W^*_\mathcal{H}) \leq 0 \\
F(W^\mathcal{H}) - F(W^\mathcal{H}_\mathcal{H}) & \alpha_1 < \alpha_2 \cap d(W^*_\mathcal{H}) > 0,
\end{cases} \tag{3.19}$$

where $W^*_\mathcal{H}$ is defined in Proposition 13. The demand for asset 2 in region $\mathcal{H}$ is implicitly defined by

$$D^\mathcal{H}_2(\alpha_1, \alpha_2) \equiv F \left( \frac{1+g}{1+\alpha - \sigma} \right) - F \left( \frac{1+g}{1+\alpha + \sigma} \right) - D^\mathcal{H}_1(\alpha_1, \alpha_2). \tag{3.20}$$

Figure 3.2 (right panel) illustrates the decision facing the hedge fund managers in $\mathcal{H}$. For $\alpha_1 \geq \alpha_2$, the utility of asset 1 (solid) is always greater than that of asset 2 (solid red). As $\alpha_2$ rises above $\alpha_1$ (dashed red), only the funds in the middle of $\mathcal{H}$ continue to prefer asset 1. These are the funds that benefit the most from the distribution of asset 1. For all other funds, the possibility of earning a performance fee is either very high or very low, so the shape of the distribution is not as valuable to the manager as the extra expected return. As $\alpha_2$ continues to rise (dotted red), eventually all funds in $\mathcal{H}$ will prefer asset 2 to asset 1.

Corollary 3 establishes an analogous result to Corollary 2 for funds in region $\mathcal{H}$. 

91
The demand for assets 1 and 2 is determined by the distribution of fund wealth, $\tilde{W}$, and the difference between $\alpha_1$ and $\alpha_2$. For $\alpha_1 \geq \alpha_2$, all funds in region $\mathcal{H}$ will prefer asset 1, as funds in this region want to “reach” to earn a performance fee. For $\alpha_1 < \alpha_2$, if they are sufficiently far apart, all funds will prefer asset 2. If the expected returns are similar, however, the funds in the middle of region $\mathcal{H}$ will prefer asset 1 to asset 2, while those in the tails will prefer asset 2. In this case, if there is large mass in the tails, then the demand for asset 1 will be relatively small from region $\mathcal{H}$ funds.

**Figure 3.2: Hedge Fund Manager Utility**

![Diagram showing utility curves for different wealth levels and asset preferences.](image)

*Note:* This figure shows the regions in which hedge fund managers will prefer asset 1 to asset 2, and vice versa, in region $\mathcal{L}$ (left) and $\mathcal{H}$ (right) for various values of $\alpha_2$. Black (red) lines plot the utility with asset 1 (asset 2). The values of $\alpha_2$ relative to $\alpha_1$ are detailed in the top left corner of each plot. Managers prefer asset 1 when its utility is higher than that of asset 2 (i.e., the black line is above the red) and vice versa. The dashed vertical line in the left panel indicates the wealth level for which managers are indifferent between the two assets when $\alpha_1 = \alpha_2$. Intersections are plotted as black diamonds (left panel) and black dots (both panels). See Appendix A.2.2 for model parameters.
3.2.4 Equilibrium

As is clear from Corollaries 2 and 3, the equilibrium values of \( \alpha_1, \alpha_2 \) depend crucially on the distribution of \( \bar{W} \). If, for example, there is a large mass of funds in region \( \mathcal{H} \), there will be high demand on asset 1, holding expected returns equal. This demand pressure will push \( \alpha_1 \) down until markets clear. The funds in region \( \mathcal{R} \) simply prefer the higher return asset, so they act as a “buffer” to ensure market prices do not become too extreme.

**Assumption 3.** The distribution of \( \bar{W} \) contains no mass points, i.e., \( f(x) = 0 \) for all \( x \in [0, 1] \).

In characterizing conditions for equilibrium, Assumption 3 is necessary to guarantee that there are no mass points at critical points \((\bar{W}_L^*, \bar{W}_L^{**}, \bar{W}_H^*)\) in which funds in regions \( \mathcal{L} \) and \( \mathcal{H} \) are indifferent between assets. With this, we can formalize conditions under which an equilibrium exists in which the expected returns are equal.

**Theorem 1.** Given Assumptions 2 and 3, there exists an equilibrium in which \( \alpha = \alpha_1 = \alpha_2 \) if and only if

\[
\frac{1 + \theta}{3\theta} - F(\mathcal{R}) \leq F(\mathcal{L}_{low} \cup \mathcal{H}) \leq \frac{1 + \theta}{3\theta},
\]

(3.21)

where \( \mathcal{L}_{low} \equiv \left( \frac{\gamma}{1+\alpha+\sigma}, \frac{\gamma}{1+\alpha} \right) \).

Theorem 1 establishes the demand for asset 1 and formalizes the notion that funds in region \( \mathcal{R} \) act as a demand buffer. When expected returns are equal, funds in \( \mathcal{R} \) are indifferent between assets, while all funds in \( \mathcal{L}_{low} \) and \( \mathcal{H} \) strictly prefer asset 1 to asset 2. The residual supply (netting out mutual fund demand), \( \frac{1+\theta}{3\theta} \), must be met by either the funds that strictly prefer asset 1 \((\mathcal{L}_{low} \cup \mathcal{H})\) or some combination of these funds and a portion of funds that are indifferent \((\mathcal{R})\).
**Theorem 2.** Given Assumptions 2 and 3 and some \( \alpha_2 \). If \( F (L_{low} \cup H) > \frac{1+\theta}{3\theta} \), then, if an equilibrium exists, \( \alpha_1 < \alpha_2 \). Conversely, if \( F (L_{low} \cup H) + F (L) < \frac{1+\theta}{3\theta} \), then, if an equilibrium exists, \( \alpha_1 > \alpha_2 \).

Theorem 2 completes the characterization of the model’s equilibrium. If the demand for asset 1 from hedge funds is too large (i.e., the mass in \( L_{low} \cup H \) is too great), the expected return of the risky asset will be lower than the expected return of the less risky asset. In addition, hedge funds will hold more of these risky assets than mutual funds. This is because those funds are either concerned with earning a performance fee or taking risk to surpass the liquidation threshold \( \gamma \). As such, they will become relatively risk-loving: holding expected returns equal, they will prefer the risky security, driving its expected return down.

If, on the other hand, the demand for asset 1 from hedge funds is too small (i.e., too many hedge funds in \( R \cup (L \setminus L_{low}) \)), the opposite pattern will result: equilibrium expected returns for the risky asset will be higher than that of the less risky asset. This is due, in part, to the fact that hedge funds are relatively risk averse in this case. They are either indifferent between assets or want to minimize the probability of falling below the threshold \( \gamma \), so will prefer the less risky asset.

Note that, in general, it is not guaranteed that an equilibrium exists in this economy. This is a result of the fact that region \( R \) funds will discretely shift to the higher return security. Thus, if there is a large mass in this region and the condition in Theorem 1 fails, setting \( \alpha_1 \neq \alpha_2 \) will lead to the entire mass of region \( R \) funds switching to the higher return asset. This discontinuity can result in a failure of the market clearing condition.

Incorporating hedge fund incentives into an equilibrium model of expected returns reveals the importance of two parameters, \( \theta \) and \( F (\cdot) \), in determining the expected return of risky and less risky securities. Figure 3.3 examines the impact these param-
Figure 3.3: Comparative Statics

\[ \alpha_1 / \alpha_2 \]

\[ F(H) / \theta \]

Note: This figure shows the equilibrium risk premium (left panel) and hedge fund holdings (right) over the entire state space for the total mass of hedge funds \( \theta \) and the fraction of hedge funds in \( \mathcal{H} \), \( F(H) / \theta \). The risk premium is expressed as the ratio of \( \alpha_1 \) to \( \alpha_2 \), while hedge fund holdings have been normalized by \( \theta \) for ease of comparison. See Appendix A.2.2 for model parametrization.

Parameters have on the risk premium (left panel) and asset allocations between hedge funds and mutual funds (right) over the entire state space of these parameters. As the mass of \( \mathcal{H} \) increases (a horizontal shift in both panels), the risk premium falls, while hedge fund holdings of the riskier asset rises. As hedge funds shift into region \( \mathcal{H} \), aggregate risk aversion declines. This generates increased demand for asset 1 from hedge funds, causing a fall in its expected return. A fall in the expected return mechanically leads to a reduction in mutual fund’s allocation to asset 1, so hedge funds increase their holdings.

A rise in the mass of hedge funds (a vertical shift) reduces both the risk premium and hedge fund holdings of the risky asset. When \( \theta \) is low, hedge funds do not have much of a price impact: even if their preferences are very different from that of mutual funds, they are not a significant enough portion of the market to shift
expected returns. As a result, for the few hedge funds that exist, the risky asset seems relatively underpriced, as mutual funds are more risk averse. Thus, hedge funds hold a large portion of the risky asset. As \( \theta \) increases, asset prices more closely reflect hedge fund preferences. As a result, the risk premium will fall, and, as a result, so too will hedge fund holdings.

According to this model, the size (\( \theta \)) and composition (\( F(\cdot) \)) of the hedge fund industry are critical to determining the risk-return tradeoff and allocations across agents. Further, to the extent these quantities vary, risk premia and allocations should vary as well. This begs the question: what is the impact of a shock to hedge fund performance on asset prices and allocations? A positive shock both shifts hedge funds towards their high-water marks (i.e., \( F(\mathcal{H}) \) increases) and increases the size of the hedge fund industry (i.e., \( \theta \) increases). This maps to a move in the northeast in Figure 3.3. This unambiguously reduces the risk premium, as hedge funds become less risk-averse on aggregate and more “marginal.” However, the impact on allocations is unclear. Holding prices fixed, hedge funds increase their holdings of the risky asset due to the reduction in aggregate risk aversion. The increase in the size of the industry, though, means that prices will not be fixed. As \( \theta \) increases, prices more reflect the preferences of hedge funds, leading to a reduction in investment in the risky security. These two have opposing effects.

The impact of a negative shock is less clear. While a negative shock will shift funds away from their HWMs (i.e., out of \( \mathcal{H} \)), a sufficiently large negative shock will shift funds below the liquidation boundary (i.e., into \( \mathcal{L}_{low} \)), as well as reduce the size of the industry (i.e., a reduction in \( \theta \)). Thus, a sufficiently large negative shock will

\[ \text{In the absence of hedge funds, the premium on asset 1 in this model is 9\%.} \]

\[ \text{The increase in the size of the hedge fund industry occurs through both direct and (unmodeled) indirect channels. Directly, positive performance increases the overall AUM of existing hedge funds. Indirectly, positive performance induces entry of new hedge funds and net inflows to existing hedge funds (Chevalier and Ellison, 1997).} \]
have a similar impact as moving southeast in Figure 3.3. While, strictly speaking, the mass of funds in region $\mathcal{H}$ will fall, this will be accompanied by a large increase in the mass of funds in region $\mathcal{L}_{low}$. Following this shock, hedge funds will increase investment in the risky asset. This is because aggregate hedge fund risk aversion decreases, but their impact on prices decreases. However, the impact on prices is ambiguous. Both aggregate hedge fund risk aversion and the size of the hedge fund industry decrease. Following this shock, there is higher demand for the risky security from fewer agents. Whether this leads to an increase or decrease in risk premia is unclear.

The remainder of this chapter tests the predictions of this model empirically. Specifically, do funds close to or very far from their high-water marks engage in selective risk-taking? What is the form of this risk-taking? Does this have an impact on equilibrium market quantities?

### 3.3 Data

This chapter makes use of hedge fund and management company data, as well as market quantities, to test the implications developed in Section 3.2. The fund-level data come from Morningstar CISDM, while the company-level analysis aggregates and merges these data with the Thomson Reuters S34 institutional holdings table. Individual security data come from CRSP and the standard risk factors come from Ken French’s website.\footnote{http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html} Betting against beta (BAB) comes from the AQR data library.\footnote{https://www.aqr.com/Insights/Datasets} Data on aggregate hedge fund AUM and performance are from BarclayHedge.
3.3.1 Morningstar CISDM

Morningstar CISDM contains a monthly time series of self-reported hedge fund net-of-fee returns and AUM at the fund level. It also contains a snapshot of fund characteristics, including fund style, fee structure, redemption and lockup notices, and manager and management company information. The data is split into “live” and “dead” funds, with funds being classified as dead if they have stopped reporting for 8-10 months.\(^\text{17}\) The data begin in January 1994 and end in December 2018.

**Table 3.1:** Number of Funds Remaining After Each Filter

<table>
<thead>
<tr>
<th>Filter</th>
<th>Live</th>
<th>Dead</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td>3,269</td>
<td>13,026</td>
<td>16,295</td>
</tr>
<tr>
<td>Report both returns and AUM</td>
<td>1,576</td>
<td>9,526</td>
<td>11,102</td>
</tr>
<tr>
<td>Continuous returns</td>
<td>1,407</td>
<td>7,769</td>
<td>9,176</td>
</tr>
<tr>
<td>Yearend assets</td>
<td>519</td>
<td>4,611</td>
<td>5,130</td>
</tr>
<tr>
<td>At least 12 obs</td>
<td>498</td>
<td>4,267</td>
<td>4,765</td>
</tr>
<tr>
<td>Report fees</td>
<td>443</td>
<td>3,805</td>
<td>4,248</td>
</tr>
</tbody>
</table>

*Note:* This table lists the number of funds remaining in the Morningstar CISDM database after iteratively applying the filters described. Funds considered to be ”dead” have ceased reporting for 8-10 months.

Overall, there are 3,269 funds in the live database and 13,026 funds in the dead database. However, many of these funds have incomplete data (return gaps, no AUM data, etc.). Several filters are imposed to eliminate these funds. First, all returns greater than 9999 are set to NA per the CISDM documentation that these observations are missing.\(^\text{18}\) Second, assets less than 10000 are set to NA, as these

---

\(^{17}\)In practice, there may be a variety of reasons for a fund to stop reporting. Most obviously, funds may stop reporting because of distress or poor return histories. However, funds that are closed to new investors may also cease reporting, as they do not benefit from publicizing their returns. Getmansky et al. (2004) estimate that 89% of funds that cease reporting in a similar database (Lipper TASS) can plausibly be attributed to liquidation. Using manually collected data of funds that cease reporting, Edelman et al. (2013) find a negligible delisting bias.

\(^{18}\)There are no greatly significant outliers after applying this filter. All results are robust to more aggressive filtering of the return data.
are likely data errors. AUM is interpolated for sequences of missing data of no more than 6 consecutive months.\textsuperscript{19} Third, after imposing these data corrections and interpolating AUM, funds most continuously report returns (i.e., no gaps between two report dates). Fourth, funds must report AUM at (i) the first reporting date, (ii) the last reporting data, and (iii) every year-end in between. Fifth, funds must have at least 12 continuous months of reporting. Sixth, funds that report 0 for both management and performance fees are dropped. In addition to these filters, funds that do not report hurdle rates or whether they have high-water marks are assumed to not have them. This reduces the sample to 320,056 fund-months and 4,248 funds, of which 443 are live. Table 3.1 displays the number of funds in the sample after each filter has been applied in turn.

Following Patton and Ramadorai (2013), I categorize these funds into ten different strategies: Security Selection, Global Macro, Relative Value, Directional Traders, Funds of Funds, Multi-Process, Emerging Markets, Fixed Income, and Other (including missing). 29\% of funds fall into the Security Selection strategy, while 23\% are funds of funds. These percentages are similar across live and defunct funds.

\subsection*{3.3.2 Thomson Reuters S34 Institutional Holdings}

As of 1978, all institutional investment managers overseeing at least $100\text{M}$ are required to disclose holdings of certain securities quarterly within 45 days of the quarter end via Form 13-F. These securities include US equities, equity options, convertible bonds, and shares of closed-end investment companies in which the manager has an investment that is greater than either $200,000 or 10,000 shares. Managers are not required to disclose short positions.

The Thomson Reuters S34 table contains the holdings from all filing institutions

\textsuperscript{19}The procedure to interpolate AUM follows Patton and Weller (2019).
since 1980. Because the holdings are disclosed at the management company level, merging this data with CISDM requires matching the management company names between the two datasets. The procedure, detailed in Appendix A.2.3, produces 485 management companies in both CISDM and Thomson Reuters.\textsuperscript{20}

After merging these databases, the last step consists of identifying the asset managers whose primary line of business is operating hedge funds through form ADV filings (see Brunnermeier and Nagel, 2004; Griffin and Xu, 2009; Jiao et al., 2016). Specifically, a company is classified as a hedge fund company if it has more than 50% of its investment listed as “[other] pooled investment vehicles” or more than 50% of its clients listed as “high net-worth individuals.” In addition, institutions must also report that they charge performance fees. If an institution has not filed a Form ADV, I manually check the website and/or other web searches to identify whether it is a hedge fund company.

Of the 1,945 management companies in the CISDM database, 485 of these are found in the Thomson Reuters data. 322 are classified as hedge funds. All results are qualitatively similar regardless of filtering based on ADV filings. Due to the limited sample size, all results are shown for the full sample of the 485 management companies.

\section*{3.4 Methodology}

This section details the construction of the main variables, including the process of imputing unobserved high-water marks, measures of fund risk, proxies for the distribution of hedge fund wealth, and market outcomes.

\textsuperscript{20}Note that not all 1,945 CISDM companies are matched due to the restriction that only investment managers overseeing $100M are required to file Form 13-F.
3.4.1 Estimation of High-Water Mark Distance

The returns reported to the CISDM database are *net-of-fee* returns. However, high-water marks and performance fees can only be computed from *gross-of-fee* returns. Two methodologies are employed to compute gross returns, high-water marks per investor, and other performance statistics: those of Agarwal et al. (2009) and Feng et al. (2011).\textsuperscript{21} Both methodologies rely on fairly similar assumptions, while the Feng et al. (2011) method allows for the computation of these quantities at a monthly frequency by accruing performance fees each month.\textsuperscript{22}

In general, the series produced by both methods are fairly similar. However, the Feng et al. (2011) methodology relies on flows computed monthly, which are known to be poorly estimated (Goetzmann et al., 2003). This results in noisier estimates of high-water marks and gross returns. As altering portfolio and risk composition is costly, the mechanisms described in Section 3.2 are likely to operate on lower frequencies. Thus, for the analysis of fund-specific behavior, the Agarwal et al. (2009) procedure is used. Given the relatively short sample (1994-2018), a higher frequency is preferred for an analysis of asset prices.\textsuperscript{23} To examine the impact on asset prices, therefore, the monthly estimation of Feng et al. (2011) is used.\textsuperscript{24}

Both procedures work by reinvesting all performance fees into the fund as the manager’s ownership, with the remaining portion assigned to outside investment.

\textsuperscript{21}See Appendix A of Agarwal et al. (2009) and the Appendix of Feng et al. (2011) for details on these procedures.

\textsuperscript{22}Among others, these procedures assume flows follow a first-in, first-out (FIFO) policy and occur at the end of each time period (yearly for Agarwal et al. (2009), monthly for Feng et al. (2011)).

\textsuperscript{23}Estimated magnitudes are similar when using an annual frequency.

\textsuperscript{24}When computing monthly gross returns, hedge fund returns are unsmoothed using the MA(2) procedure of Getmansky et al. (2004), subject to the constraint that $1 \geq \theta_0 \geq \theta_1 \geq \theta_2$. The problem of return autocorrelation is less prevalent in annualized returns, so raw returns are used when computing annual gross returns.
With a very volatile return and AUM history, it is possible for this outside investment to become negative.\textsuperscript{25} This occurs for 278 (400) funds when using the Agarwal et al. (2009) (Feng et al. (2011)) procedure. These funds are dropped from the sample. In this sense, the funds considered are restricted to those that have “well-behaved” histories.

As described in Section 3.2, the key determinant of a hedge fund’s risk appetite arising from its contract structure is its distance to high-water mark, the minimum required return that a fund must achieve before it can earn a performance fee. In practice, these high-water marks are set per investor rather than for the entire fund. This induces a slight change of notation from the previous section. Let $N_{i,j,t}$ and $H_{i,j,t}$ denote the NAV and HWM of investor $i$ in fund $j$ at time $t$, respectively. Further let $g_j$ be the fund’s hurdle rate. The minimum net return fund $i$ needs to achieve a performance fee on investor $j$’s investment at time $t$ is given by

$$R_{i,j,t}^{\text{perf}} = \frac{(1 + g_j) H_{i,j,t}}{N_{i,j,t}} - 1.$$ \hfill (3.22)

Because investors can enter the fund at different times, they can (and, in general, will) have different asset values and high-water marks. This presents an aggregation issue. While the ideal specification has the interpretation that, “if a fund achieves a return greater than some threshold, it will earn a performance fee on its entire AUM.” However, since investors have different levels of $R_{i,j,t}^{\text{perf}}$, such a specification is not possible. As an approximation to this quantity, a fund-level HWM distance variable is constructed as a value-weighted (by investor value) average of $R_{i,j,t}^{\text{perf}}$ across

\textsuperscript{25}This can occur following an 80\% outflow (over the course of several years) after a manager has accrued ownership of over 20\% of the fund.
investors. Formally, this is defined as

\[ R_{j,t}^{perf} = \frac{\sum_i N_{i,j,t} R_{i,j,t}^{perf}}{\sum_i N_{i,j,t}}. \] (3.23)

A similar value-weighting procedure (by fund AUM) is used to aggregate \( R_{j,t}^{perf} \) to the management company level.

**Table 3.2: Summary Statistics of Fund Characteristics**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>25(^{th})</th>
<th>Median</th>
<th>75(^{th})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hurdle Rate</td>
<td>4.248</td>
<td>0.069</td>
<td>0.253</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-water Mark</td>
<td>4.248</td>
<td>0.755</td>
<td>0.430</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Fee (%)</td>
<td>4.248</td>
<td>16.552</td>
<td>7.287</td>
<td>10</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Management Fee (%)</td>
<td>4.073</td>
<td>1.433</td>
<td>0.578</td>
<td>1.000</td>
<td>1.500</td>
<td>2.000</td>
</tr>
<tr>
<td>AUM ($M)</td>
<td>4.236</td>
<td>111.334</td>
<td>337.075</td>
<td>8.208</td>
<td>27.186</td>
<td>89.586</td>
</tr>
<tr>
<td>Age (years)</td>
<td>4.236</td>
<td>6.207</td>
<td>4.549</td>
<td>2.749</td>
<td>4.999</td>
<td>8.414</td>
</tr>
<tr>
<td>Manager Ownership (%)</td>
<td>3.958</td>
<td>4.172</td>
<td>6.147</td>
<td>0.295</td>
<td>1.796</td>
<td>5.402</td>
</tr>
<tr>
<td>Net Return (%)</td>
<td>4.236</td>
<td>6.459</td>
<td>12.234</td>
<td>1.630</td>
<td>5.053</td>
<td>9.300</td>
</tr>
<tr>
<td>Net Return Volatility ( \sigma^F ) (%)</td>
<td>4.236</td>
<td>11.022</td>
<td>10.505</td>
<td>4.906</td>
<td>7.998</td>
<td>13.508</td>
</tr>
<tr>
<td>Net Return Sharpe</td>
<td>4.236</td>
<td>0.539</td>
<td>0.717</td>
<td>-0.035</td>
<td>0.485</td>
<td>1.042</td>
</tr>
<tr>
<td>Gross Return Sharpe</td>
<td>3.953</td>
<td>0.647</td>
<td>0.785</td>
<td>-0.021</td>
<td>0.623</td>
<td>1.339</td>
</tr>
<tr>
<td>VW ( R^{perf} ) (%)</td>
<td>3.958</td>
<td>6.916</td>
<td>16.870</td>
<td>0.000</td>
<td>1.293</td>
<td>6.458</td>
</tr>
<tr>
<td>EW ( R^{perf} ) (%)</td>
<td>3.958</td>
<td>6.654</td>
<td>16.749</td>
<td>0.000</td>
<td>1.262</td>
<td>6.134</td>
</tr>
<tr>
<td>Min ( R^{perf} ) (%)</td>
<td>3.958</td>
<td>4.129</td>
<td>12.400</td>
<td>0.000</td>
<td>0.000</td>
<td>3.050</td>
</tr>
<tr>
<td>Max ( R^{perf} ) (%)</td>
<td>3.958</td>
<td>8.145</td>
<td>19.732</td>
<td>0.000</td>
<td>1.520</td>
<td>7.703</td>
</tr>
<tr>
<td>Net Return ( \beta^{mkt} )</td>
<td>4.246</td>
<td>0.318</td>
<td>0.464</td>
<td>0.045</td>
<td>0.224</td>
<td>0.506</td>
</tr>
<tr>
<td>Net Return ( \alpha ) (%)</td>
<td>4.246</td>
<td>2.520</td>
<td>19.371</td>
<td>-3.452</td>
<td>1.190</td>
<td>6.592</td>
</tr>
<tr>
<td>Net Return ( \sigma^e ) (%)</td>
<td>4.246</td>
<td>6.973</td>
<td>7.673</td>
<td>2.978</td>
<td>4.926</td>
<td>8.298</td>
</tr>
</tbody>
</table>

*Note: This table reports descriptive statistics of all fund characteristics in the sample. Unless otherwise stated, all statistics are of the time series average per fund. Age is the number of years since the first and last reporting. High-water Mark and Hurdle Rate are indicator variables equal to 1 if a fund has those respective provisions. Manager ownership is the percent of the fund owned by the manager. Net return and gross return are the returns net- and gross-of-fee, respectively. VW \( R^{perf} \) is the value-weighted distance to an investor’s high-water mark. EW, Min, and Max denote the equal-weighted average, minimum, and maximum over fund investors. Manager ownership, gross returns, and HWM distances are based on the algorithm of Agarwal et al. (2009). Betas, alphas, and idiosyncratic volatilities are computed using monthly returns each year with the Fama-French 3-factor model. Alpha and idiosyncratic volatility are annualized. Returns, volatility, Sharpe ratios, \( R^{perf} \), beta, alpha, and idiosyncratic volatility have been winsorized at the 5\% level.*

\( ^{26} \)Alternative specifications, such as the minimum, maximum, and equal-weighted average distance have been explored as well, and summary statistics are presented in Table 3.2.
Table 3.2 presents summary statistics of the variables described and various fund characteristics. 75.5% of funds have high-water mark provisions, and the average performance and management fees are 16.6% and 1.4%, respectively. There is a substantial right tail in fund AUM, with the mean (median) fund having an AUM of $111M ($27.2M). On average, there is roughly 6 years of data for each fund. The average value-weighted high-water mark distance is 6.9%, but this exhibits substantial variation. Equal-weighted, minimum, and maximum HWM distances behave similarly. Funds earn 6.5% and 6.9% in net and gross returns on average, with gross (net) sharpe ratios of 0.6 (0.5) on average. Funds earn a risk-adjusted annual alpha of 2.5% on average, and have minimal market exposure, with an average market beta of 0.32.

3.4.2 Measures of Risk

The model presented in Section 3.2 makes two key predictions: funds close to or very far from their HWMs will increase the risk of their portfolio, and this will take the form of investment in riskier securities. To empirically examine these predictions and further test the risk composition of funds, the analysis makes use of four risk measures. The simplest of these measures is the volatility of net returns at the fund level computed annually, denoted $\sigma^{F}_{j,t}$.

Using the holdings data of the management company, the remaining three measures examine whether and along what lines funds alter the specific securities they invest in. The first of these is the value-weighted average volatility of the securities a company is holding long. Stock volatility for each CUSIP is computed yearly using daily return data, denoted $\sigma^{V}_{k,t}$. This is aggregated to an average long volatility

104
weighted by position size given by

$$\bar{\sigma}^L_{j,t} \equiv \sum_{k \in \Psi_j} \theta_{j,k,t} \sigma_{k,t}, \quad (3.24)$$

where $\theta_{j,k,t}$ is the proportion of the long portfolio that is invested in security $k$ (so $\sum_k \theta_{j,k,t} = 1$) defined for all securities owned by company $j$, $\Psi_j$. This captures the average riskiness of the securities in a given company’s long portfolio. Importantly, this measure is independent of leverage: the weights $\theta_{j,k,t}$ sum to 1 for each company $j$ in year $t$. Simple leveraging of an existing portfolio up or down based on these incentives would not be captured by this variable.

Two variations of this measure are worth discussion. The first simply considers the change in average long volatility, $\Delta \bar{\sigma}^L_{j,t} \equiv \bar{\sigma}^L_{j,t} - \bar{\sigma}^L_{j,t-1}$. This can help determine whether funds closer to or very far from their HWMs increase the average long volatility relative to previous years. However, this measure cannot determine whether funds chase volatility; that is, do funds that are relatively risk-loving invest in securities that they expect to be riskier in the future? Assuming conditional volatility is a martingale (so that the expectation of future volatility is equal to current volatility, $\mathbb{E}[\sigma_{k,t+1} | \mathcal{H}_t] = \sigma_{k,t}$) allows for a simple definition of the “stale” change in volatility, $\Delta^S \bar{\sigma}^L_{j,t}$. This is given by

$$\Delta^S \bar{\sigma}^L_{j,t} \equiv \sum_{k \in \Psi_j} \mathbb{E}[\sigma_{k,t} | \mathcal{H}_{t-1}] \theta_{j,k,t} - \sigma_{k,t-1} \theta_{j,k,t-1} = \sum_{k \in \Psi_j} \sigma_{k,t-1} \Delta \theta_{j,k,t}, \quad (3.25)$$

27This assumption seems to be reasonably innocuous. Using a variety of volatility forecasting models, Engle and Patton (2001) estimate the autocorrelation parameter of the daily conditional volatility of the Dow Jones Industrial Index to be between 0.93 and 0.95. Patton (2011) make this assumption to compare various volatility models. Correcting for the errors-in-variables problem using the instrumental variables approach of Hansen and Lunde (2014), the average autocorrelation of monthly volatility for all stocks in the CRSP universe (with at least 60 months of daily data) is 0.82 (25th and 75th percentiles of 0.72 and 0.94, respectively).
where $H_t$ is the entire history of previous returns. Simply, this is the average change in long position weighted by lagged volatility. A value greater than 0 indicates that companies selected into securities that had a higher volatility during the previous period. Under the assumption that conditional volatility is a martingale, this doubles as indicating that companies selected into securities that are expected to be more volatile in the next period.

The final two measures are the average market beta and idiosyncratic volatility in the fund’s long portfolio. For each stock in the CRSP universe, these are computed via the regression

$$R_{i,t} - r_t^f = \alpha_{i,t} + \beta_{i,t}^{mkt} mkt_t + \beta_{i,t}^{smb} smb_t + \beta_{i,t}^{hml} hml_t + \epsilon_{i,t}$$ (3.26)

using daily data for each year that stock $i$ appears in the sample.\textsuperscript{28} Idiosyncratic volatility is defined as $\sigma_{i,t}^* \equiv \sqrt{\mathbb{E} [\epsilon_{i,t}^2]}$. $\beta_{i,t}^{mkt}$ and $\sigma_{i,t}^*$ are aggregated to the company level via a value-weighted average by position size, and $\sigma_{i,t}^*$ is annualized.

**Table 3.3:** Summary Statistics of Company Characteristics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>N</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>25\textsuperscript{th}</th>
<th>Median</th>
<th>75\textsuperscript{th}</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUM ($M$)</td>
<td>482</td>
<td>334.167</td>
<td>839.709</td>
<td>32.957</td>
<td>99.699</td>
<td>301.429</td>
</tr>
<tr>
<td>Long Portfolio Size ($M$)</td>
<td>482</td>
<td>3,984.608</td>
<td>29,174.470</td>
<td>115.672</td>
<td>292.724</td>
<td>1,349.734</td>
</tr>
<tr>
<td>Long Return (%)</td>
<td>482</td>
<td>12.841</td>
<td>16.597</td>
<td>6.809</td>
<td>11.848</td>
<td>17.680</td>
</tr>
<tr>
<td>Average Long Volatility $\bar{\sigma}^L$ (%)</td>
<td>482</td>
<td>38.179</td>
<td>13.756</td>
<td>30.610</td>
<td>35.830</td>
<td>44.057</td>
</tr>
<tr>
<td>Long Portfolio $\beta^{mkt}$</td>
<td>482</td>
<td>1.023</td>
<td>0.170</td>
<td>0.954</td>
<td>1.032</td>
<td>1.107</td>
</tr>
<tr>
<td>Long Portfolio $\alpha$ (%)</td>
<td>482</td>
<td>5.395</td>
<td>9.221</td>
<td>1.827</td>
<td>5.040</td>
<td>8.319</td>
</tr>
<tr>
<td>Long Portfolio $\sigma^*$ (%)</td>
<td>482</td>
<td>32.539</td>
<td>12.185</td>
<td>25.131</td>
<td>30.438</td>
<td>38.451</td>
</tr>
</tbody>
</table>

*Note:* This table reports descriptive statistics of all hedge fund management companies in the sample. Unless otherwise stated, all statistics are of the time series average per company. AUM is the sum of the AUM of all funds underneath a given management company. Long portfolio size is the market value of a company’s long portfolio. Average long volatility is the value-weighted average volatility of the assets in a company’s long portfolio. Market beta, alpha, and idiosyncratic volatility is computed with respect to the Fama-French three-factor model. Alpha and idiosyncratic volatility are annualized.

\textsuperscript{28}Estimates are dropped for stocks that have less than 100 observations.
Table 3.3 presents summary statistics of these and other variables of the management companies in the sample. Compared to the fund characteristics, long-only returns and market beta are higher than that of fund gross returns (12.8% vs. 6.9% and 1.01 vs. 0.32), indicating funds engage in some hedging of market risk and this hedging is costly. Idiosyncratic volatility is similar between long and net returns. The average long volatility is 38.2% with a standard deviation of 13.8%. AUM is an order of magnitude smaller than long portfolio size due to an aggregation issue: the long portfolio size reflects the holdings of all hedge fund and non-hedge fund activities of a given management company, while the AUM reflects only the size of reporting hedge funds. In practice, not all management companies solely operate hedge funds, nor do all hedge funds within a given management company report to CISDM and satisfy the filters applied.\textsuperscript{29}

\subsection*{3.4.3 Proxies of Hedge Fund Wealth Distribution and Market Quantities}

The remainder of the analysis concerns the equilibrium impact of hedge funds on aggregate market quantities. To address this, it is imperative to properly define aggregate measures capturing the incentive of hedge funds to pursue risk-seeking strategies. As described in Section 3.2, the hedge funds that are hypothesized to pursue excessive risk taking are those close to or very far from their HWMs. Funds in the middle region have little incentive to do so.

To obtain a higher frequency time series, the Feng et al. (2011) procedure is used to obtain monthly gross returns. Two simple nonparametric proxies are employed:

\textsuperscript{29}This makes the analysis of fund leverage difficult, if not impossible. As data on company AUM is only partially observed, while holdings are completely observed, leverage will be grossly overstated. When examining company leverage, there is a very fat tail, largely consisting of unrealistic values.
$OTM_t$ and $ITM_t$. $OTM_t$ (out of the money) computes the percent of funds in the left tail of the HWM distance distribution (i.e., funds whose performance fee is deep out of the money), while $ITM_t$ (in the money) computes the percent in the right tail. Both are constructed as follows. First, the number of funds at or above their HWMs are computed at each time, denoted $N_t^0$. Then, deciles of $R_{j,t}^{perf}$ are computed using the full sample of funds over time, with each fund having weight $\frac{1}{N_t}$, where $N_t$ is the total number of funds at time $t$. These deciles are then used to determine the number of funds falling into each decile per period, denoted $N_t^d$ for $d = 1, \ldots, 10$. Finally, $N_t^0$ is divided by 10 to compare with counts in other deciles.

In general, the counts $N_t^d$ will be highly seasonal. Given that a fund is at its HWM in January $(m = 1)$, the probability of remaining at or above its HWM over the next $m$ months is approximately given by the probability of achieving a positive return over that time period. If (log) returns are iid normal, this is given by

$$P \left( R_{j,m}^{perf} \leq 0 \mid R_{j,1}^{perf} = 0 \right) \approx \Phi \left( \frac{\sqrt{m} \mu_j}{\sigma_j} \right), \quad (3.27)$$

where $\mu_j$ and $\sigma_j$ are fund $j$’s mean return and standard deviation, respectively. For $\mu_j > 0$, the probability that a given fund is above its HWM ($R_{j,m}^{perf} < 0$) is increasing in the number of months $m$. To account for this seasonality, funds are assumed to be homogeneous and $z = \frac{\mu}{\sigma}$ is assumed to be a latent parameter. Denoting by $\tilde{N}_m^d$
the mean number of funds in decile $d = 0, \ldots, 10$ in month $m = 1, \ldots, 12$, $z$ can be estimated by minimizing the mean squared error

$$MSE(z) = \frac{1}{12} \sum_{m=1}^{12} \left( \frac{N_m^0}{\sum_{d=0}^{10} N_m^d} - \Phi(\sqrt{m}z) \right)^2. \quad (3.28)$$

The resulting estimate of $\hat{z} = 0.093$ has a mean absolute deviation of 2.1%, and a maximum absolute deviation of 7.5%. Finally, $N_t^0$ is normalized by $\frac{1-\Phi(\sqrt{m}\hat{z})}{\Phi(\sqrt{m}\hat{z})}$ to adjust for seasonality. The adjusted counts fail to reject the null of no seasonality at the 10% level.\footnote{Regressions of each count $N_m^d$ for $d = 0, \ldots, 10$ on a set of month fixed effects yields a minimum $F$-statistic of 1.46 across the 11 variables, with an associated $p$-value of 0.145.}

**Figure 3.4:** Proxies of Hedge Fund Wealth Distribution

![Graph showing ITM and OTM proxies](image)

*Note:* This plot presents the time series of nonparametric proxies $OTM_t$ (red, dashed) and $ITM_t$ (black, solid). $OTM_t$ is the normalized percent of funds in the bottom decile of the HWM distance distribution, while $ITM_t$ is the normalized percent of funds in the top decile. Both series have been centered and scaled to have mean zero and unit variance. Shaded regions indicate NBER recessions.

Using the normalized counts, the fraction of funds in the bottom decile is given by

$$OTM_t \equiv \frac{N_t^{10}}{\sum_{d=0}^{10} N_t^d}, \quad (3.29)$$

the year. Thus, estimating $z$ from the data will drastically overestimate the probability of a fund being at or above its HWM.
where, as before, $N_t^d$ is the (deseasoned) number of funds in decile $d = 0, \ldots, 10$ at time $t$. Similarly, the fraction of funds in the top decile is given by

$$ITM_t \equiv \frac{N_t^0}{\sum_{d=0}^{10} N_t^d}. \quad (3.30)$$

These measures map to the model described in different ways. $OTM_t$ maps most easily to the mass of funds in region $L_{low}$, while $ITM_t$ captures the mass in $H$. While funds in both regions should be relatively less risk averse, as discussed at the end of Section 3.2, there are reasons to expect asymmetric behavior. An increase in $ITM_t$ decreases aggregate risk aversion while increasing the size of the hedge fund industry. An increase in $OTM_t$, on the other hand, decreases aggregate risk aversion while decreasing the size of the industry. Thus, both variables are considered separately.

Figure 3.4 presents the time series of the proxies described in Eqs. (3.29) and (3.30). Due to the normalizations discussed above, the level of these variables do not have much of an interpretation, so they have been centered and scaled to have mean zero and unit variance. Generally speaking, $ITM_t$ is prone to higher frequency fluctuations, while $OTM_t$ has rare but severe spikes. It is interesting to note that, even while the hedge fund sector was booming in the pre-2000 period, there were still instances of large spikes in $OTM_t$. The correlation between the two measures is -0.31.

Using these two proxies for the aggregate hedge fund wealth distribution, the volatility of hedge funds’ underlying portfolios are examined. If funds are adjusting their long volatility based on HWM distance individually, they too should do so in aggregate. Aggregate long volatility is calculated as the value-weighted (based on AUM) average of $\bar{\sigma}_{f,t}^L$ and is denoted $\bar{\sigma}_t^L$. Further, this measure is normalized by the value-weighted volatility of all securities in the CRSP universe, denoted $\bar{\sigma}_t$. This
measure is greater than 1 if hedge funds are holding securities that are, on average, riskier than the average CRSP security.

If risk composition is adjusted through investing in securities with different market betas, the market risk premia should respond as well. To test this, the slope of the security market line and the betting against beta (BAB) anomaly are examined. Estimates of the security market line follow the traditional methodology of Black et al. (1972). Specifically, betas are estimated using daily data each month for each stock (with at least 17 observations) in the CRSP universe. Stocks are then sorted into 10 deciles based on their lagged beta. Value-weighted daily excess returns for each decile are then computed, and betas of each portfolio are computed monthly from 1994 to 2018. This results in 298 months × 10 portfolios = 2980 betas. The slope of the security market line is given by a regression of portfolio returns on estimated betas.

Figure 3.5 plots the time series of both of these variables. Focusing on the left panel, hedge funds hold securities that are riskier than the average CRSP security throughout the entire sample (1996-2018). The magnitude exhibits significant time variation. In 1996, the average security held by hedge funds was 70% more volatile than the average CRSP security, compared to only 7.9% more volatile in 2018. Interestingly, hedge funds invested in less risky securities throughout the financial crisis, providing some evidence of a market timing effect.

The slope of the security market line has a moderate correlation of 0.42 (p-value 0.04) with the market return. However, the security market line is clearly too flat over time. Throughout the sample, the mean slope is actually negative (-4.5%), compared to a mean market return of 7.5%. Nevertheless, the analysis is concerned with whether the security market flattens as hedge funds demand riskier securities,
**Figure 3.5:** Aggregate Hedge Fund Risk and SML Estimates

![Graph showing aggregate hedge fund risk and SML estimates](image)

*Note:* This plot presents the ratio of quarterly average long volatility of the entire hedge fund sector normalized by the average volatility of CRSP securities (left panel) and yearly estimates of the security market line along with the market return (right panel). Average long volatility of the entire hedge fund sector is computed as the weighted average (by AUM) of each management company’s average long volatility. The average volatility of CRSP securities is computed as the value-weighted average of yearly standard deviations of daily returns. The security market line is estimated by sorting firms into deciles based on monthly betas of the previous month estimated using daily data. Value-weighted daily excess returns and betas for each decile are then computed and over the following month. Slope estimates are the estimates of the coefficient of these betas on excess returns in each year. Shaded regions indicate NBER recessions.

not whether the security market line and the market return are consistent.\(^{36}\)

### 3.5 Results

This section is organized into two parts. Individual and company level responses to HWM distance are presented first, while the impact of these responses on aggregate quantities follows.

