Spring 2014

Essays on mutual fund performance, ambiguity aversion, and high frequency trading

Lin Tong
University of Iowa

ESSAYS ON MUTUAL FUND PERFORMANCE, AMBIGUITY AVERSION, AND HIGH FREQUENCY TRADING

by

Lin Tong

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Business Administration
in the Graduate College of
The University of Iowa

May 2014

Thesis Supervisor: Associate Professor Ashish Tiwari

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Lin Tong

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Business Administration at the May 2014 graduation.

Thesis committee: ______________________________
Ashish Tiwari, Thesis Supervisor

____
Wei Li

____
Erik Lie

____
Yiming Qian

____
Tong Yao

ACKNOWLEDGEMENTS

I would like to thank my advisor, Ashish Tiwari, for his guidance and encouragement. I have learned a great deal from him during my time at Iowa and could never repay his generosity. In addition, I would like to thank the other members of my thesis committee: Wei Li, Erik Lie, Yiming Qian, and Tong Yao. They have all provided invaluable support throughout my graduate studies.

The faculty and graduate students at the University of Iowa have also been responsible for a large part of my academic development, and I am grateful to everyone.

Finally, this dissertation would not have been possible without the love and support of my parents, Yanyan and Xueliang.

ABSTRACT

In this dissertation, I address a range of topics in the context of mutual fund performance and high frequency trading.

The first chapter provides novel evidence on the role of ambiguity aversion in determining the response of mutual fund investors to historical fund performance information. It presents a model of ambiguity averse investors who receive multiple performance-based signals of uncertain precision about manager skill. A key implication of the model is that when investors receive multiple signals of uncertain quality, they place a greater weight on the worst signal. There is strong empirical support for this prediction in the data. Fund flows display significantly higher sensitivity to the worst performance measure even after controlling for fund performance at multiple horizons, performance volatility, flow-performance convexity, and a host of other relevant explanatory variables. This effect is particularly pronounced in the case of retail funds in contrast to institutional funds. The results suggest that fund investor behavior is best characterized as reflecting both Bayesian learning and ambiguity aversion.

The second chapter combines data on high frequency trading (HFT) activities of a randomly selected sample of 120 stocks and data on institutional trades, I find that HFT increases the trading costs of traditional institutional investors. An increase of one standard deviation in the intensity of HFT activities increases institutional execution shortfall costs by a third. Further analysis suggests that HFT

represents an opportunistic and extra-expensive source of liquidity when demand and supply among institutional investors are imbalanced. Moreover, the impact on institutional trading costs is most pronounced when high frequency (HF) traders engage in directional strategies (e.g., momentum ignition and order anticipation). I perform various analyses to rule out an alternative explanation that HF traders are attracted to stocks that have high trading costs. First, HFT is most prevalent in liquid stocks. Second, the results are robust to controls for stable stock liquidity characteristics and events that might jointly affect HFT and trading costs. Third, an analysis of the HFT behavior around the temporary short selling ban in September 2008 highlights the opportunistic nature of liquidity provision by HF traders. Finally, Granger causality tests show that intensive HFT activity significantly contributes to institutional trading costs, but not vice versa.

The third chapter analyzes the implications of the tournament-like competition in the mutual fund industry using a framework that addresses the risk-taking incentives facing fund managers. The theoretical model presented in this chapter suggests that the increase in the *activeness* of the interim loser manager's portfolio is directly related to the magnitude of the performance gap at the interim stage, and to the strength of the investor (cash flow) response to the relative performance rankings of the funds (i.e., the strength of the tournament effect). The empirical evidence based on quarterly Active Share data for a sample of domestic stock funds, is consistent with the key predictions of the model.

TABLE OF CONTENTS

LIST OF TABLES . . . vii

LIST OF FIGURES . . . ix

CHAPTER

1 INVESTMENT DECISIONS UNDER AMBIGUITY: EVIDENCE FROM MUTUAL FUND INVESTOR BEHAVIOR . . . 1

1.1 Introduction . . . 1
1.2 Ambiguity Aversion . . . 7
1.3 The Model and Its Empirical Implications . . . 9
1.3.1 The Model . . . 9
1.3.2 Bayesian Benchmark . . . 13
1.3.2.1 Bayesian Benchmark with No Signal Precision Uncertainty . . . 14
1.3.2.2 Bayesian Benchmark with Signal Precision Uncertainty . . . 14
1.3.3 Distinction with Other Behavioral Biases . . . 16
1.3.4 Empirical Predictions . . . 18
1.4 Data and Methodology . . . 23
1.4.1 Data . . . 23
1.4.2 Empirical Methodology . . . 25
1.5 Empirical Results . . . 28
1.5.1 Ambiguity Aversion . . . 28
1.5.2 Strategy Changes as a Proxy for Fund Ambiguity . . . 30
1.5.3 Flow Volatility as a Proxy for Fund Ambiguity . . . 33
1.5.4 Family Size as a Proxy for Fund Ambiguity . . . 35
1.5.5 The Role of Advertising . . . 37
1.6 Contrast between the Response of Ambiguity Averse Investors and Bayesian Investors . . . 40
1.7 Robustness . . . 45
1.8 Concluding Remarks . . . 49

2 A BLESSING OR A CURSE? THE IMPACT OF HIGH FREQUENCY TRADING ON INSTITUTIONAL INVESTORS . . . 64

2.1 Introduction . . . 64
2.2 Related Literature . . . 71
2.3 Data and Descriptive Statistics . . . 73

2.3.1 Measuring HFT 73
2.3.2 Measuring institutional trading cost 75
2.3.3 Sample descriptive statistics 77
2.3.4 Determinants of HFT 78
2.4 Impact of HFT on Institutional Trading Costs 79
2.4.1 HFT, liquidity, and trading costs: sorted portfolios 79
2.4.2 Multivariate analysis 81
2.4.3 Impact of HFT across firm size 83
2.4.4 Direction of causality 84
2.5 Further Analysis of HFT activities 91
2.5.1 Robustness: Timing delay costs and trade-level regressions 91
2.5.2 When and how does HFT impact institutional trading costs 93
2.6 Conclusions 99

3 MUTUAL FUND TOURNAMENTS AND FUND ACTIVE SHARE . 118

3.1 Introduction 118
3.2 Active Share 123
3.3 The Model 125
3.3.1 Model Structure 125
3.3.2 Analysis of the Model 127
3.4 Empirical Evidence 134
3.4.1 Data and Sample Description 134
3.4.2 Evidence of Performance Chasing Behavior by Investors . 136
3.4.3 Do Funds Alter their Active Share in Response to Above Incentives? 138
3.4.4 Evidence Using Other Activeness Measures 140
3.5 Discussion and Conclusion 142

APPENDIX. TECHNICAL DETAILS 149

REFERENCES 165

LIST OF TABLES

Table

1.1 Summary Statistics of the Equity Mutual Fund Sample 51

1.2 Ambiguity Aversion of Retail and Institutional Investors 52

1.3 Strategy Changes as a Proxy for Fund Ambiguity Level 53

1.4 Flow Volatility as a Proxy for Fund Ambiguity Level 55

1.5 Fund Family Size as a Proxy for Fund Ambiguity Level 56

1.6 Impact of Fund Advertising on Funds' Ambiguity 57

1.7 Contrast between the Response of Ambiguity Averse Investors and Bayesian Investors 59

1.8 Robustness Test–Morningstar and Peformance Since Inception 61

1.9 Robustness Test–Flow-Performance Convex Relationship 62

2.1 Summary statistics 102

2.2 Determinants of HFT 103

2.3 HFT's impact on Execution Shortfall 104

2.4 HFT's impact on execution shortfall across stock size 105

2.5 HFT's impact on execution shortfall on event days and no-event days . . 106

2.6 Granger causality 107

2.7 HFT's impact on timing delay costs 108

2.8 Trade-level analysis of HFT's impact on execution shortfall 109

2.9 HFT and institutional buy-sell imbalances 110

2.10 HFT's impact on execution shortfall when institutional trading is imbalanced 111

2.11 Impact of HFT strategies on execution shortfall 112

3.1 Summary Statistics of the Equity Mutual Fund Sample 143

3.2 Determinants of Quarterly Fund Cash Flows 144

3.3 Changes in Fund Active Share and Past Performance 145

3.4 Changes in Fund Tracking Error and Past Performance 146

3.5 Changes in Fund Self-reported Active Share and Past Performance . . . 147

LIST OF FIGURES

Figure

1.1 Difference in Flow-Performance Sensitivity to Signal $\mathbf{s_1}$ Relative to $\mathbf{s_2}$ across Difference Signal Realizations 63

2.1 Relation between HFT intensity and liquidity. 113

2.2 Relation between liquidity and execution shortfall 114

2.3 Relation between HFT intensity and execution shortfall 115

2.4 Execution shortfall around the Short-selling Ban of September 18, 2008 . 116

2.5 HFT activity around the Short-selling Ban of the September 18, 2008 . . 117

3.1 One Period Model 148

CHAPTER 1
INVESTMENT DECISIONS UNDER AMBIGUITY: EVIDENCE FROM MUTUAL FUND INVESTOR BEHAVIOR

1.1 Introduction

The technological advances of recent decades and the resulting reduction in the cost of information have made *information overload*, a term popularized by futurist Alvin Toffler, a reality. Investors today operate in a world with increasing complexity requiring them to process large amounts of information while making decisions. However, information quality can often be difficult to judge for investors. As argued by Epstein and Schneider (2008), when faced by information signals of unknown quality, investors treat the signals as being ambiguous. In this situation, investors do not update their beliefs in Bayesian fashion. Instead, they act as if they entertain multiple probability distributions when processing the signals. There is now considerable experimental evidence documenting that investors are ambiguity averse (e.g., Bossaerts et al. (2010), Ahn et al. (2011)). Understanding how information ambiguity impacts investor choices is clearly important. In this study we provide novel evidence on this issue by examining the response of mutual fund investors to historical fund performance information.

Mutual funds offer an appealing setting in which to study the role of ambiguity on investor decisions for a number of reasons. First, mutual funds represent a very substantial component of U.S. household portfolios. Second, the well-documented phenomenon of performance-chasing by fund investors suggests a natural link between performance-related information and the investment/divestment decisions of investors. Third, funds typically make available performance statistics including relative rankings measured at various horizons (e.g., 1-year, 3-year, 5-year, and since-inception) which serve as multiple signals to investors about fund manager skill. These performance statistics are readily available in fund prospectuses and related marketing materials. Each of the multiple signals reflects the performance of a particular fund relative to the available pool of funds, albeit over different time horizons. In this sense, the signals are comparable and studying the response of investors to various signals allows for a natural test of decision making under ambiguity in a non-experimental setting. Fourth, Condie (2008) shows that it is difficult to use asset price data to assess the significance of ambiguity aversion since the impact of ambiguity averse investors on prices is empirically indistinguishable from that of expected utility maximizers with potentially biased beliefs. The mutual fund setting by contrast, allows us to directly assess the impact of ambiguity aversion on investor decisions.

Our study extends the extant literature by examining the impact of ambiguity about manager skill on investor decisions. We adopt the standard distinction made in the literature between risk and uncertainty following Knight (1921). Knight considered risky events as those that could be described by known probability distributions versus uncertain events for which the probability distributions were not known. As famously demonstrated by Ellsberg (1961), individuals are averse to the ambiguity that characterizes decisions under conditions of uncertainty. It is reasonable to believe that individual investors are faced with considerable ambiguity when it comes to their fund investment decision. Investors face a dizzying array of choices. For example, according to the 2012 Investment Company Fact Book, there were more than 4,500 U.S. equity mutual funds in existence at the end of 2011. Investors are also subjected to a barrage of performance statistics on the funds. While these performance data provide signals of the fund manager skill, the investors clearly face a great deal of uncertainty about the quality of the signals. Past performance is at best a noisy signal of managerial skill. How do ambiguity averse investors interpret and respond to such signals? The goal of this study is to provide some answers to this question as a way to further our understanding of mutual fund investor behavior.

We present a simple model of ambiguity averse investors who receive multiple performance-based signals of uncertain precision about manager skill or fund alpha. The model relies on the framework of Epstein and Schneider (2008),Klibanoff et al. (2005), and Ju and Miao (2012). Investors in the model are risk-neutral yet averse to the ambiguity regarding manager skill or alpha. Given the uncertainty about the quality of multiple signals, investors do not update their beliefs in standard Bayesian fashion but rather they behave as if they have multiple conditional distributions in mind for the future fund performance. Intuitively, ambiguity averse investors prefer to make a fund choice that is more robust across the multiple distributions. A key implication of the model is that when investors receive multiple signals of uncertain quality, they place a greater weight on the worst signal. In practical terms this implies that ambiguity averse investors are more sensitive to the worst-case scenario when evaluating funds. We find strong empirical support for this prediction in the data. Specifically, we examine the sensitivity of fund flows to past performance measured over multiple time horizons: 1 year, 3 years, and 5 years. We find that fund flows display significantly higher sensitivity to the worst performance measure even after controlling for performance volatility and a host of other relevant explanatory variables. This heightened sensitivity holds regardless of whether fund performance is measured by raw return or the Carhart 4-factor alpha. This effect is particularly pronounced in the case of retail funds whose investors are likely to face a higher degree of uncertainty regarding the quality of performance-related signals they observe, compared to the institutional fund investors.[1]

We use a number of fund characteristics as proxies for the degree of ambiguity about fund performance/manager skill. These include the fund's investment strategy shifts, return volatility, fund flow volatility, family size, and marketing effort/expenditure. We consistently find that in cases with higher degree of ambiguity as captured by our proxy measures, fund flows display significantly higher sensitivity to the worst performance measure. Our results are robust to the use of additional controls including the convexity in the flow-performance relation, Morningstar fund ratings, and fund performance since inception.

We next examine the implications of the potential differences in the ambiguity aversion of retail and institutional investors. The latter are typically viewed as being

[1]Instead of using performance measured at different time horizons as signals, in unreported tests we also consider various performance measures (including the CAPM alpha, Fama-French 3-factor alpha, Carhart 4-factor alpha, raw return, and Morningstar fund rating) over the identical time horizon (1-, 3-, or 5-year) as performance signals and obtain qualitatively similar results. These results are available upon request.

relatively sophisticated investors with a better understanding of the fund industry and who are therefore better able to interpret performance-related signals. Consequently, such investors face less ambiguity and may update their beliefs after observing the signals in a manner consistent with Bayes' rules. In a recent paper, Huang et al. (2012) show that investor fund flow sensitivity to past performance is decreasing with fund return volatility. Our analysis confirms their findings. An increase in fund return volatility implies that past performance is a less precise signal of skill or ability and hence investors rationally moderate their response to the signal. The dampening effect of return volatility on flow-performance sensitivity is more pronounced for institutional investors, which is consistent with the notion of such investors being more sophisticated.