\(^{36}\)In slight abuse of notation, I refer to a flattening of the SML as a decrease in the slope estimate.
3.5.1 Fund-Specific Results

The premise that hedge fund managers’ compensation incentives are relevant for equilibrium outcomes and portfolio allocation relies on the hypothesis that hedge fund performance exhibits significant variation both in the time series and cross section. Figure 3.6 plots the $25^{th}$, $50^{th}$, and $75^{th}$ quantiles of $R_{ij,t}^{perf}$ through time as well as the fraction of funds below their high-water marks ($R_{ij,t}^{perf} > 0$). Clearly, there is significant variation through time: 59.4% of funds were below their HWMs at the peak of the financial crisis, compared with 14.3% in 2003.\textsuperscript{37} Similarly, there is a wide variation in the cross sectional distribution. The interquartile range reaches a peak of 33.6% in 2008, compared to 4.8% in 2004. This time series is consistent with the explosion of the industry following widespread outperformance during the early 2000s, as well as the large crash in the hedge fund industry during the financial crisis.

The results in Section 3.2 suggest that the relationship between high-water mark distance and volatility is non-monotonic: funds in the left and right tails of the distribution are likely to exhibit the highest volatility. To capture this, funds are assigned into deciles or quintiles at time $t$ based on their high-water mark distance at the end of year $t - 1$, $R_{i,t-1}^{perf}$, denoted $Q_{i,t}$. Importantly, these deciles are formed using information only until the end of year $t - 1$, so do not incorporate any year $t$ information. Table 3.4 presents the HWM distance of each decile during the pre- and post-formation period (year $t - 1$ and $t$, respectively), as well as the volatility and change in volatility over that time. By construction, there is a monotonic relationship in the pre-formation HWM distance, and this relationship is reasonably maintained in

\textsuperscript{37}The probability that a fund is below its HWM is increasing in the time the fund has been in existence. This can explain the general upward trend in the percent of funds below their HWMs through time.
Figure 3.6: HWM Distance Through Time

Note: This plots the median (red solid), 25th, and 75th (red dashed) distance to a fund’s HWM through time, as well as the fraction of funds at their high-water marks (black). This plot is based on a sample of 2,747 funds from Morningstar CISDM. The distance to the fund’s HWM is the minimum required return for the fund to earn a performance fee on the average investors investment, estimated using the procedure of Agarwal et al. (2009). Shaded regions indicate NBER recessions.

the post-formation period. Funds in Decile 10 are extremely far from their HWMs; on average (median), they need to achieve a net return of 105% (77.8%) in order to reach their HWMs. These funds are clearly candidates for liquidation.

Volatility presents an interesting pattern. Volatility is high for funds in Decile 0 and then increases monotonically for funds in Deciles 3-10. Indeed, funds in Decile 0 have higher volatilities on average than those in Deciles 1-6. Despite this monotonic pattern in volatility of Deciles 3-10, funds in high deciles decrease their volatility significantly, while those in Decile 0 (1) increase their volatility by 0.75 (0.09) per-

---

38 Deciles 0, 2, and 5 are notable exceptions.

39 This implies the average (median) fund in Decile 10 has a value that is 48.8% (56.2%) of its high-water mark. These are certainly extreme values, but not out of the realm of possibility. For example, in 2008, Citadel, a hedge fund overseeing more than $30B in AUM, experienced losses in excess of 50% (i.e., a HWM distance of over 100%) in its two main funds, the Kensington and Wellington funds. This was followed by a 10-month ban on investor withdrawals, before high-water marks were eventually cleared over 3 years later, on January 17, 2012 (https://www.wsj.com/articles/citadels-ken-griffin-leaves-2008-tumble-far-behind-1438655887; https://dealbook.nytimes.com/2012/01/20/citadel-clears-its-high-water-mark/).


percentage points. This can be explained through mean reversion: fund distance to HWM increases as a consequence of high volatility the previous period. This high volatility is unlikely to be sustained, so volatility naturally decreases in the subsequent period. For this reason, it is important to consider the lagged level of return volatility when evaluating risk-taking.

<table>
<thead>
<tr>
<th>Table 3.4: Volatility by Distance Decile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decile</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

*Note:* This table lists the mean distance to a fund’s HWM in the pre-formation period (year $t - 1$), distance to a fund’s HWM in the post-formation period (year $t$), volatility, and level change in volatility per decile. Funds are assigned to deciles based on the distance to their HWMs at the end of the previous year. Decile 0 are for funds that are at their high watermarks, while funds in Decile 10 are furthest away from their HWM as of year-end $t - 1$. The top 5% of HWM distance has been winsorized.

Using these quantile assignments, the general specification to test for selective risk-taking in HWM distance quantiles is given by

$$y_{i,t} = \alpha + \sum_{q=0}^{n} \beta_q I [Q_{i,t} = q] + \gamma X_t + \delta Y_i + \lambda Z_{i,t} + \epsilon_{i,t}, \quad (3.31)$$

where $y_{i,t}$ is the variable of interest ($F_{j,t}$, $L_{j,t}$, $\beta_{j,t}$, $\sigma_{j,t}$, etc.). For analysis at the fund (company) level, deciles (quintiles) are used, and decile 5 (quintile 3) is omitted Eq. (3.31).\(^{40}\) $X_t$, $Y_i$, and $Z_{i,t}$ feature sets of time, fund, and time-fund specific controls.

\(^{40}\)Quintiles are used for company analyses due to the largely reduced sample size (4,248 hedge funds
respectively.

For analysis at the fund level, these sets of controls are given by

\[ X_t = \{ MktVol_t, MktRet_t \} \]
\[ Y_i = \{ MgmtFee_i, PerfFee_i, Strategy_i \} \]
\[ Z_{i,t} = \{ \log(AUM_{i,t-1}), Flow_{i,t-1}, \sigma^F_{i,t-1}, Age_{i,t}, NetReturn_{i,t-1}, MgrOwnership_{i,t-1} \} , \]

(3.32)

where \( MktVol_t \) is the yearly volatility of daily market returns, \( MktRet_t \) is the annualized market return in year \( t \), \( MgmtFee_i \) and \( PerfFee_i \) are the management and performance fees charged by funds (respectively), \( Strategy_i \) is a set of fund strategy fixed effects, \( AUM_{i,t-1} \) is the assets under management of fund \( i \) at \( t - 1 \), \( Flow_{i,t-1} \) is the flows into the fund between \( t - 2 \) and \( t - 1 \),\(^{41} \) \( Age_{i,t} \) is number of years fund \( i \) appears in the data between inception and time \( t \), \( NetReturn_{i,t-1} \) is the net return in year \( t - 1 \), and \( MgrOwnership_{i,t-1} \) is the percent of the fund owned by the fund manager at the end of year \( t - 1 \). For analysis at the company level, \( Strategy_i \) is omitted from \( Y_i \), and \( Age_{i,t} \) and \( MgrOwnership_{i,t-1} \) are replaced by lagged long volatility \( \sigma^L_{i,t-1} \), lagged market beta \( \beta^{mkt}_{j,t-1} \), and lagged idiosyncratic volatility \( \sigma^F_{i,t-1} \) in \( Z_{i,t} \). Controls that are statistically insignificant across all specifications within a table are omitted for brevity. \( p \)-values are computed under the assumption of a single hypothesis test (Harvey et al., 2016).

Table 3.5 presents estimates of Eq (3.31) with fund volatility \( \sigma^F_{j,t} \) as the dependent variable for funds that have high-water mark provisions.\(^{42} \) The coefficients on the

\[^{41}\text{Following Goetzmann et al. (2003), } Flow_{i,t} = \frac{AUM_{i,t} - AUM_{i,t-1}(1 + NetReturn_{i,t})}{AUM_{i,t-1}}. \]
\[^{42}\text{Funds without high-water mark provisions are always at a HWM distance of 0. As a result, they would always feature in Decile 0. Because Decile 0 already contains 63.9\% of the entire sample, parameters for this decile should be relatively well estimated without including these funds. Funds} \]
Table 3.5: Fund Return Volatility Regressions

<table>
<thead>
<tr>
<th></th>
<th>Fund Return Volatility $\sigma^F$</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decile 0</td>
<td>0.684*</td>
<td>0.341</td>
<td>0.858**</td>
<td>0.761**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.380)</td>
<td>(0.377)</td>
<td>(0.385)</td>
<td>(0.386)</td>
<td></td>
</tr>
<tr>
<td>Decile 1</td>
<td>0.415</td>
<td>0.493</td>
<td>0.174</td>
<td>0.226</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.505)</td>
<td>(0.498)</td>
<td>(0.516)</td>
<td>(0.511)</td>
<td></td>
</tr>
<tr>
<td>Decile 2</td>
<td>0.272</td>
<td>0.315</td>
<td>0.343</td>
<td>0.384</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.505)</td>
<td>(0.498)</td>
<td>(0.498)</td>
<td>(0.493)</td>
<td></td>
</tr>
<tr>
<td>Decile 3</td>
<td>0.017</td>
<td>0.060</td>
<td>0.302</td>
<td>0.326</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.510)</td>
<td>(0.503)</td>
<td>(0.495)</td>
<td>(0.491)</td>
<td></td>
</tr>
<tr>
<td>Decile 4</td>
<td>0.337</td>
<td>0.320</td>
<td>0.604</td>
<td>0.615</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.504)</td>
<td>(0.496)</td>
<td>(0.492)</td>
<td>(0.487)</td>
<td></td>
</tr>
<tr>
<td>Decile 6</td>
<td>0.186</td>
<td>0.121</td>
<td>0.381</td>
<td>0.360</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.498)</td>
<td>(0.491)</td>
<td>(0.486)</td>
<td>(0.481)</td>
<td></td>
</tr>
<tr>
<td>Decile 7</td>
<td>0.640</td>
<td>0.531</td>
<td>0.376</td>
<td>0.335</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.508)</td>
<td>(0.501)</td>
<td>(0.495)</td>
<td>(0.490)</td>
<td></td>
</tr>
<tr>
<td>Decile 8</td>
<td>1.128**</td>
<td>0.871*</td>
<td>0.177</td>
<td>0.048</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.508)</td>
<td>(0.501)</td>
<td>(0.497)</td>
<td>(0.493)</td>
<td></td>
</tr>
<tr>
<td>Decile 9</td>
<td>1.489***</td>
<td>1.146**</td>
<td>0.693</td>
<td>0.504</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.517)</td>
<td>(0.510)</td>
<td>(0.514)</td>
<td>(0.510)</td>
<td></td>
</tr>
<tr>
<td>Decile 10</td>
<td>3.768***</td>
<td>3.098***</td>
<td>0.968*</td>
<td>0.619</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.535)</td>
<td>(0.529)</td>
<td>(0.562)</td>
<td>(0.560)</td>
<td></td>
</tr>
<tr>
<td>Log(AUM)</td>
<td>$-0.217^{***}$</td>
<td>$-0.195^{***}$</td>
<td>$-0.344^{***}$</td>
<td>$-0.305^{***}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.041)</td>
<td>(0.099)</td>
<td>(0.101)</td>
<td></td>
</tr>
<tr>
<td>$\sigma^F_{t-1}$ (%)</td>
<td>0.599***</td>
<td>0.625***</td>
<td>0.195***</td>
<td>0.209***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.079***</td>
<td>0.070***</td>
<td>0.083**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.024)</td>
<td>(0.033)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NetReturn$_{t-1}$ (%)</td>
<td>0.066*</td>
<td>-0.004</td>
<td>-0.001</td>
<td>-0.009**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td></td>
</tr>
<tr>
<td>MgrOwnership$_{t-1}$ (%)</td>
<td>-0.007</td>
<td>-0.004</td>
<td>-0.050***</td>
<td>-0.050***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.011)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>Mgmt Fee (%)</td>
<td>0.469***</td>
<td>0.456***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.139)</td>
<td>(0.138)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Market Volatility (%)</td>
<td>0.088***</td>
<td></td>
<td>0.228***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.011)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Market Return (%)</td>
<td>-0.090***</td>
<td></td>
<td>-0.031***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This table presents regression coefficients of decile dummy variables based on the lag value-weighted HWM distance on a fund’s volatility over the next year. Decile 0 contains funds that are at their HWMs (distance of 0), while Decile 10 contains funds are furthest from their HWMs. Insignificant controls in all specifications are omitted. All variables are contemporaneous unless otherwise stated. The sample is annually from 1996 to 2018. Standard errors are presented in parentheses. *p<0.1; **p<0.05; ***p<0.01 under the assumption of a single hypothesis test.

Strategy FE?         | Y | Y | N | N
Time FE?             | N | Y | N | Y
Fund FE?             | N | N | Y | Y
Observations         | 13,906 | 13,906 | 14,231 | 14,231
$R^2$                | 0.492 | 0.507 | 0.676 | 0.683

Note: Without HWM provisions are thus omitted to prevent introducing heterogeneity that may arise if funds strategically select into having these provisions.
Decile $d$ indicator measures the impact of being in decile $d$ relative to being in Decile 5. The table presents broadly consistent evidence regarding the risk taking of hedge funds at or far below their HWMs. With the exception of column (2), the volatility of funds in Decile 0 is statistically significantly larger than their Decile 5 counterparts by 0.68, 0.86, and 0.76 percentage points, respectively. These funds are at their high-water marks, so their incentive to increase the volatility of their returns is high. These results are consistent with the notion that managers close to their high-water marks are encouraged to take risk more so than funds below.

A second result from these tests show that the effect at the bottom of the HWM distance distribution is even more pronounced: in specifications (1) and (2), funds in Deciles 8-10 have greater volatilities than the median fund, and this effect monotonically increases as the distance increases. Funds in Decile 8 have volatilities that are 1.13pp higher than the median fund, compared with 1.49 and 3.77pp for Decile 9 and 10 funds, respectively. The effect is also present in specifications (2) and (3) for Decile 10 funds, with a 3.10 and 0.97pp greater volatility than the median fund, respectively. Taken together, this provides evidence that liquidation-threatened funds (i.e., those funds at the bottom of the HWM distance distribution) are likely to gamble to avoid liquidation.

This provides evidence of the basic mechanism of the model: when funds are close to or extremely far from their HWMs, they have an incentive to increase return volatility to earn a performance fee or avoid liquidation, respectively. However, this

---

43Note that, as $\sigma_{d-1}^F$ is included as a control, the coefficients on the decile indicators are identical to a regression in which $\Delta\sigma_{d,t}^F$ is used as the dependent variable. Thus, these estimates can also be interpreted as the excess change in volatility for Decile $d$ funds relative to the median fund.

44An alternate viewpoint, as in Agarwal et al. (2009), is to view the performance fee as a collection of call options held by the manager for each of their investors. For funds with a HWM distance of 0, these are a collection of at-the-money call options. According to the Black-Scholes option pricing theory, the vega is highest for at-the-money call options, indicating that a manager’s incentive to deviate should be greatest when all investors are at their high-water mark.
### Table 3.6: Average Long Volatility Regressions

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>( \bar{\sigma}^L )</th>
<th>( \Delta \bar{\sigma}^L )</th>
<th>( \Delta^S \bar{\sigma}^L )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Quintile 0</td>
<td>2.102*** (0.780)</td>
<td>3.256*** (0.931)</td>
<td>1.511* (0.776)</td>
</tr>
<tr>
<td>Quintile 1</td>
<td>1.626* (0.977)</td>
<td>1.821** (0.853)</td>
<td>1.160 (1.168)</td>
</tr>
<tr>
<td>Quintile 2</td>
<td>0.657 (0.993)</td>
<td>0.576 (0.868)</td>
<td>0.228 (1.186)</td>
</tr>
<tr>
<td>Quintile 4</td>
<td>0.772 (1.002)</td>
<td>0.855 (1.197)</td>
<td>1.553 (1.978)</td>
</tr>
<tr>
<td>Quintile 5</td>
<td>2.696*** (1.022)</td>
<td>2.894*** (0.901)</td>
<td>3.974*** (1.221)</td>
</tr>
<tr>
<td>( \sigma_{t-1}^F ) (%)</td>
<td>0.054** (0.027)</td>
<td>0.045* (0.024)</td>
<td>-0.038 (0.032)</td>
</tr>
<tr>
<td>NetReturn_{t-1} (%)</td>
<td>-0.000 (0.010)</td>
<td>0.011 (0.010)</td>
<td>0.002 (0.012)</td>
</tr>
<tr>
<td>( \bar{\sigma}^L_{t-1} ) (%)</td>
<td>-0.183*** (0.044)</td>
<td>0.086* (0.045)</td>
<td></td>
</tr>
<tr>
<td>Long ( \beta_{t-1}^{mkt} )</td>
<td>1.640 (1.111)</td>
<td>-0.840 (0.990)</td>
<td>-7.180*** (1.268)</td>
</tr>
<tr>
<td>Long ( \sigma_{t-1}^L ) (%)</td>
<td>0.743*** (0.046)</td>
<td>0.621*** (0.045)</td>
<td>-0.416*** (0.020)</td>
</tr>
<tr>
<td>Performance Fee (%)</td>
<td>0.093* (0.054)</td>
<td>0.104** (0.047)</td>
<td>0.165** (0.064)</td>
</tr>
<tr>
<td>Market Volatility (%)</td>
<td>0.882*** (0.036)</td>
<td>0.522*** (0.041)</td>
<td>0.563*** (0.039)</td>
</tr>
<tr>
<td>Market Return (%)</td>
<td>-0.120*** (0.018)</td>
<td>-0.293*** (0.020)</td>
<td>0.024 (0.019)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time FE?</th>
<th>N</th>
<th>Y</th>
<th>N</th>
<th>Y</th>
<th>N</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>1,709</td>
<td>1,709</td>
<td>1,709</td>
<td>1,709</td>
<td>1,709</td>
<td>1,709</td>
</tr>
<tr>
<td>R²</td>
<td>0.741</td>
<td>0.806</td>
<td>0.584</td>
<td>0.728</td>
<td>0.296</td>
<td>0.506</td>
</tr>
</tbody>
</table>

**Note:** This table presents regression coefficients of quintile dummy variables based on the company-level lag value-weighted HWM distance on the level of, change in, and stale change in average volatility of the company’s long portfolio over the next year. For the latter two variables (change, stale change in long volatility), the lagged level of average long volatility is dropped as a control. Quintile 0 contains companies that are at their HWMs (distance of 0), while Decile 10 contains that companies are furthest from their HWMs. Insignificant controls in all specifications are omitted. All variables are contemporaneous unless otherwise stated. The sample is annually from 1996 to 2018. Standard errors are presented in parentheses. *p<0.1; **p<0.05; ***p<0.01 under the assumption of a single hypothesis test.
evidence cannot speak to the source of this increased volatility. There are two basic ways to increase return volatility: increasing leverage or investing in riskier securities. While fund leverage is unobserved, the holdings data provide a comprehensive snapshot of a management company’s long portfolio, which can be used to determine whether funds are achieving higher return volatility through the latter channel.\textsuperscript{45}

Further examining whether funds alter their portfolio allocation in response to HWM distance, Table 3.6 considers the impact of company quintiles (relative to quintile 3) on the average volatility of stocks in the company’s long portfolio. Here, the pattern is clear: companies in Quintiles 0 and 5 (i) invest in securities that are on average riskier, (ii) increase their investment in risky securities, and (iii) select into previously riskier securities relative to the median company. Companies in Quintile 0 (5) invest in securities that are 2.06pp (2.65pp) more volatile than than those in the median company. Similarly, they increase the average security’s volatility by 3.27pp (3.98pp) and invest in securities that are expected to be 2.90pp (2.20pp) more volatile the previous year relative to the median company.

Taken together, these three findings are consistent with the theory presented in Section 3.2. Not only do companies in Quintiles 0 and 5 invest in higher volatility securities, but also chase volatile securities. As conditional volatility approximately follows a random walk, higher investment in securities that were risky in the previous period serves as a viable strategy for companies seeking higher volatility securities in the current period.\textsuperscript{46} Indeed, the similarity in point estimates between specifications (3) and (5) and specifications (4) and (6) reinforces this notion: the change in fund

\textsuperscript{45}In theory, leverage could be observed by taking the ratio of the total value of a company’s long portfolio and the AUM of fund under that company. However, the resulting estimates are noisy due to the facts that (i) companies may have non-hedge fund business and (ii) not every hedge fund in a given company may self-report to Morningstar. The resulting estimates are leverage are thus biased upwards and poorly estimated.

\textsuperscript{46}See Engle and Patton (2001); Patton (2011); Hansen and Lunde (2014).
Table 3.7: Sources of Risk Regressions

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Long Market Beta $\beta^{mkt}$</th>
<th>Long Idiosyncratic Vol $\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Quintile 0</td>
<td>0.030**</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Quintile 1</td>
<td>0.008</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Quintile 2</td>
<td>0.009</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Quintile 4</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Quintile 5</td>
<td>0.027</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Log(AUM$_{t-1}$)</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\sigma^2_{t-1}$ (%)</td>
<td>0.001*</td>
<td>0.001*</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.000)</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_{t-1}$ (%)</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Long $\beta^{mkt}_{t-1}$</td>
<td>0.647***</td>
<td>0.644***</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>Long $\sigma^2_{t-1}$ (%)</td>
<td>−0.002*</td>
<td>−0.002*</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Performance Fee (%)</td>
<td>−0.000</td>
<td>−0.000</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Market Volatility (%)</td>
<td>0.003***</td>
<td>0.478***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.034)</td>
</tr>
<tr>
<td>Market Return (%)</td>
<td>−0.001***</td>
<td>−0.175***</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Time FE?</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>1,709</td>
<td>1,709</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.461</td>
<td>0.496</td>
</tr>
</tbody>
</table>

Note: This table presents regression coefficients of quintile dummy variables based on the company-level lag value-weighted HWM distance on long market beta and idiosyncratic volatility over the next year. Quintile 0 contains companies that are at their HWMs (distance of 0), while Decile 10 contains that companies are furthest from their HWMs. Insignificant controls in all specifications are omitted. All variables are contemporaneous unless otherwise stated. The sample is annually from 1996 to 2018. Standard errors are presented in parentheses. *p<0.1; **p<0.05; ***p<0.01 under the assumption of a single hypothesis test.
volatility can be heavily attributable to selecting into securities that were riskier in
the previous period.

Having established that companies in Quintiles 0 and 5 hold riskier portfolios, Ta-
ble 3.7 examines the characteristics funds in these quintiles are targeting. Specifically,
do they increase their exposure to market beta or invest in securities with greater
residual volatility?\textsuperscript{47} Interestingly, it seems as though companies in Quintiles 0 and 5
behave differently. Quintile 0 companies increase their average market exposure (by
0.03 relative to Quintile 3 companies), while Quintile 5 companies increase their av-
erage idiosyncratic volatility (by 0.41pp relative to the median company). This increase
in market beta of Quintile 0 companies accounts for 0.48pp (0.029×16.65%, the aver-
age market volatility in the sample) of the documented 3.27pp rise in fund volatility,
while the rise in idiosyncratic volatility accounts for roughly 10.4% (0.413/3.982) of
the increase in volatility of Quintile 5 companies.

This presents an interesting avenue for future research: why would these hedge
fund companies increase their exposure differentially? One possible explanation is
skill. In theory, funds simply interested in increasing their volatility might as well
invest in securities with higher market beta, as these securities are compensated for
this systematic risk.\textsuperscript{48} On the other hand, less skilled managers do not distinguish
between market risk factors and idiosyncratic volatility and simply invest in securities
with higher overall volatility, leading to increased exposure to both market risk and
idiosyncratic volatility.\textsuperscript{49}

\textsuperscript{47}The model in Section 3.2 is a stylized version of the problem facing hedge funds and cannot speak
to this decomposition.

\textsuperscript{48}The BAB anomaly of Frazzini and Pedersen (2014) contradicts the basic premise that higher
market beta stocks exhibit higher returns, while Ang et al. (2006) establish that idiosyncratic
volatility is negatively correlated with expected returns.

\textsuperscript{49}While not significant, the point estimates of Quintile 5 for long $\beta^{mkt}$ are positive and similar
to those of Quintile 0, in addition to the positive and significant coefficients for idiosyncratic
volatility.
3.5.2 Equilibrium Results

The previous section established that funds at and very far from their HWMs tend to have more volatile returns and invest in riskier securities. In this section, the impact this behavior has on aggregate hedge fund holdings and risk premia is examined. Ceteris paribus, as the hedge fund wealth distribution concentrates in the tails of the distribution (that is, as many hedge funds become close to or very far from their HWMs), their demand for risky securities rises, reducing the risk premium while hedge funds increase their allocations in risky securities. As discussed in Section 3.2, however, these effects can be confounded by changes in the size of the hedge fund industry. As a result, in order to understand these effects, it is first necessary to understand how the size of the hedge fund industry responds to the equilibrium proxies $OTM_t$ and $ITM_t$.

Table 3.8 examines this relationship. Hedge fund AUM and flows are much more sensitive to $OTM_t$ than to $ITM_t$; a one standard deviation rise in $ITM_t$ is associated with a 5% decline in the size of the hedge fund industry, over 4% of which is due to net outflows. Interestingly, hedge fund AUM does not respond particularly strongly to $ITM_t$; as more hedge funds hit their HWMs, there is no significant impact on the size of the hedge fund industry. These results suggest that the effect of hedge fund incentives may be confounded by changes in the size of the hedge fund industry during periods when $OTM_t$ is high, but such confounding is likely to be muted for $ITM_t$.

The theory suggests that when there are many hedge funds close to and very far from their HWMs, hedge funds should invest in riskier securities. Having shown such effects exist on an individual fund level, this section turns to the question of whether aggregate hedge fund volatility can be predicted by these aggregate proxies of the
Table 3.8: Hedge Fund AUM and Flows

<table>
<thead>
<tr>
<th></th>
<th>Aggregate AUM (%)</th>
<th>Aggregate Flows (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.055***</td>
<td>0.063***</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>OTM&lt;sub&gt;t&lt;/sub&gt;</td>
<td>-0.047***</td>
<td>-0.053***</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>ITM&lt;sub&gt;t&lt;/sub&gt;</td>
<td>0.004</td>
<td>-0.016*</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>mkt&lt;sub&gt;t&lt;/sub&gt;</td>
<td>0.109</td>
<td>0.245**</td>
</tr>
<tr>
<td></td>
<td>(0.095)</td>
<td>(0.107)</td>
</tr>
<tr>
<td>ΔTED&lt;sub&gt;t&lt;/sub&gt;</td>
<td>-0.052**</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>TED&lt;sub&gt;t−1&lt;/sub&gt;</td>
<td>-0.022</td>
<td>-0.048*</td>
</tr>
<tr>
<td></td>
<td>(0.021)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>Observations</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td>R²</td>
<td>0.357</td>
<td>0.149</td>
</tr>
</tbody>
</table>

Note: This table presents regression coefficients of two equilibrium proxies, OTM<sub>t</sub> and ITM<sub>t</sub> on aggregate hedge fund assets under management (AUM) and aggregate flows. These proxies capture the mass of hedge funds in the bottom and top deciles of HWM distance, respectively. Controls include the market excess return, the change in the TED spread, and the lag TED spread. The sample is quarterly from 1997Q2 to 2018Q4. Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01 under the assumption of a single hypothesis test.

HWM distance distribution. To test this, the following model is used:

\[
\frac{\bar{\sigma}_t^L}{\sigma_t} = \alpha + \delta X_t + \rho \frac{\bar{\sigma}_{t-1}^L}{\sigma_{t-1}} + \beta^{mkt} mkt_t + \beta^{smb} smb_t + \beta^{hml} hml + \gamma TED_t + \lambda TED_{t-1} + \epsilon_t,
\]

(3.33)

where \( X_t \in \{OTM_t, ITM_t\} \) have been centered and scaled to have mean zero and unit variance. The dependent variable is the ratio of the average volatility of a stock held long by the hedge fund sector to the average volatility of stocks in the CRSP universe. Fama-French factors are used as controls to account for a market timing effect (e.g., Brandt et al., 2017), while the lag and change in TED spread are included to account for the level of and changes in funding liquidity (Brunnermeier...
Table 3.9: Average Hedge Fund Volatility Results

| Dependent variable: Hedge Fund Volatility, $\frac{\sigma^L_t}{\sigma_t}$ |
|-----------------------------|-----------------|-----------------|
| (1)                         | (2)             | (3)             |
| Intercept                   | 0.199***        | 0.170***        | 0.173**         |
|                             | (0.066)         | (0.062)         | (0.067)         |
| $OTM_t$                     | −0.008          | −0.001          |
|                             | (0.009)         | (0.009)         |
| $ITM_t$                     |                 | 0.015**         | 0.015*          |
|                             |                 | (0.007)         | (0.008)         |
| $\frac{\sigma^L_{t-1}}{\sigma_t}$ | 0.828***        | 0.844***        | 0.841***        |
|                             | (0.054)         | (0.049)         | (0.053)         |
| $mkt_t$                     | 0.092           | 0.153           | 0.149           |
|                             | (0.096)         | (0.094)         | (0.099)         |
| $smb_t$                     | −0.117          | −0.115          | −0.117          |
|                             | (0.158)         | (0.154)         | (0.156)         |
| $hml_t$                     | 0.090           | 0.104           | 0.102           |
|                             | (0.120)         | (0.116)         | (0.118)         |
| $\Delta TED_t$             | −0.008          | 0.001           | 0.000           |
|                             | (0.022)         | (0.022)         | (0.022)         |
| $TED_{t-1}$                 | 0.020           | 0.033           | 0.033           |
|                             | (0.020)         | (0.021)         | (0.021)         |
| Observations                | 98              | 98              | 98              |
| R$^2$                       | 0.774           | 0.783           | 0.783           |

Note: This table presents regression coefficients of two equilibrium proxies, $OTM_t$ and $ITM_t$, on a measure of aggregate hedge fund risk composition. The dependent variable is the ratio of the value-weighted average of the monthly volatility of stocks held long by the hedge fund sector and the value-weighted average of the monthly volatility of all CRSP stocks. Controls include the Fama-French factors and the lag and change in the TED spread. The sample is quarterly from 1994Q3 to 2018Q4. Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01 under the assumption of a single hypothesis test.

Table 3.9 presents results from this regression. As hedge funds approach their HWMs, the industry as a whole invests in securities that are riskier relative to the average CRSP security. A one standard deviation increase in $ITM_t$ is associated with hedge funds increasing the volatility of their long portfolio by 1.6%. Interestingly,
there is no such relationship with $OTM_t$. Aggregate hedge fund volatility does not vary with the percent of hedge funds in the bottom decile of HWM distance. This is due to selection among outflows. Table 3.8 established that a high value for $OTM_t$ is associated with large outflows from the hedge fund sector. Such outflows not random; rather, they are concentrated among poorly performing hedge funds. As $OTM_t$ is computed based entirely on time $t−1$ information, changes in the composition of the hedge fund sector between time $t−1$ and $t$ are not incorporated into this measure. Exit among the funds in the bottom decile may thus be driving this null result.

The final prediction of this model is that hedge funds have an impact on risk premia through their time-varying effective risk aversion. To test this, the following augmentation of the security market line is used:

$$R_{d,t} - r^f_t = \alpha + \delta \beta^{mkt}_{d,t} X_t + \gamma \beta^{mkt}_{d,t} + \epsilon_{d,t},$$  \hspace{1cm} (3.34)

where $R_{d,t}$ is the return of the $d^{th}$ beta-sorted decile at time $t$. Under the CAPM, the security market line should be linear in $\beta_{d,t}$ with an intercept equal to 0 and a slope equal to the market return (i.e., $\alpha = \delta = 0$ and $\gamma = mkt$). However, if hedge funds are encouraged to take risk (that is, many funds at or very far from their HWMs), ceteris paribus, this should reduce risk premia. However, following a negative shock that leads to fund outflows (Table 3.8), hedge funds become “less marginal,” so the reduction in hedge fund size may have the opposite impact on risk premia.

Table 3.10 presents the estimates of this model. As found in Black et al. (1972), the security market line is far too flat: the mean market premium in this sample is 7.8% compared to an estimated slope of -4.8% per annum. Higher beta securities seem to have lower expected returns, a finding that is clearly at odds with the CAPM.

\footnote{In untabulated results, funds in the bottom decile of the distribution are associated with net outflows exceeding 17% relative to the median fund.}
Table 3.10: Security Market Line Slope Estimates

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.010***</td>
<td>0.011***</td>
<td>0.010***</td>
<td>0.013***</td>
<td>0.014***</td>
<td>0.013***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$\beta_{d,t}^{mkt}$</td>
<td>-0.004*</td>
<td>-0.005**</td>
<td>-0.004*</td>
<td>-0.007***</td>
<td>-0.009***</td>
<td>-0.007***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>$\beta_{d,t}^{mkt} \times OTM_t$</td>
<td>0.007***</td>
<td>0.007***</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\beta_{d,t}^{mkt} \times ITM_t$</td>
<td>-0.004**</td>
<td>-0.002</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$\beta_{d,t}^{mkt} \times OTM_t \times HFSIZE_t$</td>
<td>0.002</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$\beta_{d,t}^{mkt} \times ITM_t \times HFSIZE_t$</td>
<td>-0.010**</td>
<td>-0.010**</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>$OTM_t$</td>
<td>-0.007***</td>
<td>-0.004</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>$ITM_t$</td>
<td></td>
<td>0.005**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$HFSIZE_t$</td>
<td></td>
<td>-0.009*</td>
<td></td>
<td>-0.009*</td>
<td>-0.010*</td>
<td>-0.010*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>$OTM_t \times HFSIZE_t$</td>
<td></td>
<td></td>
<td>-0.010**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$ITM_t \times HFSIZE_t$</td>
<td></td>
<td></td>
<td></td>
<td>0.016***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.005)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{d,t}^{mkt} \times HFSIZE_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.010**</td>
<td>0.010**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Observations</td>
<td>3,040</td>
<td>3,040</td>
<td>3,040</td>
<td>3,040</td>
<td>3,040</td>
<td>3,040</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.001</td>
<td>0.007</td>
<td>0.002</td>
<td>0.002</td>
<td>0.013</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Note: This table presents regression results of various augmentations of the security market line. Each month, stocks are sorted into deciles based on their previous months’ market betas. These portfolios are then value-weighted, and betas and returns are computed over the following month. A security market line is then estimated using these betas and their interaction with two equilibrium proxies, $OTM_t$ and $ITM_t$, as well as $HFSIZE_t$, an indicator equal to one if total hedge fund AUM is above its full sample median. $OTM_t$ and $ITM_t$ have been centered and scaled to have zero mean and unit variance. The sample is monthly from 1994:3 to 2018:12. Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01 under the assumption of a single hypothesis test.

Instead, the model suggests that returns to beta depends on aggregate hedge fund risk aversion.
Table 3.11: Betting Against Beta Regression Results

<table>
<thead>
<tr>
<th>Dependent variable:</th>
</tr>
</thead>
<tbody>
<tr>
<td>BABt</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(1)</td>
</tr>
<tr>
<td>(2)</td>
</tr>
<tr>
<td>(3)</td>
</tr>
<tr>
<td>(4)</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>OTMt</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ITMt</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>OTMt × HFSizet</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ITMt × HFSizet</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>HFSizet</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>mkt_t</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>smb_t</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>hml_t</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>ΔTEDt</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Observations</td>
</tr>
<tr>
<td>R²</td>
</tr>
</tbody>
</table>

Note: This table presents regression results of two equilibrium proxies, OTMt and ITMt, on betting against beta (BAB). These proxies capture the mass of hedge funds in the bottom and top deciles of HWM distance, respectively. Controls include the market excess return and the change in the TED spread. HFSizet is an indicator equal to one if total hedge fund AUM is above its full sample median. OTMt and ITMt have been centered and scaled to have zero mean and unit variance. The sample is monthly from 1994:4 to 2018:12. Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01 under the assumption of a single hypothesis test.

Indeed, that is what seems to occur. The security market line flattens as ITMt increases. This is a product of the results in Table 3.7, which showed that only hedge funds at their HWMs seemed to increase their market exposure relative to other funds. Further, interacting $\beta_{mkt}^{d,t} \times X_t$ with $HFSizet$, an indicator equal to one if aggregate hedge fund AUM is above its median, shows that the effect is present only
when hedge funds are large. The magnitudes are quite large, with a one standard
deviation increase in $ITM_t$ further reducing the slope of the SML by 6 percentage
points per annum.

Interestingly, the SML steepens with $OTM_t$. There are two explanations for
this. First, from the results in Table 3.7, hedge funds far from their HWMs do
not significantly increase the average beta of their long portfolio, instead increasing
volatility through idiosyncratic exposure. Second, a high value of $OTM_t$ is associated
with large outflows, and these outflows are concentrated among poorly performing
hedge funds. This not only reduces the ability of hedge funds to impact aggregate
asset prices, but also changes the preferences of the aggregate hedge fund sector due
to selection among the outflows.

Consistent with the relationship with the security market line, the status of the
hedge fund wealth distribution can also explain betting against beta (BAB, Frazzini
and Pedersen, 2014). BAB is constructed as the (ex-ante) market neutral difference
of returns between the lowest and highest beta decile stocks. As such, as funds hit
their HWMs ($ITM_t$ increases), the returns of low beta securities increase relative
to high beta securities. Interpreted in the context of the model, as $ITM_t$ increases,
the aggregate hedge fund industry becomes relatively less risk averse, increasing the
demand for high risk securities. This increased demand pushes their expected return
down relative to low risk securities, and this effect is most prominent when the hedge
fund sector is large (i.e., $HFSize_t = 1$, column (4)). Similarly, as funds become very
far from their HWMs, the returns to low beta securities decreases relative to high
beta securities. This occurs as (i) funds in the bottom deciles of the distribution do
not increase the average market beta of their portfolio and (ii) large and significant
outflows in the hedge fund industry among funds concentrated in the bottom deciles.

This evidence suggests that hedge fund incentives play a crucial role in explaining
asset prices and holdings. A large mass of funds at their HWMs is associated with hedge funds investing in securities that are riskier than the average security. This is associated with a flattening of the security market line (i.e., a reduction in the CAPM-implied market risk premium) and an increase in betting against beta, reflecting this increased demand for high beta securities. A large mass of funds far below their HWMs is associated with large outflows from the hedge fund sector, a steepening of the security market line, and a decrease in BAB.

### 3.6 Conclusion

This paper examines a specific source of variation in the hedge fund manager’s optimization problem and explores the implications of this variation on aggregate asset prices and asset allocations. The prototypical hedge fund contract structure features two key nonlinearities: managers are rewarded via a performance fee for returns in excess of a high-water mark and are subject to liquidation risk following poor performance. I embed these two features of the hedge fund’s problem into a heterogeneous agent model, and explore the implications both empirically and theoretically.

These nonlinearities affect manager risk appetites and have consequences for equilibrium asset prices and asset allocations. Empirically, I find that funds closest to and furthest from their HWMs are relatively risk-loving: they increase the volatility of their returns. This is accomplished in part by investing in on average riskier securities and in securities that are expected to be riskier in the future. Further, as the distribution of hedge fund wealth shifts over time, so too do aggregate hedge fund risk composition and the returns of risky securities. As hedge funds approach their high-water marks, aggregate risk-taking in the hedge fund sector increases. This increased demand for high volatility and high beta securities lowers the market risk.
premium.

This evidence is consistent with a static model in which hedge fund managers maximize their expected compensation subject to liquidation risk and a leverage constraint. Hedge funds close to their HWMs and below the liquidation threshold are relatively risk-loving and invest in risky securities, while funds just above the liquidation threshold are risk-averse and invest in less risky securities. The total mass of hedge funds and shape of the wealth distribution are crucial to determining the equilibrium effect of this contract structure. A small mass of hedge funds results in a minimal price impact, which then leads to increased hedge fund investment in risky securities, as they seem relatively underpriced. A large mass of funds in the tails of the distribution causes a high demand for the risky security, which pushes its expected return below the less risky security.

This presents several avenues of future research. First, embedding the features of this model into a dynamic general equilibrium model can further refine the model’s predictions and help quantitatively explain the observed phenomenon. Second, the model is silent on leverage. In theory, funds can increase their volatility simply by leveraging up an existing portfolio. Empirically, the ability to examine the leverage channel is limited due to incomplete observation of hedge fund returns and AUM. Third, I hope to further test some of the predictions from the model. Among these is whether mutual funds take positions opposite to hedge funds: when hedge funds invest in riskier securities, do mutual funds respond by investing in less risky securities? Are returns to idiosyncratic volatility affected by a large mass of funds at the bottom of the distribution? Why do different parts of the distribution increase exposure to different sources of volatility?
Chapter 4

Conclusion

In this dissertation, I show that institutional investors differ drastically in the portfolios they hold and strategies they pursue and do so in response to the incentives generated by their organizational structures.

In the first chapter, I first document the substantial heterogeneity in portfolios across different investor classes, and explain this with an information choice model in which investors have differing levels of skill. I then estimate measures of skill for each investor class and show that investors form portfolios in a manner that is consistent with their skill, both in terms of the strategies they pursue and the characteristics of their portfolio. I then turn to the equilibrium impact of their positions, and show that an investor’s ability to “arbitrage away” mispricings depends crucially on their overall skill level.

In the second chapter, I build and test a model that studies the relationship between an investor’s compensation contract and her dynamic risk preferences. The key finding is that hedge funds exhibit time-varying risk preferences depending on the likelihood of earning a performance fee. In testing this model empirically, I show that hedge funds both increase the volatility of their portfolios and invest in riskier assets when incentivized to do so. I further show that these actions have equilibrium price impacts; when hedge funds herd into this behavior, risky assets tend to underperform.
Appendices

A.1 Chapter 1 Appendix

A.1.1 Additional Results

This section contains additional empirical results omitted from the main text.