High volatility is naturally associated with a high degree of uncertainty. Traditionally, uncertainty has been viewed as being the result of low signal precision. However, for ambiguity averse agents, high volatility could also imply a high degree of ambiguity surrounding the signal. Hence, the impact of volatility on the flow-performance sensitivity is likely to reflect not only the impact of signal precision, but also the ambiguity aversion of the investors.

Interestingly, we find that an increase in performance volatility, while dampening the flow-performance sensitivity in aggregate, also leads to an increase in flow sensitivity to the worst performance signal at the margin for both groups of investors, although the increase is more pronounced for retail investors. This suggests that investor behavior is best characterized as reflecting both Bayesian updating and ambiguity aversion with the two groups of investors displaying interesting differences in their response behavior. Institutional investors appear to behave in a manner more consistent with Bayesian updating whereas ambiguity aversion appears to play a relatively bigger role in the fund investment decisions of retail investors. Our results are consistent with the notion that retail investors are less sophisticated than institutional investors.[2]

Our study makes a number of contributions to the mutual fund literature and more generally to the evolving literature on the role of ambiguity in asset pricing. One, to our knowledge, it is the first study that explores the impact of ambiguity on fund investor decisions. Two, the paper complements recent findings on how fund investors respond to information about past fund performance. Three, the paper extends recent results in the literature on the impact of uncertainty in addition to risk, on expected asset returns. For example, when investors receive information of uncertain quality, theoretical models imply that aversion to ambiguity not only induces ambiguity premia and skewness in returns (Epstein and Schneider (2008)), but also results in non-participation (Easley and O'Hara (2009)), portfolio inertia and excess volatility (Illeditsch (2011)). However, there is limited empirical evidence in the literature on ambiguity-aversion behavior in asset markets. A recent exception is the study by Anderson et al. (2009), which provides empirical evidence of an uncertainty-return trade-off in equity markets.

---

[2]For example, Odean (1999), Barber and Odean (2000, 2001, 2002), and Bailey et al. (2011) provide evidence on the role of behavioral biases in the investment decisions of retail investors.

The rest of the paper is organized as follows. Section 1.2 reviews the literature on ambiguity aversion. Section 1.3 presents our model of ambiguity averse fund investors and derives testable implications. The data and empirical methodology are described in Section 1.4 while Section 1.5 presents the main empirical results. Section 1.6 presents the results of tests contrasting the response of Bayesian fund investors with that of ambiguity averse investors. Section 1.7 presents the results of robustness tests and concluding remarks are presented in Section 1.8.

1.2 Ambiguity Aversion

The distinction between risk and uncertainty was highlighted by Knight (1921) who defined uncertain events as those for which the probability distribution of outcomes is unknown.[3] The work of Ellsberg (1961) famously provided evidence of individual aversion to ambiguity or uncertainty (in contrast to risk). Subsequently, a large theoretical literature has evolved to formally develop models that accommodate ambiguity averse behavior and its implications. For example, Gilboa and Schmeidler (1989) propose an axiomatic framework of ambiguity aversion. They constructed an atemporal model in which preferences are represented by max-min expected utility over multiple possible distributions. Epstein and Schneider (2003) provide axiomatic foundation for intertemporal multiple-priors utility in discrete time.

[3] According to Knight (1921), "The practical difference between risk and uncertainty, is that in the former the distribution of the outcome in a group of instances is known (either through calculation a priori or from statistics of past experiences), while in the case of uncertainty this is not true, the reason being that it is impossible to form a group of instances, because the situation dealt with is in a high degree unique." (p. 103).

Epstein and Wang (1994) analyze the asset pricing implications of ambiguity aversion while Chen and Epstein (2002) extend the framework to continuous time. More recently this ambiguity aversion framework has subsequently been applied to explain some of the well known phenomena in asset markets. For example, Leippold et al. (2008) incorporate both learning and ambiguity in a Lucas exchange economy. The model is able to match the observed equity premium, the interest rate, and the stock return volatility, under empirically plausible parameter values. Epstein and Schneider (2008) study ambiguity averse investor behavior when processing information of uncertain quality. They find that aversion to ambiguity induces ambiguity premia and skewness in returns. Easley and O'Hara (2009) find that ambiguity aversion on the part of some traders can lead to non-participation in asset markets. Illeditsch (2011) finds that when investors receive a signal with unknown precision, ambiguity aversion causes portfolio inertia and excess volatility. In an experimental setting, Bossaerts et al. (2010) find evidence of heterogeneity in investor attitudes towards ambiguity. Moreover, they show that there is a wide range of prices for which a sufficiently ambiguity averse investor will avoid an ambiguous portfolio. In contrast to the theoretical and/or experimental studies, there is limited empirical evidence regarding the ambiguity averse behavior of investors. An exception is the study by Anderson et al. (2009) that provides empirical evidence of an uncertainty-return trade-off in equity markets.

Our paper contributes to the ambiguity literature in two aspects. First, we extend the extant theoretical framework to a multiple signals setting. We provide an answer to the question of how ambiguity aversion impacts investor decisions when facing multiple signals with unknown quality. Second, we provide empirical evidence based on the behavior of mutual fund investors that is consistent with the model's implications. To our knowledge ours is the first attempt at using an ambiguity aversion framework to study the response of fund investors to fund performance based signals.

1.3 The Model and Its Empirical Implications

In this section, we build a model to analyze the important features of the flow-performance sensitivity for an ambiguity averse mutual fund investor and derive testable empirical predictions.

1.3.1 The Model

Assume there is a population of investors, each with 1 unit of capital to invest with a fund. The investor decides on whether to fully invest her unit of capital with a particular fund. Her decision is based on her opportunity cost of capital, denoted hereafter by k, which is assumed to differ across the investors. The investors, indexed by k, are otherwise assumed to be identical. We assume that k has support in $[0, \infty)$, with cumulative distribution function denoted by $F(k)$.

The fund's return, R, is governed by

$$R = \mu + \alpha + \epsilon, \tag{1.1}$$

where α denotes the fund manager skill, μ is the market risk premium given the fund's risk profile, and $\epsilon \sim N\left(0, \sigma_\epsilon^2\right)$ represents the noisy component of the fund's return.

The risk-free rate is normalized to zero. We note that the managerial skill, α, is not directly observable by the investor. The investor is assumed to have knowledge about the distribution of skill in the population of the fund managers, i.e., investor knows a priori: $\alpha \sim N\left(\mu_\alpha, \sigma_\alpha^2\right)$. The investor, at the time of making investing decision, observes two signals, $s_{1,2}$, about managerial ability α, $s_i = (\alpha - \mu_\alpha) + \eta_i$, with

$$\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right). \tag{1.2}$$

[4]

We define the signal-to-noise ratio by $H_i = \sigma_\alpha^2/\sigma_i^2$, which captures the precision of the signal. Through the standard Bayesian updating, we have the posterior distribution of α, after observing the two signals, as

$$N\left(\mu_\alpha + a^T s, \sigma_\alpha^2/H\right) \equiv N\left(\mu_p, \sigma_p^2\right), \tag{1.3}$$

where $a = \left(H_1/H, H_2/H\right)^T$, and $H = 1 + H_1 + H_2$.

What differentiates our model from the standard Bayesian model is that we assume the investor is ambiguous about the signal precision. To capture the investor's attitude about this ambiguity, we adopt the framework in Klibanoff et al. (2005) and Ju and Miao (2012). We model investors as being risk neutral but ambiguity averse. Specifically, the investor's utility is

$$U(c) = E_\theta\left[-\exp\left(-\frac{1}{\gamma}E_\pi(c)\right)\right], \tag{1.4}$$

where c is the investment payoff on a state space S, and π is a probability measure on S. Importantly, the investor is uncertain about the "right" probability and considers a

[4]There may be some potential interest in analyzing the ambiguity of the signal correlation. But this feature is turned off here by assuming the noise terms in the two signals are uncorrelated.

set of multiple probability measures denoted by Δ. The parameter θ in Equation (1.4) denotes the investor's subjective prior over Δ, and in effect measures the subjective relevance of a particular probability measure, π. Due to the concavity of the negative exponential function, the investor represented by Equation (1.4) manifests ambiguity aversion. To capture the uncertainty faced by the investor, we assume that the probability measure θ in the definition of her utility function, is such that H_1 and H_2 are independent with a 50% probability of being equal to either h or l, with $h > l$. In this context, the degree of ambiguity faced by the investor is directly related to the difference between h and l. In the limiting case, as this differences converges to zero, we have the standard Bayesian learning framework.

From the above assumption, we have that $E_\pi(c) = \mu + \mu_\alpha + a^T s$. Investor k, by investing with the fund, achieves the following utility:

$$U(c) = -\frac{1}{4} e^{\frac{-(\mu+\mu_\alpha)}{\gamma}} \left[e^{\frac{-h(s_1+s_2)}{\gamma(1+2h)}} + e^{\frac{-l(s_1+s_2)}{\gamma(1+2l)}} + e^{\frac{-(hs_1+ls_2)}{\gamma(1+h+l)}} + e^{\frac{-(ls_1+hs_2)}{\gamma(1+h+l)}} \right].$$

The investor will invest with the fund if and only if the above utility is above her reservation level of utility, $-e^{-k/\gamma}$. We note the following propositions regarding fund flows.

Proposition 1. *The amount of capital under the management of the fund is $F(k^\star)$, where*

$$k^\star = \mu + \mu_\alpha - \frac{\gamma}{4} \log \left[e^{\frac{-h(s_1+s_2)}{\gamma(1+2h)}} + e^{\frac{-l(s_1+s_2)}{\gamma(1+2l)}} + e^{\frac{-(hs_1+ls_2)}{\gamma(1+h+l)}} + e^{\frac{-(ls_1+hs_2)}{\gamma(1+h+l)}} \right]. \tag{1.5}$$

Flow-performance sensitivity is captured by $dF(k^\star)/ds_i$, with

$$\frac{dF(k^\star)}{ds_i} = \frac{1}{4}F' e^{\frac{(k^\star-\mu-\mu_\alpha)}{\gamma}} \left[\frac{h}{1+2h}e^{-\frac{h(s_1+s_2)}{\gamma(1+2h)}} + \frac{l}{1+2l}e^{-\frac{l(s_1+s_2)}{\gamma(1+2l)}} \right. \\ \left. + \frac{l}{1+h+l}\left(e^{-\frac{(hs_1+ls_2)}{\gamma(1+h+l)}} + e^{-\frac{(ls_1+hs_2)}{\gamma(1+h+l)}}\right) + \frac{h-l}{1+h+l}e^{-\frac{(h-l)s_i+l(s_1+s_2)}{\gamma(1+h+l)}}\right]. \tag{1.6}$$

Proof: Let $U(c) = -e^{-k^\star/\gamma}$ and solve for $k^\star$ to get Equation (1.5). Then taking the derivative of $F(k^\star)$ with respect to s_1 and s_2 respectively, yields Equation (1.6).

This leads to the following proposition regarding the fund flow-performance sensitivity.

Proposition 2. *The flow-performance sensitivity is higher for the signal with relatively lower realized value:*

$$\frac{dF(k^\star)}{ds_1} > \frac{dF(k^\star)}{ds_2} \tag{1.7}$$

if and only if $s_1 < s_2$, i.e., in relative terms, the signal, s_1, conveys bad news, while the signal, s_2, conveys good news.[5]

Proof: As a consequence of Proposition 1, we have

$$\frac{dF(k^\star)}{ds_1} - \frac{dF(k^\star)}{ds_2} = \frac{1}{4}F' e^{\frac{(k^\star-\mu-\mu_\alpha)}{\gamma}} \left(\frac{h-l}{1+h+l}\right)\left[e^{\frac{-(hs_1+ls_2)}{\gamma(1+h+l)}} - e^{\frac{-(ls_1+hs_2)}{\gamma(1+h+l)}}\right]. \tag{1.8}$$

When $h > l$, we have $hs_1 + ls_2 < ls_1 + hs_2$, thus the above expression is always positive.

[5]Remark: In the model, the shape of the flow-performance relation (i.e., whether or not it is convex) depends on the specification of the cumulative distribution function $F(k)$. On the one hand, this implies that the model is unable to explain why the flow-performance relation has a specific functional form, because such a relation is driven by direct assumption. On the other hand, the model is flexible enough to allow for such a relation. Our point is that our key result, namely, that the flow-performance sensitivity is higher for the signal with relatively lower realized value, is logically consistent and potentially complementary to the convexity in the flow-performance relation.

Thus, the model implies that ambiguity averse investors' fund flow is more responsive to the worst signal. In other words, in the population of ambiguity averse investors that face multiple signals, we expect to observe heightened flow sensitivity towards the signal that conveys bad news.

Denote $(h-l)/2$ by δ. The higher the parameter value for δ, the wider the gap between the two possible parameter values for the signal precision, and therefore the higher the ambiguity the investor faces. As a shorthand, we denote $e^{\delta(s_2-s_1)/\gamma(1+h+l)}$ by X. Clearly, $X > 1$ if $s_1 < s_2$. We have the following corollary.

Corollary 1. *The extra sensitivity to the worst signal is higher when the level of ambiguity is higher, keeping the average precision of the signal constant (i.e., $(h+l)/2$ is fixed). That is:*

$$\frac{d}{d\delta}\left(\frac{dF}{ds_1}-\frac{dF}{ds_2}\right) > 0 \tag{1.9}$$

if $s_1 < s_2$.

Proof: Taking the derivative of Equation (1.8) with respect to δ yields:

$$\frac{d}{d\delta}\left(\frac{dF}{ds_1}-\frac{dF}{ds_2}\right) = \frac{1}{2(1+h+l)X} F' e^{\left[\frac{k^{\star}-\mu-\mu_{\alpha}}{\gamma}-\frac{(h+l)(s_1+s_2)}{2(1+h+l)\gamma}\right]} \left[\left(X^2-1\right)+\delta\frac{s_2-s_1}{1+h+l}\left(X^2+1\right)\right]$$

$$>0.$$

1.3.2 Bayesian Benchmark

Next, we contrast our model with the standard Bayesian benchmark case with risk averse utility.