Attention Allocation Robustness

The main body of the paper asserts that variation in factor timing ability is the primary driver of differences in attention capacities across different types of investors. This section examines the robustness of these findings to various adjustments, including using controlling for different investment universes, allowing for time-varying fundamentals using Kalman filtering, and using even and odd quarters separately to estimate the measures.

Within-Investment Universe Attention  Different investors may focus on different universes of securities when making investment decisions. For example, the universe of securities considered by a small-cap mutual fund differ markedly from those of a value mutual fund. To control for the fact that some investors may have a much broader mandate than others, I recompute the main measures of attention capacity and allocations after controlling for each investor type’s aggregate investor universe. The investment universe is constructed in a similar way to the instrument
for market equity used in the demand system estimation. Specifically, an investor’s exposure is defined as
\[ q^c_{i,t} = \sum_{s=1}^{S_t} w_{i,s,t}^c d^c_{s,t} - \bar{d}^c_{m_i,t}. \]  
(A.1)

The sole difference between this definition and that in Eq. (2.36) is the baseline exposure \( \bar{d}^c_{m_i,t} \). In the main text, this was taken to be common across investors and equal simply to the market-weighted average decile. To adjust for heterogeneous investment universes, the baseline exposure is instead taken to be common within an investor type and equal to
\[ \bar{d}^c_{m,t} = \sum_{s=1}^{S_t} \bar{w}_{m,s,t}^c d^c_{s,t}, \]  
(A.2)

where the baseline weights \( \bar{w}_{m,s,t} \) are given by
\[ \bar{w}_{m,s,t} = \frac{\sum_{i=1}^{N_{m,t}} 1 \left[ w_{i,s,t} > 0 \right]}{\sum_{s=1}^{S_t} \sum_{i=1}^{N_{m,t}} 1 \left[ w_{i,s,t} > 0 \right]}. \]  
(A.3)

In words, the baseline weights for type \( m \) investors are proportional to the fraction of investors that invest in each security. As a result, stocks that are invested in by a greater fraction of type \( m \) investors receive higher weights.

Table A.1 presents the attention proxies after controlling for investment universes. Hedge funds continue to exhibit both significantly positive attention capacity and timing ability; in fact, the point estimates are slightly larger after controlling for investment universes. However, the estimates for long-term investors and brokers are no longer significant after controlling. Instead, long-term investors have positive (though not statistically significant) capacity, while brokers continue to exhibit negative capacity. These results further distinguish hedge funds from other investors in their overall capabilities and their ability to factor time.

134
Table A.1: Estimated Attention Proxies Relative to Investment Universe

<table>
<thead>
<tr>
<th>Attention Proxy</th>
<th>$F_{Skill_i}$</th>
<th>$F_{Timing_i}$</th>
<th>$F_{Selection_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>2.023***</td>
<td>99.8%</td>
<td>1.882***</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>0.642</td>
<td>77.8%</td>
<td>0.486</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>0.492</td>
<td>66.1%</td>
<td>0.300</td>
</tr>
<tr>
<td>Private Banking</td>
<td>0.750</td>
<td>87.7%</td>
<td>0.611</td>
</tr>
<tr>
<td>Long-Term</td>
<td>0.533</td>
<td>69.0%</td>
<td>0.391</td>
</tr>
<tr>
<td>Broker</td>
<td>-0.315</td>
<td>42.7%</td>
<td>-0.613</td>
</tr>
</tbody>
</table>

*Note:* This table presents alternative estimates of attention capacity by investor type controlling for differences in investors’ investment universes. Specifically, the “baseline” exposure $\hat{\epsilon}_t$ is determined per investor group, rather than simply taken as the market-weighted average score. This baseline is measured as the exposure that results from a portfolio using the fraction of investors of a given type that invest in each security as weights. The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with the unpermuted factor returns. *p<0.1; **p<0.05; ***p<0.01.

**Time-Varying Fundamental**  The model and main empirical results assumed the fundamental component of returns associated with each characteristic was constant over time. However, fluctuations in the fundamental may occur due to time-varying risk aversion, structural breaks, etc. To account for this possibility, the fundamental component of returns is extracted for each factor using a Kalman filter. Specifically, the fundamental is assumed to follow a random walk:

$$f_t^c = \mu_t^c + \epsilon_t^c \quad \text{and} \quad \mu_t^c = \mu_{t-1}^c + \eta_t^c. \quad (A.4)$$

Then, the decomposition of attention capacity is given by

$$F_{Skill_i} = C \left[ q_{i,t}; f_t^c \right]$$

$$= C \left[ q_{i,t}; \mu_t^c \right] + C \left[ q_{i,t}; \epsilon_t^c \right]. \quad (A.5)$$

Table A.2 presents the results. With a time-varying fundamental, the covariance is essentially shifted from the idiosyncratic component ($F_{Timing_i}$) to the fundamen-
tal ($F_{Selection_i}$). As a result, covariance with the fundamental is able to explain a more significant fraction of investors’ total attention capacity. The main results, however, are broadly consistent. Most of hedge funds overall capacity is due to their ability to time, while the opposite is true for long-term investors, brokers, and short-sellers. However, there is more significant variation in investors’ factor selecting ability. Hedge funds, investment advisors, and mutual funds are all able to significantly select factors cross sectionally, while households and short-sellers exhibit significantly negative factor selecting ability.

**Table A.2: Estimated Attention Proxies Using Kalman Filter**

<table>
<thead>
<tr>
<th></th>
<th>$F_{Skill_i}$</th>
<th></th>
<th>$F_{Timing_i}$</th>
<th></th>
<th>$F_{Selection_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>1.712**</td>
<td>99.4%</td>
<td>1.089</td>
<td>93.7%</td>
<td>0.623***</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>0.213</td>
<td>92.3%</td>
<td>0.046</td>
<td>63.4%</td>
<td>0.167***</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>0.145</td>
<td>83.0%</td>
<td>0.095</td>
<td>74.2%</td>
<td>0.050**</td>
</tr>
<tr>
<td>Household</td>
<td>0.061</td>
<td>58.6%</td>
<td>0.145</td>
<td>70.1%</td>
<td>-0.084**</td>
</tr>
<tr>
<td>Private Banking</td>
<td>-0.166</td>
<td>27.2%</td>
<td>-0.116</td>
<td>32.8%</td>
<td>-0.050</td>
</tr>
<tr>
<td>Long-Term</td>
<td>-0.310**</td>
<td>1.3%</td>
<td>-0.336***</td>
<td>0.5%</td>
<td>0.026</td>
</tr>
<tr>
<td>Broker</td>
<td>-0.733***</td>
<td>0.2%</td>
<td>-0.699***</td>
<td>0.2%</td>
<td>-0.034</td>
</tr>
<tr>
<td>Short-Seller</td>
<td>-1.955**</td>
<td>2.3%</td>
<td>-1.217</td>
<td>9.4%</td>
<td>-0.738***</td>
</tr>
</tbody>
</table>

*Note:* This table presents alternative estimates of attention capacity by investor type accounting for time-variation in the expected component of returns. Expected returns are estimated at each point using a Kalman filter of the form $f_i^t = \mu_i^t + \epsilon_i^t$, where $\mu_i^t = \mu_i^{t-1} + \eta_i^t$. The time series and cross-sectional covariances are subsequently estimated as $\mathbb{C}[d_i^t, \epsilon_i^t]$ and $\mathbb{C}[d_i^t, \mu_i^t]$, respectively. The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with the unpermuted factor returns. *p<0.1; **p<0.05; ***p<0.01.

**Even/Odd Estimation** One may worry how well-estimated the estimates of attention capacities and allocations are. To alleviate this concern, Table A.3 shows the estimates for overall attention capacity estimated using even and odd quarters separately. The estimates are both quantitatively and qualitatively similar for all investor types across the two subsamples, indicating that these measures are reasonably well-estimated. The decomposition of attention capacity into $F_{Timing}$ and $F_{Selection}$
(untabulated) are similarly well-estimated.

Table A.3: Estimated Attention Capacity Using Even and Odd Quarters

<table>
<thead>
<tr>
<th></th>
<th>Even Quarters</th>
<th></th>
<th>Odd Quarters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
<td>Percentile</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>1.907*</td>
<td>97.0%</td>
<td>1.516**</td>
<td>97.6%</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>0.069</td>
<td>61.4%</td>
<td>0.356</td>
<td>67.6%</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>0.230</td>
<td>84.0%</td>
<td>0.060</td>
<td>25.4%</td>
</tr>
<tr>
<td>Household</td>
<td>-0.013</td>
<td>50.4%</td>
<td>0.133</td>
<td>88.6%</td>
</tr>
<tr>
<td>Private Banking</td>
<td>-0.593</td>
<td>6.4%</td>
<td>0.260</td>
<td>34.2%</td>
</tr>
<tr>
<td>Long-Term</td>
<td>-0.319*</td>
<td>4.3%</td>
<td>-0.300***</td>
<td>0.1%</td>
</tr>
<tr>
<td>Broker</td>
<td>-0.405</td>
<td>10.8%</td>
<td>-1.061***</td>
<td>0.0%</td>
</tr>
<tr>
<td>Short-Seller</td>
<td>-2.322*</td>
<td>3.5%</td>
<td>-1.589**</td>
<td>2.0%</td>
</tr>
</tbody>
</table>

Note: This table presents estimates of attention capacity by investor type estimated using even quarters (month-ends June and December) and odd quarters (month-ends March and September) separately. $F_{Skill_i}$ is the covariance between an investor’s portfolio and factor returns. The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with unpermuted factor returns. All covariances have been multiplied by 1,000 for readability. *p<0.1; **p<0.05; ***p<0.01.

Effects of Shorting  The final concern relates to the omission of short positions when computing investor’s exposures. To quantify the impact of this mismeasurement, I use the CRSP mutual fund holdings data, which began reporting short positions in 2006. I compute exposures using both long and short positions $d_{i,t}^{c,net}$ and compare them with exposures using only long positions $d_{i,t}^{c,long}$. I then run regressions at the aggregate fund-type level of the form:

$$d_{i,t}^{c,net} = \alpha + \beta_1 d_{i,t}^{c,long} + \beta_2 s_{i,t} + \beta_3 d_{i,t}^{c,long} \times s_{i,t} + \epsilon_{i,t},$$

(A.6)

where $s_{i,t}$ is the fraction of the portfolio held short. The resulting estimates are presented in Table A.4. Column (1) presents results from the univariate regression. As can be seen, exposures using long-only data underestimate exposures by approximately 5-6%. However, the long exposures are able to capture a high degree of the
variance in investors’ net exposures, with an $R^2$ of over 0.94. When introducing the interaction term in Column (2), there is some evidence that, among investors who short more aggressively, their exposures are underestimated by a greater amount. However, the additional explanatory power that comes from this interaction term is minimal, with the $R^2$ increasing by 0.01. As a result, it seems as though the long side of investors’ portfolios reasonably well captures their overall exposures.

**Table A.4: Effect of Short Positions, Aggregate Portfolios**

<table>
<thead>
<tr>
<th></th>
<th>Net Exposures $d_{i,t}^{net}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>$d_{i,t}^{c,long}$</td>
<td>1.055***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>$s_{i,t}$</td>
<td>0.216***</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
</tr>
<tr>
<td>$d_{i,t}^{c,long} \times s_{i,t}$</td>
<td>0.770***</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.042***</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
</tr>
<tr>
<td>Observations</td>
<td>12,507</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.942</td>
</tr>
</tbody>
</table>

*Note:* This table presents estimates of a regression of net exposures on exposures using only the long side of mutual funds’ portfolio, as well as the interaction between the long exposures and the fraction of the portfolio that is held short ($s_{i,t}$). The regression is run at the aggregate mutual fund type level. Standard errors are in parentheses. *p<0.1; **p<0.05; ***p<0.01.

**Persistence and Timing Horizon**

This section studies whether the attention capacities are persistent across the 20-year sample period and over what horizon is timing ability realized. To examine persistence, the sample is split into an early (1999-2008) and late (2008-2018) period. Attention capacities are estimated separately over these two periods. Table A.5 presents this comparison. Interestingly, there is significant attenuation in the
estimates for hedge funds, long-term investors, and brokers in the second half of the sample. Despite this, hedge funds still exhibit marginally significant attention capacity in the later period. Generally speaking, however, the signs and relative significance levels are qualitatively similar across the two halves of the sample.

Table A.5: Estimated Attention Capacity Using First and Second Half of Sample

<table>
<thead>
<tr>
<th></th>
<th>First Half</th>
<th></th>
<th>Second Half</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
<td>Percentile</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>2.608**</td>
<td>97.9%</td>
<td>0.315</td>
<td>91.5%</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>0.140</td>
<td>68.4%</td>
<td>0.182*</td>
<td>96.2%</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>0.005</td>
<td>50.1%</td>
<td>0.162</td>
<td>94.4%</td>
</tr>
<tr>
<td>Household</td>
<td>0.063</td>
<td>58.5%</td>
<td>0.123</td>
<td>51.2%</td>
</tr>
<tr>
<td>Private Banking</td>
<td>0.027</td>
<td>52.5%</td>
<td>-0.275*</td>
<td>3.6%</td>
</tr>
<tr>
<td>Long-Term</td>
<td>-0.602**</td>
<td>1.6%</td>
<td>-0.003</td>
<td>17.6%</td>
</tr>
<tr>
<td>Broker</td>
<td>-1.380***</td>
<td>0.0%</td>
<td>-0.198</td>
<td>11.1%</td>
</tr>
<tr>
<td>Short-Seller</td>
<td>-1.669</td>
<td>6.1%</td>
<td>-1.657**</td>
<td>1.9%</td>
</tr>
</tbody>
</table>

Note: This table presents estimates of attention capacity by investor type estimated using the first half of the data (until December 2008) and second half of the data (beginning March 2009) separately. \( F_{Skill_i} \) is the covariance between an investor’s portfolio and factor returns. The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with unpermuted factor returns. All covariances have been multiplied by 1,000 for readability. *p<0.1; **p<0.05; ***p<0.01.

To examine the horizon of investors’ factor timing ability, I next split up \( FTiming \) into timing ability over the next month, second month, and third month. For example, for the quarter ending in December 1999, I estimate timing ability separately for January 2000, February 2000, and March 2000, assuming the portfolio stays constant over the entire quarter. Table A.6 presents this decomposition. Interestingly, hedge funds timing ability does not realize until two months after portfolio formation. On the other hand, negative timing ability for private banking and long-term investors realize immediately, and long-term investors’ ability persists over the subsequent months.
Table A.6: Factor Timing Horizon

<table>
<thead>
<tr>
<th></th>
<th>Month $t$</th>
<th>Month $t+1$</th>
<th>Month $t+2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>0.047</td>
<td>58.4%</td>
<td><strong>0.990</strong></td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>-0.043</td>
<td>24.3%</td>
<td>0.078</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>-0.021</td>
<td>36.9%</td>
<td>-0.056</td>
</tr>
<tr>
<td>Household</td>
<td>0.075</td>
<td>74.3%</td>
<td>-0.165</td>
</tr>
<tr>
<td>Private Banking</td>
<td><strong>-0.314</strong></td>
<td>0.3%</td>
<td>-0.052</td>
</tr>
<tr>
<td>Long-Term</td>
<td><strong>-0.165</strong></td>
<td>0.3%</td>
<td><strong>-0.176</strong></td>
</tr>
<tr>
<td>Broker</td>
<td>0.149</td>
<td>90.0%</td>
<td><strong>-0.615</strong></td>
</tr>
<tr>
<td>Short-Seller</td>
<td>0.010</td>
<td>50.0%</td>
<td>-0.242</td>
</tr>
</tbody>
</table>

Note: This table presents the horizon of factor timing ability by investor type in the months following portfolio formation. The first column shows the time series covariance between exposures at the beginning of the quarter ($month t \in \{1, 4, 7, 10\}$) and returns in the same month. The second column shows the time series covariance between exposures at the beginning of the quarter and returns over the following month ($month t \in \{2, 5, 8, 11\}$). The third column shows the time series covariance between exposures at the beginning of the quarter and returns in the last month of the quarter ($month t \in \{3, 6, 9, 12\}$). The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with unpermuted factor returns. All covariances have been multiplied by 1,000 for readability. *p<0.1; **p<0.05; ***p<0.01.

Attention Allocations Within Investor Types

The main body of the paper studies allocation differences across investor classes. However, the basic theory can also be tested within investor classes. To do so, investors within an investor class are split into terciles based on $F_{Skill_i}$. Table A.7 presents estimates of the resulting covariance of the aggregate portfolio within each investor class-covariance tercile.\(^1\) By construction, the relationship is monotonic; the estimated attention capacities are increasing as one moves along the terciles.

This also provides a rough outline of the fraction of investors in each investor class that are high and low capacity. Among hedge funds, both the second and third terciles have positive and significant estimates of overall capacity, indicating that, on an aggregated basis, 2/3 of hedge funds provide some sort of added value.

\(^1\)Note that, as dividing investors into terciles requires portfolios at the individual investor level, it is not possible to carry out this analysis for aggregate investor classes. As a result, households and short-sellers are omitted.
Table A.7: Estimated Investor Attention Capacity, \( F_{Skill_i} \)

<table>
<thead>
<tr>
<th></th>
<th>Low Estimate</th>
<th>Low Percentile</th>
<th>2 Estimate</th>
<th>2 Percentile</th>
<th>High Estimate</th>
<th>High Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hedge Fund</td>
<td>-5.036***</td>
<td>0.0%</td>
<td>2.702***</td>
<td>100.0%</td>
<td>4.514***</td>
<td>100.0%</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>-1.423***</td>
<td>0.0%</td>
<td>0.054</td>
<td>70.1%</td>
<td>2.853***</td>
<td>100.0%</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>-0.781***</td>
<td>0.0%</td>
<td>0.148</td>
<td>83.1%</td>
<td>2.755***</td>
<td>100.0%</td>
</tr>
<tr>
<td>Private Banking</td>
<td>-1.776***</td>
<td>0.0%</td>
<td>-0.260</td>
<td>26.4%</td>
<td>2.536***</td>
<td>100.0%</td>
</tr>
<tr>
<td>Long-Term</td>
<td>-3.822***</td>
<td>0.0%</td>
<td>-0.192</td>
<td>5.5%</td>
<td>1.131***</td>
<td>99.9%</td>
</tr>
<tr>
<td>Broker</td>
<td>-2.048***</td>
<td>0.0%</td>
<td>-1.054***</td>
<td>0.0%</td>
<td>3.310***</td>
<td>100.0%</td>
</tr>
</tbody>
</table>

Note: This table presents estimates of attention capacity by investor type. Within each type, investors are split into terciles each date based on estimated attention capacity \( F_{Skill_i} \) and are aggregated into a single portfolio. Attention capacity is captured by \( F_{Skill_i} \), the covariance between an investor’s portfolio and factor returns. The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with unpermuted factor returns. All covariances have been multiplied by 1,000 for readability. *p<0.1; **p<0.05; ***p<0.01.

This is only 1/3 for brokers. Additionally, there is substantial variation across terciles in all investor classes. The bottom (top) tercile is negative (positive) and highly significant. However, it is important to note that there may be significant estimation error in the estimates of individual investors’ covariances given the low frequency and short sample.\(^2\)

Further breaking down this covariance into its time series (\( FT_{iming} \)) and cross sectional (\( FS_{election} \)) components in Tables A.8 and A.9 highlights that almost all the meaningful variation in investors’ ability comes from the time series dimension. Investors’ ability to factor time is the distinguishing feature in determining their overall ability. To emphasize this, only long-term investors exhibit significant variation in their ability to cross sectionally select factors across terciles. Further, while the investors’ factor selecting ability is generally monotonic across terciles, it is not always so, and exhibits relatively little variation.

Finally, in unreported results, the sample has been similarly divided based on

\(^2\)A minimum of 8 consecutive quarters of reporting is required for inclusion in this analysis.
Table A.8: Estimated Investor Timing Ability, $F_{Timing_i}$

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th></th>
<th>High</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
<td>Percentile</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>$-4.175^{***}$</td>
<td>0.0%</td>
<td>$2.589^{***}$</td>
<td>100.0%</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>$-1.326^{***}$</td>
<td>0.0%</td>
<td>$-0.044$</td>
<td>33.8%</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>$-0.681^{***}$</td>
<td>0.0%</td>
<td>0.112</td>
<td>83.0%</td>
</tr>
<tr>
<td>Private Banking</td>
<td>$-1.504^{***}$</td>
<td>0.0%</td>
<td>$-0.426^{*}$</td>
<td>3.8%</td>
</tr>
<tr>
<td>Long-Term</td>
<td>$-3.382^{***}$</td>
<td>0.0%</td>
<td>$-0.265^{***}$</td>
<td>0.0%</td>
</tr>
<tr>
<td>Broker</td>
<td>$-1.823^{***}$</td>
<td>0.0%</td>
<td>$-1.144^{***}$</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Note: This table presents estimates of idiosyncratic attention by investor type. Within each type, investors are split into terciles each date based on estimated attention capacity ($F_{Skill_i}$) and are aggregated into a single portfolio. Idiosyncratic attention is captured by $F_{Timing_i}$, the time series covariance between an investor’s portfolio and factor returns. The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with unpermuted factor returns. All covariances have been multiplied by 1,000 for readability. *p<0.1; **p<0.05; ***p<0.01.

Table A.9: Estimated Investor Selection Ability, $F_{Selection_i}$

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th></th>
<th>High</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
<td>Percentile</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>$-0.861$</td>
<td>17.3%</td>
<td>0.113</td>
<td>58.6%</td>
</tr>
<tr>
<td>Investment Advisor</td>
<td>$-0.098$</td>
<td>11.2%</td>
<td>0.098</td>
<td>88.4%</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>$-0.100$</td>
<td>5.7%</td>
<td>0.037</td>
<td>60.8%</td>
</tr>
<tr>
<td>Private Banking</td>
<td>$-0.272$</td>
<td>8.7%</td>
<td>0.166</td>
<td>70.4%</td>
</tr>
<tr>
<td>Long-Term</td>
<td>$-0.440^{***}$</td>
<td>0.5%</td>
<td>0.073</td>
<td>77.4%</td>
</tr>
<tr>
<td>Broker</td>
<td>$-0.225$</td>
<td>13.4%</td>
<td>0.089</td>
<td>86.0%</td>
</tr>
</tbody>
</table>

Note: This table presents estimates of fundamental attention by investor type. Within each type, investors are split into terciles each date based on estimated attention capacity ($F_{Skill_i}$) the number of previous reports filed and are aggregated into a single portfolio. Fundamental attention is captured by $F_{Selection_i}$, the cross sectional covariance between an investor’s portfolio and factor returns. The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with unpermuted factor returns. All covariances have been multiplied by 1,000 for readability. *p<0.1; **p<0.05; ***p<0.01.
investors’ age and AUM.\textsuperscript{3} Generally speaking, there are no systematic patterns for
this relationship that is consistent across investor classes. In general, one may think
that older investors are able to acquire greater information over time, and thus may
exhibit a greater ability to factor select. I do not find that this is the case. This
provides further justification of the model’s result that investors only learn about the
fundamental at the beginning of the model.

Repricing

The main body focused most analysis on scaled repricing, which accounts for differ-
ences in ownership shares across investors and over time. To emphasize the impor-
tance of this normalization, Figure A.1 plots the raw measure of repricing. From this
figure, one may conclude that investment advisors affect risk premia the most, dwarf-
ing the magnitudes of other investors, including hedge funds. However, adjusting for
ownership shares in Figure A.2 paints a very different picture. On a per-dollar basis,
investment advisors hardly have any meaningful impact on prices. In fact, the mag-
nitude by which they matter is similar to that of long-term investors. Additionally,
the impact of brokers and hedge funds are magnified, reflecting the fact that they are
a relatively small percent of the market.

Portfolio Correlation Across Investors

Proposition 15 formalizes how the similarity of two investors’ portfolios, as measured
by their covariance, depends on their capacities. It states that cross sectional and time
series covariance is generally increasing in the attention capacity of either investor,
as long as the other investor has above average capacity. The intuition for this is

\textsuperscript{3}Age is determined by the length of time between the first 13F filing and the date under consid-
eration.
Figure A.1: Repricing by Investor Type

Note: This figure plots histograms of Repricing by investor type. Repricing is the average difference between an observed return and the counterfactual return that would be observed if a particular class of investors had their AUM proportionally redistributed to other investors. Repricing is defined per characteristic, with each characteristic corresponding to an observation. Units are percentage points per quarter.
Figure A.2: Scaled Repricing by Investor Type

Note: This figure plots histograms of Scaled Repricing by investor type. Scaled Repricing is the average difference between an observed return and the counterfactual return that would be observed if a particular class of investors had their AUM proportionally redistributed to other investors, expressed in percentage points per $1T of redistribution. Scaled Repricing is defined per characteristic, with each characteristic corresponding to an observation.
simple. For two investor classes \(i\) and \(j\), if \(i\) has above average attention capacity, then his portfolio will be positively correlated with realizations. An increase in \(j\)'s capacity increases both (i) the magnitude of his positions (Proposition 14) and (ii) the covariance between his positions and realizations (Proposition 3). This leads to an increase in the overall covariance of the two investors' portfolios both in the time series and cross section.

Table A.10: Time Series Correlation/Covariance Matrix

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>HF</th>
<th>HH</th>
<th>IA</th>
<th>LT</th>
<th>MF</th>
<th>PB</th>
<th>SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broker</td>
<td>-</td>
<td>0.008***</td>
<td>0.004***</td>
<td>0.002***</td>
<td>0.001***</td>
<td>0.002***</td>
<td>0.000</td>
<td>0.007***</td>
</tr>
<tr>
<td>Hedge Fund</td>
<td>0.203***</td>
<td>-0.030***</td>
<td>0.013***</td>
<td>-0.006***</td>
<td>0.011***</td>
<td>-0.012***</td>
<td>-0.028***</td>
<td></td>
</tr>
<tr>
<td>Household</td>
<td>-0.242***</td>
<td>-0.715***</td>
<td>-</td>
<td>-0.006***</td>
<td>0.002***</td>
<td>-0.005***</td>
<td>0.004***</td>
<td>-0.012***</td>
</tr>
<tr>
<td>Inv. Advisor</td>
<td>0.213***</td>
<td>0.644***</td>
<td>-0.694***</td>
<td>-</td>
<td>0.000***</td>
<td>0.003***</td>
<td>0.000***</td>
<td>-0.007***</td>
</tr>
<tr>
<td>Long-Term</td>
<td>0.161***</td>
<td>-0.300***</td>
<td>0.237***</td>
<td>0.115***</td>
<td>-</td>
<td>0.000***</td>
<td>0.004***</td>
<td>-0.002***</td>
</tr>
<tr>
<td>Mutual Fund</td>
<td>0.191***</td>
<td>0.475***</td>
<td>-0.481***</td>
<td>0.620***</td>
<td>0.034**</td>
<td>-</td>
<td>-0.001***</td>
<td>-0.002***</td>
</tr>
<tr>
<td>Private Banking</td>
<td>-0.024</td>
<td>-0.328***</td>
<td>0.223***</td>
<td>0.034**</td>
<td>0.490***</td>
<td>-0.154***</td>
<td>-</td>
<td>0.003***</td>
</tr>
<tr>
<td>Short-Seller</td>
<td>0.129***</td>
<td>-0.212***</td>
<td>-0.208***</td>
<td>-0.228***</td>
<td>-0.085***</td>
<td>-0.072***</td>
<td>0.047***</td>
<td>-</td>
</tr>
</tbody>
</table>

*Note: This table presents the time series correlation (lower triangle) and covariance (upper triangle) between the aggregate portfolios of each investor type. *p<0.1; **p<0.05; ***p<0.01.*

Table A.10 tests this prediction in the time series. In the upper triangle (covariance), it is clear that hedge funds hold positions that covary positively with investment advisors and mutual funds, two investors who exhibit some ability to factor time. Additionally, they covary negatively with long-term investors and short-sellers who exhibit below average timing ability. This seems to affirm the basic predictions from the model.

Examining the cross sectional dimension next, Figure A.3 plots the correlation of investors’ portfolios over time for hedge funds, households, investment advisors, and mutual funds. The picture paints a very stark picture. Households hold very different portfolios than all of these investors, with correlations near \(-1\) from 2005 through 2015. This may be the product of constructing households as the residual sector. On the other hand, investment advisors and mutual funds hold highly similar
portfolios in the cross section, and this has remained very stable over time. Hedge funds hold positions that are generally positively correlated with investment advisors and mutual funds, though, interestingly, this has varied quite significantly over time. For example, hedge funds and mutual funds held portfolios that exhibit a correlation of roughly 0.5 between 2005 and 2013, though this drops to near zero in the early 2000s and post-2015.

**Figure A.3:** Cross Sectional Correlations Between Investors Over Time

![Cross Sectional Correlations](image)

*Note:* This figure plots the cross sectional correlation between select investor types over time. The investor types considered are hedge funds, households, investment advisors, and mutual funds. The cross sectional correlation is computed as the correlation between their exposures at date $t$ to the 55 characteristics considered. NBER recessions are in grey.

**Attention Allocation of Mutual Funds**

Lastly, I repeat the analysis of investors’ attention allocation decisions for mutual funds. The mutual fund data and classifications come from FactSet. There are several caveats in the data. First, the classifications are not constant across time, and there is large variation in the fraction of funds belonging to a classification at a
given date. This is particularly true in the early part of the sample, where the number of funds in a given classification drops near zero seemingly randomly. Second, the classifications are not mutually exclusive. For example, all funds considered belong to the domestic equity group, while funds can simultaneously be classified as large and index funds, for example.

**Table A.11: Estimated Attention Proxies, All Characteristics**

<table>
<thead>
<tr>
<th></th>
<th>Attention Proxy</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>F Skill&lt;sub&gt;i&lt;/sub&gt;</strong></td>
<td><strong>FTiming&lt;sub&gt;i&lt;/sub&gt;</strong></td>
<td><strong>F Selection&lt;sub&gt;i&lt;/sub&gt;</strong></td>
</tr>
<tr>
<td></td>
<td>Estimate</td>
<td>Percentile</td>
<td>Estimate</td>
</tr>
<tr>
<td>DomEq</td>
<td>1.089***</td>
<td>100.0%</td>
<td>0.984***</td>
</tr>
<tr>
<td>ETF</td>
<td>0.255</td>
<td>76.8%</td>
<td>0.229</td>
</tr>
<tr>
<td>Index</td>
<td>−0.150</td>
<td>19.2%</td>
<td>−0.185</td>
</tr>
<tr>
<td>Growth</td>
<td>1.982***</td>
<td>99.5%</td>
<td>1.468***</td>
</tr>
<tr>
<td>Value</td>
<td>1.995**</td>
<td>98.9%</td>
<td>2.005***</td>
</tr>
<tr>
<td>Small</td>
<td>2.335</td>
<td>93.9%</td>
<td>2.052**</td>
</tr>
<tr>
<td>Large</td>
<td>0.772**</td>
<td>99.3%</td>
<td>0.647**</td>
</tr>
<tr>
<td>ESG</td>
<td>0.802</td>
<td>92.0%</td>
<td>0.747*</td>
</tr>
</tbody>
</table>

*Note: This table presents estimates of attention capacity and allocations by investor type. F Skill<sub>i</sub> is the covariance between an investor’s portfolio and factor returns, while FTiming<sub>i</sub> and F Selection<sub>i</sub> decompose this covariance into time series and cross sectional covariance, respectively. The percentile is determined by randomly permuting investors’ scores 1,000 times and computing the covariance of the permuted scores with unpermuted factor returns. All covariances have been multiplied by 1,000 for readability. *p<0.1; **p<0.05; ***p<0.01.*

Nevertheless, meaningful takeaways can still be made. Table A.11 presents estimates of mutual funds’ capacities and allocations by type. Overall, domestic equity mutual funds tend to provide added value, and this is concentrated in their ability to factor time. This effect is concentrated primarily among funds that specialize in size and book-to-market strategies. Interestingly, ESG funds also display an ability to factor time, though the magnitude is generally lower.

As one would expect, passive funds (index and ETFs) display no ability to time. As these funds are passive, they should not allocate attention to idiosyncratic payoffs

---

4While some ETFs are active, the vast majority tend to be passive funds.
because their mandate forbids them from drastically updating their portfolio over
time. As a result, their time series variation is very low, as is their covariance between
portfolios and idiosyncratic returns.

Across all of the mutual fund classifications considered, there is no systematic
ability to select factors. This is consistent with the results in the main text. Investors
who have above average attention capacities devote these resources to factor timing
rather than factor selection.

A.1.2 Inter-Asset Attention Allocation

The main body of the paper considered the attention allocation problem with a single
risk factor. The case with an arbitrary number of ex-ante identical and independent
risk factors is equivalent; in this case, investors will allocate an equal amount of
attention across factors. In this section, the model is generalized to \( N \) independent,
non-identically distributed risk factors. The payoff of each risk factor is given by Eq.
(2.2). Letting the subscript \( j \) denote the quantities of risk factor \( j \), define \( K_{i,j} \) as the
total attention allocated to a risk factor \( j \): \( K_{i,j} = K_{\mu,i,j} + TK_{\epsilon,i,j} \). The inter-asset
attention allocation problem chooses \( K_{i,j} \) given the intra-asset attention allocation:

\[
\max_{K_{i,j}} \quad \frac{T}{2} \sum_{j=1}^{N} \hat{\sigma}_{f,i,j}^{-1} \left\{ 2b_j \hat{\sigma}_{\mu,i,j} + 2c_j \hat{\sigma}_{\epsilon,i,j} + v_j + a_j^2 \right\} - \frac{TN}{2}
\]

s.t. \( K_{m_i} \geq \sum_{j=1}^{N} K_{i,j} \) \hspace{1cm} (A.7)

\( K_{i,j} \geq 0 \quad \forall j = [1, \ldots, N] \)

\( K_{\mu,i,j} = K_{\mu,i,j}^* \quad \text{and} \quad K_{\epsilon,i,j} = K_{\epsilon,i,j}^* \quad \forall j = [1, \ldots, N] \).

The second constraint extends the no-forgetting constraint to the inter-asset attention
problem, while the last constraint ensures that investors optimally allocate attention
between the fundamental and idiosyncratic components for each asset.

Generally, this problem does not yield closed form solutions. However, for a specific parametrization of fundamental and idiosyncratic volatilities, the solution can be characterized by a waterfilling equilibrium à la Kacperczyk et al. (2016). The specific parametrization is that the variance ratio is equal to $\sqrt{T}$ for all risk factors.

**Assumption 4.** The variance ratio $\sigma_{i,j}/\sigma_{\mu,j} = \sqrt{T}$ for all risk factors $j$.

Given Assumption 4, the objective in the investor’s inter-asset attention allocation problem (A.7) is linear in $K_{i,j}$. This is because $b_j = c_j$ for all assets by Proposition 1. Further, the aggregate quantities can be simplified:

$$\sigma_{f,j}^{-1} = \frac{(\sqrt{T} + 1)^2}{\bar{K}_j + \sigma_{\mu,j}^{-1} (\sqrt{T} + 1)}$$

(A.8)

$$b_j = c_j = \frac{\bar{K}_j}{\bar{K}_j + \sigma_{\mu,j}^{-1} (\sqrt{T} + 1)}$$

The new objective becomes

$$\mathbb{E}[U_i] = \frac{T}{2} \sum_{j=1}^{N} 2\bar{K}_j + \left(K_{i,j} + \sigma_{\mu,j}^{-1} (\sqrt{T} + 1)\right) \left(\sqrt{T} + 1\right) \left(\sigma_{\mu,j}^3 + \gamma^2 \left(s_j^2 + \sigma_s\right) \left(\sqrt{T} + 1\right)\right) \frac{\bar{K}_j + \sigma_{\mu,j}^{-1} (\sqrt{T} + 1)}{\bar{K}_j + \sigma_{\mu,j}^{-1} (\sqrt{T} + 1)},$$

(A.9)

where $\bar{K}_j = \sum_m \psi_m K_{m,j}$ is the average attention allocated to a given asset. Note that this is linear in $K_{i,j}$ and can be written as

$$\mathbb{E}[U_i] = \sum_{j=1}^{N} \tau_j + v_j K_{i,j},$$

(A.10)
where

\[
\tau_j \equiv \frac{2\bar{K}_j + \left(\sigma^{-1}_{\mu,j} \left(\sqrt{T} + 1\right)\right) \left(\sqrt{T} + 1\right) \left(\sigma^{-3}_{\mu,j} + \gamma^2 \left(\bar{s}_j^2 + \sigma_s\right) \left(\sqrt{T} + 1\right)\right)}{\bar{K}_j + \sigma^{-1}_{\mu,j} \left(\sqrt{T} + 1\right)}
\]

\[
u_j \equiv \frac{\left(\sqrt{T} + 1\right) \left(\sigma^{-3}_{\mu,j} + \gamma^2 \left(\bar{s}_j^2 + \sigma_s\right) \left(\sqrt{T} + 1\right)\right)}{\bar{K}_j + \sigma^{-1}_{\mu,j} \left(\sqrt{T} + 1\right)}.
\tag{A.11}
\]

The resulting attention allocation problem is linear in the attention allocated to each risk factor. This means that investors will allocate attention only to the factors with the largest marginal benefit (i.e., the factors with the highest value of \(\nu_j\)). This will generally be the assets with the highest variance; however, the specific marginal utility of attention depends on the actions of all other investors in the economy. This is known as a “waterfilling” equilibrium: only collective attention decisions are pinned down, while those of individual investors are not. This is similar to the attention allocation problem in Kacperczyk et al. (2016).

Without Assumption 4, the attention allocation equilibrium can be examined numerically. Figure A.4 plots the amount of attention allocated to risk factor 1 in two agent, two factor economy as a function of the overall volatility of factor 1 relative to that of factor 2. That is, a value of 0.5 on the x-axis indicates that factor 1 is half as volatile as factor 2 (the variance ratio is the same in both assets). It is clear that the amount of attention is increasing in total volatility; as factor 1 becomes more volatile, the total attention allocated to it by both agents increases. When the factors are identical, attention is divided evenly among the two factors. On a relative basis, the higher capacity investor allocates more to the less risky asset, consistent with the intuition in Proposition 2.
Figure A.4: Inter-Asset Attention Allocation

This figure illustrates the inter-asset attention allocation. The x-axis varies the ratio of total volatility \((\sigma_{\mu,j} + \sigma_{\epsilon,j})\) across the risk factors. The left panel displays the total amount of attention allocated to factor 1 by each of the investors, while the right panel displays the amount of attention allocated to factor 1 as a fraction of each investor’s total attention capacity. The parameters are listed in Table A.16.

Table A.12: Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T)</td>
<td>2</td>
<td>Time horizon</td>
</tr>
<tr>
<td>(\sigma_{\mu})</td>
<td>0.4</td>
<td>Fundamental volatility</td>
</tr>
<tr>
<td>(\sigma_{\epsilon})</td>
<td>0.4</td>
<td>Idiosyncratic volatility</td>
</tr>
<tr>
<td>(\bar{s})</td>
<td>0.2</td>
<td>Average supply</td>
</tr>
<tr>
<td>(\sigma_s)</td>
<td>0.1</td>
<td>Supply volatility</td>
</tr>
<tr>
<td>(\psi)</td>
<td>(0.2, 0.8)</td>
<td>Mass of investor types</td>
</tr>
<tr>
<td>(K_m)</td>
<td>(1, 0.5)</td>
<td>Attention capacity of investor types</td>
</tr>
<tr>
<td>(\gamma)</td>
<td>0.3</td>
<td>Risk aversion coefficient</td>
</tr>
<tr>
<td>(r)</td>
<td>0</td>
<td>Risk-free rate</td>
</tr>
</tbody>
</table>

This table lists the parameter values used in model simulations.

A.1.3 Proofs and Derivations

This Appendix first presents useful definitions and notation used in the proofs. The solution to the model is then derived. Next, additional theoretical results are stated...
and, finally, all the proofs of all theoretical results are stated.

Useful Notation

Average posterior precision: \( \bar{\sigma}_{f,t}^{-1} = \int_0^1 \hat{\sigma}_{f,i,t}^{-1} di = \sum_m \psi_m \hat{\sigma}_{f,m,t}^{-1} \).

Sum of precisions shorthand: \( \delta_i \equiv \hat{\sigma}_{\mu,i}^{-1} + T \hat{\sigma}_{\epsilon,i}^{-1} = \bar{K}_m + \sigma^{-1}_\mu + T \sigma^{-1}_\epsilon \).