1.3.2.1 Bayesian Benchmark with No Signal Precision Uncertainty

Consider first the simplest case of a risk averse investor with exponential utility and risk aversion coefficient a, i.e., $U(c) = -e^{ac}$, who receives two performance-related signals with known precision. Specifically, consider the case when $h = l \equiv H$. Similar to Proposition 1, the amount of capital under management of the fund is $F(k^\star)$, where

$$k^* = \mu + \mu_\alpha + \frac{H}{1+2H}(s_1 + s_2) - \frac{1}{2}a\left(\sigma_\epsilon^2 + \frac{\sigma_\alpha^2}{1+2H}\right). \tag{1.10}$$

The flow-performance sensitivity is captured by $dF(k^\star)/ds_i$, where

$$\frac{dF(k^\star)}{ds_1} = \frac{dF(k^\star)}{ds_2} = F' \times \left(\frac{H}{1+2H}\right). \tag{1.11}$$

From the above expression, it is clear that in this case the flow-performance sensitivity depends only on signal precision H, and is independent of the level of the signal realization. This implies we would not expect to observe additional flow sensitivity to the worst signal in the population of Bayesian investors who do not face signal precision uncertainty.

1.3.2.2 Bayesian Benchmark with Signal Precision Uncertainty

We next consider the case where risk averse investors are uncertain about the precision of the realized signals, s_1 and s_2. Specifically, the precision of each signal may take on one of two values: h (high) and l (low).

Note that we assume $s_1 < s_2$ as in the ambiguity aversion model. In this case, the investor faces the following possible scenarios:

- Both signal realizations are negative, i.e., $s_1 < s_2 < 0$.

- Both signal realizations are positive, i.e., $0 < s_1 < s_2$.
- The signal realizations are opposite in sign, i.e., $s_1 < 0$, and $s_2 > 0$.

To explore these cases we numerically analyze the investor flow sensitivity to the respective performance related signals over a range of plausible values of the signals. We calibrate the key parameters for this analysis based on the sample of equity mutual funds described in detail in Section IV. In particular, we set the investor's prior regarding the average fund manager skill, μ_α, equal to zero. The value of the parameter σ_α^2 is set equal to the cross-sectional variance of the Carhart 4-factor alphas across all funds in the sample. The appendix provides additional details of the numerical analysis.

Intuitively, the Bayesian investor's sensitivity to a signal is a function of the inferred precision of the observed signal. As a result, the further away a signal is from the investor's prior regarding the signal, the lower the signal's implied precision. Recall, that the investor's prior beliefs about the distribution of the signals are centered on zero. Accordingly, if both signal realizations are negative then the investor attaches a lower precision to the signal further away from zero, i.e., the more negative (worst) signal s_1. On the other hand, if both signal realizations are positive then the investor attaches a lower precision to the more positive signal, i.e., s_2.

Figure 1 shows the difference in the flow-performance sensitivity of the signal, s_1, relative to the signal, s_2, as a function of the respective signal realizations. Graph A displays the case when both signals have negative realizations, i.e., $s_1 < s_2 < 0$. As is evident, in this case the flow-performance sensitivity to the worst signal, s_1, is consistently lower across the entire range of signal realizations. This is in stark contrast to our earlier result for the case with ambiguity aversion.

Graph B shows the difference in the flow-performance sensitivity of signal s_1 relative to the signal, s_2, when both signals have positive realizations, i.e., $0 < s_1 < s_2$. In this case the flow-performance sensitivity to the worst signal, s_1, is higher over a certain range of signal realizations and it is lower outside of this range. This result is also different from the case with ambiguity aversion where the flow-performance sensitivity is consistently higher for the worst signal.

Graph C shows the difference in flow-performance sensitivity when $s_1 < 0 < s_2$. Consistent with intuition, the flow-performance sensitivity is always lower for the signal further away from zero. As a result, there is a certain range of signal realizations in which the flow-performance to the worst signal is higher. However, outside this range it is lower.

In summary, the results of this subsection confirm that the ambiguity aversion model leads to predictions that differ from the Bayesian benchmark case with respect to the investor flow-performance sensitivity to the worst signal.

1.3.3 Distinction with Other Behavioral Biases

In the behavioral finance literature, a number of alternatives have been suggested as departures from the traditional rational agent (Savage utility) paradigm. Examples include loss-averse preferences as well as behavioral biases such as overconfidence.[6] It is worth noting that the hypothesis developed in Proposition 2 is uniquely attributable to the existence of ambiguity aversion on the part of investors. In particular, neither loss aversion, nor overconfidence will lead to a differential sensitivity of investor fund flows to the lower realization signal in the absence of ambiguity aversion. To illustrate this, we next formally examine the implications of loss aversion and overconfidence respectively, on the behavior of investors who face multiple noisy signals. For this analysis we abstract from the effect of ambiguity aversion by imposing the restriction that the two signals have equal precision, i.e., $H_1 = H_2 = H$.

First, consider the case of loss-averse utility preferences. The case with habit formation utility follows in similar fashion. Assume, as in Kahneman and Tversky (1979), the investor displays loss aversion in her utility function but she is not ambiguous about the signal precision. Thus, given the posterior distribution of α, we assume the utility function takes the following form:

$$U(c) = E_{\pi}\left(\alpha \mathbb{1}[0,\infty)\right) + \lambda E_{\pi}\left(\alpha \mathbb{1}(-\infty,0)\right), \tag{1.12}$$

where $\lambda > 1$. Writing the above equation in explicit form, we have:

$$\begin{aligned} U(c) = &\int_0^{\infty} \frac{\alpha}{\sqrt{2\pi\sigma_p^2}} e^{\frac{-(\alpha-\mu_p)^2}{2\sigma_p^2}} d\alpha + \lambda \int_{-\infty}^{0} \frac{\alpha}{\sqrt{2\pi\sigma_p^2}} e^{\frac{-(\alpha-\mu_p)^2}{2\sigma_p^2}} d\alpha \\ = &\frac{\sigma_p(1-\lambda)}{\sqrt{2\pi}} e^{\frac{-\mu_p^2}{2\sigma_p^2}}, \end{aligned} \tag{1.13}$$

where α_p and μ_p are given in Equation (1.3) with $H_1 = H_2 \equiv H$. Thus, the flow-

[6]Bailey et al. (2011) provide evidence of the impact of behavioral biases including overconfidence, on the decisions of mutual fund investors.

performance sensitivity under the assumption of loss aversion is:

$$\frac{dF(k^\star)}{ds_1} = \frac{dF(k^\star)}{ds_2} = \frac{(\lambda - 1)H\mu_p^2}{\sigma_p(1+2H)\sqrt{2\pi}} F' e^{\frac{-\mu_p^2}{2\sigma_p^2}}, \tag{1.14}$$

which is independent of whether or not $s_1 < s_2$.

Second, we note that overconfidence cannot by itself result in the asymmetric sensitivity to signals with different realizations. Overconfidence is the belief on the part of the investor that a certain signal is more precise than it actually is. For example, suppose that an overconfident investor receives a signal, s_1, from a private channel, while another signal, s_2, is publicly available. She is confident that her private signal, s_1, is always more reliable than the public signal, s_2. As a result, when making investment decisions, the investor allocates additional capital to the fund whenever s_1 is sufficiently positive, and conversely she withdraws money from the fund whenever the signal, s_1, is negative. The public signal, s_2, on the other hand, regardless of its realization, will have a lesser influence on the investor's fund investment decisions. Simply put, under the assumption of overconfidence, regardless of whether $s_2 > s_1$, or $s_2 \leq s_1$, we always have:

$$\frac{dF(k^\star)}{ds_1} > \frac{dF(k^\star)}{ds_2}. \tag{1.15}$$

Obviously under this setting of overconfidence, we cannot arrive at Proposition 2.

1.3.4 Empirical Predictions

In order to develop testable hypotheses, it is important to first clarify why the mutual fund industry provides a perfect setting to test our model. First, mutual funds represent a significant proportion of the U.S. household assets.[7] These investors span all age, income groups, and wealth levels and are thus representative of the population of individual investors. Another favorable feature about the mutual fund industry is that it has separate share classes for individual investors and institutional investors, which allows us to study the behavioral differences between the two types of investors.

Second, the mutual fund flow-performance relationship has been well documented in the literature. Previous studies (e.g., Ippolito (1992); Gruber (1996); Chevalier and Ellison (1997); Sirri and Tufano (1998)) show that mutual fund investors make investment or redemption decisions relying on past fund performance. This relationship allows us to directly observe the investors' response to performance related information. Third, mutual funds make their past performance statistics readily available to investors. Such statistics include the performance figures for each share class over the previous 1 year, 3 years, 5 years, as well as the entire period since inception. However, the degree to which past performance is informative of fund manager skill is unknown to investors. Also, these performance data have different realizations since a fund's performance may fluctuate over time. Thus, these performance statistics over different time horizons serve as multiple comparable signals with unknown quality and different realizations from which investors learn about fund manager skill. Fourth, the comprehensive data on fund flows and past performance make it possible for us to study investors' response to multiple signals under ambiguity aversion in a

[7] According to the 2012 Investment Company Institute Fact Book, 44% of all U.S. households owned mutual funds.

non-experimental setting.

In this subsection, we develop several testable hypotheses from the model concerning the impact of ambiguity aversion on mutual fund investor flow-performance sensitivity. We have the following hypotheses:

Hypothesis 1. *When a fund's past performance is measured over multiple time horizons, fund flows display additional sensitivity to the minimum performance measure in the presence of ambiguity averse investors.*

According to Proposition 2, given signals with unknown quality, ambiguity averse investors' fund flow response is more sensitive to the worst signal, i.e., the signal with the worst realization. In the mutual fund industry, fund investors are routinely provided with performance statistics measured over past 1 year, 3 years, and 5 years, from which they try to learn about fund manager skill. Proposition 2 says that ambiguity averse investors' fund flows will display additional sensitivity towards the worst signal, i.e., the minimum performance over multiple time horizons in this setting.

Hypothesis 2. *Individual investors show stronger ambiguity aversion than institutional investors, as measured by a higher marginal sensitivity to the minimum performance measure.*

The above hypothesis reflects the notion that individual (retail) investors are less sophisticated compared to institutional investors. As a result, they may be less confident about how much the past performance is indicative about fund manager skill. They may therefore be subject to greater ambiguity in terms of interpreting such information. On the other hand, institutional investors, who are believed to be much more sophisticated, may hold more confident beliefs about the precision of signals when they look at past performance measures. As a result, we expect to see stronger ambiguity aversion in the sample of retail investors compared to institutional investors.

Hypothesis 3. *Funds that change their investment strategy more aggressively and/or frequently are characterized by a higher degree of ambiguity. Investor fund flows in such funds will display a higher marginal sensitivity to the minimum performance measure.*

Fund investment strategy changes can be viewed as a proxy for a fund's ambiguity level. If a fund constantly switches its' investment strategy, e.g., from a passive diversification strategy to active factor timing or stock selection, investors may find it hard to evaluate its relative performance and to form a concrete expectation of future return. Thus, the more aggressively and/or frequently a fund changes its strategy, the more ambiguous it is from the investors' perspective.

Hypothesis 4. *Funds with more volatile cash flows are characterized by a higher degree of ambiguity. Investor fund flows in such funds will display a higher marginal sensitivity to the minimum performance measure.*

Fund inflows reflect a general positive view among investors about the fund's future prospects and outflows are indicative of a negative view about the fund. Thus, flow volatility can be viewed as a proxy for the degree of variability or uncertainty of investor opinion with respect to the fund's prospects. In this sense, highly volatile investor fund flows imply greater uncertainty about the fund's future performance. Hence, funds with more volatile flows are likely to be the ones that are more ambiguous to investors.

Hypothesis 5. *Funds that belong to a smaller family appear more ambiguous to investors. Investor fund flows to such funds will display a higher marginal sensitivity to the minimum performance measure.*

Since the 1990s, there has been a sharp increase in multiple share classes that belong to the same fund family. If a fund belongs to a family with a large asset base, it is more recognizable and enjoys the reputation built up by the entire fund family. On the other hand, if a fund belongs to a small and little known family, investors may be more conservative when making their investment decisions. It may also be harder for them to rely on past performance of such funds in drawing inference about manager skill. In other words, funds belonging to smaller families may appear more ambiguous to investors.

Hypothesis 6. *Funds with greater marketing expenditures appear to be less ambiguous to investors. Investor fund flows to such funds will display a lower marginal sensitivity to the minimum performance measure.*

We hypothesize that funds that spend more on marketing are less ambiguous to investors since investors are likely to be more familiar with funds that advertise more.[8] In most advertisements, funds highlight their past performance and services through magazines, TV programs, etc., as discussed by Jain and Wu (2000). A higher visibility (due to increased advertising) may lead to a greater degree of confidence among investors regarding the quality of past performance as signals for manager skill. In other words, funds with higher advertising related expenditures may be less ambiguous to investors in general.

1.4 Data and Methodology

1.4.1 Data

We use data from the Center for Research in Security Prices (CRSP) Survivorship Bias Free Mutual Fund Database, which includes information on the funds' total net assets, returns and characteristics. We focus on actively managed U.S. equity mutual funds, thus, we also exclude index funds and funds that are closed to new investors. To be consistent with prior studies, we exclude sector funds, international funds, bond funds, and balanced funds from our analysis. We classify funds into five categories based on their objective codes:[9] aggressive growth, growth, growth and

---

[8]Marketing related expenses including 12b-1 fees have been employed as empirical proxies for investor search and participation costs in studies by Sirri and Tufano (1998), and Huang et al. (2007).

[9]We categorize funds according to the following criteria. First, funds with Lipper objective codes G, LCGE, MCGE, MLGE, SCGE, with Wiesenberger objective codes G, G-S, S-G, GRO, LTG, SCG, or with Strategic Insight objective codes GRO, SCG are classified as growth funds. Second, funds with Wiesenberger objective code AGG or with Strategic Insight objective code AGG are classified as aggressive growth. Third, funds with Lipper objective code GI, with Wiesenberger objective codes G-I-S, G-S-I, I-G, I-G-S, GCI, G-I, I-S-G, S-G-I, S-I-G, GRI, or with Strategic Insight objective code GRI are classified as growth and income funds. Fourth, funds with Lipper objective codes EI, EIEI, I, with Wiesenberger objective codes I, I-S, IEQ, ING, or with Strategic Insight objective codes

income, income and others. We also classify funds into retail shares or institutional shares.