Bound notation: Many of the proofs rely on ensuring that \( \zeta \) does not become too large. To see how this can be guaranteed, consider the definition of \( \zeta \) after substituting for \( v \) and \( a \):

\[
\zeta = \frac{2 (T - 1) (c - b)}{(1 - b)^2 \sigma_\mu + (1 - c)^2 \sigma_\epsilon + \gamma^2 \sigma_f^2 (s^2 + \sigma_s)}.
\]  

(A.12)

As clear from the definition of \( \zeta \), this can be ensured by assuming \( s^2 + \sigma_s \) is sufficiently large, as in the limit as \( s^2 + \sigma_s \to \infty \), \( \zeta \to 0 \). A more natural restriction is on the maximum attention capacity, \( \bar{K} = \max_m K_m \). To see this, define \( \bar{\ell}(x) \) (\( \ell(x) \)) to be the minimum values of \( b + c \) such that the maximum (minimum) value of \( \zeta \) is less (greater) than \( x \), respectively:

\[
\bar{\ell}(x) = \min \left\{ \ell \leq 2 : x \geq \arg \max_{b+c \leq \ell} \zeta \right\}
\]  

(A.13)

\[
\ell(x) = \min \left\{ \ell \leq 2 : x \geq \arg \min_{b+c \leq \ell} \zeta \right\}.
\]  

(A.14)

Finally, define \( \ell(x) \) to be the minimum value of \( b + c \) such that the maximum value of \( |\zeta| \) is less than \( x \):

\[
\ell(x) = \min \left\{ \ell \leq 2 : x \geq \arg \max_{b+c \leq \ell} |\zeta| \right\}
\]  

(A.15)
The price coefficients $b$ and $c$ can be bounded by the action of the investor with the
greatest attention capacity. For some $K_\mu$,

\[ b + c \leq \frac{K_\mu}{\sigma^{-1}_\mu + K_\mu} + \frac{K - K_\mu}{T\sigma^{-1}_\epsilon + \bar{K} - K_\mu}. \] (A.16)

This bound is concave in $K_\mu$. Thus, any bound on $b+c$ can be guaranteed by imposing
a bound on the maximum attention capacity $\bar{K}$. This is done by finding the value
of $K_\mu$ that maximizes this bound, then setting the resulting bound equivalent to the
bound on $b + c$. First, the derivative with respect to $K_\mu$ is

\[ \frac{\sigma^{-1}_\mu}{(\sigma^{-1}_\mu + K_\mu)^2} - \frac{T\sigma^{-1}_\epsilon}{(T\sigma^{-1}_\epsilon + \bar{K} - K_\mu)^2}. \] (A.17)

Setting this to zero and solving for $K_\mu$ yields two roots

\[ K_\mu = \frac{\bar{K}\sigma_\mu \sigma_\epsilon^2 + 2T\sigma_\mu \sigma_\epsilon \pm (\bar{K}\sigma_\mu \sigma_\epsilon + \sigma_\epsilon)^2 \sqrt{T\sigma_\mu \sigma_\epsilon}}{\sigma_\mu \sigma_\epsilon^2 - T\sigma_\mu^2 \sigma_\epsilon}. \] (A.18)

The higher root is greater than $\bar{K}$, which cannot occur due to the no-forgetting
constraints.\(^5\) Evaluating the bound at the lower root yields the bound

\[ b + c \leq \frac{2\bar{K}\sigma_\mu \sigma_\epsilon + T\sigma_\mu + \sigma_\epsilon - 2\sqrt{T\sigma_\mu \sigma_\epsilon}}{\bar{K}\sigma_\mu \sigma_\epsilon + T\sigma_\mu + \sigma_\epsilon}. \] (A.19)

Therefore, a sufficient condition to ensure that $b + c \leq \ell$ is

\[ \bar{K} \leq \frac{(\ell - 1) (T\sigma_\mu + \sigma_\epsilon) + 2\sqrt{T\sigma_\mu \sigma_\epsilon}}{(2 - \ell) \sigma_\mu \sigma_\epsilon} \equiv g(\ell) \] (A.20)

\(^5\)The lower root may be negative. However, as the bound is concave, enforcing the no-forgetting
constraint only yields a tighter bound on $b + c$. 

154
Model Solution

The solution is derived in three steps. First, prices are computed using the market clearing condition. Then, the ex-ante expected utility is derived. Finally, the attention allocation problem is solved. We proceed in turn.

Prices: Plugging optimal demand (2.14) into the market clearing condition (2.9) yields

$$\int_0^1 \gamma^{-1} \hat{\sigma}_{f,t}^{-1} \left( \hat{f}_{i,t} - r p_t \right) \, di = \bar{s} + s_t. \tag{A.21}$$

Solving for prices $p_t$ yields

$$p_t = \frac{1}{r} \left[ -\gamma \hat{\sigma}_{f,t} (\bar{s} + s_t) + \left( \hat{\sigma}_{f,t} \int_0^1 \hat{\sigma}_{f,i,t} \hat{\sigma}_{\mu,i,t} \sum_{s=1}^t K_{\mu,i,s} \, di \right) \mu + \left( \hat{\sigma}_{f,t} \int_0^1 \hat{\sigma}_{f,i,t} \hat{\sigma}_{\epsilon,i,t} K_{\epsilon,i,t} \, di \right) \epsilon_i \right]$$

$$= \frac{1}{r} \left[ a_t + b_t \mu + c_t \epsilon_i + d_t s_t \right], \tag{A.22}$$

where

$$a_t = -\gamma \hat{\sigma}_{f,t} \bar{s}$$

$$b_t = \hat{\sigma}_{f,t} \int_0^1 \hat{\sigma}_{f,i,t} \hat{\sigma}_{\mu,i,t} \sum_{s=1}^t K_{\mu,i,s} \, di$$

$$c_t = \hat{\sigma}_{f,t} \int_0^1 \hat{\sigma}_{f,i,t} \hat{\sigma}_{\epsilon,i,t} K_{\epsilon,i,t} \, di$$

$$d_t = -\gamma \hat{\sigma}_{f,t}.$$  

The price coefficients $b_t$ and $c_t$ are weighted averages of $\hat{\sigma}_{\mu,i,t} \sum_{s=1}^t K_{\mu,i,s}$ and $\hat{\sigma}_{\epsilon,i,t} K_{\epsilon,i,t}$, respectively, with weights $\hat{\sigma}_{f,i,t}^{-1}$. As the attention allocation is symmetric across agents of the same type, denote by the subscript $i = m$ the relevant quantity of type $m$.
investors. Then, $b_t$ and $c_t$ are given by

$$
\begin{align*}
    b_t &= \bar{\sigma}_{f,t} \sum_m \left( \psi_m \bar{\sigma}_{f,m,t}^{-1} \hat{\sigma}_{\mu,m,t} \sum_{s=1}^{t} K_{\mu,m,s} \right) \\
    c_t &= \bar{\sigma}_{f,t} \sum_m \left( \psi_m \bar{\sigma}_{f,m,t}^{-1} \hat{\sigma}_{\epsilon,m,t} K_{\epsilon,m,t} \right),
\end{align*}
$$

\text{(A.24)}

in which case $b_t$ and $c_t$ are weighted averages of the type $m$ quantity with weights $\psi_m \bar{\sigma}_{f,m,t}^{-1}$.

Finally, Lemma 2 and Corollary 1 assert that (i) posterior variances $\hat{\sigma}_{\mu,i,t}$, $\hat{\sigma}_{\epsilon,i,t}$, and $\hat{\sigma}_{f,i}$ are constant, (ii) the posterior mean of the fundamental $\hat{\mu}_{i,t}$ is constant, and (iii) the price coefficients $a_t$, $b_t$, $c_t$, and $d_t$ are constant. As a result, time subscripts will be dropped where applicable. The resulting price is given by

$$
p_t = \frac{1}{r} [a + b\mu + c\epsilon_t + d\sigma_t].
$$

\text{(A.25)}

**Ex-ante expected utility:** Simplifying the expected utility $\mathbb{E} [u_{i,t}]$,

$$
\begin{align*}
    \mathbb{E} [u_{i,t}] &= \mathbb{E} \left[ \mathbb{E}_{i,t} [W_{i,t}] - \frac{\gamma}{2} \mathbb{V}_{i,t} [W_{i,t}] \right] \\
    &= \mathbb{E} \left[ \mathbb{E}_{i,t} \left[ \hat{\sigma}_{f,i} \left( \hat{f}_{i,t} - rp_t \right) (f_t - rp_t) \right] - \frac{1}{2} \mathbb{V}_{i,t} \left[ \hat{\sigma}_{f,i} \left( \hat{f}_{i,t} - rp_t \right) (f_t - rp_t) \right] \right] \\
    &= \mathbb{E} \left[ \hat{\sigma}_{f,i} \left( \hat{f}_{i,t} - rp_t \right)^2 - \frac{1}{2} \hat{\sigma}_{f,i}^2 \left( \hat{f}_{i,t} - rp_t \right)^2 \mathbb{V}_{i,t} [f_t - rp_t] \right] \\
    &= \frac{1}{2} \mathbb{E} \left[ \hat{\sigma}_{f,i}^{-1} \left( \hat{f}_{i,t} - rp_t \right)^2 \right] \\
    &= \frac{1}{2} \mathbb{E} \left[ m_t^2 \right],
\end{align*}
$$

\text{(A.26)}

where $m_t = \hat{\sigma}_{f,i}^{-1/2} \left( \hat{f}_{i,t} - rp_t \right)$.

Note that $m_t$ is normally distributed, as $\hat{f}_{i,t}$ and $p_t$ are linear functions of normally distributed random variables. Thus, $m_t^2$ is a non-central $\chi^2$ distributed random

156
variable. Its expectation is given by
\[
\mathbb{E} [m_t^2] = \mathbb{V} [m_t] + \mathbb{E} [m_t]^2.
\] (A.27)

To compute \( \mathbb{E} [m_t^2] \), first consider realized returns
\[
f_t - r p_t = (1 - b) \mu + (1 - c) \epsilon_t - ds_t - a,
\] (A.28)
which is the sum of three mean-zero independent normal random variables and a constant. Therefore, \( f_t - r p_t = v^{-1/2} \delta_t - a \), where \( \delta_t \) is an iid standard normal random variable and \( v \) is
\[
v = (1 - b)^2 \sigma_\mu + (1 - c)^2 \sigma_\epsilon + d^2 \sigma_s.
\] (A.29)

By the law of iterated expectations, \( \mathbb{E} [m_t] = -\hat{\sigma}_{f,t}^{-1} a \).

Next, the variance of investor \( i \)'s expected returns can be decomposed into the variance of investor \( i \)'s forecast error, the variance of realized returns, and the covariance between the two:
\[
\mathbb{V} \left[ \hat{f}_{i,t} - r p_t \right] = \mathbb{V} \left[ \hat{f}_{i,t} - f_t + f_t - r p_t \right]
= \mathbb{V} \left[ \hat{f}_{i,t} - f_t \right] + \mathbb{V} [f_t - r p_t] + 2 \mathbb{C} \left[ \hat{f}_{i,t} - f_t, f_t - r p_t \right]
\] (A.30)
Evaluating each component in turn, first consider the variance of the forecast error

\[ \mathbb{V} \left[ \hat{f}_{i,t} - f_t \right] = \mathbb{V} \left[ \hat{\sigma}_{\mu,i} \hat{K}_{\mu,i} \theta_{\mu,i,1} + \hat{\sigma}_{\epsilon,i} \hat{K}_{\epsilon,i} \theta_{\epsilon,i,t} - \mu - \epsilon_t \right] \]

\[ = \mathbb{V} \left[ \hat{\sigma}_{\mu,i} \hat{K}_{\mu,i} (\mu + \xi_{\mu,i,1}) + \hat{\sigma}_{\epsilon,i} \hat{K}_{\epsilon,i} (\epsilon_t + \xi_{\epsilon,i,t}) - \mu - \epsilon_t \right] \]

\[ = \mathbb{V} \left[ \hat{\sigma}_{\mu,i} \left( K_{\mu,i} - \hat{\sigma}_{\mu,i}^{-1} \right) \mu + \hat{\sigma}_{\mu,i} K_{\mu,i} \xi_{\mu,i,1} + \hat{\sigma}_{\epsilon,i} \left( K_{\epsilon,i} - \hat{\sigma}_{\epsilon,i}^{-1} \right) \epsilon_t + \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} \xi_{\epsilon,i,t} \right] \]

\[ = \mathbb{V} \left[ -\hat{\sigma}_{\mu,i} \sigma_{\mu}^{-1} \mu + \hat{\sigma}_{\mu,i} K_{\mu,i} \xi_{\mu,i,1} - \hat{\sigma}_{\epsilon,i} \sigma_{\epsilon}^{-1} \epsilon_t + \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} \xi_{\epsilon,i,t} \right] \]

\[ = \hat{\sigma}_{\mu,i} \sigma_{\mu}^{-1} + \hat{\sigma}_{\mu,i}^{2} K_{\mu,i} \sigma_{\epsilon}^{-1} \sigma_{\epsilon}^{-1} + \hat{\sigma}_{\epsilon,i}^{2} K_{\epsilon,i} \sigma_{\epsilon}^{-1} + \hat{\sigma}_{\epsilon,i} \sigma_{\epsilon}^{-1} \]

\[ = \hat{\sigma}_{\mu,i} + \hat{\sigma}_{\epsilon,i}, \quad (A.31) \]

where the second line follows from the signal structure (2.5) and the fifth line follows from independence of the random variables.

The variance of realized returns \( f_t - r p_t \) is \( v \), calculated above. The covariance is given by

\[ \mathbb{C} \left[ \hat{f}_{i,t} - f_t, f_t - r p_t \right] = \mathbb{C} \left[ -\hat{\sigma}_{\mu,i} \sigma_{\mu}^{-1} \mu, (1 - b) \mu \right] + \mathbb{C} \left[ -\hat{\sigma}_{\epsilon,i} \sigma_{\epsilon}^{-1} \epsilon_t, (1 - c) \epsilon_t \right] \]

\[ = -\hat{\sigma}_{\mu,i} (1 - b) - \hat{\sigma}_{\epsilon,i} (1 - c). \quad (A.32) \]

Putting these terms together, the variance is

\[ \mathbb{V} \left[ \hat{f}_{i,t} - r p_t \right] = (2b - 1) \hat{\sigma}_{\mu,i} + (2c - 1) \hat{\sigma}_{\epsilon,i} + v. \quad (A.33) \]

Thus, the expected utility is given by

\[ \mathbb{E} \left[ u_{i,t} \right] = \frac{1}{2} \left( \mathbb{V} \left[ m_t \right] + \mathbb{E} \left[ m_t \right] \right)^2 \]

\[ = \frac{1}{2} \hat{\sigma}_{f,i}^{-1} \left\{ (2b - 1) \hat{\sigma}_{\mu,i} + (2c - 1) \hat{\sigma}_{\epsilon,i} + v + a^2 \right\} \quad (A.34) \]

\[ = \frac{1}{2} \hat{\sigma}_{f,i}^{-1} \left\{ 2b \hat{\sigma}_{\mu,i} + 2c \hat{\sigma}_{\epsilon,i} + v + a^2 \right\} - \frac{1}{2}, \]

158
which is the equation (2.17).

**Attention allocation:** Plugging the expected utility (2.17) into the attention allocation problem (2.8) and using Lemma 2 and Corollary 1 yields the problem:

\[
\max_{\{K_{\mu,i}, K_{\epsilon,i}\}} \frac{1}{2} T \hat{\sigma}_{f,i}^{-1} \left\{ 2b\hat{\sigma}_{\mu,i} + 2c\hat{\sigma}_{\epsilon,i} + v + a^2 \right\} - \frac{1}{2} T
\]

s.t. \( K_{\mu,i} \geq K_{\mu,i} + T K_{\epsilon,i} \) \hspace{1cm} (A.35)

\( K_{\mu,i,t}, K_{\epsilon,j,t} \geq 0 \ \forall j \in \{1, \ldots, N\}, \ t \in \{1, \ldots, T\} \).

The first order conditions are given by

\[
[K_{\mu,i}] : \frac{1}{2} T \hat{\sigma}_{\mu,i}^2 \hat{\sigma}_{f,i}^{-2} \left\{ 2(c - b) \hat{\sigma}_{\epsilon,i} + v + a^2 \right\} - \lambda_i + \nu_{\mu,i} = 0
\]
\[
[K_{\epsilon,i}] : \frac{1}{2} T \hat{\sigma}_{\epsilon,i}^2 \hat{\sigma}_{f,i}^{-2} \left\{ 2(b - c) \hat{\sigma}_{\mu,i} + v + a^2 \right\} - T\lambda_i + \nu_{\mu,i} = 0
\]

(A.36)

If the no-forgetting constraints are not binding, combining these FOCs yield

\[
2 (b - c) \left( \hat{\sigma}_{\mu,i}^{-1} + T \hat{\sigma}_{\epsilon,i}^{-1} \right) + (v + a^2) \left( \hat{\sigma}_{\mu,i}^{-2} - T \hat{\sigma}_{\epsilon,i}^{-2} \right) = 0.
\]

(A.37)

To solve this equation for fundamental and idiosyncratic attention, it is necessary that the information capacity constraint binds. Lemma 6 establishes that this holds under all parametrizations.

**Lemma 6.** The information capacity constraint binds for all investors in any optimal attention allocation: \( K_{m_i} = K_{\mu,i} + T K_{\epsilon,i} \ \forall i. \)

Using the definitions of \( \hat{\sigma}_{\mu,i}^{-1} \) and \( \hat{\sigma}_{\epsilon,i}^{-1} \) in Eqs. (2.11) and (2.12) and substituting...
\( K_{\mu,i} = K_{m_i} - TK_{\epsilon,i}, \) this condition is

\[
\begin{align*}
2 (b - c) \left( K_{m_i} + \sigma_{\mu}^{-1} + T \sigma_{\epsilon}^{-1} \right) + \\
(v + a^2) \left( T (T - 1) K_{\epsilon,i}^2 - 2T \left( K_{m_i} - \sigma_{\mu}^{-1} - \sigma_{\epsilon}^{-1} \right) K_{\epsilon,i} + \left( K_{m_i} + \sigma_{\mu}^{-1} \right)^2 - T \sigma_{\epsilon}^{-2} \right) &= 0,
\end{align*}
\]

(A.38)

which is quadratic in \( K_{\epsilon,i} \). Applying the quadratic formula yields

\[
K_{\epsilon,i} = \frac{K_{m_i} + \sigma_{\mu}^{-1} + \sigma_{\epsilon}^{-1} \pm \sqrt{T^{-1} \left( K_{m_i} + \sigma_{\mu}^{-1} + T \sigma_{\epsilon}^{-1} \right) \left( K_{m_i} + \sigma_{\mu}^{-1} + T \sigma_{\epsilon}^{-1} + \frac{2(T-1)(c-b)}{v+a^2} \right)}}{T - 1}.
\]

(A.39)

Note that the positive root has \( K_{\epsilon,i} \geq \frac{K_{m_i}}{T-1} \), which implies \( K_{\mu,i} \leq -\frac{K_{m_i}}{T-1} < 0 \), violating the no-forgetting constraint for fundamental attention. Thus, the lower root (as stated in Eq. (2.21)) is the only candidate interior solution. Fundamental attention is \( K_{\mu,i} = K_{m_i} - TK_{\epsilon,i} \) and is given in Eq. (2.20). To confirm that this solution (when feasible) is optimal, it suffices to show that the objective function in Eq. (A.35) is strictly concave in \( K_{\mu,i} \) and \( K_{\epsilon,i} \). The second derivative with respect to each is

\[
\begin{align*}
\frac{\partial^2 \mathbb{E} [U_i]}{\partial K_{\mu,i}^2} &= -2\hat{\sigma}_{\mu,i}^3 \hat{\sigma}_{\epsilon,i} \hat{\sigma}_{\epsilon,i}^{-1} \{2 (c - b) \hat{\sigma}_{\epsilon,i} + v + a^2 \} < 0 \\
\frac{\partial^2 \mathbb{E} [U_i]}{\partial K_{\epsilon,i}^2} &= -2\hat{\sigma}_{\mu,i} \hat{\sigma}_{\epsilon,i} \hat{\sigma}_{\epsilon,i}^{-1} \{2 (b - c) \hat{\sigma}_{\mu,i} + v + a^2 \} < 0,
\end{align*}
\]

(A.40)

where both inequalities hold under Assumption 1 (see Lemma 2). Thus, if the interior solution is feasible, it is optimal.

**Additional Results**

In this section, technical results useful in proofs and additional results about the attention allocation are presented. The first result imposes bounds on the values of \( \zeta \) for which no-forgetting constraints do not bind for investor \( i \).
Lemma 7. Given Assumption 1, if the no-forgetting constraint for investor \(i\) does not bind, then

\[
\zeta \in \left( \frac{(T-1) \left( T \sigma_{\mu}^{-2} - (K_{m} + T \sigma_{\epsilon}^{-1}) \right)}{T \left( K_{m} + \sigma_{\mu}^{-1} + T \sigma_{\epsilon}^{-1} \right)}, \frac{(T-1) \left( (K_{m} + \sigma_{\mu}^{-1})^{2} - T \sigma_{\epsilon}^{-2} \right)}{K_{m} + \sigma_{\mu}^{-1} + T \sigma_{\epsilon}^{-1}} \right)
\]  

(A.41)

Additional results concerning the aggregate variance of a given investor type and the covariance of two investors’ portfolios are below.

Proposition 14. Suppose Assumption 1 holds and attention capacity is sufficiently uniformly bounded. The cross sectional portfolio variance of the representative investor is increasing in attention capacity if \(K_{m} > \bar{K}_{\mu}\) and decreasing otherwise. Time series portfolio variance of the representative investor is increasing if \(K_{m} \geq \bar{K}_{\epsilon}\) and, for sufficiently small \(\sigma_{s}\), decreasing otherwise.

Proposition 15. Suppose Assumption 1 holds and attention capacity is sufficiently uniformly bounded. Cross sectional portfolio covariance \(\mathbb{C}^{CS}[q_{i,t}, q_{j,t}]\) is increasing in investor \(i\)’s attention capacity if \(K_{m_{j}} > \bar{K}_{\mu}\) and decreasing otherwise. Further, cross sectional portfolio covariance is negative only if \(K_{m_{i}} < \bar{K}_{\mu} < K_{m_{j}}\) or vice versa.

Similarly, time series portfolio covariance \(\mathbb{C}^{TS}[q_{i,t}, q_{j,t}]\) is increasing in investor \(i\)’s attention capacity if \(K_{m_{j}} \geq \bar{K}_{\epsilon}\) and, for sufficiently small \(\sigma_{s}\), decreasing otherwise.

The next result simply shows that both fundamental and idiosyncratic attention are increasing in attention capacity. While Lemma 6 ensures the information capacity constraint is binding, it does not ensure that neither type of attention decreases.

Proposition 16. Given Assumption 1, fundamental and idiosyncratic attention are weakly increasing in attention capacity, with at least one strict inequality. Both inequalities are strict if the no-forgetting constraints do not bind.
The last result compares attention allocations across economies by varying the volatility of fundamental and idiosyncratic components. Fundamental attention rises when fundamental volatility rises, and the same pattern holds for idiosyncratic attention.

**Proposition 17.** Given Assumption 1, for attention capacity sufficiently uniformly bounded, fundamental attention is increasing in the variance of the fundamental \( \sigma_\mu \) and decreasing in the variance of the idiosyncratic component \( \sigma_\epsilon \).

**Proofs**

**Proof of Lemma 1:**

*Proof.* Since the costs to learning from prices are equivalent to those of obtaining a private signal, the decision can be recast as two independent problems:

1. An investor has the option of (i) learning from prices about \( \mu \) (and thus obtaining an unbiased signal with precision \( K_{p,\mu,t} \)) and a conditionally independent signal with precision \( \tilde{K}_{\mu,i,t} \), or (ii) not learning from prices and obtaining a conditionally independent signal with precision \( K_{\mu,i,t} = K_{p,\mu,t} + \tilde{K}_{\mu,i,t} \).

2. An investor has the option of (i) learning from prices about \( \epsilon \) (and thus obtaining an unbiased signal with precision \( K_{p,\epsilon,t} \)) and a conditionally independent signal with precision \( \tilde{K}_{\epsilon,i,t} \), or (ii) not learning from prices and obtaining a conditionally independent signal with precision \( K_{\epsilon,i,t} = K_{p,\epsilon,t} + \tilde{K}_{\epsilon,i,t} \).

Note that the problems are independent because investors update beliefs independently of other risk factors. If, in both scenarios, investor \( i \)'s utility under option (ii) is higher than under option (i), the investor never chooses to learn from prices.
Recall ex-ante expected utility is given by Eq. (2.17)

\[ \mathbb{E} [u_{i,t}] = \frac{1}{2} \left\{ \hat{\sigma}^{-2}_{f,i,t} \mathbb{V} \left[ \hat{f}_{i,t} - r_{p_t} \right] + \hat{\sigma}^{-1}_{f,i,t} \mathbb{E} \left[ \hat{f}_{i,t} - r_{p_t} \right]^2 \right\} \]  

(A.42)

The only term that is different between options (i) and (ii) is the variance of investor \( \hat{i} \)'s beliefs about returns \( \mathbb{V} \left[ \hat{f}_{i,t} - r_{p_t} \right] \). This is because either option yields equally informative posteriors and unbiased beliefs. Hence, the posterior variance \( \hat{\sigma}_{f,i} = \hat{\sigma}_{\mu,i} + \hat{\sigma}_{\epsilon,i} \) and a priori expectations \( \mathbb{E} \left[ \hat{f}_{i,t} - r_{p_t} \right] \) are identical under the two options. Further, there are no downstream effects of the signal choice in subsequent periods. For scenario (1), while the signals in time \( t \) affect the beliefs about \( \mu \) in all future periods, prices are independent conditional on \( \mu \) across time (the signal from prices is correlated with the idiosyncratic component \( \epsilon_t \) and the supply shock \( s_t \), both of which are \( iid \)). As a result, the two scenarios are equivalent for expected utility in all future periods. For scenario (2), the signals do not change downstream beliefs as \( \epsilon_t \) is \( iid \), so the decision has no effect on future utility.

**Scenario 1: Learning about \( \mu \).** Because prices are of the linear form \( p_t = \frac{1}{r} [a_t + b_t \mu + c_t \epsilon_t + d_t s_t] \), an unbiased signal for \( \mu \) can be obtained via the transformation

\[ \theta_{p,\mu,t} = b_t^{-1} (r_{p_t} - a_t) = \mu + b_t^{-1} c_t \epsilon_t + b_t^{-1} d_t s_t, \]  

(A.43)

which has precision \( K_{p,\mu,t} = b_t^2 (c_t^2 \sigma_\epsilon + d_t^2 \sigma_s)^{-1} \).

Let \( \hat{\mu}_{i,t} \) denote the posterior expectation of \( f_t \) for an investor who learns about \( \mu \) from prices. Using the variance decomposition (A.30), first consider the forecast
error
\[ \mathbb{V} [\hat{f}_{\mu, i, t} - f_t] = \mathbb{V} \left[ \dot{\hat{\sigma}}_{\mu, i, t} \left( \sum_{s=1}^{t-1} K_{\mu, i, s} \theta_{\mu, i, s} + \tilde{K}_{\mu, i, t} \dot{\theta}_{\mu, i, t} + K_{p, \mu, t} \theta_{p, \mu, t} \right) + \dot{\sigma}_{e, i, t} K_{e, i, t} \theta_{e, i, t} - \mu - \epsilon_t \right] \]
\[ = \mathbb{V} \left[ \dot{\hat{\sigma}}_{\mu, i, t} \left( \sum_{s=1}^{t-1} K_{\mu, i, s} + \tilde{K}_{\mu, i, t} + K_{p, \mu, t} - \dot{\sigma}_{\mu, i, t}^{-1} \right) \mu \right] + \mathbb{V} \left[ \dot{\sigma}_{e, i, t} \left( K_{e, i, t} - \dot{\sigma}_{e, i, t}^{-1} \right) \epsilon_t \right] + \mathbb{V} \left[ \dot{\sigma}_{e, i, t} K_{e, i, t} \theta_{e, i, t} \right] \]
\[ = \dot{\sigma}_{\mu, i, t}^{2} \sigma_{\mu}^{-1} + \dot{\sigma}_{e, i, t}^{2} \sigma_{e}^{-1} + \dot{\sigma}_{\mu, i, t}^{2} \left( \sum_{s=1}^{t-1} K_{\mu, i, s} + \tilde{K}_{\mu, i, t} + K_{p, \mu, t} \right) + \dot{\sigma}_{e, i, t}^{2} K_{e, i, t} \]
\[ = \dot{\sigma}_{\mu, i, t} + \dot{\sigma}_{e, i, t}, \]  \hspace{1cm} (A.44)

where the third inequality follows from the fact that \( \dot{\sigma}_{\mu, i, t}^{-1} = \sigma_{\mu}^{-1} + \sum_{s=1}^{t} K_{\mu, i, s} \) and \( K_{\mu, i, t} = K_{p, \mu, t} + \tilde{K}_{\mu, i, t} \). Note that this variance is equivalent to the forecast error variance (A.31).

The variance of realized returns is equal in both scenarios and given by (A.29).

Finally, the covariance between investor \( i \)'s forecast error and realized returns is

\[ \mathbb{C} \left[ \hat{f}_{\mu, i, t} - f_t, f_t - r p_t \right] = \mathbb{C} \left[ -\dot{\hat{\sigma}}_{\mu, i, t} \sigma_{\mu}^{-1} \mu, \left( 1 - b_t \right) \mu \right] + \mathbb{C} \left[ \left( \dot{\sigma}_{\mu, i, t} K_{p, \mu, t} b_t^{-1} c_t - \dot{\sigma}_{e, i, t} \sigma_{e}^{-1} \right) \epsilon_t, \left( 1 - c_t \right) \epsilon_t \right] + \mathbb{C} \left[ \dot{\sigma}_{\mu, i, t} K_{p, \mu, t} b_t^{-1} d_t s_t, -d_t s_t \right] \]
\[ = - \left( 1 - b_t \right) \dot{\sigma}_{\mu, i, t} + \left( 1 - c_t \right) \left( \dot{\sigma}_{\mu, i, t} \sigma_{e} K_{p, \mu, t} b_t^{-1} c_t - \dot{\sigma}_{e, i, t} \right) \]
\[ - \dot{\sigma}_{\mu, i, t} K_{p, \mu, t} b_t^{-1} d_t^2 \sigma_s. \]  \hspace{1cm} (A.45)

Putting this together,

\[ \mathbb{V} \left[ \hat{f}_{\mu, i, t} - r p_t \right] = (2b_t - 1) \dot{\sigma}_{\mu, i, t} + (2c_t - 1) \dot{\sigma}_{e, i, t} + v_t + 2 \left( \dot{\sigma}_{\mu, i, t} \sigma_{e} K_{p, \mu, t} b_t^{-1} c_t \right) \left( 1 - c_t \right) - 2 \dot{\sigma}_{\mu, i, t} K_{p, \mu, t} b_t^{-1} d_t^2 \sigma_s. \]  \hspace{1cm} (A.46)
The difference between the variance from not learning from prices (Eq. (A.33)) and this is
\[
\mathbb{V}[\hat{f}_{i,t} - r_{pt}] - \mathbb{V}[\hat{f}_{\mu,i,t} - r_{pt}] = 2\hat{\sigma}_{\mu,i,t}K_{p,\mu,t}b_t^{-1}[-\sigma_c(1 - c_t) + d_t^2\sigma_s].
\] (A.47)

It is clear that for sufficiently large \(\sigma_s\), this difference is positive. To derive a sufficient condition to guarantee this, note that \(c_t(1 - c_t)\) achieves a maximum of \(\frac{1}{4}\) at \(c_t = \frac{1}{2}\). Further, from the definition of prices (A.23), \(d_t = \gamma\bar{\sigma}_{f,t}\). Posterior variance is decreasing in attention capacity, so \(\bar{\sigma}_{f,t}\) is less than the posterior variance of the investor with the greatest attention capacity. Further, \(\hat{\sigma}_{f,i,t}\) is convex over the feasible region, so achieves a minimum. Let \(\bar{K}\) denote the maximum attention capacity. Then, minimizing \(\bar{\sigma}_{f,i,t}\) yields the problem
\[
\min_{K_{\epsilon,t}} \left(\sigma_{\mu}^{-1} + \bar{K} - TK_{\epsilon,t}\right)^{-1} + \left(\sigma_{\epsilon}^{-1} + K_{\epsilon,t}\right)^{-1}.
\] (A.48)

Taking first order conditions and solving yields roots
\[
K_{\epsilon,t} = \frac{\bar{K} + \sigma_{\mu}^{-1} - \sqrt{T}\sigma_{\epsilon}^{-1}}{T + \sqrt{T}}
\] (A.49)
\[
K_{\epsilon,t} = \frac{\bar{K} + \sigma_{\mu}^{-1} + \sqrt{T}\sigma_{\epsilon}^{-1}}{T - \sqrt{T}}.
\]

The second root is greater than \(\bar{K}/T\), violating the no-forgetting constraint. Thus, \(\hat{\sigma}_{f,t}\) achieves a minimum at \(K_{\epsilon,t} = \frac{\bar{K} + \sigma_{\mu}^{-1} - \sqrt{T}\sigma_{\epsilon}^{-1}}{T + \sqrt{T}}\). Evaluating the posterior variance at this minimum yields a bound on the average posterior variance:
\[
\bar{\sigma}_{f,t} \geq \frac{(\sqrt{T} + 1)^2}{\bar{K} + \sigma_{\mu}^{-1} + T\sigma_{\epsilon}^{-1}}.
\] (A.50)
Thus, a sufficient condition is
\[
\sigma_s \geq \frac{\sigma^* \left( \bar{K} + \sigma^{-1} + T \sigma^{-1} \right)^2}{4 \gamma^2 \left( \sqrt{T} + 1 \right)^4}.
\]  \tag{A.51}

For all \( \sigma_s \) greater than this lower bound, the utility from obtaining a private signal is greater than that from obtaining a signal from prices. Therefore, investors will not choose to learn from prices about \( \mu \).

**Scenario 2: Learning about \( \epsilon_t \).** As prices are of the form \( p_t = \frac{1}{v} [a_t + b_t \mu + c_t \epsilon_t + d_t s_t] \), an unbiased signal for \( \epsilon_t \) can be obtained via the transformation
\[
\theta_{p,\epsilon,t} = c_t^{-1} (r p_t - a_t) = \epsilon_t + c_t^{-1} b_t \mu + c_t^{-1} d_t s_t,
\]  \tag{A.52}
which has precision \( K_{p,\epsilon,t} = c_t^2 \left( b_t^2 \sigma + d_t^2 \sigma_s \right)^{-1} \).

Let \( \tilde{f}_{\epsilon,i,t} \) denote the posterior expectation of \( f_t \) for an investor who learns about \( \epsilon_t \) from prices. Analogous calculations yield
\[
\mathbb{V} \left[ \tilde{f}_{\epsilon,i,t} - r p_t \right] = (2b_t - 1) \hat{\sigma}_{\epsilon,\mu,t} + (2c_t - 1) \hat{\sigma}_{\epsilon,i,t} + \nu_t + 2 \left( \hat{\sigma}_{\epsilon,i,t} \sigma_{\mu} K_{p,\epsilon,t} c_t^{-1} b_t \right) (1 - b_t) - 2 \hat{\sigma}_{\epsilon,i,t} K_{p,\epsilon,t} c_t^{-1} d_t^2 \sigma_s. \tag{A.53}
\]
The difference between the variance from not learning from prices (Eq. (A.33)) and this is
\[
\mathbb{V} \left[ f_{i,t} - r p_t \right] - \mathbb{V} \left[ \tilde{f}_{\epsilon,i,t} - r p_t \right] = 2 \hat{\sigma}_{\epsilon,i,t} K_{p,\epsilon,t} c_t^{-1} \left( -\sigma_{\mu} b_t (1 - b_t) + d_t^2 \sigma_s \right). \tag{A.54}
\]
Following the same steps as above, a sufficient condition that ensures this difference
is positive is
\[ \sigma_s \geq \frac{\sigma_\mu (\bar{K} + \sigma_\mu^{-1} + T \sigma_\epsilon^{-1})^2}{4\gamma^2 \left(\sqrt{T} + 1\right)^4}. \]  
(A.55)

Thus, for all \( \sigma_s \geq \max\{\sigma_\mu, \sigma_!\}(\bar{K} + \sigma_\mu^{-1} + T \sigma_\epsilon^{-1})^2\), an investor never chooses to learn from prices about either \( \mu \) or \( \epsilon_t \), completing the proof. \( \square \)

**Proof of Lemma 2:**

*Proof. To prove (i), note that, if

\[ \frac{\partial \mathbb{E} [u_{i,t}]}{\partial K_{\mu,i,t}} = 2 (c_t - b_t) \dot{\sigma}_{\epsilon,i,t} + v_t + a_t^2 > 0 \quad \forall i, t, \]  
(A.56)

then the marginal utility of fundamental attention at time 1 is strictly greater than that in all future periods:

\[ \frac{\partial \mathbb{E} [U_i]}{\partial K_{\mu,i,1}} > \frac{\partial \mathbb{E} [U_i]}{\partial K_{\mu,i,t}} \quad \forall i, t \geq 2, \]  
(A.57)

meaning the no-forgetting constraint must bind for fundamental attention at all times \( t \geq 2 \) (i.e., \( \nu_{\mu,i,t} = 0 \) for all \( t \geq 2 \)).

Suppose by way of contradiction that \( \frac{\partial \mathbb{E} [u_{i,t}]}{\partial K_{\mu,i,t}} \leq 0 \) for some investor \( i \) and time \( t \). This means \( b_t > c_t \). Then,

\[ \frac{\partial \mathbb{E} [u_{i,t}]}{\partial K_{\mu,i,t}} = 2 (c_t - b_t) \dot{\sigma}_{\epsilon,i,t} + (1 - b_t)^2 \sigma_\mu + (1 - c_t)^2 \sigma_\epsilon + \gamma^2 \dot{\sigma}_{f,i,t}^2 (\text{bars}^2 + \sigma_s) \]  
(A.58)

\[ > 2 (c_t - b_t) \sigma_\epsilon + (1 - b_t)^2 \sigma_\mu + (1 - c_t)^2 \sigma_\epsilon, \]

which is convex in \( b_t \). Solving for the roots yields

\[ b_t = 1 + \sigma_\mu^{-1} \sigma_\epsilon \pm \sigma_\mu^{-1} \sqrt{\sigma_\epsilon ((1 - c_t^2) \sigma_\mu + \sigma_\epsilon)}. \]  
(A.59)
From Eq. (A.23), $b_t \in \left[0, \frac{\bar{K}}{\sigma^{-1}_\mu + \bar{K}}\right]$, where $\bar{K}$ is the maximum attention capacity. As the upper root is greater than 1, this is not feasible. The marginal utility $\frac{\partial \mathcal{E}[u_{i,t}]}{\partial K_{\mu,i,t}}$ is negative by assumption, meaning

\begin{equation}
 b_t > 1 + \sigma^{-1}_\mu \sigma_e - \sigma^{-1}_\mu \sqrt{\sigma_e \left((1 - c_t^2) \sigma_\mu + \sigma_e \right)}, \tag{A.60}
\end{equation}

Using the fact that $b_t \leq \frac{\bar{K}}{\sigma^{-1}_\mu + \bar{K}}$, this implies

\begin{equation}
 \frac{\bar{K}}{\sigma^{-1}_\mu + \bar{K}} > 1 + \sigma^{-1}_\mu \sigma_e - \sigma^{-1}_\mu \sqrt{\sigma_e \left((1 - c_t^2) \sigma_\mu + \sigma_e \right)} \\
\Rightarrow \quad \bar{K} > \frac{1 + \sigma^{-1}_\mu \sigma_e - \sigma^{-1}_\mu \sqrt{\sigma_e \left((1 - c_t^2) \sigma_\mu + \sigma_e \right)}}{\sqrt{\sigma_e \left((1 - c_t^2) \sigma_\mu + \sigma_e \right)} - \sigma_e} \tag{A.61}
\end{equation}

where the third inequality follows from the fact that the bound is increasing in $c_t$ and $c_t < 1$. Note that this contradicts Assumption 1, meaning $\frac{\partial \mathcal{E}[u_{i,t}]}{\partial K_{\mu,i,t}} > 0$ for all investors and dates.

Finally, (ii) follows from the fact that the posterior variance is identical for all time periods because attention to the fundamental is only allocated at $t = 1$. Thus, the optimization problem is identical across time periods, meaning that the resulting idiosyncratic attention must be identical across time.

**Proof of Corollary 1:**

*Proof.* The price coefficient $b_t$ is dependent on the posterior variance $\hat{\sigma}_{f,i,t}$ and the sum of fundamental attention until time $t$. Both of these quantities are constant;
therefore, \( b_t \) is constant across time. Similarly, \( c_t \) depends on the posterior variance and the per-period idiosyncratic attention, both of which are constant. Therefore, \( c_t \) is constant across time. Finally, \( a_t \) and \( d_t \) depend only on the average posterior variance \( \bar{\sigma}_{f,t} \). As the posterior variance is constant across time for all investors, the average posterior must also be constant, guaranteeing \( a_t \) and \( d_t \) do not vary across time.

\( \square \)

**Proof of Lemma 3:**

*Proof.* First, substituting for \( a \) and \( v \), the price term \( \zeta \) is given by

\[
\zeta = \frac{2 (T - 1) (c - b)}{(1 - b)^2 \sigma_\mu + (1 - c)^2 \sigma_\epsilon + \gamma^2 \bar{\sigma}_f^2 (\bar{s}^2 + \sigma_s)},
\]

(A.62)

which is clearly decreasing in \( \bar{s}^2 + \sigma_s \) and approaches zero in the limit as \( \bar{s}^2 + \sigma_s \to \infty \).

**Case 1:** \( \frac{\sigma_{\mu}^{-1} - \sqrt{T} \sigma_{\epsilon}^{-1}}{K_{m_i}} \in (-1, 0] \). Then, \( b \geq c \) by Proposition 1. Idiosyncratic attention must then be positive:

\[
K_{t,i} \geq \frac{\left(\sqrt{T} - 1\right) \left(K_{m_i} + \sigma_{\mu}^{-1} - \sqrt{T} \sigma_{\epsilon}^{-1}\right)}{\sqrt{T} (T - 1)} > 0.
\]

(A.63)

Taking the limit as \( \bar{s}^2 + \sigma_s \to \infty \) (i.e., \( \zeta \to \infty \)),

\[
K_{\mu,i} \to \frac{\left(\sqrt{T} - 1\right) \left(K_{m_i} - \sqrt{T} \left(\sigma_{\mu}^{-1} - \sqrt{T} \sigma_{\epsilon}^{-1}\right)\right)}{T - 1} > 0.
\]

(A.64)

Thus, for sufficiently large \( \bar{s}^2 + \sigma_s \), fundamental attention is positive, meaning that neither no-forgetting constraints are binding.

**Case 2:** \( \frac{\sigma_{\mu}^{-1} - \sqrt{T} \sigma_{\epsilon}^{-1}}{K_{m_i}} \in \left[0, \frac{1}{\sqrt{T}}\right) \). By Proposition 1, \( b < c \). Note then that funda-
mental attention must be positive:

\[
K_{\mu,i} \geq \frac{\sqrt{T} - 1}{T - 1} \left( K_{m_i} - \sqrt{T} \left( \sigma^{-1}_\mu - \sqrt{T} \sigma^{-1}_\epsilon \right) \right) > 0.
\] (A.65)

Following above, taking the limit as \( s^2 + \sigma_s \to \infty, \)

\[
K_{\epsilon,i} = \frac{\left(1 - \frac{1}{\sqrt{T}}\right)(K_{m_i} + \sigma^{-1}_\mu - T \sigma^{-1}_\epsilon)}{T - 1} > 0.
\] (A.66)

Again, for sufficiently large \( s^2 + \sigma_s, \) idiosyncratic attention, so neither no-forgetting constraints are bidning.

\[ \square \]

**Proof of Proposition 1:**

*Proof.* From the definition of prices (A.23),

\[
b - c = \bar{\sigma}_f \int_0^1 \hat{\sigma}_{i,j}^{-1} \left( \hat{\sigma}_{\mu,j} K_{\mu,j} - \hat{\sigma}_{\epsilon,j} K_{\epsilon,j} \right) di = \bar{\sigma}_f \int_0^1 \hat{\sigma}_{i,j}^{-1} \left( \sigma^{-1}_\epsilon K_{\mu,j} - \sigma^{-1}_\mu K_{\epsilon,j} \right) di.
\] (A.67)

This is a weighted average of \( \sigma^{-1}_\epsilon K_{\mu,j} - \sigma^{-1}_\mu K_{\epsilon,j}. \) As a result, if \( \sigma^{-1}_\epsilon K_{\mu,j} - \sigma^{-1}_\mu K_{\epsilon,j} > 0 \) for all investors, \( b > c. \) Similarly, if \( \sigma^{-1}_\epsilon K_{\mu,j} - \sigma^{-1}_\mu K_{\epsilon,j} < 0 \) for all investors, \( b < c. \)

**Case 1:** \( \sigma_\epsilon / \sigma_\mu > \sqrt{T}. \) First, suppose BWOC that \( b > c. \) Then, \( K_{\mu,j} \) is bounded above and \( K_{\epsilon,j} \) is bounded below:

\[
K_{\mu,j} < \frac{\sqrt{T} K_{m_i}}{T + \sqrt{T}} \quad \text{and} \quad K_{\epsilon,j} > \frac{K_{m_i}}{T + \sqrt{T}}.
\] (A.68)

\[ ^6 \text{The case for which } b < c \text{ is symmetric and omitted for brevity.} \]
Computing the difference $\sigma^{-1}_\epsilon K_{\mu,i} - \sigma^{-1}_\mu K_{\epsilon,i}$ yields

$$\sigma^{-1}_\epsilon K_{\mu,i} - \sigma^{-1}_\mu K_{\epsilon,i} < \sigma^{-1}_\epsilon \frac{\sqrt{T} K_{m_i}}{T + \sqrt{T}} - \sigma^{-1}_\mu \frac{K_{m_i}}{T + \sqrt{T}}$$

$$= \frac{-K_{m_i}}{T + \sqrt{T}} \left( \sigma^{-1}_\mu - \sqrt{T} \sigma^{-1}_\epsilon \right)$$

$$= 0,$$

which implies $b < c$, contradicting the original assertion.

**Case 2:** $\sigma_\epsilon / \sigma_\mu < \sqrt{T}$. Suppose BWOC $b \leq c$. The same bounds as above apply with opposite inequalities:

$$K_{\mu,i} > \frac{\sqrt{T} K_{m_i}}{T + \sqrt{T}} \quad \text{and} \quad K_{\epsilon,i} < \frac{K_{m_i}}{T + \sqrt{T}}.$$ 

(A.70)

meaning that $\sigma^{-1}_\epsilon K_{\mu,i} - \sigma^{-1}_\mu K_{\epsilon,i}$ must be positive:

$$\sigma^{-1}_\epsilon K_{\mu,i} - \sigma^{-1}_\mu K_{\epsilon,i} > \sigma^{-1}_\epsilon \frac{\sqrt{T} K_{m_i}}{T + \sqrt{T}} - \sigma^{-1}_\mu \frac{K_{m_i}}{T + \sqrt{T}}$$

$$= \frac{-K_{m_i}}{T + \sqrt{T}} \left( \sigma^{-1}_\mu - \sqrt{T} \sigma^{-1}_\epsilon \right)$$

$$= 0,$$

implying that $b > c$, which contradicts the original assumption.

**Case 3:** $\sigma_\epsilon / \sigma_\mu = \sqrt{T}$. This case is symmetric to Case 2 with opposite inequalities.

\[\square\]

**Proof of Proposition 2:**

Proof. As the information capacity constraint binds (Lemma 6), dividing the constraint by $K_{m_i}$ yields

$$1 = \frac{K_{\mu,i}}{K_{m_i}} + T \frac{K_{\epsilon,i}}{K_{m_i}}.$$ 

(A.72)

171
It is clear that the relative attentions $\frac{K_{\mu,i}}{K_{m_i}}$ and $\frac{K_{\epsilon,i}}{K_{m_i}}$ must have opposite relationships with $K_{m_i}$ (i.e., their derivatives have opposite signs). This proof considers relative idiosyncratic attention $\frac{K_{\epsilon,i}}{K_{m_i}}$.

**Case 1:** $\sigma_{\epsilon}/\sigma_{\mu} = \sqrt{T}$. By Proposition 1, $b = c$, which implies that $\zeta = 0$. Then, idiosyncratic attention is given by

$$K_{\epsilon,i} = \frac{K_{m_i}}{T + \sqrt{T}},$$

which is clearly proportional to $K_{m_i}$. Thus, relative idiosyncratic (and relative fundamental) attention is independent of the attention capacity $K_{m_i}$.

**Case 2:** $\sigma_{\epsilon}/\sigma_{\mu} < \sqrt{T}$. By Proposition 1, this implies $\zeta < 0$. Let $K_{m_j} = \alpha K_{m_i}$ for some $\alpha \in (0, 1)$. It is clear that $K_{\epsilon,i}, K_{\epsilon,j} > 0$.

First, suppose that the no-forgetting constraint binds for the fundamental attention of investor $i$: $K_{\mu,i} = 0$. By Proposition 16, this implies that the no-forgetting constraint also binds for the fundamental attention of investor $j$, meaning the relative attention allocations are identical.

Consider the difference in relative idiosyncratic attention if the no-forgetting constraint of investor $i$ does not bind:

$$\frac{K_{\epsilon,i}}{K_{m_i}} - \frac{K_{\epsilon,j}}{\alpha K_{m_i}} \geq (\alpha K_{m_i} (T - 1))^{-1} \left\{ \sqrt{T^{-1} (\delta_j^2 + \delta_j \zeta)} - \left[ (1 - \alpha) (\sigma_{\mu}^{-1} + \sigma_{\epsilon}^{-1}) + \alpha \sqrt{T^{-1} (\delta_i^2 + \delta_i \zeta)} \right] \right\},$$

(A.74)

where the inequality follows from the fact that the no-forgetting constraint may bind for the fundamental attention of investor $j$ but not for investor $i$.  

172
Setting this to zero yields a quadratic:

\[ 0 = \sqrt{T^{-1}} (\delta_j^2 + \delta_j \zeta) - \left(1 - \alpha \right) (\sigma^{-1}_\mu + \sigma^{-1}_\epsilon) + \alpha \sqrt{T^{-1}} (\delta_j^2 + \delta_j \zeta) \]

\[ = T^{-1} (\delta_j^2 + \delta_j \zeta) - \left(1 - \alpha \right) (\sigma^{-1}_\mu + \sigma^{-1}_\epsilon) + \alpha \sqrt{T^{-1}} (\delta_j^2 + \delta_j \zeta)^2 \]

\[ = \left( T^{-1} K_{mi}^2 - (\sigma^{-1}_\mu + \sigma^{-1}_\epsilon)^2 - T^{-1} (\delta_j^2 + \delta_j \zeta) + 2 (\sigma^{-1}_\mu + \sigma^{-1}_\epsilon) \sqrt{T^{-1}} (\delta_j^2 + \delta_j \zeta) \right) \alpha^2 + \]

\[ + \left( 2T^{-1} K_{mi} (\sigma^{-1}_\mu + T \sigma^{-1}_\epsilon) + T^{-1} K_{mi} \zeta + 2 (\sigma^{-1}_\mu + \sigma^{-1}_\epsilon)^2 - 2 (\sigma^{-1}_\mu + \sigma^{-1}_\epsilon) \sqrt{T^{-1}} (\delta_j^2 + \delta_j \zeta) \right) \alpha + \]

\[ + \left( T^{-1} (\sigma^{-1}_\mu + T \sigma^{-1}_\epsilon)^2 + T^{-1} (\sigma^{-1}_\mu + T \sigma^{-1}_\epsilon) \zeta - (\sigma^{-1}_\mu + \sigma^{-1}_\epsilon) \right) \]

(A.75)

with roots \( \alpha = 1 \) and

\[ \alpha \equiv \frac{(\sigma^{-1}_\mu + T \sigma^{-1}_\epsilon) \zeta - (T - 1) (\sigma^{-2}_\mu - T \sigma^{-2}_\epsilon)}{K_{mi}^2 - T (\sigma^{-1}_\mu + \sigma^{-1}_\epsilon)^2 - (\delta_j^2 + \delta_j \zeta) + 2T (\sigma^{-1}_\mu + \sigma^{-1}_\epsilon) \sqrt{T^{-1}} (\delta_j^2 + \delta_j \zeta)}. \]  

(A.76)

Now, note that \( \alpha \) is equal to zero for \( \zeta^* \equiv \frac{(T-1)(\sigma^{-2}_\mu - T \sigma^{-2}_\epsilon)}{\sigma^{-1}_\mu + T \sigma^{-1}_\epsilon} \) and has asymptotes

\[ \zeta \equiv \frac{(\sqrt{T} + 1) (\sigma^{-1}_\mu + \sqrt{T} \sigma^{-1}_\epsilon) (2K_{mi} - (\sqrt{T} - 1) (\sigma^{-1}_\mu - \sqrt{T} \sigma^{-1}_\epsilon))}{\delta_j} \]

\[ \tilde{\zeta} \equiv \frac{(\sqrt{T} - 1) (\sigma^{-1}_\mu - \sqrt{T} \sigma^{-1}_\epsilon) (2K_{mi} + (\sqrt{T} + 1) (\sigma^{-1}_\mu + \sqrt{T} \sigma^{-1}_\epsilon))}{\delta_j}. \]

(A.77)

Given the assumption that \( \sigma_j/\sigma_\mu < \sqrt{T} \), the roots and asymptotes are ordered \( \zeta < \tilde{\zeta} < \zeta^* < 0 \). Thus, the lower bound \( \alpha \) must be of the same sign for all \( \zeta \in [\zeta^*, 0] \). For \( \zeta = 0 \),

\[ \alpha = - \frac{(\sqrt{T} + 1) (\sigma^{-1}_\mu + \sqrt{T} \sigma^{-1}_\epsilon)}{2K_{mi} - (\sqrt{T} - 1) (\sigma^{-1}_\mu - \sqrt{T} \sigma^{-1}_\epsilon)} < 0 \]

(A.78)

because \( \sigma_j/\sigma_\mu < \sqrt{T} \). Thus, for \( \zeta \in [\zeta^*, 0] \), the sign of \( \frac{K_{e,i}}{K_{mi}} - \frac{K_{e,j}}{\alpha K_{mi}} \) must be the same for all \( \alpha \in (0, 1) \). For \( \zeta = 0 \), this difference is

\[ \frac{K_{e,i}}{K_{mi}} - \frac{K_{e,j}}{\alpha K_{mi}} = - \left( \alpha K_{mi} \left( T + \sqrt{T} \right) \right)^{-1} (1 - \alpha) (\sigma^{-1}_\mu - \sqrt{T} \sigma^{-1}_\epsilon) > 0 \]

(A.79)
again because $\sigma_i/\sigma_\mu < \sqrt{T}$ and $\alpha \in (0, 1)$.

Thus, for all $\zeta \in [\zeta^*, 0]$, relative idiosyncratic attention is increasing in attention capacity (or, equivalently, relative fundamental attention is decreasing in attention capacity) for all $\alpha \in (0, 1)$. To ensure $\zeta \in [\zeta^*, 0]$, it suffices to restrict the attention capacities of all investors by

$$\bar{K} \leq g (\ell (\zeta^*)),$$  \hspace{1cm} (A.80)

where the functions $\ell$ and $g$ are defined in Eqs. (A.14) and (A.16), respectively.

**Case 3:** $\sigma_i/\sigma_\mu > \sqrt{T}$. Again, let $K_{m_j} = \alpha K_{m_i}$ for some $\alpha \in (0, 1)$. From Proposition 1, $\zeta > 0$, implying that $K_{\mu,i}, K_{\mu,j} > 0$.

First, suppose the no-forgetting constraint binds for the idiosyncratic attention of investor $i$: $K_{\mu,i} = 0$. This implies that the no-forgetting constraint for the idiosyncratic attention of investor $j$ also binds by Proposition 16, meaning the relative allocations are identical. Further, note that if the no-forgetting constraint for idiosyncratic attention binds for investor $j$ but not for investor $j$, the statement is satisfied.

Given $\sigma_i/\sigma_\mu > \sqrt{T}$, the roots and asymptotes are ordered either $\zeta \leq 0 < \zeta \leq \zeta^*$ or $0 < \zeta \leq \zeta^*$. If $\zeta \leq 0$, then

$$K_{m_i} \geq \frac{\left(\sqrt{T} - 1\right)\left(\sigma_\mu^{-1} - \sqrt{T}\sigma_i^{-1}\right)}{2}$$  \hspace{1cm} (A.81)

and the lower root $\alpha$ must be of the same sign for all $\zeta \in [0, \zeta^*)$. For $\zeta = 0$,

$$\alpha = -\frac{\left(\sqrt{T} + 1\right)\left(\sigma_\mu^{-1} + \sqrt{T}\sigma_i^{-1}\right)}{2K_{m_i} - \left(\sqrt{T} - 1\right)\left(\sigma_\mu^{-1} - \sqrt{T}\sigma_i^{-1}\right)} < 0$$  \hspace{1cm} (A.82)

because of the lower bound on $K_{m_i}$. Thus, for $\zeta \in [0, \zeta^*)$, the sign of $K_{\mu,i}/K_{m_i} - \frac{K_{\mu,j}}{\alpha K_{m_i}}$
must be the same for all \( \alpha \in (0, 1) \). For \( \zeta = 0 \), this difference is negative:

\[
\frac{K_{\epsilon,i}}{K_{m_i}} - \frac{K_{\epsilon,j}}{\alpha K_{m_i}} = -\left(\alpha K_{m_i}\left(T + \sqrt{T}\right)\right)^{-1} (1 - \alpha) \left(\sigma^{-1}_\mu - \sqrt{T}\sigma^{-1}_\epsilon\right) < 0 \tag{A.83}
\]

because \( \sigma_\epsilon/\sigma_\mu > \sqrt{T} \). Thus, if \( \zeta < 0 \) and the maximum attention capacity is bounded by

\[
\bar{K} \leq g\left(\bar{\ell}\left(\bar{\zeta}\right)\right), \tag{A.84}
\]

relative idiosyncratic attention is decreasing in attention capacity (relative fundamental attention is increasing in attention capacity).

Finally, if \( \zeta > 0 \), the lower bound \( \alpha \) must be negative, as the denominator is positive, while the numerator is increasing in \( \zeta \) and equals 0 for \( \zeta^* > \zeta \). Thus, for \( \zeta \in [0, \zeta] \), the difference \( \frac{K_{\epsilon,i}}{K_{m_i}} - \frac{K_{\epsilon,j}}{\alpha K_{m_i}} \) is negative for all \( \alpha \in (0, 1) \). A condition that guarantees \( \zeta < \zeta \) is

\[
\bar{K} \leq g\left(\bar{\ell}\left(\bar{\zeta}\right)\right). \tag{A.85}
\]

Thus, for all values of \( K_{m,i} \), there exist bounds on the maximum attention capacity that guarantee fundamental attention is increasing in attention capacity, completing the proof.

\[\square\]

**Proof of Proposition 3:**

**Proof.** The derivative of fundamental covariance (2.23) with respect to attention capacity is

\[
\frac{\partial C^{CS}[q_{i,t}, f_t]}{\partial K_{m_i}} = \gamma^{-1}\sigma_\mu \left\{ \hat{\sigma}^{-1}_{f,i} \frac{\partial}{\partial K_{m_i}} \left( \hat{\sigma}_{\mu,i} K_{\mu,i} - b \right) + \left( \hat{\sigma}_{\mu,i} K_{\mu,i} - b \right) \frac{\partial \hat{\sigma}^{-1}_{f,i}}{\partial K_{m_i}} \right\}
\]

\[
= \gamma^{-1}\sigma_\mu \hat{\sigma}^{-1}_{f,i} \left\{ \hat{\sigma}^2_{\mu,i} \sigma^{-1}_{\mu,i} K'_{\mu,i} + \left( \hat{\sigma}_{\mu,i} K_{\mu,i} - b \right) \hat{\sigma}^{-1}_{f,i} \left( \sigma^2_{\mu,i} K'_{\mu,i} + \hat{\sigma}^2_{f,i} K'_{f,i} \right) \right\} \quad (\text{A.86})
\]
and the derivative of idiosyncratic covariance (2.24) is

$$\frac{\partial \text{CTS} \left[ q_{i,t}, f_t \right]}{\partial K_{m_i}} = \gamma^{-1} \sigma_\epsilon \left\{ \frac{\partial -1}{\partial K_{m_i}} \left( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c \right) \right\} + \left( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c \right) \frac{\partial \sigma^{-1}_{f,i}}{\partial K_{m_i}} \right\}$$

$$= \gamma^{-1} \sigma_\epsilon \left[ \frac{\partial^2}{\partial K_{m_i}} \hat{\sigma}^{-1}_{f,i} K''_{\epsilon,i} + \frac{\partial}{\partial K_{m_i}} \left( \hat{\sigma}^{-1}_{f,i} \left( \hat{\sigma}_{\mu,i} K''_{\mu,i} + \hat{\sigma}^2_{\epsilon,i} K_{\epsilon,i} \right) \right) \right],$$

(A.87)

where $K'_{\mu,i}$ and $K'_{\epsilon,i}$ are the derivatives of fundamental and idiosyncratic attention wrt attention capacity $K_{m_i}$, respectively. By Proposition 16, these derivatives are both positive. Further, recall that $b$ and $c$ are simply weighted averages of $\hat{\sigma}_{\mu,i} K_{\mu,i}$ and $\hat{\sigma}_{\epsilon,i} K_{\epsilon,i}$, respectively. Thus, for all investors with above average fundamental (idiosyncratic) attention, $\hat{\sigma}_{\mu,i} K_{\mu,i} > b \left( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} > c \right)$, respectively, which guarantees the derivatives are positive.

For investors with below average fundamental and idiosyncratic attention, the derivatives are positive if

$$f_\mu (\zeta) \equiv \hat{\sigma}^2_{\mu,i} \sigma^{-1}_{\mu,i} K'_{\mu,i} + \left( \hat{\sigma}_{\mu,i} K_{\mu,i} - b \right) \hat{\sigma}^{-1}_{f,i} \left( \hat{\sigma}^2_{\mu,i} K''_{\mu,i} + \hat{\sigma}^2_{\epsilon,i} K_{\epsilon,i} \right) > 0 \quad (A.88)$$

and

$$f_\epsilon (\zeta) \equiv \hat{\sigma}^2_{\epsilon,i} \sigma^{-1}_{\epsilon,i} K'_{\epsilon,i} + \left( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c \right) \hat{\sigma}^{-1}_{f,i} \left( \hat{\sigma}^2_{\mu,i} K''_{\mu,i} + \hat{\sigma}^2_{\epsilon,i} K_{\epsilon,i} \right) > 0. \quad (A.89)$$

Taking a first-order approximation of these quantities, $f (\zeta) = f (0) + f' (0) \zeta + o (\zeta)$. Evaluating $f_\mu (0)$,

$$f_\mu (0) = \sigma^{-1}_{\mu,i} \delta^{-2}_{i} \left( \sqrt{T} + 1 \right) + (1 - b) \delta^{-1}_{i} - \sigma^{-1}_{\mu,i} \delta^{-2}_{i} \left( \sqrt{T} + 1 \right)$$

$$= (1 - b) \delta^{-1}_{i} \quad (A.90)$$

$$> 0,$$
and \( f_\epsilon (0), \)

\[
f_\mu (0) = \sqrt{T} \sigma_\epsilon^{-1} \delta_i^{-2} \left( \sqrt{T} + 1 \right) + (1 - c) \delta_i^{-1} - \sqrt{T} \sigma_\epsilon^{-1} \delta_i^{-2} \left( \sqrt{T} + 1 \right)
\]

\[
= (1 - c) \delta_i^{-1}
\]

\[
> 0.
\]

The strict inequalities follow from the fact that \( b, c < 1 \) from Eq. (A.23). Thus, for \( \zeta \) sufficiently close to zero, \( f_\mu (\zeta), f_\epsilon (\zeta) > 0, \) meaning that both fundamental and idiosyncratic covariances are increasing in attention capacity. Define by \( \zeta_\mu^* \) and \( \zeta_\epsilon^* \) the minimum absolute values of \( \zeta \) for which the fundamental and idiosyncratic covariances, respectively, are decreasing in attention capacity:

\[
\zeta_\mu^* \equiv \inf \left\{ |\zeta| : \frac{\partial C^{CS} [g_{i,t}, f_t]}{\partial K_{m_i}} \leq 0 \right\}
\]

(A.92)

\[
\zeta_\epsilon^* \equiv \inf \left\{ |\zeta| : \frac{\partial C^{TS} [g_{i,t}, f_t]}{\partial K_{m_i}} \leq 0 \right\}.
\]

(A.93)

Then, for all \( \zeta \in \left( - \min \{ \zeta_\mu^*, \zeta_\epsilon^* \} , \min \{ \zeta_\mu^*, \zeta_\epsilon^* \} \right) \), both fundamental and idiosyncratic covariance are increasing in attention capacity. A bound on the maximum attention capacity that ensures this is

\[
\bar{K} \leq g \left( \ell \left( \min \{ \zeta_\mu^*, \zeta_\epsilon^* \} \right) \right),
\]

(A.94)

where the functions \( \ell \) and \( g \) are defined in Eqs. (A.15) and (A.16).

Finally, fundamental (Eq. (2.23)) and idiosyncratic covariance (Eq. (2.24)) are equal to zero when

\[
K_{\mu,i} = \frac{b}{1 - b} \sigma_\mu^{-1} \quad \text{and} \quad K_{\epsilon,i} = \frac{c}{1 - c} \sigma_\epsilon^{-1},
\]

(A.95)
respectively. Let fundamental and idiosyncratic attention be an implicit function of attention capacity:

\[ K_{\mu,i} = k_\mu (K_{m,i}) \quad \text{and} \quad K_{\epsilon,i} = k_\epsilon (K_{m,i}). \quad (A.96) \]

Then, the attention capacities for which fundamental and idiosyncratic covariances are zero are

\[ \bar{K}_\mu \equiv k_\mu^{-1} \left( \frac{b}{1 - b \sigma^{-1}_\mu} \right) \quad \text{and} \quad \bar{K}_\epsilon \equiv k_\epsilon^{-1} \left( \frac{c}{1 - c \sigma^{-1}_\epsilon} \right). \quad (A.97) \]

\[ \square \]

**Proof of Proposition 4:**

*Proof.* The proof considers the ratio of fundamental to idiosyncratic covariance. The resulting conditions are identical. The ratio of fundamental to idiosyncratic covariance is

\[ \frac{C^{CS}[q_{i,t}, f_i]}{C^{TS}[q_{i,t}, f_i]} = \frac{\sigma_\mu \hat{\sigma}_{\mu,i} K_{\mu,i} - b}{\sigma_\epsilon \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c}. \quad (A.98) \]

The derivative with respect to the attention capacity \( K_{m,i} \) is

\[ \frac{\partial}{\partial K_{m,i}} \frac{C^{CS}[q_{i,t}, f_i]}{C^{TS}[q_{i,t}, f_i]} = \frac{\sigma_\mu \hat{\sigma}_{\mu,i} K'_{\mu,i} \left( 1 - \hat{\sigma}_{\mu,i} K_{\mu,i} \right) (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) - \hat{\sigma}_{\epsilon,i} K'_{\epsilon,i} \left( 1 - \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} \right) (\hat{\sigma}_{\mu,i} K_{\mu} - b)}{(\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c)^2}. \quad (A.99) \]

The derivative is positive if

\[ f(\zeta) \equiv \hat{\sigma}_{\mu,i} K'_{\mu,i} \left( 1 - \hat{\sigma}_{\mu,i} K_{\mu,i} \right) (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) - \hat{\sigma}_{\epsilon,i} K'_{\epsilon,i} \left( 1 - \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} \right) (\hat{\sigma}_{\mu,i} K_{\mu} - b). \quad (A.100) \]

**Case 1:** \( \sigma_\epsilon / \sigma_\mu = \sqrt{T} \). Then, \( b = c \) by Proposition 1, implying \( K_{\mu,i} = \frac{K_{m,i}}{\sqrt{T+1}} \) and \( K_{\epsilon,i} = \frac{\sqrt{T-1} K_{m,i}}{\sqrt{T+1}} \). Simple calculations show \( \hat{\sigma}_{\mu,i} K_{\mu,i} = \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} = \frac{K_{m,i}}{K_{m,i} + (T+\sqrt{T}) \sigma^{-1}_\epsilon}. \)
Solving for the covariance ratio yields
\[ \frac{CCS[q_{i,t}, f_t]}{CTS[q_{i,t}, f_t]} = \frac{\sigma_\mu}{\sigma_\epsilon}, \] (A.101)
which is clearly independent of attention capacity.

**Case 2:** \( \sigma_\epsilon / \sigma_\mu < \sqrt{T} \). First, note that, as \( \frac{K_{\mu,i}}{K_{m_i}} \) is decreasing in \( K_{m_i} \), \( \tilde{K}_\epsilon > \tilde{K}_\mu \). For \( K_{m_i} > \tilde{K}_\mu \), both \( (\hat{\sigma}_{\mu,i} K_\mu - b) \) and \( (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) \) are positive. Similarly, for \( K_{m_i} \leq \tilde{K}_\epsilon \), both \( (\hat{\sigma}_{\mu,i} K_\mu - b) \) and \( (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) \) are negative. For \( \tilde{K}_\mu < K_{m_i} < \tilde{K}_\epsilon \), \( (\hat{\sigma}_{\mu,i} K_\mu - b) \) is positive, while \( (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) \) is negative, meaning \( f(\zeta) \) is negative. The proof considers the case \( K_{m_i} > \tilde{K}_\mu \). It is clear that, in the case of \( K_{m_i} \leq \tilde{K}_\mu \), all inequalities hold with the opposite sign.

By Proposition 2, \( \frac{\partial K_{\mu,i}/K_{m_i}}{\partial K_{m_i}} < 0 \). This can be rewritten as
\[ \frac{K_{\mu,i}' K_{m_i} - K_{\mu,i}}{K_{m_i}^2} < 0 \Rightarrow K_{\mu,i}' < \frac{K_{\mu,i}}{K_{m_i}}. \] (A.102)

Similarly, \( K_{\epsilon,i}' > \frac{K_{\epsilon,i}}{K_{m_i}} \). Thus,
\[ f(\zeta) > K_{m_i}^{-1} \left\{ \hat{\sigma}_{\mu,i} K_{\mu,i} (1 - \hat{\sigma}_{\mu,i} K_{\mu,i}) (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) - \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} (1 - \hat{\sigma}_{\epsilon,i} K_{\epsilon,i}) (\hat{\sigma}_{\mu,i} K_{\mu,i} - b) \right\}. \] (A.103)

Further, from the proof of Proposition 1, for \( \sigma_\epsilon / \sigma_\mu < \sqrt{T} \), it must be that \( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} > \hat{\sigma}_{\mu,i} K_{\mu,i} \). Note that, if both \( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} \) and \( \hat{\sigma}_{\mu,i} K_{\mu,i} \) are less than \( \frac{1}{2} \), it then follows that
\[ \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} (1 - \hat{\sigma}_{\epsilon,i} K_{\epsilon,i}) > \hat{\sigma}_{\mu,i} K_{\mu,i} (1 - \hat{\sigma}_{\mu,i} K_{\mu,i}) \]. Both are guaranteed to be less than \( \frac{1}{2} \) for \( K_{m_i} < \min \{ \sigma_\mu^{-1}, \sigma_\epsilon^{-1} \} \). If this is the case, then
\[ f(\zeta) > K_{m_i}^{-1} \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} (1 - \hat{\sigma}_{\epsilon,i} K_{\epsilon,i}) \left\{ (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) - (\hat{\sigma}_{\mu,i} K_{\mu,i} - b) \right\}. \] (A.104)

Recall that \( b \) and \( c \) are weighted averages of \( \hat{\sigma}_{\mu,i} K_{\mu,i} \) and \( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} \), respectively. Thus, if \( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - \hat{\sigma}_{\mu,i} K_{\mu,i} \) is decreasing in attention capacity, \( (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) - (\hat{\sigma}_{\mu,i} K_{\mu,i} - b) \)
is negative for $K_{m_i} > K_{\mu}$.

Define $h (\zeta) \equiv \frac{\partial}{\partial K_{m_i}} (\dot{\sigma}_{e,i} K_{e,i} - \dot{\sigma}_{\mu,i} K_{\mu,i})$. Note that $h (\zeta) = h (0) + h' (0) \zeta + o (\zeta)$. Evaluating $h (0)$,

$$h (0) = \delta_i^{-2} \left( \sqrt{T} + 1 \right) \left( \sigma_\mu^{-1} - \sqrt{T} \sigma_\varepsilon^{-1} \right) < 0. \quad (A.105)$$

Thus, for $\zeta$ sufficiently close to zero, $\dot{\sigma}_{e,i} K_{e,i} - \dot{\sigma}_{\mu,i} K_{\mu,i}$ is decreasing in $K_{m_i}$ if $K_{m_i} > K_{\mu}$ and increasing otherwise. Define

$$\zeta^* \equiv \sup \{ \zeta < 0 : h (\zeta) < 0 \} \quad (A.106)$$

as the maximum value of $\zeta$ (which is negative by Proposition 1) such that $\dot{\sigma}_{e,i} K_{e,i} - \dot{\sigma}_{\mu,i} K_{\mu,i}$ is decreasing in $K_{m_i}$. As long as $\zeta > \zeta^*$, $\dot{\sigma}_{e,i} K_{e,i} - \dot{\sigma}_{\mu,i} K_{\mu,i}$ is increasing in attention capacity. A sufficient condition to guarantee this is that

$$\bar{K} \leq g (\ell (\zeta^*)). \quad (A.107)$$

**Case 3:** $\sigma_i / \sigma_\mu > \sqrt{T}$. Now, note that, as $K_{m_i} / K_{m_i}$ is increasing in $K_{m_i}$, $K_{\mu} > K_{e}$. For $K_{m_i} > K_{e}$, both $(\dot{\sigma}_{\mu,i} K_{\mu} - b)$ and $(\dot{\sigma}_{e,i} K_{e,i} - c)$ are positive. Similarly, for $K_{m_i} \leq K_{\mu}$, both $(\dot{\sigma}_{\mu,i} K_{\mu} - b)$ and $(\dot{\sigma}_{e,i} K_{e,i} - c)$ are negative. For $K_{e} < K_{m_i} < K_{\mu}$, $(\dot{\sigma}_{\mu,i} K_{\mu} - b)$ is negative, while $(\dot{\sigma}_{e,i} K_{e,i} - c)$ is positive, meaning $f (\zeta)$ is positive. The proof considers the case $K_{m_i} > K_{e}$. It is clear that, in the case of $K_{m_i} \leq K_{e}$, all inequalities hold with the opposite sign.

Following the same steps as above,

$$f (\zeta) < K_{m_i}^{-1} \dot{\sigma}_{\mu,i} K_{\mu,i} \left( 1 - \dot{\sigma}_{\mu,i} K_{\mu,i} \right) \left\{ (\dot{\sigma}_{e,i} K_{e,i} - c) - (\dot{\sigma}_{\mu,i} K_{\mu,i} - b) \right\}. \quad (A.108)$$

180
It suffices to show that $\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - \hat{\sigma}_{\mu,i} K_{\mu,i}$ is increasing in attention capacity. Evaluating $h(0)$,
\[ h(0) = \delta_i^{-2} \left( \sqrt{T} + 1 \right) \left( \sigma_{\mu}^{-1} - \sqrt{T} \sigma_{\epsilon}^{-1} \right) > 0 \quad (A.109) \]
because $\sigma_{\epsilon}/\sigma_{\mu} > \sqrt{T}$. Defining $\zeta^{**} \equiv \inf \{ \zeta > 0 : h(\zeta) > 0 \}$, a bound on the maximum attention capacity that ensures $f(\zeta) > 0$ for $K_{m_i} > \hat{K}_\epsilon$ and $f(\zeta) < 0$ otherwise is
\[ K \leq g(\hat{\zeta}(\zeta^{**})) , \quad (A.110) \]

To complete the proof, note that total covariance is simply the sum of fundamental and idiosyncratic covariances. As these are both increasing in attention capacity, it follows that total covariance must be increasing as well. \hfill \Box

**Proof of Proposition 5:**

*Proof.* Evaluating portfolio dispersion yields
\[ \mathbb{V} [q_{i,t} - q_{m,t}] = (T - 1)^2 \left[ (T + 1) K_{m_i} + \frac{T (\sigma_{\mu}^{-1} + \sigma_{\epsilon}^{-1}) \zeta}{\delta_i + \zeta} - \frac{(2K_{m_i} + \zeta) \sqrt{T} \delta_i}{\sqrt{\delta_i + \zeta}} \right] . \quad (A.111) \]
The derivative with respect to attention capacity is
\[ \frac{\partial \mathbb{V} [q_{i,t} - q_{m,t}]}{\partial K_{m_i}} = (T - 1)^{-2} \left[ T + 1 - \frac{T (\sigma_{\mu}^{-1} + \sigma_{\epsilon}^{-1}) \zeta}{(\delta_i + \zeta)^2} - \frac{T (4 (\delta_i^2 + \delta_i \zeta) + (2K_{m_i} + \zeta) \zeta)}{2 (\delta_i + \zeta) \sqrt{T} (\delta_i^2 + \delta_i \zeta)} \right] . \quad (A.112) \]
Defining this derivative as a function $f$ of $\zeta$, note that it can be written has
\[ \frac{\partial \mathbb{V} [q_{i,t} - q_{m,t}]}{\partial K_{m_i}} \equiv f(\zeta) = f(0) + f'(0) \zeta + o(\zeta) . \quad (A.113) \]
Evaluating $f(0)$,
\[ f(0) = \left( \sqrt{T} + 1 \right)^{-2} > 0 . \quad (A.114) \]
Thus, as $\zeta \to 0$, portfolio dispersion is increasing in attention capacity. As a result, for $\zeta$ sufficiently close to zero, the derivative of dispersion must be positive.

Define $\zeta^*$ to be the minimum absolute value of $\zeta$ such that the derivative is negative

$$
\zeta^* = \inf \left\{ |\zeta| : \frac{\partial V [q_{t,t} - q_{m,t}]}{\partial K_{m_i}} \leq 0 \right\}.
$$

(A.115)

Thus, for all $\zeta \in (-\zeta^*, \zeta^*)$, portfolio dispersion is increasing in attention capacity. A bound on the maximum attention capacity that ensures this condition is

$$
\bar{K} \leq g (\ell (\zeta^*)).
$$

(A.116)

Proof of Proposition 6:

Proof. The derivative of the cross-sectional portfolio variance (Eq. (2.26)) is

$$
\frac{\partial \psi^{CS} [q_{i,t}]}{\partial K_{m_i}} = \gamma^2 \left\{ \frac{\partial \hat{\sigma}^{-2}_{f,i}}{\partial K_{m_i}} \left[ (\hat{\sigma}_{\mu,i} K_{\mu,i} - b)^2 \sigma_{\mu} + \hat{\sigma}_{\mu,i}^2 K_{\mu,i} \right] \right.
\hat{\sigma}^{-2}_{f,i} \left[ 2 (\hat{\sigma}_{\mu,i} K_{\mu,i} - b) \sigma_{\mu} \frac{\partial \hat{\sigma}_{\mu,i} K_{\mu,i}}{\partial K_{m_i}} + \frac{\partial \hat{\sigma}_{\mu,i}^2 K_{\mu,i}}{\partial K_{m_i}} \right] \right\}.
$$

(A.17)

Now, note that

$$
\frac{\partial \hat{\sigma}^{-2}_{f,i}}{\partial K_{m_i}} = 2 \hat{\sigma}^{-3}_{f,i} \left( \hat{\sigma}_{\mu,i}^2 K'_{\mu,i} + \hat{\sigma}_{\phi,i}^2 K'_{\phi,i} \right)
$$

$$
\frac{\partial \hat{\sigma}_{\mu,i} K_{\mu,i}}{\partial K_{m_i}} = \hat{\sigma}_{\mu,i} \sigma_{\mu}^{-1} K'_{\mu,i}
$$

$$
\frac{\partial \hat{\sigma}_{\mu,i}^2 K_{\mu,i}}{\partial K_{m_i}} = \hat{\sigma}_{\mu,i}^3 (\sigma_{\mu}^{-1} - K_{\mu,i} K'_{\mu,i})
$$

(A.18)

where $K'_{\mu,i}$ and $K'_{\phi,i}$ are the derivatives of fundamental and idiosyncratic attention
wrt attention capacity, respectively. Using these, the derivative becomes

\[
\frac{\partial V_{CS}[q_{i,t}]}{\partial K_{m,i}} = \gamma^{-2} \hat{\sigma}_{f,i}^{-2} \left\{ 2 \hat{\sigma}_{f,i}^{-1} \left( \hat{\sigma}_{\mu,i}^2 K_{\mu,i} + \hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i} \right) \right\} + \left( \hat{\sigma}_{\mu,i} K_{\mu,i} - b \right)^2 \sigma_{\mu} + \hat{\sigma}_{\mu,i}^2 K_{\mu,i} \right\}.
\]

Note that \( K_{\mu,i}, K_{\epsilon,i} \geq 0 \) by Proposition 16, explaining the above signs. Focusing on the last term,

\[
2 \left( \hat{\sigma}_{\mu,i} K_{\mu,i} - b \right) + \hat{\sigma}_{\mu,i} \left( \sigma_{\mu}^{-1} - K_{\mu,i} \right) = \hat{\sigma}_{\mu,i} \left( K_{\mu,i} + \sigma_{\mu}^{-1} \right) - 2b = 1 - 2b,
\]

which is clearly positive for \( b \leq \frac{1}{2} \). A condition that guarantees this is that the maximum attention capacity is less than \( \sigma_{\mu}^{-1} \):

\[
\bar{K} \leq \sigma_{\mu}^{-1}.
\]

With this condition, cross sectional variance is unambiguously increasing.