We primarily study the period from January 1993 through December 2011, since the CRSP database does not report 12b-1 fees until 1992 and institutional funds begin to mushroom in the 1990s. However, as a robustness check we confirm that our results are qualitatively unchanged when we extend the sample to the period: January 1985 through December 2011. We examine fund flows and other characteristics at the quarterly frequency. Consistent with prior studies, we define quarterly net flow into a fund as

$$Flow_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1}(1 + R_{i,t})}{TNA_{i,t-1}}, \tag{1.16}$$

where $R_{i,t}$ denotes fund i's return during quarter t, and $TNA_{i,t}$ is the fund's total net asset value at the end of quarter t. Thus, our definition of flows reflects the percentage growth of the fund's assets in quarter t. To prevent the potential impact of extreme values of flows resulting from the errors associated with mutual fund mergers and splits in CRSP mutual fund database, we filter out the top and bottom 1% tails of the net flow data. To further guard against this issue, we delete records of funds from our analysis before their total net asset value first hits the $3 million mark.

Table 1.1 reports the summary statistics of mutual funds characteristics. We note that since we are interested in studying the fund flow behavior of retail as well as institutional investors we treat each fund share class as an individual fund, consistent

ING are classified as income funds. Fifth, all the other actively managed equity funds in our sample are classified as others.

with the research design employed by Huang et al. (2007, 2012).[10] In 1993, there are 707 distinct fund share classes in our sample and 63 of them are open only to institutional investors. In 2011, the number of funds in our sample grows to 4,242 with 785 institutional funds. In total, our sample includes 7,020 distinct fund share classes and 216,366 fund share class-quarters. In an average quarter, the sample includes 2,847 funds with average total net assets (TNA) of $678.06 million and an average net flow of 1.51%. Following Sirri and Tufano (1998) we measure total expense as the expense ratio plus one-seventh of the front-end load. The 12b-1 fees are the part of fund expenses that cover distribution expenses and sometimes shareholder service expenses. Distribution expenses include marketing, advertising, and compensation paid for brokers who sell the funds. As may be seen from Table 1.1, the 12b-1 fees for retail funds are 0.57%, which is nearly three times of that for institutional funds.

1.4.2 Empirical Methodology

We formally analyze the relationship between fund flows and performance measured over multiple time horizons when controlling for other factors. We estimate the following model using 76 quarters of fund-level data over the period 1993 to 2011 to test our baseline hypothesis:

$$\begin{aligned} Flow_{i,t} =& a + b_1 Perf\_1yr_{i,t} + b_2 Perf\_3yr_{i,t} + b_3 Perf\_5yr_{i,t} + cMin\_Rank_{i,t} \\ &+ Controls + \epsilon_{i,t}. \end{aligned} \tag{1.17}$$

[10] As emphasized in Huang et al. (2007), since our focus is to study fund flows, treating each fund share class separately will not lead to the double-counting problem. Also, since most of our tests require a separation between retail shares and institutional shares, we conduct all tests at the fund share class level.

Following Sirri and Tufano (1998), the variables $Perf\_1yr$, $Perf\_3yr$, and $Perf\_5yr$ represent fractional performance ranks ranging from 0 to 1 based on fund i's performance during past 12 months, 36 months, and 60 months, respectively. The variable Min_Rank is defined as:

$$Min\_Rank_{i,t} = Min\left(Perf\_1yr_{i,t}, Perf\_3yr_{i,t}, Perf\_5yr_{i,t}\right). \tag{1.18}$$

Thus, the coefficient c in Equation (1.17) captures the additional flow sensitivity to performance measured over the particular horizon during which the fund had the worst performance ranking. This coefficient is the focus of our tests. As we discuss below, this additional sensitivity is significant in both economic and statistical sense in the population of retail investors.

Since it is unclear which measure of performance a typical investor would focus on when evaluating funds, we consider two alternative measures. The first measure is the average monthly raw return measured over a specified time horizon, i.e., 12 months, 36 months or 60 months. The second measure is the fund's factor-adjusted performance using the Carhart (1997) 4-factor model. In order to estimate the 4-factor model, we first calculate fund i's factor loadings in quarter t by regressing the past 60 months' excess returns on the four factors:

$$R_{i,\tau} - R_{f,\tau} = \alpha_i + \beta_i^{MKT} MKT_\tau + \beta_i^{SMB} SMB_\tau + \beta_i^{HML} HML_\tau + \beta_i^{UMD} UMD_\tau + \epsilon_{i,\tau}, \tag{1.19}$$

where $R_{i,\tau}$ is the return for fund i and $R_{f,\tau}$ is the one-month T-bill rate in month τ. The market factor, MKT_τ, represents the monthly excess market return. The factors

SMB_{τ}, HML_{τ}, and UMD_{τ} represent the monthly returns on the size, value, and momentum factor mimicking portfolios, respectively. We obtain the factor returns from Kenneth French's website. We then calculate the fund's factor adjusted alpha each month using the monthly fund excess returns, and the factor loadings estimated as above. We compute the average of these monthly alphas over distinct horizons of 12 months, 36 months, and 60 months, respectively.

To obtain fractional performance ranks ($Perf\_1yr$, $Perf\_3yr$, and $Perf\_5yr$) ranging from 0 to 1, we apply different approaches to rank funds based on the two performance measures. In the case of the raw return measure, we rank funds every quarter within fund objective categories based on their average raw returns over each of the three lagged time horizons. For rankings based on the 4-factor alphas, we rank all funds each quarter according to their Carhart (1997) alphas over each of the three time horizons considered.[11]

The control variables employed in Equation (1.17) include a number of fund characteristics that have been shown to affect fund flows. In particular, we control for the logarithm of one plus fund age, previous quarter's flow, fund size as measured by the natural logarithm of fund total net asset in the previous quarter, volatility of monthly raw returns during the prior 12 months, and the lagged total expense ratio. Finally, following Sirri and Tufano (1998) we also include the category flow, defined

[11]For an average fund during the sample period 1993-2011, the correlation between its 1- and 3-year raw return-based ranks is 0.46, between 1- and 5-year ranks is 0.32, and between the 3- and 5-year ranks is 0.52. The correlation between the 1- and 3-year Carhart (1997) 4-factor alpha-based ranks is 0.49, between the 1- and 5-year ranks is 0.33, and between 3- and 5-year ranks is 0.55.

as the percentage quarterly net asset growth of the fund's objective category.

We estimate the model in Equation (1.17) by conducting a cross-sectional linear regression each quarter and reporting the time-series means and the related Newey-West t-statistics of the coefficients following Fama and MacBeth (1973). We follow this approach for all of the analysis throughout the paper.

1.5 Empirical Results

1.5.1 Ambiguity Aversion

In this subsection, we test Hypotheses 1 and 2 by studying investor ambiguity aversion behavior within subsamples of retail funds and institutional funds respectively. Our model implies that in the presence of ambiguity averse investors, fund flows will display additional sensitivity to a fund's minimum performance ranking. This marginal sensitivity is in addition to the general response of flows to past performance measures.