The derivative of time series variance (Eq. (2.27)) is

\[
\frac{\partial V_{TS}[q_{i,t}]}{\partial K_{m,i}} = \gamma^{-2} \hat{\sigma}_{f,i}^{-2} \left\{ \left( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c \right)^2 \sigma_{\epsilon} + \hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i} \right\} + \hat{\sigma}_{f,i}^{-2} \left\{ 2 \left( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c \right) \sigma_{\epsilon} \frac{\partial \hat{\sigma}_{\epsilon,i} K_{\epsilon,i}}{\partial K_{m,i}} + \frac{\partial \hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i}}{\partial K_{m,i}} \right\}.
\]
Similar to above, the derivatives are

\[
\frac{\partial \hat{\sigma}_{f,i}^{-2}}{\partial K_{m_i}} = 2 \hat{\sigma}_{f,i}^{-3} \left( \hat{\sigma}_{\mu,i}^2 K_{\mu,i} + \hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i}' \right)
\]

\[
\frac{\partial \hat{\sigma}_{\epsilon,i} K_{\epsilon,i}}{\partial K_{m_i}} = \hat{\sigma}_{\epsilon,i} \sigma_{\epsilon}^{-1} K_{\epsilon,i}'
\]

\[
\frac{\partial \hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i}}{\partial K_{m_i}} = \hat{\sigma}_{\epsilon,i}^3 (\sigma_{\epsilon}^{-1} - K_{\epsilon,i}) K_{\epsilon,i}'
\]

which simplifies the derivative of time series variance:

\[
\frac{\partial \gamma_{TS} [g_{i,t}]}{\partial K_{m_i}} = \gamma^{-2} \hat{\sigma}_{f,i}^{-2} \left\{ 2 \hat{\sigma}_{f,i}^{-1} \left( \hat{\sigma}_{\mu,i}^2 K_{\mu,i}' + \hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i}' \right) \left[ (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) \sigma_{\epsilon} + \hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i} \right] \right\}^{(+)}
\]

\[
\hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i}' \left\{ 2 (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) + \hat{\sigma}_{\epsilon,i} (\sigma_{\epsilon}^{-1} - K_{\epsilon,i}) \right\}^{(+)}
\]

where the signs rely on the fact that $K_{\mu,i}'$, $K_{\epsilon,i}' \geq 0$. The last term can be simplified to

\[
2 (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c) + \hat{\sigma}_{\epsilon,i} (\sigma_{\epsilon}^{-1} - K_{\epsilon,i}) = \hat{\sigma}_{\epsilon,i} (K_{\epsilon,i} + \sigma_{\epsilon}^{-1}) - 2c
\]

\[
= 1 - 2c,
\]

which is positive for $c \leq \frac{1}{2}$. To guarantee this, a sufficient condition is

\[
\hat{\sigma} \leq \sigma_{\epsilon}^{-1}.
\]

Thus, for $\hat{\sigma} \leq \min \{ \sigma_{\mu}^{-1}, \sigma_{\epsilon}^{-1} \}$, both cross sectional and time series variance of investors’ portfolios are increasing in attention capacity.

\[\square\]

Proof of Proposition 7:
Proof. Investor autocorrelation is given by
\[
\rho_i = \frac{(\hat{\sigma}_{\mu,i} K_{\mu,i} - b)^2 \sigma_{\mu} + \hat{\sigma}_{\mu,i}^2 K_{\mu,i}}{(\hat{\sigma}_{\mu,i} K_{\mu,i} - b)^2 \sigma_{\mu} + \hat{\sigma}_{\mu,i}^2 K_{\mu,i} + (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c)^2 \sigma_{\epsilon} + \hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i} + \sigma^2_s)}
\]
\[
= \frac{\gamma^2 \hat{\sigma}_{f,i}^2 \nabla_C S[q_{i,t}]}{\gamma^2 \hat{\sigma}_{f,i}^2 \nabla_C S[q_{i,t}] + \gamma^2 \hat{\sigma}_{f,i}^2 \nabla_T S[q_{i,t}]}.
\]
Taking the derivative with respect to attention capacity yields
\[
\frac{\partial \rho_i}{\partial K_{m_i}} = \frac{\partial \gamma^2 \hat{\sigma}_{f,i}^2 \nabla_C S[q_{i,t}]}{\partial K_{m_i}} \frac{\gamma^2 \hat{\sigma}_{f,i}^2 \nabla_T S[q_{i,t}]}{\gamma^2 \hat{\sigma}_{f,i}^2 \nabla_C S[q_{i,t}] + \gamma^2 \hat{\sigma}_{f,i}^2 \nabla_T S[q_{i,t}]}.
\]
(1.27)

From the proof of Proposition 15, this can be simplified:
\[
\frac{\partial \rho_i}{\partial K_{m_i}} = \frac{\gamma^2 \hat{\sigma}_{f,i}^2 K_{\mu,i}^T (1 - 2b) \nabla_T S[q_{i,t}] - \hat{\sigma}_{f,i}^2 K_{\mu,i}^T (1 - 2c) \nabla_C S[q_{i,t}]}{\gamma^2 \hat{\sigma}_{f,i}^2 \nabla_C S[q_{i,t}] + \gamma^2 \hat{\sigma}_{f,i}^2 \nabla_T S[q_{i,t}]}.
\]
(1.28)

The denominator is obviously positive, so the derivative is positive iff the numerator, defined as a function of $\zeta$, is positive:
\[
f (\zeta) = \gamma^2 \hat{\sigma}_{f,i}^2 K_{\mu,i}^T (1 - 2b) \nabla_T S[q_{i,t}] - \hat{\sigma}_{f,i}^2 K_{\mu,i}^T (1 - 2c) \nabla_C S[q_{i,t}] .
\]
(1.29)

Note that $f (\zeta) = f (0) + f' (0) \zeta + o (\zeta)$. Evaluating $f (0)$,
\[
f (0) = \frac{\sqrt{T} + 1}{\sigma_{\epsilon}^2} \left((1 - 2b) \nabla_T S[q_{i,t}] - \sqrt{T} (1 - 2c) \nabla_C S[q_{i,t}] \right)
\]
\[
eq \frac{\sqrt{T} + 1}{\sigma_{\epsilon}^2} \left((1 - 2b) \left[ (\hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c)^2 \sigma_{\epsilon} + \hat{\sigma}_{\epsilon,i}^2 K_{\epsilon,i} + \sigma^2_s \right] - \sqrt{T} (1 - 2c) \left[ (\hat{\sigma}_{\mu,i} K_{\mu,i} - b)^2 \sigma_{\mu} + \hat{\sigma}_{\mu,i}^2 K_{\mu,i} \right] \right)
\]
\[
eq \frac{\sqrt{T} + 1}{\sigma_{\epsilon}^2} \sigma_{\epsilon} \left((1 - c)^2 (1 - 2b) \sigma_{\epsilon}^{-1} - (1 - b)^2 (1 - 2c) \sqrt{T} \sigma_{\epsilon}^{-1} + \sigma^2_s \right).
\]
(1.30)

Now, suppose $\sigma_{\epsilon}/\sigma_{\mu} \geq \sqrt{T}$. Note that $b < c$ by Proposition 1. Further, assume the maximum attention capacity is bounded such that $b, c \leq \frac{1}{2}$ (i.e., $K \leq \min \{ \sigma_{\epsilon}^{-1} \}$.}
185
Then,
\[
\begin{align*}
    f(0) &> \frac{\sqrt{T} + 1}{\delta_i^2} \sigma(1 - c)^2 (1 - 2b) - (1 - b)^2 (1 - 2c) \\
    &= \frac{\sqrt{T} + 1}{\delta_i^2} \sigma\left\{ c^2 (1 - 2b) - b^2 (1 - 2c) \right\} \tag{A.132} \\
    &\geq 0,
\end{align*}
\]

where the bracketed term is positive because \( b < c \). Thus, for \( \zeta \) sufficiently close to zero, \( f(\zeta) > 0 \), meaning autocorrelation is increasing in attention capacity. Let \( \zeta^* \) be the maximum value such that \( f(\zeta) \geq 0 \):
\[
    \zeta^* \equiv \sup \{ \zeta : f(\zeta) \geq 0 \}. \tag{A.133}
\]

A bound on the maximum attention capacity that ensures this is
\[
    K \leq g\left(\tilde{\ell}(\zeta^*)\right). \tag{A.134}
\]

Next, suppose \( \sigma/\sigma_\mu < \sqrt{T} \). Then, \( b > c \) by Proposition 1. For all \( \sigma_s < \sigma_s^* \), the numerator when \( \zeta = 0 \) is negative (i.e., \( f(0) < 0 \)), where
\[
    \sigma_s^* \equiv \frac{(1 - b)^2 (1 - 2c) \sqrt{T} \sigma^{-1} - (1 - c)^2 (1 - 2b) \sigma^{-1}}{d^2 (1 - 2b) \sigma_\mu^{-1} \sigma^{-1}} > 0. \tag{A.135}
\]

Thus, for \( \sigma_s < \sigma_s^* \), the numerator of the derivative is positive, meaning that, for \( \zeta \) sufficiently close to zero, \( f(\zeta) < 0 \). Let \( \zeta^{**} \) be the minimum value for which \( f(\zeta) < 0 \):
\[
    \zeta^{**} \equiv \inf \{ \zeta : f(\zeta) \leq 0 \}. \tag{A.136}
\]
This is guaranteed for for the maximum attention capacity bounded by:

\[ \bar{K} \leq g(\ell(\zeta^{**})). \tag{A.137} \]

\[ \square \]

**Proof of Proposition 8:**

*Proof.* The derivative of investor’s average holdings in Eq. (2.28) with respect to the realized fundamental \( \mu \) is given by

\[ \frac{\partial \mathbb{E}^{TS}[\bar{q}_{m,t}]}{\partial \mu} = \gamma^{-1} \hat{\sigma}_{f,m}^{-1} (\hat{\sigma}_{\mu,m} K_{\mu,m} - b). \tag{A.138} \]

Note that this is clearly positive if and only if \( \hat{\sigma}_{\mu,m} K_{\mu,m} - b > 0 \) This condition is equivalent to \( K_{m} > \bar{K}_{\mu} \), where \( \bar{K}_{\mu} \) is defined in Eq. (A.97). \( \square \)

**Proof of Proposition 9:**

*Proof.* The price impact is given by

\[ \Delta^{m} \mathbb{E}^{TS}[f_{t} - r_{p_{t}}] = \psi_{m}^{-1} [a_{m} - a + (b_{m} - b) \mu]. \tag{A.139} \]

Counterfactual prices assume investor \( i \) only holds the market. As a result, his information is not reflected in asset prices. This means that the counterfactual average posterior precision is

\[ \hat{\sigma}_{f,-m}^{-1} = \sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'}^{-1} \frac{1}{1 - \psi_{m}}, \tag{A.140} \]

which is the average posterior precision over all investors not of type \( m \). Similarly, the
counterfactual coefficient $b_{-m}$ is the weighted average over all type $m' \neq m$ investors:

$$
\begin{align*}
\bar{b}_{-m} &= \frac{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'} \hat{\sigma}_{\mu,m'} K_{\mu,m'}}{1 - \psi_m} \\
&= \frac{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'} \hat{\sigma}_{\mu,m'} K_{\mu,m'}}{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,-m}}. \\
\end{align*} \\
(A.141)
$$

The difference between $b_{-m}$ and $b$ is

$$
\begin{align*}
b_{-m} - b &= \frac{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'} \hat{\sigma}_{\mu,m'} K_{\mu,m'}}{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,-m}} - \hat{\sigma}_f \sum_m \psi_m \hat{\sigma}_{f,m} K_{\mu,m} \\
&= \frac{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'} \hat{\sigma}_{\mu,m'} K_{\mu,m'}}{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,-m}} - \frac{\sum_m \psi_m \hat{\sigma}_{f,m} K_{\mu,m}}{\sum_m \psi_m \hat{\sigma}_{f,m}} \\
&= \frac{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'} \hat{\sigma}_{\mu,m'} K_{\mu,m'}}{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,-m}} \left( \frac{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'} \hat{\sigma}_{\mu,m'} K_{\mu,m'}}{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'}} - \hat{\sigma}_{\mu,m} K_{\mu,m} \right). \\
(A.142)
\end{align*}
$$

Similarly, computing $a_{-m} - a$ yields

$$
\begin{align*}
\bar{a}_{-m} - a &= \gamma \theta \left( \hat{\sigma}_f - \hat{\sigma}_{f,-m} \right) \\
&= \gamma \theta \hat{\sigma}_f \sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'} \hat{\sigma}_{\mu,m'} K_{\mu,m'} \left( \frac{1}{1 - \psi_m} - \psi_m \hat{\sigma}_{f,m'} \right) \\
&= \gamma \theta \hat{\sigma}_f \sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'} \hat{\sigma}_{\mu,m'} K_{\mu,m'} \left( \frac{1}{1 - \psi_m} - \hat{\sigma}_{f,m'} \right). \\
(A.143)
\end{align*}
$$

Putting this together, the counterfactual price impact in Eq. (2.29) is

$$
\begin{align*}
\Delta^m \mathbb{E} \left[ f_t - r_{pt} \right] &= \frac{\psi_m \hat{\sigma}_{f,m}^{-1}}{\sum_m \psi_m \hat{\sigma}_{f,m}^{-1}} \left( \frac{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'} \hat{\sigma}_{\mu,m'} K_{\mu,m'}}{\sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'}} - \hat{\sigma}_{\mu,m} K_{\mu,m} \right) \mu + \\
&\quad \gamma \theta \hat{\sigma}_f \sum_{m' \neq m} \psi_{m'} \hat{\sigma}_{f,m'} \hat{\sigma}_{\mu,m'} K_{\mu,m'} \left( \frac{1}{1 - \psi_m} - \hat{\sigma}_{f,m}^{-1} \right). \\
(A.144)
\end{align*}
$$
Now, assuming the average supply \( \bar{s} = 0 \), this is given by

\[
\Delta^m \mathbb{E}[f_t - r p_t] = \frac{\psi_m \hat{\sigma}^{-1}_{f,m}}{\sum_m \psi_m \hat{\sigma}^{-1}_{f,m}} \left( \frac{\sum_{m' \neq m} \psi_m \hat{\sigma}^{-1}_{f,m} \hat{\sigma}_{\mu,m'}^{-1} K_{\mu,m'}}{\sum_{m' \neq m} \psi_m \hat{\sigma}^{-1}_{f,m'}} \right) \hat{\sigma}_{\mu,m} - \hat{\sigma}_{\mu,m} K_{\mu,m} \right) \mu. \tag{A.145}
\]

Note that \( b - b_{-m} \) is positive if and only if \( \hat{\sigma}_{\mu,m} K_{\mu,m} - b > 0 \). To see this, note that the first term is the weighted average of \( \hat{\sigma}_{\mu,m} K_{\mu,m} \) excluding type \( m \) investors. Thus, if \( \hat{\sigma}_{\mu,m} K_{\mu,m} \) is above average, it must be above the average when excluding type \( m \) from the average. Letting

\[
\kappa_m = \frac{\psi_m \hat{\sigma}^{-1}_{f,m}}{\sum_m \psi_m \hat{\sigma}^{-1}_{f,m}} \left( \frac{\sum_{m' \neq m} \psi_m \hat{\sigma}^{-1}_{f,m} \hat{\sigma}_{\mu,m'}^{-1} K_{\mu,m'}}{\sum_{m' \neq m} \psi_m \hat{\sigma}^{-1}_{f,m'}} \right) \hat{\sigma}_{\mu,m} K_{\mu,m} \right) \mu. \tag{A.146}
\]

yields the expression for price impact \( \Delta^m \mathbb{E}[f_t - r p_t] = \kappa_m \mu \).

Finally, for \( \hat{\sigma}_{\mu,m} K_{\mu,m} > b \), the derivative of \( \kappa_m \) with respect to \( K_m \) yields

\[
\frac{\partial \kappa_m}{\partial K_m} = \frac{\psi_m \frac{\partial \hat{\sigma}^{-1}_{f,m}}{\partial K_m} \sum_{m' \neq m} \psi_m \hat{\sigma}^{-1}_{f,m'}}{\left( \sum_m \psi_m \hat{\sigma}^{-1}_{f,m} \right)^2} \left( \frac{\sum_{m' \neq m} \psi_m \hat{\sigma}^{-1}_{f,m} \hat{\sigma}_{\mu,m'}^{-1} K_{\mu,m'}}{\sum_{m' \neq m} \psi_m \hat{\sigma}^{-1}_{f,m'}} \right) \hat{\sigma}_{\mu,m} K_{\mu,m} \right) \mu. \tag{A.147}
\]

This is negative because the posterior precision is increasing in attention capacity, \( \hat{\sigma}_{\mu,m} K_{\mu,m} > b \) by assumption, and \( \hat{\sigma}_{\mu,m} K_{\mu,m} \) increasing in attention capacity by Proposition 16. Thus, \( \kappa_m \) is negative and decreasing for \( K_m \geq \bar{K}_\mu \) and positive otherwise.

\[ \square \]

**Proof of Proposition 10:**

189
Proof. The volatility impact of type $m$ investors is

$$\Delta^m VTS[f_t - r_p] = \psi^{-1}_m \left( \Delta^S VTS[f_t - r_p] - \Delta^S VTS[f_t - r_{p-m,t}] \right)$$

$$= \psi^{-1}_m \left( ((1-c)^2 - (1-c-m)^2) \sigma_c + (d^2 - d_{m}^2) \sigma_s \right). \quad (A.148)$$

Following the same steps as the proof of Proposition 9, it can be shown that

$$c - c_m = \frac{\psi_m \hat{\sigma}_{f,m}^{-1}}{\sum_m \psi_m \hat{\sigma}_{f,m}'^{-1}} \left( \hat{\sigma}_{\epsilon,m} K_{\epsilon,m} - \frac{\sum_{m'} \psi_{m'} \hat{\sigma}_{f,m'}^{-1} \hat{\sigma}_{\epsilon,m'}^{-1} K_{\epsilon,m'}}{\sum_{m'} \psi_{m'} \hat{\sigma}_{f,m'}^{-1}} \right) \quad (A.149)$$

and

$$d - d_m = \gamma \hat{\sigma}_{f,m} \psi_m \left( \frac{\hat{\sigma}_{f,m}^{-1} - \sum_{m'} \psi_{m'} \hat{\sigma}_{f,m'}^{-1}}{1 - \psi_m} \right). \quad (A.150)$$

It is clear that, for $\hat{\sigma}_{\epsilon,m} K_{\epsilon,m} < d > c > c_m$ for the same reason as in the previous proof. Similarly, for $\hat{\sigma}_{f,m}^{-1} - \hat{\sigma}_{f,m}^{-1} > 0$, $d > d_m$. If these two conditions are met, as $0 \leq c, c_m < 1$ and $d, d_m \leq 0$, it therefore follows that

$$(1-c)^2 - (1-c_m)^2 < 0 \quad \text{and} \quad d^2 - d_{m}^2 < 0, \quad (A.151)$$

meaning that investor type $m$’s impact on volatility is negative. A sufficient condition to ensure these two conditions are met is $K_m \geq \max \{ K_m, K_{\epsilon} \}$.

Similarly, if $\hat{\sigma}_{\epsilon,m} K_{\epsilon,m} - c < 0$ and $\hat{\sigma}_{f,m}^{-1} - \hat{\sigma}_{f,m}^{-1} < 0$, the volatility impact of type $m$ investors is positive. A sufficient condition to ensure this is $K_m \leq \min \{ K_{\mu}, K_{\epsilon} \}$.

Finally, to show that the volatility impact is decreasing in $K_m$, taking the derivative yields

$$\frac{\partial \Delta^m VTS[f_t - r_p]}{\partial K_m} = \frac{\partial^c}{\partial K_m} \left[ (1-c_m) \frac{\partial c_m}{\partial K_m} - (1-c) \frac{\partial c}{\partial K_m} \right] \sigma_c + \left( d \frac{\partial d}{\partial K_m} - d_m \frac{\partial d_m}{\partial K_m} \right) \sigma_s. \quad (A.152)$$

By construction, the coefficients $c_m$ and $d_m$ are independent of $K_m$. Thus, this
simplifies to
\[ \frac{\partial \Delta m VTS [f_t - r p_t]}{\partial K_m} = 2\psi^{-1}_m \left[ d \frac{\partial d}{\partial K_m} \sigma_s - (1 - c) \frac{\partial c}{\partial K_m} \sigma_e \right]. \] (A.153)

From previous results, it is clear that \( \frac{\partial d}{\partial K_m} \geq 0 \) and \( \frac{\partial c}{\partial K_m} \). Combined with the fact that \( d \leq 0 \) and \( c < 1 \), this implies that the volatility impact is decreasing in attention capacity.

\[ \Box \]

Proof of Lemma 6:

Proof. The proof relies on showing that, if the marginal utilities are negative for both fundamental and idiosyncratic attention, the marginal utility of fundamental attention at a future date must be negative, violating the conditions of optimality.

The first order conditions are given by
\[
[K_{\mu,i,t}] = \frac{1}{2} \sum_{s=t}^{T} \hat{\sigma}_{\mu,i,s} \hat{\sigma}_{f,i,s}^{-2} \left\{ 2(c_s - b_s) \hat{\sigma}_{e,i,s} + v_s + a_s^2 \right\} - \lambda_i + \nu_{\mu,i,t}
\]
\[
[K_{\epsilon,i,t}] = \frac{1}{2} \hat{\sigma}_{\epsilon,i,t} \hat{\sigma}_{f,i,t}^{-2} \left\{ 2(b_t - c_t) \hat{\sigma}_{\mu,i,t} + v_t + a_t^2 \right\} - \lambda_i + \nu_{\epsilon,i,t}.
\] (A.154)

Suppose BWOC that, for some investor \( i \), the information capacity constraint is not binding. This implies that, for some interior allocation \( \sum_{t=1}^{T} (K_{\mu,i,t} + K_{\epsilon,i,t}) < K_{m_i} \), the gradient of expected utility is weakly negative, \( \nabla \mathbb{E} [U_i] \leq 0 \).

First, note that the gradient of expected utility cannot be strictly positive, as this implies that both \( b_t - c_t \) and \( c_t - b_t \) are less than zero, which is clearly a contradiction. Thus, the only case in which such a solution exists is when the gradient is \textit{exactly} zero, \( \nabla \mathbb{E} [U_i] = 0 \). As this implies \( \lambda_i = 0 \), this is precisely the case for which the no-forgetting constraints are not binding for any time period \( \nu_{\mu,i,t} = \nu_{\epsilon,i,t} = 0 \).

As both marginal utilities are zero, it thus follows the marginal utility of idiosyn-
cratic attention that
\[ c_t - b_t = \frac{v_t + a_t^2}{2\sigma_{\mu,i,t}} > 0. \] (A.155)

This, in turn, implies
\[ \frac{\partial E[u_{i,t}]}{\partial K_{\mu,i,t}} = \sigma_{\mu,i,t}^2 \hat{\sigma}_{f,i,t}^{-2} \{ 2(c_t - b_t) \hat{\sigma}_{\epsilon,i,t} + v_t + a_t^2 \} > 0. \] (A.156)

For \( t < T \), note that the time \( t+1 \) marginal utility of fundamental attention is simply the time \( t \) marginal utility less (A.156):

\[ \frac{\partial E[U_i]}{\partial K_{\mu,i,t+1}} = \frac{\partial E[U_i]}{\partial K_{\mu,i,t}} - \frac{\partial E[u_{i,t}]}{\partial K_{\mu,i,t}} = 0, \] (A.157)

where the first term is 0 by supposition and the second is positive by (A.156).

For \( t = T \), Eq. (A.156) directly implies that \( c_T > b_T \) and thus

\[ \frac{\partial E[U_i]}{\partial K_{\mu,i,T}} = \sigma_{\mu,i,T}^2 \hat{\sigma}_{f,i,T}^{-2} \{ 2(c_T - b_T) \hat{\sigma}_{\epsilon,i,T} + v_T + a_T^2 \} > 0. \] (A.158)

Thus, for any date \( t \), marginal utilities cannot all be weakly negative, contradicting the original assertion. Thus, the information capacity constraint must always bind.

\[ \square \]

**Proof of Lemma 7:**

*Proof.* The no-forgetting constraints do not bind for investor \( i \) if both fundamental (2.20) and idiosyncratic (2.21) attention are positive. Suppose first BWOC that the no-forgetting constraints do not bind for investor \( i \) and \( \zeta \leq \frac{(T-1)(T\sigma_{\mu}^2 - (K_{m_i} + T\sigma_{\epsilon}^{-1})^2)}{T(K_{m_i} + \sigma_{\mu}^{-1} + T\sigma_{\epsilon}^{-1})}. \)
Then, as fundamental attention is clearly increasing in $\zeta$, it follows that

$$
K_{\mu,i} \leq \frac{- (K_{m_i} + T \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon) + \sqrt{T \left( K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon \right) \left( K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon + \frac{(T-1)(T\sigma^{-2}_\mu - (K_{m_i} + T \sigma^{-1}_\varepsilon)^2) - \sigma^{-1}_\mu}{T(K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon)} \right)}}{T-1} \\
= \frac{- (K_{m_i} + T \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon) + \sqrt{(K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon)^2}}{T-1} \\
= 0,
$$

(A.159)

where the second line follows from simplifying the square root term. Thus, fundamental attention is weakly negative, violating the assumption that the no-forgetting constraint on fundamental attention does not bind.

Next, suppose BWOC that the no-forgetting constraints do not bind for investor $i$ and $\zeta \geq \frac{(T-1)(K_{m_i} + \sigma^{-1}_\mu)^2 - T \sigma^{-2}_\varepsilon}{K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon}$. Idiosyncratic attention is decreasing in $\zeta$, so

$$
K_{\varepsilon,i} \leq \frac{K_{m_i} + \sigma^{-1}_\mu + \sigma^{-1}_\varepsilon - \sqrt{T-1 \left( K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon \right) \left( K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon - \frac{(T-1)(K_{m_i} + \sigma^{-1}_\mu)^2 - T \sigma^{-2}_\varepsilon}{K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon} \right)}}{T-1} \\
= \frac{K_{m_i} + \sigma^{-1}_\mu + \sigma^{-1}_\varepsilon - \sqrt{(K_{m_i} + \sigma^{-1}_\mu + \sigma^{-1}_\varepsilon)^2}}{T-1} \\
= 0,
$$

(A.160)

where the second line follows from simplifying the square root term.

If $\zeta \geq \frac{(T-1)(K_{m_i} + \sigma^{-1}_\mu)^2 - T \sigma^{-2}_\varepsilon}{K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon}$, idiosyncratic attention is weakly negative, violating the assumption that the no-forgetting constraint on idiosyncratic attention does not bind.

Taking these two parts together, it follows that, if the no-forgetting constraints do not bind for investor $i$, then

$$
\zeta \in \left( \frac{(T-1) \left( T \sigma^{-2}_\mu - (K_{m_i} + T \sigma^{-1}_\varepsilon)^2 \right)}{T(K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon)} , \frac{(T-1) \left( (K_{m_i} + \sigma^{-1}_\mu)^2 - T \sigma^{-2}_\varepsilon \right)}{K_{m_i} + \sigma^{-1}_\mu + T \sigma^{-1}_\varepsilon} \right)
$$

(A.161)

193
Proof of Proposition 14:

Proof. First consider the derivative of representative cross sectional variance (Eq. (A.55)):

$$\frac{\partial \gamma^{CS}[q_{m,t}]}{\partial K_m} = 2\gamma^{-2}\sigma_\mu \hat{\sigma}_{f,m}^{-1} (\hat{\sigma}_{\mu,m} K_{\mu,m} - b) \frac{\partial \gamma^{-1}(\hat{\sigma}_{\mu,m} K_{\mu,m} - b)}{\partial K_m} \sigma_\mu$$

$$= 2\gamma^{-2}\sigma_\mu \hat{\sigma}_{f,m}^{-1} (\hat{\sigma}_{\mu,m} K_{\mu,m} - b) \left(\frac{\partial \gamma^{-1}(\hat{\sigma}_{\mu,m} K_{\mu,m} - b)}{\partial K_m} \right)$$

(A.162)

where $\mathcal{C}^{CS}[q_{m,t}, f_t]$ is positive by Proposition 3. Thus, cross sectional variance is increasing in $K_m$ iff $\hat{\sigma}_{\mu,m} K_{\mu,m} > b$ and decreasing otherwise. Recall from the definition of $\bar{K}_\mu$ in Eq. (A.97) that for all $K_m > \bar{K}_\mu$, $\hat{\sigma}_{\mu,m} K_{\mu,m} > b$ and vice versa. Thus, for all $K_m > \bar{K}_\mu$, cross sectional representative variance is increasing in attention capacity and decreasing otherwise.

Next, the derivative of representative time series variance (Eq. (A.56)) is

$$\frac{\partial \gamma^{TS}[q_{m,t}]}{\partial K_m} = 2\gamma^{-2}\hat{\sigma}_{f,m}^{-1} \left\{ (\hat{\sigma}_{\epsilon,m} K_{\epsilon,m} - c) \frac{\partial \hat{\sigma}_{f,m}^{-1}(\hat{\sigma}_{\epsilon,m} K_{\epsilon,m} - c)}{\partial K_m} + \sigma_\epsilon \frac{\partial \hat{\sigma}_{f,m}^{-1}}{\partial K_m} \right\}$$

$$= 2\gamma^{-1}\hat{\sigma}_{f,m}^{-1} (\hat{\sigma}_{\epsilon,m} K_{\epsilon,m} - c) \frac{\partial \gamma^{-1}\hat{\sigma}_{f,m}^{-1}(\hat{\sigma}_{\epsilon,m} K_{\epsilon,m} - c)}{\partial K_m} \sigma_\epsilon + 2\gamma^{-2}\sigma_\epsilon^2 \hat{\sigma}_{f,m}^{-1} \frac{\partial \hat{\sigma}_{f,m}^{-1}}{\partial K_m}$$

$$= 2\gamma^{-1}\hat{\sigma}_{f,m}^{-1} (\hat{\sigma}_{\epsilon,m} K_{\epsilon,m} - c) \left(\frac{\partial \gamma^{-1}\hat{\sigma}_{f,m}^{-1}(\hat{\sigma}_{\epsilon,m} K_{\epsilon,m} - c)}{\partial K_m} \right) + 2\gamma^{-2}\sigma_\epsilon^2 \hat{\sigma}_{f,m}^{-1} \left(\frac{\partial \hat{\sigma}_{f,m}^{-1}}{\partial K_m} \right)$$

(A.163)

It is clear that, for $\hat{\sigma}_{\epsilon,m} K_{\epsilon,m} \geq c$ (i.e., $K_m \geq \bar{K}_\epsilon$ in Eq. (A.97)), the time series variance of the representative investor is increasing in $K_m$. Finally, note that the second term can be made arbitrarily small as $\sigma_\epsilon \to 0$. Thus, the representative time series portfolio variance is decreasing in attention capacity if $K_m < \bar{K}_\epsilon$ for sufficiently small $\sigma_\epsilon$. 

□
Proof of Proposition 15:

Proof. Without loss of generality, the derivatives with respect to investor $i$’s attention capacity are considered (those wrt investor $j$’s are symmetric). First, note that the cross sectional covariance can be written as

$$ C^CS [q_{i,t}, q_{j,t}] = \gamma^{-2} \hat{\sigma}_{f,i}^{-1} \hat{\sigma}_{f,j}^{-1} \left( \hat{\sigma}_{\epsilon,i} K_{\mu,i} - b \right) \left( \hat{\sigma}_{\epsilon,j} K_{\mu,j} - b \right) \quad (A.164) $$

$$ = \sigma^{-1} C^CS [q_{i,t}, f_t] C^CS [q_{j,t}, f_t]. $$

The derivative with respect to investor $i$’s attention capacity is

$$ \frac{\partial C^CS [q_{i,t}, q_{j,t}]}{\partial K_{m_i}} = \sigma^{-1} C^CS [q_{i,t}, f_t] \frac{\partial C^CS [q_{j,t}, f_t]}{\partial K_{m_i}}, \quad (A.165) $$

where the bracketed term is positive by Proposition 3. Thus, it is clear that the cross sectional covariance is increasing in investor $i$’s attention capacity iff $C^CS [q_{i,t}, f_t] > 0$, which only occurs if $K_{m_i} > \bar{K}_mu$, where $\bar{K}_\mu$ is given in Eq. (A.97).

Further, note that the cross sectional covariance is negative only if either $\hat{\sigma}_{\mu,i} K_{\mu,i} < b$ or $\hat{\sigma}_{\mu,j} K_{\mu,j} < b$, but not both. This is equivalently written as $K_{m_i} < \bar{K}_\mu < K_{m_j}$.

Next, the time series covariance is given by

$$ C^{TS} [q_{i,t}, q_{j,t}] = \gamma^{-2} \hat{\sigma}_{f,i}^{-1} \hat{\sigma}_{f,j}^{-1} \left\{ \left( \hat{\sigma}_{\epsilon,i} K_{\epsilon,i} - c \right) \left( \hat{\sigma}_{\epsilon,j} K_{\epsilon,j} - c \right) + d^2 \sigma_s \right\} $$

$$ = \sigma^{-1} C^{TS} [q_{i,t}, f_t] C^{TS} [q_{j,t}, f_t] + \gamma^{-2} \hat{\sigma}_{f,i}^{-1} \hat{\sigma}_{f,j}^{-1} d^2 \sigma_s. \quad (A.166) $$

The derivative with respect to $K_{m_i}$ is

$$ \frac{\partial C^{TS} [q_{i,t}, q_{j,t}]}{\partial K_{m_i}} = \sigma^{-1} C^{TS} [q_{i,t}, f_t] \frac{\partial C^{TS} [q_{j,t}, f_t]}{\partial K_{m_i}} + \gamma^{-2} \hat{\sigma}_{f,i}^{-1} d^2 \sigma_s \frac{\partial \hat{\sigma}_{f,i}}{\partial K_{m_i}}, \quad (A.167) $$
where \( \frac{\partial \mathcal{C}^{TS}[q_{i,t}, f_t]}{\partial K_{m_i}} \) is positive by Proposition 3. As before, it is clear the time series covariance is increasing in investor \( i \)'s attention capacity if \( \mathcal{C}^{TS}[q_{j,t}, f_t] \geq 0 \), which occurs only if \( K_{m_j} \geq \bar{K}_e \).

Finally, note that, as the second term approaches zero in the limit as \( \sigma_s \to 0 \), the second term can be made arbitrarily small. Thus, if investor \( j \) has below average attention \( K_{m_j} < \bar{K}_e \), for sufficiently small \( \sigma_s \), time series portfolio covariance is decreasing.

\[ \square \]

**Proof of Proposition 16:**

*Proof.* First, note that, if the no-forgetting constraint binds for either the fundamental or idiosyncratic component, an increase in attention capacity either (i) only increases attention to the component for which the no-forgetting constraint does not bind or (ii) increase the attention to both, so the no-forgetting constraints no longer bind (by the concavity of the objective function). In either case, attention is weakly increasing in attention capacity.

Next, consider the case in which the no-forgetting constraints do not bind. The derivative of fundamental attention is

\[
\frac{\partial K_{\mu,i}}{\partial K_{m_i}} = \frac{2T(2\delta_i + \zeta)}{2\sqrt{T(\delta_i^2 + \delta_i \zeta)}} - \frac{1}{T - 1}, \tag{A.168}
\]

which is convex in \( \zeta \) and has roots \( \zeta = \frac{2\delta_i \left(-(T-1) \pm \sqrt{(T-1)}\right)}{T} \).

As \( T \geq 2 \), these roots are imaginary. The derivative is convex, which implies that the derivative is strictly positive for all values of \( \zeta \).

196
Finally, the derivative of idiosyncratic attention is given by

\[ \frac{\partial K_{e,i}}{\partial K_{m,i}} = \frac{1 - \frac{2\delta_i + \zeta}{2\sqrt{T(\delta_i^2 + \delta_i \zeta)}}}{T - 1}, \]

which is concave in \( \zeta \) and has roots \( \zeta = 2\delta_i \left( T - 1 \pm \sqrt{T(T - 1)} \right) \). As the derivative is concave, it must be positive for all values of \( \zeta \) between these two roots. The proof rests on showing that the bounds specified in Lemma 7 are tighter than the bounds necessary to make this derivative positive, ensuring that, for all possible values of \( \zeta \) for which the no-forgetting constraint binds, idiosyncratic attention is increasing in attention capacity.

The difference between the lower root and the lower bound in Lemma 7 is

\[ 2\delta_i \left( T - 1 - \sqrt{T(T - 1)} \right) - \frac{(T - 1) \left( T\sigma_{\mu}^{-2} - (K_{m,i} + T\sigma_{\epsilon}^{-1})^2 \right)}{T\delta_i} \]

\[ = \frac{2\delta_i^2 T \left( T - 1 - \sqrt{T(T - 1)} \right) - (T - 1) \left( T\sigma_{\mu}^{-2} - (K_{m,i} + T\sigma_{\epsilon}^{-1})^2 \right)}{T\delta_i} \]

\[ = \frac{(T\delta_i)^{-1} \left\{ \left( T(T - 1) - 2T\sqrt{T(T - 1)} \right) \sigma_{\mu}^{-2} \right.}{T\delta_i} \]

\[ + \left. \left( \frac{2T(T - 1) - 2T\sqrt{T(T - 1)} + T - 1}{\sigma_{\mu}^{-2}} \right) \left( K_{m,i} + T\sigma_{\epsilon}^{-1} \right) \right\} \]

\[ + 4 \left( \frac{T(T - 1) - T\sqrt{T(T - 1)}}{\sigma_{\mu}^{-1}} \right) \left( K_{m,i} + T\sigma_{\epsilon}^{-1} \right) \]

\[ < 0, \]

where the signs in the brackets all follow from the fact that \( T \geq 2 \). Hence, the resulting inequality is negative, meaning that, for the values below which idiosyncratic attention is decreasing in attention capacity, the no-forgetting constraints bind.

Now, consider the difference between the upper root and the upper bound in
Lemma 7:

\[ 2\delta_i \left( T - 1 + \sqrt{T(T-1)} \right) - \frac{(T-1) \left( (K_{m_i} + \sigma^{-1}_\mu)^2 - T\sigma^{-2}_\epsilon \right)}{\delta_i} \]

\[ = \frac{2\delta_i^2 \left( T - 1 + \sqrt{T(T-1)} \right) - (T-1) \left( (K_{m_i} + \sigma^{-1}_\mu)^2 - T\sigma^{-2}_\epsilon \right)}{\delta_i} \]

\[ = \delta_i^{-1} \left\{ \left( T - 1 + 2\sqrt{T(T-1)} \right) \left( (K_{m_i} - \sigma^{-1}_\mu)^2 + T\sigma^{-2}_\epsilon \right) + 2\left( T - 1 + \sqrt{T(T-1)} \right) T\sigma^{-1}_\epsilon \left( K_{m_i} + \sigma^{-1}_\mu \right) \right\} \]

\[ > 0, \]

where the signs of the bracketed terms are because \( T \geq 2 \). As a result, for all values above which idiosyncratic attention is increasing in attention capacity, the no-forgetting constraints bind.

Combining these two results, for all feasible values of \( \zeta \) for which the no-forgetting constraints do not bind, idiosyncratic attention is increasing in attention capacity. \( \square \)

**Proof of Proposition 17:**

*Proof.* The proof considers the derivatives with respect to fundamental and idiosyncratic precisions \( \sigma^{01}_\mu \) and \( \sigma^{-1}_\epsilon \). Noting that \( \zeta \) is an implicit function of these precisions, the derivative of fundamental attention wrt \( \sigma^{-1}_\mu \) is

\[ \frac{\partial K_{\mu,i}}{\partial \sigma^{-1}_\mu} = \frac{T(2\delta_i + \zeta)}{\sqrt{T(\delta_i^2 + \delta_i \zeta)}} - 2T \]

\[ + \frac{T\delta_i \frac{\partial \zeta}{\partial \sigma^{-1}_\mu}}{2(T-1)} \frac{\sqrt{T(\delta_i^2 + \delta_i \zeta)}}{2(T-1)} \sqrt{T(\delta_i^2 + \delta_i \zeta)}. \]

(A.172)

The first term is the change in fundamental attention holding prices fixed (i.e., \( \zeta \) not changing), while the second term accounts for the impact a change in \( \zeta \) has
on fundamental attention. Note that the first term is convex in \( \zeta \) and has roots
\[
\zeta = 2\delta_i \left( T - 1 \pm \sqrt{T(T-1)} \right).
\]
As in the proof of Proposition 16, these roots are outside the bounds specified in Lemma 7, ensuring that the first term is negative for all feasible values of \( \zeta \).

The derivative is thus negative if the change in prices does not "undo" this effect. Technically, if the absolute value of the first term is greater than the second term, the derivative will be negative. Note that, in the limit, as the maximum attention \( \bar{K} \to 0 \), \( \frac{\partial \zeta}{\partial \sigma_\mu} \to 0 \). Thus, there exists some uniform bound on attention capacity that ensures the derivative is negative.

Next, the derivative of fundamental attention with respect to idiosyncratic attention is
\[
\frac{\partial K_{\epsilon,i}}{\partial \sigma_{\epsilon}^{-1}} = \frac{T^2(2\delta_i + \zeta)}{\sqrt{T(\delta_i^2 + \delta_i \zeta)}} - \frac{2T}{2(T-1)} + \frac{T\delta_i \frac{\partial \zeta}{\partial \sigma_\mu}}{2(T-1) \sqrt{T(\delta_i^2 + \delta_i \zeta)}}.
\] (A.173)

Again, the first term is the change that occurs holding prices fixed, while the second accounts for a change in the value of \( \zeta \). The first term is convex in \( \zeta \) with roots
\[
\zeta = \frac{2\delta_i (-T - 1) \pm \sqrt{-(T-1)}}{T},
\]
which are clearly imaginary. As a result, the first term is positive for all values of \( \zeta \).

The entire derivative is therefore positive if the effect of \textit{zeta} is not too large. As the maximum attention \( \bar{K} \to 0 \), \( \zeta \) cannot change too much: \( \frac{\partial \zeta}{\partial \sigma_\mu} \to 0 \). Thus, for attention capacity sufficiently uniformly bounded, this derivative is positive. \( \square \)

### A.1.4 Factor Universe

When constructing these factors, the usual rules apply. For accounting data, a minimum of six months is required between the Compustat data date and use in the portfolio formation. Only common equity (share codes 10 and 11) listed on the NYSE, AMEX, and NASDAQ (exchange codes 1, 2, 3) with non-missing return,
price, and shares outstanding data are used. The bottom 5% of stocks by market capitalization are omitted. All portfolios are rebalanced monthly unless otherwise stated.
Table A.13: Anomaly Characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abbrev.</th>
<th>Publication</th>
<th>Sample Period</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>size</td>
<td>Fama and French (1993)</td>
<td>1963-1990</td>
<td>Log market value at end of June. ( size_t = \log (prc_{\text{t,Jun}} \times shrout_{\text{t,Jun}}) ).</td>
</tr>
<tr>
<td>Value (annual)</td>
<td>value</td>
<td>Fama and French (1993)</td>
<td>1963-1990</td>
<td>Book equity as of December ( t - 1 ) divided by market equity in June. ( value_t = \frac{prc_{\text{t,Dec}} \times shrout_{\text{t,Dec}}}{prc_{\text{t-1,Dec}} \times shrout_{\text{t-1,Dec}}} ). The book value of equity is defined as the book value of shareholders’ equity plus deferred taxes and investment tax credit (( txditec ), if available) minus book value of preferred stock. Shareholder’s equity is defined (based on availability) as: (i) shareholders’ equity (seq), (ii) common equity (ceq) plus carrying value of preferred stock (upstk), or (iii) total assets (at) minus total liabilities (lt). Redemption value (psktrv), liquidation value (psklv), or carrying value (upstk) is used for the book value of preferred stock.(^7)</td>
</tr>
<tr>
<td>Operating Profitability</td>
<td>oprof</td>
<td>Fama and French (2015)</td>
<td>1963-2013</td>
<td>Revenue minus COGS minus SG&amp;A and interest expense (if available) divided by book equity. ( oprof_t = \frac{\text{rev}_t - \text{cogs}_t - \text{sga}_t - \text{int}_t}{\text{seq}_t} ).</td>
</tr>
<tr>
<td>Gross Profitability</td>
<td>gprof</td>
<td>Novy-Marx (2013)</td>
<td>1963-2010</td>
<td>Revenue minus COGS divided by total assets. ( gprof_t = \frac{\text{rev}_t - \text{cogs}_t}{\text{at}_t} ).</td>
</tr>
<tr>
<td>Value Profitability</td>
<td>vprof</td>
<td>Novy-Marx (2013)</td>
<td>1963-2010</td>
<td>Sum of ranks of value and gross profitability, ( vprof_t = \text{rank}(value_t) + \text{rank}(gprof_t) ).</td>
</tr>
</tbody>
</table>

\(^7\)This definition comes from Ken French’s website, https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/variable_definitions.html
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abbrev.</th>
<th>Publication</th>
<th>Original Period</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Piotroski’s F-score</td>
<td>fscore</td>
<td>Piotroski (2000)</td>
<td>1976-1996</td>
<td>Sum of financial indicator variables. $f_{score_t} = 1[roa_t &gt; 0] + 1[\Delta roa_t &gt; 0] + 1[cfo_t &gt; 0] + 1[\Delta margin_t &gt; 0] + 1[\Delta turn_t &gt; 0] + 1[\Delta leverage_t &lt; 0] + 1[\Delta liquid_t &gt; 0] + 1[eqiss_t &lt; 0]$, where $roa_t = \frac{\text{ibn}<em>{t}}{\text{at}</em>{t-1}}$, $cfo_t = \frac{\text{concf}<em>{t}}{\text{at}</em>{t-1}}$, $accrual_t = roa_t - cfo_t$, $margin_t = 1 - \frac{\text{cost}<em>{t}}{\text{at}</em>{t-1}}$, $turn_t = \frac{\text{rev}<em>{t}}{\text{at}</em>{t-1}}$, $leverage_t = \frac{\text{dltit}<em>{t}}{\text{at}</em>{t}}$, $liquid_t = \frac{\text{at}<em>{t}}{\text{leit}</em>{t}}$, $eqiss_t = \text{scstk}<em>{t} - \text{prstkcc}</em>{t}$.</td>
</tr>
<tr>
<td>Debt Issuance</td>
<td>debtiss</td>
<td>Spiess and Affleck-Graves (1999)</td>
<td>1975-1989</td>
<td>Indicator if long term debt issued. $\text{debtiss}_t = 1[\text{dlltis}_t \geq 0]$.</td>
</tr>
<tr>
<td>Share Repurchases</td>
<td>repurch</td>
<td>Ikeberry et al. (1995)</td>
<td>1980-1990</td>
<td>Binary variable if common or preferred shares repurchased. $\text{repurch}_t = 1[\text{prstkcc}_t &gt; 0]$.</td>
</tr>
<tr>
<td>Share issuance (annual)</td>
<td>eqissa</td>
<td>Pontiff and Woodgate (2008)</td>
<td>1970-2003</td>
<td>Percent change in adjusted shares. $eqissa = \log(\text{adjshrout}<em>t) - \log(\text{adjshrout}</em>{t-1})$, where $\text{adjshrout}<em>t = \frac{\text{shrout}</em>{t}}{\prod_{i=1}^{t-1}(1 + \text{adjshr}_{i})}$.</td>
</tr>
<tr>
<td>Accruals</td>
<td>accruals</td>
<td>Sloan (1996)</td>
<td>1962-1991</td>
<td>$\text{accruals}_t = (\Delta ct + \Delta ch_t) - (\Delta lct - \Delta dlc_t - \Delta txp_t) - dp_t$.</td>
</tr>
<tr>
<td>Asset Growth</td>
<td>growth</td>
<td>Cooper et al. (2008)</td>
<td>1968-2003</td>
<td>Percent change in total assets. $\text{growth}_t = \frac{\text{at}<em>t}{\text{at}</em>{t-1}}$.</td>
</tr>
<tr>
<td>Asset Turnover</td>
<td>aturn</td>
<td>Soliman (2008)</td>
<td>1984-2002</td>
<td>Sales scaled by total assets. $\text{aturn}_t = \frac{\text{sales}_t}{\text{at}_t}$.</td>
</tr>
<tr>
<td>Change in Asset Turnover</td>
<td>daturn</td>
<td>Soliman (2008)</td>
<td>1984-2002</td>
<td>$\text{daturn}_t = \Delta \text{aturn}_t$.</td>
</tr>
<tr>
<td>Profit Margin</td>
<td>pmargins</td>
<td>Soliman (2008)</td>
<td>1984-2002</td>
<td>Operating income divided by sales. $\text{pmargins}_t = \frac{\text{oiapd}_t}{\text{sales}_t}$.</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abbrev.</th>
<th>Publication</th>
<th>Original Period</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in Profit Margin</td>
<td>dpmargins</td>
<td>Soliman (2008)</td>
<td>1984-2002</td>
<td>(dpmargins_t = \Delta p\text{margins}_t)</td>
</tr>
<tr>
<td>Return on Noncurrent Operating Assets</td>
<td>rnoa</td>
<td>Soliman (2008)</td>
<td>1984-2002</td>
<td>Profit margin times asset turnover. (rnoa_t = p\text{margins}_t \times a\text{turn}_t).</td>
</tr>
<tr>
<td>Change in Return on Noncurrent Operating Assets</td>
<td>drnoa</td>
<td>Soliman (2008)</td>
<td>1984-2002</td>
<td>(drnoa_t = \Delta rnoa_t).</td>
</tr>
<tr>
<td>Gross Margins</td>
<td>gmargins</td>
<td>Novy-Marx (2013)</td>
<td>1963-2010</td>
<td>Gross profitability scaled by sales. (gmargins = \frac{gp_{t}}{\text{sales}_{t}})</td>
</tr>
<tr>
<td>Value-Momentum</td>
<td>valmom</td>
<td>Novy-Marx (2013)</td>
<td>1963-2010</td>
<td>Sum of ranks of value and momentum. (valmom_t = \text{rank}(\text{value}<em>{t}) + \text{rank}(\text{mom6m}</em>{t}))</td>
</tr>
<tr>
<td>Value-Momentum-Profitability</td>
<td>valmomgpro</td>
<td>Novy-Marx (2013)</td>
<td>1963-2010</td>
<td>Sum of ranks of value momentum, and gross profitability. (valmom_t = \text{rank}(\text{value}<em>{t}) + \text{rank}(\text{mom6m}</em>{t}) + \text{rank}(\text{gprof}_{t}))</td>
</tr>
<tr>
<td>Dividend Yield</td>
<td>divyld</td>
<td>Naranjo et al. (1998)</td>
<td>1963-1994</td>
<td>Dividends over previous year divided by market value. (\text{divyld}<em>t = \frac{\text{Div}</em>{t-1,\text{dec}} \times \text{shrou}<em>{t-1,\text{dec}}}{\text{prc}</em>{t-1,\text{dec}} \times \text{shrou}_{t-1,\text{dec}}}).</td>
</tr>
<tr>
<td>Earnings/Price</td>
<td>pe</td>
<td>Basu (1977)</td>
<td>1956-1971</td>
<td>Market value divided by income before extraordinary items. (\text{ep}<em>{t} = \frac{\text{prc}</em>{t-1,\text{dec}} \times \text{shrou}<em>{t-1,\text{dec}}}{\text{ib}</em>{t}}).</td>
</tr>
<tr>
<td>Cash Flow/Market Value of Equity</td>
<td>cfp</td>
<td>Lakonishok et al. (1994)</td>
<td>1963-1990</td>
<td>Net income plus depreciation and amortization scaled by market value of equity. (\text{cfp}<em>{t} = \frac{\text{ib}</em>{t} + \text{dp}<em>{t}}{\text{prc}</em>{t,\text{dec}} \times \text{shrou}_{t,\text{dec}}}).</td>
</tr>
<tr>
<td>Sales Growth</td>
<td>sgrowth</td>
<td>Lakonishok et al. (1994)</td>
<td>1963-1990</td>
<td>Percent change in net sales. (sgrowth = \frac{\text{sales}<em>{t}}{\text{sales}</em>{t-1}} - 1).</td>
</tr>
</tbody>
</table>

Continued on next page
Table A.13 – Continued from previous page

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abbrev.</th>
<th>Publication</th>
<th>Original Period</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net Operating Assets</td>
<td>noa</td>
<td>Hirshleifer et al. (2004)</td>
<td>1964-2002</td>
<td>Operating assets minus liabilities scaled by total assets. ( noa = \frac{\text{operating assets} - \text{total liabilities}}{\text{total assets}} ).</td>
</tr>
<tr>
<td>Investment</td>
<td>inv</td>
<td>Chen et al. (2011)</td>
<td>1972-2010</td>
<td>Change in PP&amp;E plus change in Inventories divided by lagged assets. ( \text{inv}_t = \frac{\Delta \text{PP&amp;E}<em>t - \Delta \text{invent}}{\text{at}</em>{t-1}} ).</td>
</tr>
<tr>
<td>Return on Assets</td>
<td>roa</td>
<td>Chen et al. (2011)</td>
<td>1972-2010</td>
<td>Income before extraordinary items divided by one-quarter lagged total assets. ( \text{roa}_t = \frac{\text{income before extraordinary items}}{\text{total assets}} ).</td>
</tr>
<tr>
<td>Return on Assets (annual)</td>
<td>roaa</td>
<td>Chen et al. (2011)</td>
<td>1972-2010</td>
<td>Income before extraordinary items divided by one-year lagged total assets. ( \text{roaa}_t = \frac{\text{income before extraordinary items}}{\text{total assets}} ).</td>
</tr>
<tr>
<td>Return on Book Equity</td>
<td>roe</td>
<td>Chen et al. (2011)</td>
<td>1972-2010</td>
<td>Income before extraordinary items divided by one-quarter lagged book equity. ( \text{roe} = \frac{\text{income before extraordinary items}}{\text{book equity}} ). The book value of equity is defined as the book value of shareholders’ equity plus deferred taxes and investment tax credit (txditeq, if available) minus book value of preferred stock. Shareholder’s equity is defined (based on availability) as (i) shareholders’ equity (seqq), (ii) common equity (ceqq) plus carrying value of preferred stock (upstkq), or (iii) total assets (atq) minus total liabilities (ltq). Redemption value (pstkq) or carrying value (upstkq) is used for the book value of preferred stock.</td>
</tr>
<tr>
<td>Return on Market Equity</td>
<td>rme</td>
<td>Chen et al. (2011)</td>
<td>1972-2010</td>
<td>Income before extraordinary items divided by one-quarter lagged market equity ( \text{rme}_t = \frac{\text{income before extraordinary items}}{\text{market equity}} ).</td>
</tr>
</tbody>
</table>

\(^8\)This definition is the quarterly analogy to definition of book value on Ken French’s website.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abbrev.</th>
<th>Publication</th>
<th>Original Period</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investment-to-Capital</td>
<td>invcap</td>
<td>Xing (2008)</td>
<td>1964-2003</td>
<td>Capital expenditures divided by PP&amp;E. ( \frac{\text{invcap}_t}{\text{ppent}_t} ).</td>
</tr>
<tr>
<td>Investment Growth</td>
<td>invgrowth</td>
<td>Xing (2008)</td>
<td>1964-2003</td>
<td>Percent change in investment. ( \frac{\text{invgrowth}<em>t}{\text{cap}</em>{t-1}} ).</td>
</tr>
<tr>
<td>Leverage</td>
<td>lev</td>
<td>Bhandari (1988)</td>
<td>1948-1981</td>
<td>Total assets minus common equity divided by market value. ( \text{lev} = \frac{\text{atl}<em>t - \text{eqit}}{\text{prc}</em>{t-1}} ).</td>
</tr>
<tr>
<td>Return on Equity (annual)</td>
<td>roea</td>
<td>Haugen and Baker (1996)</td>
<td>1979-1993</td>
<td>Income before extraordinary items divided by one-year lagged book equity. ( \text{roea} = \frac{\text{ibr}<em>t}{\text{be}</em>{t-1}} ). Book equity is defined as in Fama and French (1993).</td>
</tr>
<tr>
<td>Sales-to-Price</td>
<td>sp</td>
<td>Barbee et al. (1996)</td>
<td>1979-1991</td>
<td>Ratio of sales to price. ( \text{sp} = \frac{\text{sales}<em>{t-1,dec} \times \text{shrount}</em>{t-1,dec}}{\text{prc}_{t-1,dec}} ).</td>
</tr>
<tr>
<td>Growth in LTNOA</td>
<td>ltnoagrowth</td>
<td>Fairfield et al. (2003)</td>
<td>1964-1993</td>
<td>Growth in net operating assets minus accruals. ( \text{ltnoagrowth} = \text{grnoat} - \text{acc}_t ). Net operating assets defined as ( \text{noat} = \frac{\text{rect}_t + \text{inret}_t + \text{aco}_t + \text{ppent}_t + \text{intan}_t + \text{ao}_t - \text{ap}_t - \text{loa}_t - \text{lo}_t}{\text{at}<em>t} ). ( \text{grnoat} = \text{noat} - \text{noat}</em>{t-1} ). Accruals are defined as ( \text{acc}_t = \Delta \text{rect}_t + \Delta \text{inret}_t + \Delta \text{aco}_t - \Delta \text{ap}_t - \Delta \text{loa}_t - \Delta \text{lo}_t ).</td>
</tr>
<tr>
<td>Momentum (6m)</td>
<td>mom6m</td>
<td>Jegadeesh and Titman (1993)</td>
<td>1965-1989</td>
<td>Sum of 6 previous months returns skipping the most recent month. ( \text{mom6m} = \sum_{i=2}^{7} \text{rt}_{i-1} ).</td>
</tr>
<tr>
<td>Momentum (12m)</td>
<td>mom12m</td>
<td>Jegadeesh and Titman (1993)</td>
<td>1965-1989</td>
<td>Sum of 12 previous months returns skipping the most recent month. ( \text{mom12m} = \sum_{i=2}^{13} \text{rt}_{i-1} ).</td>
</tr>
<tr>
<td>Momentum-Reversal</td>
<td>momrev</td>
<td>Jegadeesh and Titman (1993)</td>
<td>1965-1989</td>
<td>Sum of returns from 19 months prior to 14 months prior. ( \text{momrev} = \sum_{i=14}^{19} \text{rt}_{i-1} ).</td>
</tr>
<tr>
<td>Industry Momentum</td>
<td>indmom</td>
<td>Moskowitz and Grinblatt (1999)</td>
<td>1963-1995</td>
<td>Ranking of industry returns based on two-digit SIC codes. ( \text{indmom} = \text{rank} \left( \sum_{i=1}^{6} \text{rt}_{i-1} \right) ).</td>
</tr>
</tbody>
</table>

Continued on next page
Table A.13 – Continued from previous page

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abbrev.</th>
<th>Publication</th>
<th>Original Period</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Momentum-Reversal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Short Interest          | short   | Dechow et al. (2001)                 | 1976-1993       | Average monthly returns in the same month over the past 5 years. seasonal = !5
|                         |         |                                      |                 |                                                                               |
| Long-term Reversal      | ltrev   | DeBondt and Thaler (1985)            | 1926-1982       | Cumulative returns from 5 years prior to 1 year prior. ltrev = \sum_{t=13}^{60} r_{t-t}. |
| Value (monthly)         | valuem  | Asness and Frazzini (2013)           | 1950-2011       | Ratio of quarterly book value (defined as in Chen et al. (2011)) to monthly market value. valuem = \frac{sh_{t}}{vm_{t}}. |
| Standardized Unexpected | s ebenfalls Fostet al. (1984)       | 1974-1981       | Change in income before extraordinary items (quarterly) divided by its standard deviation over the past two years skipping the most recent quarter. sœur = \frac{ibqt - ibqt-1}{\sigma_{ibqt-2t-8}}. |
| Earnings                |         |                                      |                 |                                                                               |
| Short-term Reversal     | strev   | Jegadeesh (1990)                     | 1934-1987       | Return in the previous month. strev = r_{t-1} |
| Idiosyncratic Volatility| ivol    | Ang et al. (2006)                    | 1986-2000       | Standard deviation of residuals from regressions of daily stock returns on the Fama and French (1993) three-factor model with a window of one month. A minimum of 17 daily observations required. ivol = sd (r_{i,t} - (\beta_i^{mkt}mkt_t + \beta_i^{smb}smb_t + \beta_i^{hml}hml_t)). |
| Beta Arbitrage          | beta    | Fama and MacBeth (1973)              | 1926-1968       | Beta with respect to the CRSP equal-weighted index over the past 60 months. beta_t = \beta_{t-60:t-1}. |
| Seasonality             | season  | Heston and Sadka (2008)              | 1965-2002       | Average monthly returns in the same month over the past 5 years. season = \sum_{t=1}^{5} r_{t-1,m}. |

Continued on next page
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abbrev.</th>
<th>Publication</th>
<th>Original Period</th>
<th>Construction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry Reversals</td>
<td>indrev</td>
<td>Da et al. (2014)</td>
<td>1982-2009</td>
<td>Stocks’ return the prior month relative to its industry’s prior month return. ( indrev = r_{t-1} - r_{t-1}^{ind} ).</td>
</tr>
<tr>
<td>Industry Reversals (Low Volatility)</td>
<td>indrevlv</td>
<td>Da et al. (2014)</td>
<td>1982-2009</td>
<td>Stocks’ return the prior month relative to its industry’s prior month return only if the stock’s idiosyncratic volatility is below the median NYSE’s idiosyncratic volatility. ( indrev = (r_{t-1} - r_{t-1}^{ind}) 1 [ivol_{i,t} &lt; iVol_{NYSE,t}] ).</td>
</tr>
<tr>
<td>Composite Issuance</td>
<td>compiss</td>
<td>Daniel and Titman (2006)</td>
<td>1968-2003</td>
<td>Percent change in market equity from 5 years prior to 1 year prior minus returns over that time. ( compiss = \log \left( \frac{\text{prc}<em>{t-13} \times \text{shrout}</em>{t-13}}{\text{prc}<em>{t-60} \times \text{shrout}</em>{t-60}} \right) - \sum_{t=13}^{60} r_{t-1} ).</td>
</tr>
<tr>
<td>Price</td>
<td>price</td>
<td>Blume and Huisic (1973)</td>
<td>1932-1971</td>
<td>Log of stock price. ( price = \log (\text{prc}_i) ).</td>
</tr>
<tr>
<td>Firm Age</td>
<td>age</td>
<td>Barry and Brown (1984)</td>
<td>1926-1980</td>
<td>Log of number of months a firm has been listed in CRSP. ( age = \log (1 + nummonths_i) ).</td>
</tr>
<tr>
<td>Share Volume</td>
<td>shvol</td>
<td>Datar et al. (1998)</td>
<td>1962-1991</td>
<td>Average number of shares traded over previous three months divided by shares outstanding. ( shvol = \frac{1}{3} \sum_{t=1}^{3} \frac{\text{volume}<em>{t-1}}{\text{shrout}</em>{t}} ).</td>
</tr>
</tbody>
</table>
Table A.14: Anomaly Summary Statistics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abbrev.</th>
<th>Average Return</th>
<th>Volatility</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accruals</td>
<td>accruals</td>
<td>-0.010</td>
<td>0.082</td>
<td>-0.124</td>
</tr>
<tr>
<td>Firm Age</td>
<td>age</td>
<td>0.019</td>
<td>0.110</td>
<td>0.169</td>
</tr>
<tr>
<td>Asset Turnover</td>
<td>aturn</td>
<td>0.027</td>
<td>0.175</td>
<td>0.156</td>
</tr>
<tr>
<td>Beta Arbitrage</td>
<td>beta</td>
<td>0.026</td>
<td>0.274</td>
<td>0.096</td>
</tr>
<tr>
<td>Cash Flow/Market Value of Equity</td>
<td>cfp</td>
<td>0.086</td>
<td>0.259</td>
<td>0.331</td>
</tr>
<tr>
<td>Composite Issuance</td>
<td>compiss</td>
<td>0.012</td>
<td>0.197</td>
<td>0.061</td>
</tr>
<tr>
<td>Change in Asset Turnover</td>
<td>daturn</td>
<td>0.037</td>
<td>0.148</td>
<td>0.251</td>
</tr>
<tr>
<td>Debt Issuance</td>
<td>debtiss</td>
<td>-0.041</td>
<td>0.229</td>
<td>-0.180</td>
</tr>
<tr>
<td>Dividend Yield</td>
<td>divyld</td>
<td>-0.010</td>
<td>0.182</td>
<td>-0.054</td>
</tr>
<tr>
<td>Change in Profit Margins</td>
<td>dpmargins</td>
<td>0.060</td>
<td>0.253</td>
<td>0.235</td>
</tr>
<tr>
<td>Change in Return on Noncurrent Assets</td>
<td>drnoa</td>
<td>0.037</td>
<td>0.177</td>
<td>0.212</td>
</tr>
<tr>
<td>Equity Issuance (annual)</td>
<td>eqissa</td>
<td>0.045</td>
<td>0.131</td>
<td>0.342</td>
</tr>
<tr>
<td>Piotroski’s F-Score</td>
<td>fscore</td>
<td>0.007</td>
<td>0.163</td>
<td>0.045</td>
</tr>
<tr>
<td>Gross Margins</td>
<td>gmargins</td>
<td>-0.083</td>
<td>0.226</td>
<td>-0.365</td>
</tr>
<tr>
<td>Gross Profitability</td>
<td>gprof</td>
<td>0.059</td>
<td>0.188</td>
<td>0.314</td>
</tr>
<tr>
<td>Asset Growth</td>
<td>growth</td>
<td>0.029</td>
<td>0.144</td>
<td>0.204</td>
</tr>
<tr>
<td>Industry Momentum</td>
<td>indmom</td>
<td>0.050</td>
<td>0.159</td>
<td>0.316</td>
</tr>
<tr>
<td>Industry Momentum-Reversal</td>
<td>indmomrev</td>
<td>0.021</td>
<td>0.119</td>
<td>0.177</td>
</tr>
<tr>
<td>Industry Relative Reversals</td>
<td>indrrev</td>
<td>0.013</td>
<td>0.185</td>
<td>0.072</td>
</tr>
<tr>
<td>Industry Relative Reversals (Low Volatility)</td>
<td>indrrevlv</td>
<td>0.000</td>
<td>0.127</td>
<td>0.003</td>
</tr>
<tr>
<td>Investment</td>
<td>inv</td>
<td>0.013</td>
<td>0.115</td>
<td>0.110</td>
</tr>
<tr>
<td>Investment-to-Capital</td>
<td>invcap</td>
<td>0.010</td>
<td>0.263</td>
<td>0.038</td>
</tr>
<tr>
<td>Investment Growth</td>
<td>invgrowth</td>
<td>0.062</td>
<td>0.187</td>
<td>0.334</td>
</tr>
<tr>
<td>Idiosyncratic Volatility</td>
<td>ivol</td>
<td>-0.021</td>
<td>0.281</td>
<td>-0.074</td>
</tr>
<tr>
<td>Leverage</td>
<td>lev</td>
<td>-0.021</td>
<td>0.255</td>
<td>-0.084</td>
</tr>
<tr>
<td>Growth in LTNOA</td>
<td>ltnoagrowth</td>
<td>-0.008</td>
<td>0.112</td>
<td>-0.071</td>
</tr>
<tr>
<td>Long-term Reversal</td>
<td>ltrev</td>
<td>0.028</td>
<td>0.241</td>
<td>0.117</td>
</tr>
<tr>
<td>Momentum (12m)</td>
<td>mom12m</td>
<td>0.042</td>
<td>0.309</td>
<td>0.135</td>
</tr>
<tr>
<td>Momentum (6m)</td>
<td>mom6m</td>
<td>0.062</td>
<td>0.272</td>
<td>0.228</td>
</tr>
<tr>
<td>Momentum-Reversal</td>
<td>momrev</td>
<td>0.068</td>
<td>0.242</td>
<td>0.281</td>
</tr>
<tr>
<td>Net Operating Assets</td>
<td>noa</td>
<td>0.073</td>
<td>0.116</td>
<td>0.632</td>
</tr>
</tbody>
</table>

Continued on next page
Table A.14 – Continued from previous page

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Abbrev.</th>
<th>Average Return</th>
<th>Volatility</th>
<th>Sharpe Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Profitability</td>
<td>oprof</td>
<td>0.022</td>
<td>0.242</td>
<td>0.089</td>
</tr>
<tr>
<td>Earnings/Price</td>
<td>pe</td>
<td>-0.025</td>
<td>0.096</td>
<td>-0.258</td>
</tr>
<tr>
<td>Profit Margin</td>
<td>pmargins</td>
<td>-0.053</td>
<td>0.215</td>
<td>-0.248</td>
</tr>
<tr>
<td>Price</td>
<td>price</td>
<td>0.067</td>
<td>0.338</td>
<td>0.199</td>
</tr>
<tr>
<td>Share Repurchases</td>
<td>repurch</td>
<td>0.003</td>
<td>0.183</td>
<td>0.017</td>
</tr>
<tr>
<td>Return on Noncurrent</td>
<td>rnoa</td>
<td>-0.006</td>
<td>0.271</td>
<td>-0.021</td>
</tr>
<tr>
<td>Operating Assets</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Return on Assets</td>
<td>roa</td>
<td>0.087</td>
<td>0.237</td>
<td>0.367</td>
</tr>
<tr>
<td>Return on Assets (annual)</td>
<td>roaa</td>
<td>0.035</td>
<td>0.228</td>
<td>0.153</td>
</tr>
<tr>
<td>Return on Equity</td>
<td>roe</td>
<td>0.107</td>
<td>0.238</td>
<td>0.450</td>
</tr>
<tr>
<td>Return on Equity (annual)</td>
<td>roea</td>
<td>0.070</td>
<td>0.212</td>
<td>0.331</td>
</tr>
<tr>
<td>Return on Market Equity</td>
<td>rome</td>
<td>0.198</td>
<td>0.254</td>
<td>0.779</td>
</tr>
<tr>
<td>Seasonality</td>
<td>season</td>
<td>-0.065</td>
<td>0.201</td>
<td>-0.324</td>
</tr>
<tr>
<td>Sales Growth</td>
<td>sgrowth</td>
<td>0.034</td>
<td>0.185</td>
<td>0.184</td>
</tr>
<tr>
<td>Short Interest</td>
<td>short</td>
<td>0.024</td>
<td>0.127</td>
<td>0.187</td>
</tr>
<tr>
<td>Share Volume</td>
<td>shvol</td>
<td>-0.044</td>
<td>0.201</td>
<td>-0.217</td>
</tr>
<tr>
<td>Size</td>
<td>size</td>
<td>0.014</td>
<td>0.206</td>
<td>0.069</td>
</tr>
<tr>
<td>Sales-to-Price</td>
<td>sp</td>
<td>0.049</td>
<td>0.202</td>
<td>0.241</td>
</tr>
<tr>
<td>Short-term Reversal</td>
<td>strev</td>
<td>0.000</td>
<td>0.207</td>
<td>0.001</td>
</tr>
<tr>
<td>Standardized Unex-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pected Earnings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value-Momentum</td>
<td>valmom</td>
<td>0.021</td>
<td>0.144</td>
<td>0.147</td>
</tr>
<tr>
<td>Value-Momentum-</td>
<td>valmomgprof</td>
<td>0.086</td>
<td>0.140</td>
<td>0.619</td>
</tr>
<tr>
<td>Profitability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value (annual)</td>
<td>value</td>
<td>0.019</td>
<td>0.182</td>
<td>0.105</td>
</tr>
<tr>
<td>Value (monthly)</td>
<td>valuem</td>
<td>0.034</td>
<td>0.231</td>
<td>0.149</td>
</tr>
<tr>
<td>Value Profitability</td>
<td>vprof</td>
<td>0.057</td>
<td>0.117</td>
<td>0.484</td>
</tr>
</tbody>
</table>

Note: This table presents the average annualized returns, volatility, and Sharpe ratio for the factor universe. Factors are formed as decile long-short value-weighted portfolios based on a set of price and accounting variables. All variables are cross-sectionally orthogonalized to a security’s market beta to ensure the resulting factors are ex-ante market neutral. Details about the characteristics and their construction are given in Appendix A.1.4. The sample spans 1999Q1-2018Q4.

A.1.5 Data Construction

Data on institutional investors and holdings comes from FactSet Ownership (v5).

This section details the specific tables accessed via WRDS and the process to combine
these tables into the final datasets used for this paper.

**Institutional Holdings:** FactSet Ownership compiles holdings data from several sources, including (i) Form-13F filings, (ii) institutional stakes, gathered from annual reports (10K), beneficial ownership (13D and 13G) and insider filings (Forms 3, 4, and 5), and (iii) sum of fund-level reports. The Form-13F holdings at the filer level are in table `own_inst_13f_detail_eq` and aggregated into single institutions using the “rollup ID” in table `own_ent_13f_combined_inst`. Institutional stakes are in table `own_inst_stakes_detail_eq`. Fund-level reports are in `own_fund_detail_eq` and aggregated using the table `own_ent_funds`.

These separate tables are combined using logic based on whether the institution is a 13F filer (`fds_13f_flag = 1` in table `own_ent_institutions`) and whether the required to be disclosed in 13F (`fds_13f_flag = 1` in table `own_sec_coverage`). The logic, as recommended by the FactSet documentation, is:

- **13F filers and 13F securities:** Use latest 13F position unless there is a more recent stakes-based position.

- **Non-13F filers and/or non-13F securities:** Use stakes-based position if it is within 18 months; otherwise, use sum of funds.

Values of each position are computed using the security prices in `own_sec_prices_eq`. Securities are then mapped to the CRSP universe using the CUSIP linktable `sym_cusip`. Positions that exceed 40% of a stock’s market capitalization are dropped. If the sum of ownership exceeds the sum of a stock’s market capitalization and short interest, all positions are proportionally scaled down. The household sector is constructed as the difference between the sum of a stock’s market capitalization and short interest and the sum of institutional ownership.
Investor Type Classification: FactSet classifies institutional investors into various “entity sub types,” which are used to group investors into the six investor types considered. The entity sub types are in table own_inst_type, which are described in table entity_sub_type_map. The aggregation is the same as in Kojien et al. (2020) and is repeated below for convenience:

- **Broker:** BM (Bank Investment Division), BR (Broker), MM (Market Maker), ST (Stock Borrowing/Lending)
- **Hedge Fund:** AR (Arbitrage), FH (Fund of Hedge Funds Manager), FF (Fund of Funds Manager), FS (Fund Distributor), FU (Fund), HF (Hedge Fund Manager)
- **Investment Advisor:** IA (Investment Advisor), IC (Investment Company), PP (Real Estate Manager), RE (Research Firm), SB (Subsidiary Branch)
- **Long-Term:** FO (Foundation/Endowment Manager), IN (Insurance Company), PF (Pension Fund Manager), SV (Sovereign Wealth Manager)
- **Mutual Fund:** MF (Mutual Fund Manager)
- **Private Banking:** CP (Corporate), FY (Family Office), PB (Private Banking/Wealth Management), VC (Venture Capital/Private Equity)

Brokers include banks’ own portfolio of investments, as well as the holdings resulting from brokerage and market making services. Investment advisors differ from mutual funds in that most of their assets under management come from institutional accounts they manage, as opposed to the majority coming directly from shareholders into the mutual funds managed. Private banking includes the holdings of high net worth clients in private banking divisions, as well as privately managed family offices or corporate assets.
Table A.15 contains information on the final sample of institutional investors every 5 years, including the number and AUM of each investor type, as well as the largest investor. The number of mutual funds has been relatively stable over time, though their AUM has grown considerably. Hedge funds have also grown substantially in number and size. The largest single investors tend to be investment advisors and mutual funds.

A.2 Chapter 2 Appendix

A.2.1 Proofs

Proof of Proposition 11.

Proof. Let \( k \) (\( j \)) be the higher (lower) expected return asset. First, consider the case in which falling below the liquidation threshold is certain with either asset. Then, \( H_j = H_k = 0 \) and \( L_j = L_k = 1 \) with certainty, so utilities are given by

\[
\begin{align*}
  u_j^H(\tilde{W}) &= m + i\alpha_j - A\delta \\
  u_k^H(\tilde{W}) &= m + i\alpha_k - A\delta.
\end{align*}
\]  

(A.174)

Since \( k \) is the higher expected return asset, \( \alpha_k > \alpha_j \), so \( u_k^H(\tilde{W}) > u_j^H(\tilde{W}) \). Clearly, if \( \alpha_k = \alpha_j \), the expected utilities are equal.

In all other remaining situations, the fund cannot be liquidated with asset \( k \), but may with asset \( j \). Similarly, the fund may clear its HWM with asset \( k \), but cannot with asset \( j \). In these cases, the utility with asset \( k \) can be bounded below by

\[
  u_k^H = m + i\alpha_k,
\]  

(A.175)
Table A.15: Largest Investors

<table>
<thead>
<tr>
<th>Year</th>
<th>Investor Type</th>
<th>Number</th>
<th>AUM ($B)</th>
<th>Name</th>
<th>AUM ($B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aggregate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td>Inv. Advisor</td>
<td>1303</td>
<td>4347</td>
<td>BlackRock Fund Advisors</td>
<td>491</td>
</tr>
<tr>
<td></td>
<td>Mutual Fund</td>
<td>121</td>
<td>1771</td>
<td>AllianceBernstein LP</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td>Hedge Fund</td>
<td>470</td>
<td>245</td>
<td>Mackenzie Financial Corp.</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Long-Term</td>
<td>91</td>
<td>560</td>
<td>CalPERS</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Private Banking</td>
<td>346</td>
<td>377</td>
<td>United States Trust Co., NA</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Broker</td>
<td>19</td>
<td>82</td>
<td>Morgan Stanley &amp; Co. LLC</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Household</td>
<td>6329</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short-Seller</td>
<td>267</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>Inv. Advisor</td>
<td>1520</td>
<td>4284</td>
<td>BlackRock Fund Advisors</td>
<td>603</td>
</tr>
<tr>
<td></td>
<td>Mutual Fund</td>
<td>128</td>
<td>1615</td>
<td>Wellington Management Co. LLP</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>Hedge Fund</td>
<td>782</td>
<td>339</td>
<td>Renaissance Technologies LLC</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Long-Term</td>
<td>93</td>
<td>443</td>
<td>NYS Common Retirement Fund</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Private Banking</td>
<td>500</td>
<td>266</td>
<td>SunTrust Banks, Inc.</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Broker</td>
<td>37</td>
<td>135</td>
<td>Morgan Stanley &amp; Co. LLC</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Household</td>
<td>4138</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short-Seller</td>
<td>281</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>Inv. Advisor</td>
<td>1765</td>
<td>8710</td>
<td>BlackRock Fund Advisors</td>
<td>874</td>
</tr>
<tr>
<td></td>
<td>Mutual Fund</td>
<td>120</td>
<td>4298</td>
<td>T. Rowe Price Associates, Inc.</td>
<td>416</td>
</tr>
<tr>
<td></td>
<td>Hedge Fund</td>
<td>876</td>
<td>1138</td>
<td>Renaissance Technologies LLC</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Long-Term</td>
<td>107</td>
<td>918</td>
<td>CalPERS</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>Private Banking</td>
<td>724</td>
<td>674</td>
<td>UBS Financial Services, Inc.</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Broker</td>
<td>48</td>
<td>288</td>
<td>Credit Suisse Securities</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Household</td>
<td>7508</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short-Seller</td>
<td>674</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>Inv. Advisor</td>
<td>2233</td>
<td>10610</td>
<td>BlackRock Fund Advisors</td>
<td>1330</td>
</tr>
<tr>
<td></td>
<td>Mutual Fund</td>
<td>128</td>
<td>7468</td>
<td>T. Rowe Price Associates, Inc.</td>
<td>556</td>
</tr>
<tr>
<td></td>
<td>Hedge Fund</td>
<td>1013</td>
<td>1334</td>
<td>AQR Capital Management LLC</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>Long-Term</td>
<td>121</td>
<td>1068</td>
<td>CalPERS</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Private Banking</td>
<td>1043</td>
<td>936</td>
<td>UBS Financial Services, Inc.</td>
<td>67</td>
</tr>
<tr>
<td></td>
<td>Broker</td>
<td>57</td>
<td>357</td>
<td>MUFG Securities EMEA Plc</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>Household</td>
<td>6982</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Short-Seller</td>
<td>822</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: This table presents statistics on the aggregate size of each investor type, as well as the largest investor in each type every 5 years. The number of investors is determined as the number of unique investors who filed a form 13-F effective on December 31 of that year. The AUM of the aggregate investor type and the largest investor is expressed in billions of dollars.
because liquidation is impossible \((L_k = 0)\) but a performance fee is not \((P(H_k = 1) > 0)\). Similarly, utility with asset \(j\) can be bounded above by

\[
u_j^H = m + i\alpha_j, \tag{A.176}\]

because liquidation is possible \((P(L_j = 1) > 0)\), but clearing the HWM is not \((H_j = 0)\). As \(\alpha_k > \alpha_j\), it is clear that \(u_k^H > u_j^H\).

Important Quantities for Region \(\mathcal{L}\). The probabilities of falling below the liquidation threshold \(\gamma\) are needed to derive the utilities in Eq. 3.14. These are given by

\[
P(f_1 < \varrho_1) = \frac{\varrho_1 + \sigma}{2\sigma} \quad \text{and} \quad P(f_2 < \varrho_2) = \begin{cases} \frac{(\sigma + c_2)^2}{2\sigma^2} & \varrho_2 \leq 0 \\ \frac{2\sigma^2 - (\sigma - c_2)^2}{2\sigma^2} & \varrho_2 > 0 \end{cases}. \tag{A.177}\]

Proof of Lemma 4.

Proof. Define \(\varrho \equiv \varrho_1 = \varrho_2\) and \(\alpha \equiv \alpha_1 = \alpha_2\). First, I show that \(u_1^H \left(\frac{\gamma}{1+\alpha}\right) = u_2^H \left(\frac{\gamma}{1+\alpha}\right)\). Note that, for \(\bar{W} = \frac{\gamma}{1+\alpha}\), \(\varrho = 0\) Evaluating \(u_j^H \left(\frac{\gamma}{1+\alpha}\right)\) (from Eq. (3.14)) then gives

\[
u_1^H \left(\frac{\gamma}{1+\alpha}\right) = m + i\alpha - \frac{A\delta}{2} \quad \text{and} \quad u_2^H \left(\frac{\gamma}{1+\alpha}\right) = m + i\alpha - \frac{A\delta}{2}, \tag{A.178}\]

so \(u_1^H \left(\frac{\gamma}{1+\alpha}\right) = u_2^H \left(\frac{\gamma}{1+\alpha}\right)\).

Now, suppose \(\bar{W} \in \left(\frac{\gamma}{1+\alpha + \sigma}, \frac{\gamma}{1+\alpha}\right)\). Note that this implies that \(\varrho \in (0, \sigma)\). The only difference between the utilities \(u_1^H\) and \(u_2^H\) are the probabilities of default \(P(f_1 < \varrho)\) and \(P(f_2 < \varrho)\). The manager will prefer the one with the smaller probability of
default. As \( c > 0 \), these probabilities are given by

\[
\begin{align*}
P (f_1 \leq c) &= \frac{c + \sigma}{2\sigma} \\
P (f_2 \leq c) &= \frac{2\sigma^2 - (\sigma - c)^2}{2\sigma^2}.
\end{align*}
\]

As \( \sigma > c \),

\[
P (f_2 \leq c) = \frac{2\sigma^2 - (\sigma - c)^2}{2\sigma^2} = \frac{\sigma^2 + 2\sigma c - c^2}{2\sigma^2} = \frac{1}{2} + \frac{c}{\sigma} - \frac{c^2}{2\sigma^2} > \frac{1}{2} + \frac{c}{2\sigma} = \frac{c + \sigma}{2\sigma} = P (f_1 \leq c),
\]

so \( u_1^H(\tilde{W}) > u_2^H(\tilde{W}) \).

Finally, suppose \( \tilde{W} \in (\frac{\gamma}{1+\alpha}, \frac{\gamma}{1+\alpha - \sigma}) \). Note that this implies \( c \in (-\sigma, 0) \). As with the case before, the manager chooses the asset with the lower probability of default. As \( c < 0 \), these probabilities are given by

\[
\begin{align*}
P (f_1 < c) &= \frac{c + \sigma}{2\sigma} \\
P (f_2 < c) &= \frac{(\sigma + c)^2}{2\sigma^2}.
\end{align*}
\]

As \( \sigma > -c \),

\[
P (f_2 \leq c) = \frac{(\sigma + c)^2}{2\sigma^2} = \frac{\sigma^2 + 2\sigma c + c^2}{2\sigma^2} = \frac{1}{2} + \frac{c}{\sigma} + \frac{c^2}{2\sigma^2} < \frac{1}{2} + \frac{c}{2\sigma} = \frac{c + \sigma}{2\sigma} = P (f_1 \leq c),
\]

so \( u_1^H(\tilde{W}) < u_2^H(\tilde{W}) \).

\( \square \)

**Proof of Proposition 12.**

*Proof.* Suppose \( \alpha_1 > \alpha_2 \). As utility is strictly increasing in expected returns, it follows from Lemma 4 that \( u_1^H(\tilde{W}) > u_2^H(\tilde{W}) \) for \( \tilde{W} < \frac{\gamma}{1+\alpha_2} \). Note that we can replace \( \alpha \) in Lemma 4 with \( \alpha_2 \). For any \( \tilde{W} \in (\frac{\gamma}{1+\alpha_1}, \frac{\gamma}{1+\alpha_2}) \), the probability of falling below the liquidation threshold is greater than 50% with asset 2, but less than 50% for asset 1. Thus, not only is the probability of liquidation lower with asset 1, but
the manager also receives a greater expected return on his own investment, \( ir_j \).

Suppose, instead, \( \alpha_1 < \alpha_2 \), it follows from Lemma 4 that \( u_1^H(\tilde{W}) < u_2^H(\tilde{W}) \) for \( \tilde{W} > \frac{\gamma}{1+\alpha_2} \). Note that we can replace \( \alpha \) in Lemma 4 with \( \alpha_2 \) for the same reason as above. For any \( \tilde{W} \in \left( \frac{\gamma}{1+\alpha_2}, \frac{\gamma}{1+\alpha_1} \right) \), the probability of liquidation is higher under asset 1 than asset 2, so all managers in this range should prefer asset 2.

**Proof of (1).** Return to the case where \( \alpha_1 > \alpha_2 \). The above established that \( u_1^H(\tilde{W}) > u_2^H(\tilde{W}) \) for \( \tilde{W} < \frac{\gamma}{1+\alpha_2} \). It suffices to show that \( d(\tilde{W}) \equiv u_1^H(\tilde{W}) - u_2^H(\tilde{W}) \) achieves a minimum at \( W^*_L \equiv \frac{2\gamma}{2(1+\alpha_2)-\sigma} \) for \( \tilde{W} \geq \frac{\gamma}{1+\alpha_2} \). As the difference is assumed to be positive there, it must hold for the entire range.

First, since \( \tilde{W} \geq \frac{\gamma}{1+\alpha_2} \), it follows that \( -\sigma < c_1, c_2 \leq 0 \). This means the difference in utility is given by

\[
d(\tilde{W}) = i(\alpha_1 - \alpha_2) - A\delta \left( \frac{c_1 + \sigma}{2\sigma} - \frac{(\sigma + c_2)^2}{2\sigma^2} \right)
= i(\alpha_1 - \alpha_2) + \frac{A\delta}{2\sigma^2} \left( 2\sigma c_2 + c_2^2 - \sigma c_1 \right).
\]

The derivative of this with respect to \( \tilde{W} \) is

\[
\frac{\partial}{\partial W} d(\tilde{W}) = \frac{A\delta}{2\sigma^2} \left( 2\sigma \frac{\partial c_2}{\partial W} + 2c_2 \frac{\partial c_2}{\partial W} - \sigma \frac{\partial c_1}{\partial W} \right)
= \frac{A\delta}{2\sigma^2} \left( -2\sigma \frac{c_2}{W^2} - 2\gamma \tilde{W} - \frac{\gamma}{W} \tilde{W}^2 - \frac{\gamma}{W} \tilde{W} \left( 1 + \alpha_2 \right) + \frac{\gamma}{W^2} \right)
= \frac{A\delta\gamma}{2\sigma^2 W^2} \left( -\sigma - 2\gamma \frac{W}{W^2} + 2 \left( 1 + \alpha_2 \right) \right).
\]

Setting this derivative equal to zero, it is clear that \( \tilde{W} = \frac{2\gamma}{2(1+\alpha_2)-\sigma} \).

Now, to ensure this is a minimum, taking the second derivative with respect to
\[ W, \]

\[ \frac{\partial^2}{\partial W^2} d(W) = -\frac{2A\delta\gamma}{\sigma^2 W^3} \left( -\sigma - 2 \frac{\gamma}{W} + 2 (1 + \alpha_2) \right) + \frac{2\gamma A\delta\gamma}{W^2 \sigma^2 W^2} = \frac{A\delta\gamma}{\sigma^2 W^3} \left( \sigma - \frac{\gamma}{W} - 2 (1 + \alpha_2) \right). \]

Evaluating this at \( W = W^*_L \) gives

\[ \frac{\partial^2}{\partial W^2} d(W^*_L) = \frac{3A\delta\gamma}{2\sigma^2 W^3} (\sigma - 2 (1 + \alpha_2)) > 0, \]

where the equality follows from the fact that \( W \) must be positive, and so \( 2 (1 + \alpha_2) - \sigma > 0 \). Thus, \( u_1^H(\tilde{W}) - u_2^H(\tilde{W}) \) achieves a minimum at \( \tilde{W} = W^*_L \). By assumption, \( d(W^*_L) > 0 \). If the difference is positive at its minimum, it must be positive for all values in region \( \mathcal{L} \). Thus, if this is the case, all region \( \mathcal{L} \) funds will strictly prefer asset 1 to asset 2.

**Proof of (2).** From the proof of (1), it is established that the difference between the utilities achieves a minimum at \( \tilde{W} = W^*_L \) for \( \tilde{W} < \frac{\gamma}{1 + \alpha_2} \). As both functions are continuous and \( u_1^H(\tilde{W}) > u_2^H(\tilde{W}) \) for \( \tilde{W} < \frac{\gamma}{1 + \alpha_2} \), if \( u_1^H(W^*_L) < u_2^H(W^*_L) \), then there must exist a value \( W^* \in \left( \frac{\gamma}{1 + \alpha_2}, W^*_L \right) \) such that \( u_1^H(W^*) = u_2^H(W^*) \).

To complete the proof, it suffices to show that \( u_1^H \left( \frac{\gamma}{1 + \alpha_1 - \sigma} \right) > u_2^H \left( \frac{\gamma}{1 + \alpha_1 - \sigma} \right) \).

If this holds, by continuity, there must exist value \( W^* \in \left( W^*_L, \frac{\gamma}{1 + \alpha_1 - \sigma} \right) \) such that \( u_1^H(W^*) = u_2^H(W^*) \), where, between this threshold and \( W^*_L \), the utility of asset 2 is higher and above this threshold, the utility of asset 1 is higher.

To show \( u_1^H \left( \frac{\gamma}{1 + \alpha_1 - \sigma} \right) > u_2^H \left( \frac{\gamma}{1 + \alpha_1 - \sigma} \right) \), first note that \( c_1 = -\sigma \) (that is, funds with asset 1 are liquidated with probability 0), while \( c_2 = \alpha_1 - \alpha_2 - \sigma > -\sigma \) (that is, funds with asset 2 are liquidated with positive probability). Invoking Proposition 11, it then follows that \( u_1^H \left( \frac{\gamma}{1 + \alpha_1 - \sigma} \right) > u_2^H \left( \frac{\gamma}{1 + \alpha_1 - \sigma} \right) \).
Proof of (3). Now, suppose \( \alpha_1 < \alpha_2 \). It has been established that \( u_1^H(\tilde{W}) < u_2^H(\tilde{W}) \) for \( \tilde{W} > \frac{\gamma}{1+\alpha_2} \). It suffices to show that \( u_1^H(\tilde{W}) - u_2^H(\tilde{W}) \) achieves a maximum at \( W_{\ell}^{**} \equiv \frac{2\gamma}{2(1+\alpha_2)+\sigma} \) for \( \tilde{W} \leq \frac{\gamma}{1+\alpha_2} \). As the difference is negative at that point by assumption, it must be negative over the entire range.

Since \( \tilde{W} \leq \frac{\gamma}{1+\alpha_2} \), \( 0 < \varepsilon_1, \varepsilon_2 \leq \sigma \). Thus, the difference in utility is given by

\[
d(\tilde{W}) = i (\alpha_1 - \alpha_2) - A\delta \left( \frac{\varepsilon_1 + \sigma}{2\sigma} - \frac{2\sigma^2 - (\sigma - \varepsilon_2)^2}{2\sigma^2} \right)
= i (\alpha_1 - \alpha_2) - \frac{A\delta}{2\sigma^2} (\sigma \varepsilon_1 - 2\sigma \varepsilon_2 + \varepsilon_2^2).
\]

The derivative of this with respect to \( \tilde{W} \) is

\[
\frac{\partial}{\partial \tilde{W}} d(\tilde{W}) = \frac{A\delta}{2\sigma^2} \left( \sigma \frac{\partial \varepsilon_1}{\partial \tilde{W}} - 2\sigma \frac{\partial \varepsilon_2}{\partial \tilde{W}} + 2\varepsilon_2 \frac{\partial \varepsilon_2}{\partial \tilde{W}} \right)
= \frac{A\delta \gamma}{2\sigma^2 \tilde{W}^2} \left( -\sigma + 2\frac{\gamma}{\tilde{W}} - \tilde{W} (1 + \alpha_2) \right)
= \frac{A\delta \gamma}{2\sigma^2 \tilde{W}^2} \left( \frac{2\gamma}{\tilde{W}} - \sigma - 2(1 + \alpha_2) \right).
\]

Setting this to zero yields \( \tilde{W} = W_{\ell}^{**} \).

Now, to ensure this is a maximum, the second derivative with respect to \( \tilde{W} \) is given by

\[
\frac{\partial^2}{\partial \tilde{W}^2} d(\tilde{W}) = \frac{-A\delta \gamma}{\sigma^2 \tilde{W}^3} \left( \frac{2\gamma}{\tilde{W}} - \sigma - 2(1 + \alpha_2) \right) + \frac{A\delta \gamma}{2\sigma^2 \tilde{W}^2} \left( -\frac{2\gamma}{\tilde{W}^2} \right)
= \frac{A\delta \gamma}{\sigma^2 \tilde{W}^3} \left( \sigma + 2(1 + \alpha_2) - 3\frac{\gamma}{\tilde{W}} \right).
\]
Evaluating this at \( \hat{W} = W_{L}^{**} \),
\[
\frac{\partial^{2}}{\partial \hat{W}^{2}} d(W_{L}^{**}) = \frac{A \delta \gamma}{\sigma^{2} \hat{W}^{3}} \left( \sigma + 2 (1 + \alpha_2) - 3 \frac{2 (1 + \alpha_2) + \sigma}{2} \right) \\
= -\frac{A \delta \gamma}{\sigma^{2} \hat{W}^{3}} \left( 1 + \alpha_2 \right) \left( \frac{1}{2} \right) \leq 0.
\]

Since the second derivative is negative at this point, \( d(\hat{W}) \) achieves a maximum at \( \hat{W} = W_{L}^{**} \). If the difference is negative at its maximum, then it must be negative for all \( \hat{W} \leq \frac{\gamma}{1 + \alpha_2} \). Thus, all region \( L \) funds will strictly prefer asset 2 to asset 1.

Proof of (4). The proof of (3) established that \( d(\hat{W}) \) achieves a maximum at \( \hat{W} = W_{L}^{**} \). Since \( u_{j}^{H} \) is continuous and \( u_{1}^{H}(\hat{W}) < u_{2}^{H}(\hat{W}) \) for \( \hat{W} > \frac{\gamma}{1 + \alpha_2} \), there must exist a value \( \bar{W}_{L}^{C} \in \left( W_{L}^{**}, \frac{\gamma}{1 + \alpha_2} \right) \) such that \( u_{1}^{H}(\bar{W}_{L}^{C}) = u_{2}^{H}(\bar{W}_{L}^{C}) \).

To complete the proof, it suffices to show that \( u_{1}^{H} \left( \frac{\gamma}{1 + \alpha_2 + \sigma} \right) < u_{2}^{H} \left( \frac{\gamma}{1 + \alpha_2 + \sigma} \right) \). If this holds, there must exist a value \( W_{L}^{C} \in \left( \frac{\gamma}{1 + \alpha_2 + \sigma}, W_{L}^{**} \right) \) such that \( u_{1}^{H}(W_{L}^{C}) = u_{2}^{H}(W_{L}^{C}) \) by continuity. Below this threshold, asset 2 must be preferred and, above this threshold (and below \( \bar{W}_{L}^{C} \)), asset 1 must be preferred.

To show \( u_{1}^{H} \left( \frac{\gamma}{1 + \alpha_2 + \sigma} \right) < u_{2}^{H} \left( \frac{\gamma}{1 + \alpha_2 + \sigma} \right) \), first note that \( \varsigma_{1} = \varsigma_{2} = \sigma \). Therefore, falling below the threshold \( \gamma \) is guaranteed. By Proposition 11, it then follows that the manager will strictly prefer the higher return asset (i.e., asset 2). Thus,
\[
u_{1}^{H} \left( \frac{\gamma}{1 + \alpha_2 + \sigma} \right) < u_{2}^{H} \left( \frac{\gamma}{1 + \alpha_2 + \sigma} \right),
\]
completing the proof of (4).

Finally, consider \( i = 0 \). For \( \hat{W} = W_{L}^{*} \), \( c_{1}, c_{2} \) are given by
\[
\varsigma_{1} = \alpha_2 - \alpha_1 - \frac{\sigma}{2} \quad \text{and} \quad \varsigma_{2} = -\frac{\sigma}{2}.
\]
Evaluating $d (W^*_\mathcal{L})$ gives

\[
d (W^*_\mathcal{L}) = - A\delta \left( \frac{c_1 + \sigma}{2\sigma} \right) + A\delta \frac{(\sigma + c_2)^2}{2\sigma^2}
\]
\[
= \frac{A\delta}{2\sigma^2} \left( \left( \frac{\sigma}{2} \right)^2 - \sigma \left( \alpha_2 - \alpha_1 + \frac{\sigma}{2} \right) \right)
\]
\[
= \frac{A\delta}{2\sigma} \left( \alpha_1 - \alpha_2 - \frac{\sigma}{4} \right).
\]

As $A, \delta, \sigma > 0$, $d (W^*_\mathcal{L}) \geq 0$ iff $\alpha_1 - \alpha_2 \geq \frac{\sigma}{4}$.

Now, consider $\tilde{W} = W^{**}_\mathcal{L}$. Then, $c_1, c_2$ are given by

\[
c_1 = \alpha_2 - \alpha_1 + \frac{\sigma}{2}
\]
\[
c_2 = \frac{\sigma}{2}.
\]

Evaluating $d (W^{**}_\mathcal{L})$ gives

\[
d (W^{**}_\mathcal{L}) = - A\delta \left( \frac{c_1 + \sigma}{2\sigma} \right) + A\delta \frac{2\sigma^2 - (\sigma - c_2)^2}{2\sigma^2}
\]
\[
= \frac{A\delta}{2\sigma^2} \left( 2\sigma^2 - \left( \frac{\sigma}{2} \right)^2 - \sigma \left( \alpha_2 - \alpha_1 + \frac{3\sigma}{2} \right) \right)
\]
\[
= \frac{A\delta}{2\sigma} \left( \frac{\sigma}{4} - \alpha_2 + \alpha_1 \right).
\]

As $A, \delta, \sigma > 0$, $d (W^{**}_\mathcal{L}) \leq 0$ iff $\alpha_1 - \alpha_2 \leq -\frac{\sigma}{4}$, completing the proof. \qed

**Proof of Corollary 2.**

*Proof.* The proof is a straightforward extension of Lemma 4 and Proposition 12. \qed

**Important Quantities for Region $\mathcal{H}$.** In order to characterize the utility of funds in region $\mathcal{H}$ in Eq. (3.18), the probability of clearing the HWM, $P (f_j > \bar{c}_j)$ and the
conditional expectation of the risk component, \( \mathbb{E} [f_j \mid f_j > \bar{c}_j] \). These are given by

\[
P(f_1 > \bar{c}_1) = \frac{\sigma - \bar{c}_1}{2\sigma} \\
\mathbb{E} [f_1 \mid f_1 > \bar{c}_1] = \frac{\sigma + \bar{c}_1}{2}
\]

\[
P(f_2 > \bar{c}_2) = \begin{cases} 
\frac{2\sigma^2 - (\sigma + \bar{c}_2)^2}{2\sigma^2} & \bar{c}_2 \leq 0 \\
\frac{(\sigma - \bar{c}_2)^2}{2\sigma^2} & \bar{c}_2 > 0
\end{cases} \\
\mathbb{E} [f_2 \mid f_2 > \bar{c}_2] = \begin{cases} 
\frac{(\sigma - 2\bar{c}_2)(\sigma + \bar{c}_2)}{3(2\sigma^2 - (\sigma + \bar{c}_2)^2)} & \bar{c}_2 \leq 0 \\
\frac{\sigma + 2\bar{c}_2}{3} & \bar{c}_2 > 0
\end{cases}
\]

(A.187)

**Proof of Lemma 5.**

*Proof. Define \( \bar{c} \equiv \bar{c}_1 = \bar{c}_2 \) and \( \alpha \equiv \alpha_1 = \alpha_2 \). For \( \tilde{W} \in \mathcal{H} \) (note that this maps to \( \bar{c} \in (-\sigma, \sigma) \)), manager’s utility with the two assets are given by Eq. (3.18). As \( \alpha_1 = \alpha_2 \), utility is identical with the two assets except for the expected value of the performance fee (the last component of utility). Agents will simply choose the asset that has the higher expected performance fee. If \( \bar{c} \leq 0 \), the difference between the utilities, \( d(\tilde{W}) \equiv u_1^H(\tilde{W}) - u_2^H(\tilde{W}) \), is given by

\[
d(\tilde{W}) = p (1 - i) \frac{(\sigma - \bar{c})^2}{4\sigma} - p (1 - i) \frac{(\sigma - 2\bar{c})^2}{12\sigma^2}
\]

(A.188)

which is zero when \( \bar{c} = \frac{\sigma}{2}, -\sigma \). However, as this is only the difference for \( \bar{c} \leq 0 \), this is only zero for \( \bar{c} = -\sigma \). Clearly, as \( \bar{c} \leq 0, \sigma - 2\bar{c} > 0 \), so \( d(\tilde{W}) > 0 \) for \( \bar{c} \leq 0 \). Thus, asset 1 is strictly preferred.

Now, if \( \bar{c} > 0 \), the difference is given by

\[
d(\tilde{W}) = p (1 - i) \frac{(\sigma - \bar{c})^2}{4\sigma} - p (1 - i) \frac{(\sigma - \bar{c})^3}{6\sigma^2}
\]

(A.189)
which is strictly positive as \( \bar{c} \in (0, \sigma) \) Thus, \( u_1^H(\bar{W}) > u_2^H(\bar{W}) \), so asset 1 is strictly preferred.

The second part of the lemma follows from the fact that preferences are strictly increasing in \( \alpha_j \). \( \square \)

**Proof of Proposition 13.**

*Proof.* Proving this result relies on showing that \( d(\bar{W}) \) achieves a unique maximum at \( W^*_H = \frac{2(1+g)}{2(1+\alpha_2)+\sigma - \sqrt{\sigma} \sqrt{4(\alpha_1 - \alpha_2) + \sigma}} \) over the range \( \bar{W} \in \mathcal{H} \).

*Proof of (1).* First, consider \( \bar{c}_2 > 0 \) and take the derivative of \( d \) with respect to \( \bar{W} \):

\[
\frac{\partial}{\partial \bar{W}} d(\bar{W}) = \frac{p(1-i)}{2\sigma^2} (\sigma - \bar{c}_2)^2 \frac{\partial \bar{e}_2}{\partial \bar{W}} - \frac{p(1-i)}{2\sigma} (\sigma - \bar{c}_1) \frac{\partial \bar{e}_1}{\partial \bar{W}}
\]

\[
= \frac{p(1-i)}{2\sigma^2} \frac{1+g}{\bar{W}^2} (2\sigma \bar{c}_2 - \sigma \bar{c}_1 - \bar{c}_2^2)
\]

\[
= \frac{p(1-i)}{2\sigma^2} \frac{1+g}{\bar{W}^2} \left( (\sigma - \alpha_2) + \sigma \frac{1+g - \bar{W}(1+\alpha_2)}{W} \right) - \left( \frac{1+g - \bar{W}(1+\alpha_2)}{W} \right)^2
\]

\[
= \frac{p(1-i)}{2\sigma^2} \frac{1+g}{\bar{W}^2} \left( \sigma(\alpha_1 - 2\alpha_2 - 1) + \sigma \frac{1+g}{W} - \frac{(1+g)^2}{W^2} + \frac{(1+g)(1+\alpha_2) - (1+\alpha_2)^2}{W} \right)
\]

\[
= \frac{p(1-i)}{2\sigma^2} \left( 1+g \right) \left( \bar{W}^2 \left( \sigma(\alpha_1 - 2\alpha_2 - 1) - (1+\alpha_2)^2 \right) + \bar{W}(1+g)(2(1+\alpha_2) + \sigma) - (1+g)^2 \right).
\]  

(A.190)

Setting this to zero and applying the quadratic formula gives

\[
\bar{W} = \frac{-(1+g)(2(1+\alpha_2) + \sigma) \pm (1+g) \sqrt{(2(1+\alpha_2) + \sigma)^2 + 4 \left( (\sigma - \alpha_2 - 1) - (1+\alpha_2)^2 \right)}}{2 \left( (\sigma - \alpha_2 - 1) - (1+\alpha_2)^2 \right)}
\]

\[
= \frac{1+g}{2} \frac{2(1+\alpha_2) + \sigma \pm \sqrt{\sigma(4(\alpha_1 - \alpha_2) + \sigma)}}{(1+\alpha_2)^2 - \sigma(\alpha_1 - 2\alpha_2 - 1)}.
\]  

(A.191)
Multiplying and dividing by the conjugate $2 (1 + \alpha_2)+\sigma \mp \sqrt{\sigma (4 (\alpha_1 - \alpha_2) + \sigma)}$ yields

\[
\tilde{W} = \frac{1 + g}{2} \frac{4 ((1 + \alpha_2)^2 + \sigma (1 + 2\alpha_2 - \alpha_1))}{(1 + \alpha_2)^2 - \sigma (\alpha_1 - 2\alpha_2 - 1)} \frac{2 (1 + \alpha_2) + \sigma \pm \sqrt{\sigma (4 (\alpha_1 - \alpha_2) + \sigma)}}{2 (1 + g)}.
\]

(A.192)

Note that this only holds for $\tilde{W} \in \mathcal{H}$ and $\bar{\sigma}_2 > 0$. This means that $\tilde{W} \in \left(\frac{1+g}{1+\alpha_2+\sigma}, \frac{1+g}{1+\alpha_2}\right)$. It is simple to verify that, as $\alpha_1 < \alpha_2$, both critical values are in this range. Now, taking the second derivative of $d$,

\[
\frac{\partial^2}{\partial \tilde{W}^2} d(\tilde{W}) = \frac{p (1 - i) (1 + g) (4 (1 + g)^2 - 3 (1 + g) (2 + \alpha_2) + \sigma) \tilde{W} + 2 (1 + \alpha_2)^2 + \sigma (1 + 2\alpha_2 - \alpha_1) \tilde{W}^2}{2 \sigma^2 \tilde{W}^5}.
\]

(A.193)

Now, note that, as $\tilde{W}$ is positive, $\frac{\partial^2}{\partial \tilde{W}^2} d(\tilde{W})$ is decreasing in $\tilde{W}$ if

\[
2 \left( (1 + \alpha_2)^2 + \sigma (1 + 2\alpha_2 - \alpha_1) \right) \tilde{W} - 3 (1 + g) (2 + \alpha_2) + \sigma \leq 0.
\]

(A.194)

Evaluating this for $\tilde{W} = W^*_H$ gives

\[
2 \left( (1 + \alpha_2)^2 + \sigma (1 + 2\alpha_2 - \alpha_1) \right) \tilde{W} - 3 (1 + g) (2 + \alpha_2) + \sigma = \frac{4 ((1 + \alpha_2)^2 + \sigma (1 + 2\alpha_2 - \alpha_1))}{2 (1 + \alpha_2)^2 - \sigma \sqrt{4 (\alpha_1 - \alpha_2) + \sigma}} \frac{2 (1 + \alpha_2) + \sigma \pm \sqrt{\sigma (4 (\alpha_1 - \alpha_2) + \sigma)}}{2 (1 + g)} - 3 (1 + g) (2 + \alpha_2) + \sigma \leq 0 \quad \text{(A.195)}
\]

\[
(1 + g) \frac{4 ((1 + \alpha_2)^2 + \sigma (1 + 2\alpha_2 - \alpha_1))}{2 (1 + \alpha_2)} - 6 (1 + \alpha_2)^2 + \sigma = (1 + g) \frac{4 \sigma (1 + 2\alpha_2 - \alpha_1) - 8 (1 + \alpha_2)^2 - 6 \sigma (1 + \alpha_2)}{2 (1 + \alpha_2)}
\]

\[
(1 + g) \frac{-\sigma (1 + 2\alpha_1 - \alpha_2) - 4 (1 + \alpha_2)^2}{(1 + \alpha_2)} \leq 0.
\]
where the last inequality follows from the fact that $\alpha_2 \leq 1$ and $\alpha_1 \geq 0$.

As $\frac{\partial^2}{\partial \tilde{W}^2} d(\tilde{W})$ is decreasing in $\tilde{W}$, we can bound $\frac{\partial^2}{\partial \tilde{W}^2} d(\tilde{W})$ from above at $\tilde{W} = W^*_\tilde{n}$:

$$
\frac{\partial^2}{\partial \tilde{W}^2} d(\tilde{W}) = \frac{p(1 - i) (1 + g)^3 \left( 4 - 3 \left( 2 + \frac{\sigma}{(1 + \alpha_2)} \right) + 2 \left( 1 + \frac{(1 + 2\alpha_2 - \alpha_1)}{(1 + \alpha_2)^2} \right) \right)}{2\sigma^2 W^5}
$$

$$
= \frac{p(1 - i) (1 + g)^3}{2\sigma^2 W^5 \left( 1 + \alpha_2 \right)^2} \left( -3\sigma (1 + \alpha_2) + 2\sigma (1 + 2\alpha_2 - \alpha_1) \right)
$$

$$
= \frac{p(1 - i) (1 + g)^3}{2\sigma^2 W^5 \left( 1 + \alpha_2 \right)^2} \sigma (\alpha_2 - 1 - 2\alpha_1)
$$

$$
\leq 0.
$$

As the second derivative is negative, $d(\tilde{W})$ achieves a maximum at $\tilde{W} = W^*_\tilde{n}$. Similar calculations show that the other critical point, $\tilde{W} = \frac{2(1 + g)}{2(1 + \alpha_2) + \sigma + \sqrt{\sigma^2 + 4(\alpha_1 - \alpha_2) + \sigma^2}}$, is a minimum.

Finally, in order to verify that there are no other maxima, consider $\bar{c}_2 \leq 0$. Then, the derivative of $d$ with respect to $\tilde{W}$ is given by

$$
\frac{\partial}{\partial \tilde{W}} d(\tilde{W}) = \frac{p(1 - i) \frac{\partial \bar{c}_1}{\partial \tilde{W}} (\sigma \bar{c}_1 - \sigma^2) - p(1 - i) \frac{\partial \bar{c}_2}{\partial \tilde{W}} (-\sigma^2 + 2\sigma \bar{c}_2 + \bar{c}_2^2)}{2\sigma^2 \tilde{W}^2}
$$

$$
= \frac{p(1 - i) (1 + g)}{2\sigma^2 \tilde{W}^2} \left( -\sigma \bar{c}_1 + 2\sigma \bar{c}_2 + \bar{c}_2^2 \right)
$$

$$
= \frac{p(1 - i) (1 + g)}{2\sigma^2 \tilde{W}^2} \left( -\sigma \bar{W} (1 + \alpha_1) + 2\sigma \frac{1 + g - \bar{W} (1 + \alpha_2)}{W} + \frac{1 + g - \bar{W} (1 + \alpha_2)}{W} \right)
$$

$$
= \frac{p(1 - i) (1 + g)}{2\sigma^2 \tilde{W}^2} \left( (1 + \alpha_2)^2 - \sigma (1 + 2\alpha_2 - \alpha_1) + (1 + g) \frac{\sigma - 2(1 + \alpha_2)}{\tilde{W}^2} + \frac{(1 + g)^2}{\tilde{W}^2} \right)
$$

$$
= \frac{p(1 - i) (1 + g)}{2\sigma^2} \left( \tilde{W}^2 (1 + \alpha_2)^2 - \sigma (1 + 2\alpha_2 - \alpha_1) + \tilde{W} (1 + g) (\sigma - 2(1 + \alpha_2)) + (1 + g)^2 \right)
$$

(A.197)
Applying the quadratic formula yields
\[ \tilde{W} = (1 + g) \frac{-(\sigma - 2(1 + \alpha_2)) \pm \sqrt{(\sigma - 2(1 + \alpha_2))^2 - 4\left((1 + \alpha_2)^2 - \sigma(1 + 2\alpha_2 - \alpha_1)\right)}}{2 \left((1 + \alpha_2)^2 - \sigma(1 + 2\alpha_2 - \alpha_1)\right)} \]

\[ = \frac{(1 + g) - (\sigma - 2(1 + \alpha_2)) \pm \sqrt{\sigma^2 + 4\sigma(\alpha_2 - \alpha_1)}}{2} \left((1 + \alpha_2)^2 - \sigma(1 + 2\alpha_2 - \alpha_1)\right). \]  

(A.198)

Multiplying and dividing by the conjugate of the numerator \(- (\sigma - 2(1 + \alpha_2)) \mp \sqrt{\sigma^2 + 4\sigma(\alpha_2 - \alpha_1)}\) gives
\[ \tilde{W} = \frac{2(1 + g)}{2(1 + \alpha_2) - \sigma \pm \sqrt{\sigma} \sqrt{4(\alpha_2 - \alpha_1) + \sigma}}. \]  

(A.199)

However, neither of these critical values are in the region \(\tilde{W} \in \mathcal{H} \cap \tilde{c}_2 \leq 0\). This region spans \((\frac{1 + g}{1 + \alpha_2}, \frac{1 + g}{1 + \alpha_2 - \sigma})\). To see this, first consider \(\tilde{W} = \frac{2(1 + g)}{2(1 + \alpha_2) - \sigma \pm \sqrt{\sigma} \sqrt{4(\alpha_2 - \alpha_1) + \sigma}}:\)
\[ \tilde{W} = \frac{2(1 + g)}{2(1 + \alpha_2) - \sigma \pm \sqrt{\sigma} \sqrt{4(\alpha_2 - \alpha_1) + \sigma}} \leq \frac{2(1 + g)}{2(1 + \alpha_2) - \sigma + \sqrt{\sigma} \sqrt{\sigma}} \]
\[ = \frac{1 + g}{1 + \alpha_2}. \]  

(A.200)

Similarly, for \(\tilde{W} = \frac{2(1 + g)}{2(1 + \alpha_2) - \sigma - \sqrt{\sigma} \sqrt{4(\alpha_2 - \alpha_1) + \sigma}}:\)
\[ \tilde{W} = \frac{2(1 + g)}{2(1 + \alpha_2) - \sigma - \sqrt{\sigma} \sqrt{4(\alpha_2 - \alpha_1) + \sigma}} \geq \frac{2(1 + g)}{2(1 + \alpha_2) - \sigma - \sqrt{\sigma} \sqrt{\sigma}} \]
\[ = \frac{1 + g}{1 + \alpha_2 - \sigma}. \]  

(A.201)

so neither of these correspond to a maximum.

Thus, \(\tilde{W} = W^*_\mathcal{H}\) is the unique maximum. If this value is negative, then \(u^H_1(\tilde{W}) \leq \)
for all $\tilde{W} \in \mathcal{H}$, so asset 2 preferred.

Proof of (2). The proof of (1) established that $d(\tilde{W})$ achieves a unique maximum at $\tilde{W} = W_{\mathcal{H}}^*$. In order to complete the proof of (2), it suffices to show that $d\left(\frac{1+g}{1+\alpha_1+\sigma}\right)$, $d\left(\frac{1+g}{1+\alpha_2-\sigma}\right) < 0$. As $d(W_{\mathcal{H}}^*) > 0$ by assumption, there must exist some values $W^\mathcal{H} \in \left(\frac{1+g}{1+\alpha_1+\sigma}, W_{\mathcal{H}}^*\right)$ and $\tilde{W}^\mathcal{H} \in \left(W_{\mathcal{H}}^*, \frac{1+g}{1+\alpha_2-\sigma}\right)$ such that $d(W^\mathcal{H}) = d(\tilde{W}^\mathcal{H}) = 0$. Continuity of $d(\cdot)$ guarantees the remainder of the statement.

Note that $\tilde{W} = \frac{1+g}{1+\alpha_1+\sigma} \implies \bar{c}_1 = \sigma$, $\bar{c}_2 = \sigma + \alpha_1 - \alpha_2$. Evaluating $d\left(\frac{1+g}{1+\alpha_1+\sigma}\right)$ yields

$$d\left(\frac{1+g}{1+\alpha_1+\sigma}\right) = m + i\alpha_1 - \left(m + i\alpha_2 + p(1-i) \frac{(\alpha_2 - \alpha_1)^3}{6\sigma^2}\right)$$

$$= i(\alpha_1 - \alpha_2) - p(1-i) \frac{(\alpha_2 - \alpha_1)^3}{6\sigma^2} \quad \text{(A.202)}$$

$$< 0,$$

as $\alpha_1 < \alpha_2$.

Next, consider $\tilde{W} = \frac{1+g}{1+\alpha_2-\sigma}$. This implies $\bar{c}_1 = -\sigma + \alpha_2 - \alpha_1$, $\bar{c}_2 = -\sigma$. Evaluating $d\left(\frac{1+g}{1+\alpha_2-\sigma}\right)$ yields

$$d\left(\frac{1+g}{1+\alpha_2-\sigma}\right) = m + i\alpha_1 + p(1-i) \frac{2\sigma + \alpha_1 - \alpha_2}{4\sigma} - (m + i\alpha_2 + p(1-i)\sigma)$$

$$< m + i\alpha_1 + p(1-i)\sigma - (m + i\alpha_2 + p(1-i)\sigma)$$

$$= i(\alpha_1 - \alpha_2)$$

$$\leq 0,$$

(A.203)

completing the proof of (2).
Finally, suppose $i = 0$ and consider $\tilde{W} = W^*_H$. This implies
\[
\bar{c}_1 = \frac{2(\alpha_2 - \alpha_1) + \sigma - \sqrt{\sigma} \sqrt{4(\alpha_1 - \alpha_2) + \sigma}}{2},
\]
\[
\bar{c}_2 = \frac{\sigma - \sqrt{\sigma} \sqrt{4(\alpha_1 - \alpha_2) + \sigma}}{2}.
\]
(A.204)

Now, evaluate $d(W^*_H)$,
\[
d(W^*_H) = p\left(\frac{2(\alpha_1 - \alpha_2) + \sigma + \sqrt{\sigma} \sqrt{4(\alpha_1 - \alpha_2) + \sigma}}{16\sigma}\right)^2 - p\left(\frac{\sqrt{\sigma} + \sqrt{4(\alpha_1 - \alpha_2) + \sigma}}{48\sqrt{\sigma}}\right)^3
\]
\[
= \frac{3}{48\sigma}\left(2(\alpha_1 - \alpha_2) + \sigma + \sqrt{\sigma} \sqrt{4(\alpha_1 - \alpha_2) + \sigma}\right)^2 - \frac{\sqrt{\sigma} + \sqrt{4(\alpha_1 - \alpha_2) + \sigma}}{48\sqrt{\sigma}}^3.
\]
(A.205)

Let $\Delta \alpha \equiv \alpha_1 - \alpha_2$. Setting this to zero and evaluating,
\[
3\left(2\Delta \alpha + \sigma + \sqrt{\sigma} \sqrt{4\Delta \alpha + \sigma}\right)^2 = \sqrt{\sigma} \left(\sqrt{\sigma} + \sqrt{4\Delta \alpha + \sigma}\right)^3
\]
\[
3\left(2\Delta \alpha + \sigma + \sqrt{\sigma} \sqrt{4\Delta \alpha + \sigma}\right)^2 = 2\sqrt{\sigma} \left(\sqrt{\sigma} + \sqrt{4\Delta \alpha + \sigma}\right) \left(2\Delta \alpha + \sigma + \sqrt{\sigma} \sqrt{4\Delta \alpha + \sigma}\right)
\]
\[
6\Delta \alpha + 3\sigma + 3\sqrt{\sigma} \sqrt{4\Delta \alpha + \sigma} = 2\sigma + 2\sqrt{\sigma} \sqrt{4\Delta \alpha + \sigma}
\]
\[
6\Delta \alpha + \sigma + \sqrt{\sigma} \sqrt{4\Delta \alpha + \sigma} = 0
\]
\[
6\Delta \alpha + \sigma = -\sqrt{\sigma} \sqrt{4\Delta \alpha + \sigma}
\]
\[
36\Delta \alpha^2 + 12\sigma \Delta \alpha + \sigma^2 = 4\sigma \Delta \alpha + \sigma^2
\]
\[
\Delta \alpha = -\frac{2\sigma}{9}.
\]
(A.206)

As $d(\cdot)$ is strictly increasing in $\alpha_1$, it then follows that $d(W^*_H) \leq 0$ iff $\alpha_1 - \alpha_2 \leq -\frac{2\sigma}{9}$. \qed

Proof of Corollary 3.

Proof. The proof is a straightforward extension of Lemma 5 and Proposition 13. \qed
Proof of Theorem 1.

Proof. Recall that the demand for asset 1 from hedge funds, \( \eta^H \), is bounded by

\[
F (\Omega^S) \leq \eta^H \leq F (\Omega^S) + F (\Omega^I). \tag{A.207}
\]

The fraction of mutual funds that prefer asset 1 when \( \alpha_1 = \alpha_2 \) is \( \eta^M = \frac{1}{3} \). Thus, the market clearing condition is

\[
\theta \eta^H + (2 - \theta) \frac{1}{3} = 1. \tag{A.208}
\]

This equation is satisfied iff

\[
\eta^H = \frac{1 + \theta}{3\theta}. \tag{A.209}
\]

Plugging this into the inequality above, such an equilibrium exists iff

\[
\frac{1 + \theta}{3\theta} - F (\Omega^I) \leq F (\Omega^S) \leq \frac{1 + \theta}{3\theta} \tag{A.210}
\]

When \( \alpha = \alpha_1 = \alpha_2 \), from Lemma 11, all funds in region \( R \) are indifferent between assets (i.e., \{ \tilde{W} \in R \} \subseteq \Omega^I \). Further, from Propositions 12 and 13, funds in regions \( L \) and \( H \) are only indifferent at specific points (i.e., \( \frac{\gamma}{1 + \alpha} \)). By Assumption 3, such points have no mass. As a result, \( F (\Omega^I) = F (R) \).

Similarly, Corollaries 2 and 3 establish the ranges of \( \tilde{W} \) for which funds in regions \( L \) and \( H \) strictly prefer asset 1 when \( \alpha = \alpha_1 = \alpha_2 \). These are precisely \( L_{low} \) and \( H \), respectively. Thus, \( \Omega^S = L_{low} \cup H \). Substituting these into Eq. (A.210) gives the desired result:

\[
\frac{1 + \theta}{3\theta} - F (R) \leq F (L_{low} \cup H) \leq \frac{1 + \theta}{3\theta}. \tag{A.211}
\]

\[\square\]

228
Proof of Theorem 2.

Proof. Building on the proof of Theorem 1, for some $\alpha = \alpha_1 = \alpha_2$, $F(\Omega^S) = F(L_{\text{low}} \cup H)$ and $F(\Omega^I) = F(\mathcal{R})$.

If $F(\Omega^S) = F(L_{\text{low}} \cup H) > \frac{1+\theta}{3\theta}$, then $\eta^H > \frac{1+\theta}{3\theta}$. This implies that the total demand for asset 1 by hedge funds and mutual funds is

$$\theta \eta^H + (2 - \theta) \frac{1}{3} > 1,$$

(A.212)

violating the market clearing condition. Since the utility of both hedge funds and mutual funds is strictly increasing in $\alpha_1$ and $\alpha_2$, it is clear that $F(L_{\text{low}} \cup H)$ is decreasing in $\alpha_2$ and increasing in $\alpha_1$. Thus, there cannot exist an equilibrium in which $\alpha_1 > \alpha_2$ (as this would raise the total demand of asset 1). If an equilibrium exists, therefore, it must be that $\alpha_1 < \alpha_2$.

Similarly, if $F(\Omega^S) + F(\Omega^I) = F(L_{\text{low}} \cup H) + F(\mathcal{R}) < \frac{1+\theta}{3\theta}$, then $\eta^H < \frac{1+\theta}{3\theta}$. The total demand for asset 1 by hedge funds and mutual funds is less than 1, so markets do not clear. In order for markets to clear, the expected return of asset 1 must exceed that of asset 2 by strict monotonicity of preferences. Hence, $\alpha_1 > \alpha_2$. \hfill \Box

### A.2.2 Calibration

For the calibration, a piecewise distribution is assumed for $\tilde{W}$.

**Assumption 5.** The distribution of fund value $f(\cdot)$ in regions $\mathcal{L}$ and $\mathcal{H}$ is uniform.
That is, \( f(\cdot) \) is given by

\[
f(x) = \begin{cases} 
\kappa^L & x \in \mathcal{L} \\
\kappa^H & x \in \mathcal{H} \\
\tilde{f}(x) & x \in \mathcal{R} \\
0 & \text{else}
\end{cases}
\]

for any distribution function \( \tilde{f}(x) \) and with the restriction that \( f(x) \) integrates to 1:

\[
\int_0^1 f(x) \, dx = \kappa^H \left( \min \left\{ 1, \frac{1 + g}{1 + \bar{\alpha} - \sigma} \right\} - \frac{1 + g}{1 + \bar{\alpha} + \sigma} \right) \\
+ \kappa^L \left( \frac{2\gamma \sigma}{(1 + \bar{\alpha} - \sigma)(1 + \bar{\alpha} + \sigma)} \right) \\
+ \int_{\mathcal{R}} \tilde{f}(x) \, dx \\
= 1.
\]

Assumption 5 provides a simple framework for analyzing the equilibrium properties of the model. All results do not depend on this uniformity assumption, but this significantly simplifies the comparative static calculations.
A.2.3 CISDM-Thomson Reuters Matching Procedure

Aggregating the CISDM data to the management company level and merging this to 13-F data requires (i) identifying common management companies within the CISDM data, (ii) identifying common management companies within the Thomson Reuters data, and (iii) linking these based on company name.

In order to aggregate the CISDM data to the management company level, I rely on four key variables: firm name, firm assets, date of firm assets, and a contact e-mail address. First, I assume that identical firm names correspond to the same management company, and assign each firm name a unique ID. For firms that do not match based on the firm name but do share the same domain (everything following the @ symbol in the email address), I classify these firms as identical or not based on internet searches and whether their listed assets are the same. If they match, I assign the ID to be the maximum of the two firms’ unique IDs. Finally, for firms that do not share a common domain, I compare each firm name with (i) the four closest firm...
names alphabetically and (ii) the five closest firm names based on the Levenshtein string distance metric. This procedure identifies 1,945 management companies in the sample.\textsuperscript{9} Importantly, the company ID assigned here is permanent and unique.

While Thomson Reuters provides an ID for each reporting institution, these IDs are neither permanent nor unique; the same reporting institution can have multiple IDs, and the same ID can have multiple reporting institutions. The former issue is not of particular concern, as the procedure described below allows for the mapping of multiple Thomson Reuters IDs to a single CISDM ID. The latter, however, is particularly concerning, as it may result in CISDM firm being matched to multiple management companies. The Thomson Reuters documentation states that ID reassignments are typically made after a reporting gap of one year. I manually check every observation in which (i) the listed management company has a name change and (ii) there is a reporting gap of greater than one quarter (95 days) to determine if the ID has been reassigned. Of the 891 instances, 505 are reassignments. I assign a new ID (henceforth TR ID) to these management companies based on both the provided ID and the reassignment dummy.

To complete the matching procedure, I first standardize the names in both databases by cleaning up suffixes, typos, etc.. Then, I merge based on the standardized company name. This links 397 management companies in CISDM to 418 TR IDs. For the unmatched names, I manually review each of the CISDM management company names, and search for an identifying string in the Thomson Reuters names (e.g., “1741” in “1741 Asset Management”). This matches an additional 91 CISDM companies to 100 TR IDs. Overall, there are 485 CISDM management companies that map to 518 TR IDs.

\textsuperscript{9}There are 1,984 unique company names in the data, but I identify 39 instances of duplicates.
Bibliography


Da, Zhi, Qiang Liu, and Ernst Schaumburg, “A Closer Look at the Short-Term


236


238