To test this implication, we estimate the baseline model in Equation (1.17) for retail funds and institutional funds separately. According to Hypothesis 2, we expect to see a significant coefficient, c, for the variable Min_Rank for the subsample of retail funds only. The results are reported in Table 1.2. Columns 1 and 2 report results using raw return as the performance measure and Columns 3 and 4 report results using the Carhart (1997) alpha as the performance measure.

Consistent with Hypothesis 2, we find that the coefficient for Min_Rank is positive and significant at the 1% level for for both measures of performance in the retail funds sample. However, for institutional funds, the coefficient is positive but insignificant. Focusing on the results for retail funds in Column 1 where the performance is measured in terms raw returns, the coefficient for the Min_Rank variable is 0.048. The coefficient for performance measured over the 1-year horizon is 0.05, and this is the performance horizon that has the largest affect on fund flows. This implies that for retail funds, a 14% (the average absolute change in the 1-year performance ranking for retail funds in our sample) increase in the fund's 1-year performance rank will result in a 0.7% increase in the fund's assets. However, if the 1-year performance rank happens to be the worst among the three horizons, a 14% increase in the fund's (1-year) ranking results in an inflow equal to $(0.048 + 0.05) \times 14\% = 1.37\%$ of the fund's assets. Thus, a 14% improvement in the worst performance rank results in a doubling of the fund flows enjoyed by the fund relative to the normal increase in flows from general performance improvement. The economic magnitude is quite significant given the average flow in the retail sample is 1.32%. In Column 3 when performance is measured using Carhart (1997) 4-factor alpha, we observe a similar coefficient for the Min_Rank. As seen from the results presented in Columns 2 and 4 of Table 1.2, in the case of institutional funds, even though the coefficients for Min_Rank are positive, they are not statistically significant and are much smaller in magnitude than their retail counterparts.

We note that the coefficients on the other control variables included in Equation (1.17) are consistent with previous findings in literature. The positive and significant coefficients for all three performance measures in Columns 1 through 4 conform to the performance chasing behavior as documented in Chevalier and Ellison (1997) and Sirri and Tufano (1998), among others. Sirri and Tufano (1998) also document the negative impact of volatility on flows. The positive coefficient on *Previous Quarterly Flow* confirms the persistence in fund flows. Similarly, the negative coefficients for *Total Expense* in Columns 1 through 4 are consistent with previous studies by Barber et al. (2005) and Sirri and Tufano (1998), among others.

After controlling for above factors, the significant coefficients for *Min_Rank* in Columns 1 and 3 indicate that minimum performance ranks have significant explanatory power for fund flows. In sum, Table 1.2 provides evidence for ambiguity aversion behavior among retail funds investors as shown by the significant positive coefficient for the minimum performance rank, which is consistent with Hypotheses 1 and 2. Accordingly, we focus on retail funds in conducting the tests reported for the remainder of the section.

1.5.2 Strategy Changes as a Proxy for Fund Ambiguity

Hypothesis 3 states that fund strategy changes could be viewed as a proxy for a fund's ambiguity level. Investors are likely to face a higher degree of ambiguity with regard to a fund that switches its investment strategy too aggressively/frequently. We adopt two ways to measure a fund's strategy shifts. The first measure is the fund's average absolute change in its factor loadings, while the second measure is the fund's R-squared computed from a time series regression using the entire history of fund returns.

The first proxy is motivated by Lynch and Musto (2003). Each quarter t, we compute a fund's factor loadings with respect to the four Carhart (1997) model factors over two non-overlapping 30-month periods, namely, the prior 1-30 and 31-60 month periods. We then compute the average absolute change in the factor loadings from the initial 30-month period to the most recent 30-month period as

$$LDEL_{i,t} = \frac{1}{4} \sum_{f} |\beta^{f}_{i,t,1-30} - \beta^{f}_{i,t,31-60}|, \tag{1.20}$$

for $f = MKT, HML, SML$ and UMD. A higher loading-change indicates a more aggressive shift in the fund's strategy.

The second proxy is the fund's R-squared from a time series regression of the fund's monthly returns on the four Carhart (1997) factors. A low R-squared value implies that the 4-factor model is a poor performance attribution model for the fund. A potential reason could be that the factor loadings of the fund are not constant over the sample period implying frequent shifts in the factor exposures or the investment style. Such shifts would make it harder for investors to interpret past performance related signals and contribute to an enhanced level of ambiguity in interpreting such signals. Each quarter, we divide the sample of funds into three groups, Low, Mid and $High$ based on their $LDEL$ or R-squared values. We then we apply the baseline model described in Equation (1.17) to each group and report the time-series average of the coefficients. The results are reported in Table 1.3.

Performance in Table 1.3 is measured by raw return (Column 1-3) or the Carhart (1997) 4-factor model alpha (Column 4-6). Panel A of Table 1.3 reports results when strategy changes are measured using the average absolute change in factor loadings ($LDEL$). As shown in Column 1, for the group of funds with low loading-change, an indication of less aggressive strategy shifts, the additional flow sensitivity to the minimum performance rank, as captured by the Min_Rank coefficient, is only 0.011 and it is statistically insignificant. However, for funds which shift their strategies more aggressively, as shown in Column 3, the marginal sensitivity to minimum performance is nearly six times higher at 0.061 which is significant at the 1% level. The F-test statistic rejects the null that the Min_Rank coefficients across the three groups are equal. Columns 4-6 report qualitatively similar flow-performance sensitivity results when performance is measured using the Carhart (1997) 4-factor alpha.

Panel B of Table 1.3 presents results using a fund's 4-factor model R-squared values as a proxy for the fund's ambiguity. Columns 1-3 report results when a fund's performance is measured by the ranking of its raw return. Flows of funds with low R-squared values show additional sensitivity to the minimum performance rank as seen by the Min_Rank coefficient in Column 1 (0.056), which is significant at 1% level. In Column 3, however, flows of funds with high R-squared values, an indication of relatively stable investment strategy, have marginal sensitivity to minimum performance of only 0.021 which is statistically insignificant. The F-test statistic rejects the null that the Min_Rank coefficients across the three groups are equal. Columns 4-6 report qualitatively similar flow-performance sensitivity results when performance is measured using the Carhart (1997) 4-factor alpha.

In conclusion, the results in Table 1.3 show that funds that change investment strategy more aggressively/frequently are more ambiguous to investors, as evidenced by the greater marginal sensitivity of investor fund flows to the minimum performance measure.

1.5.3 Flow Volatility as a Proxy for Fund Ambiguity

Hypothesis 4 states that the flow volatility could be used as a proxy for a fund's ambiguity level. Note that fund flows are the consequence of investors' asset allocation decisions. A net inflow is an indication of an overall positive view of the fund while a net outflow represents an overall negative view. Thus, flow volatility captures the uncertainty about the funds' future performance from the perspective of an average investor. In this sense, flow volatility is a direct measure of the fund's ambiguity level. We expect to observe stronger ambiguity aversion behavior among investors of funds with more volatile past flows. We measure flow volatility ($Flow\_Vol$) as the standard deviation of a fund's previous 12 quarters' fund flows.

To test this hypothesis, we estimate the following regression model:

$$\begin{aligned}
Flow_{i,t} =& a + b_{11} Low_Flowvol_{i,t} \times Perf_1yr_{i,t} + b_{12} Mid_Flowvol_{i,t} \times Perf_1yr_{i,t} \\
&+ b_{13} High_Flowvol_{i,t} \times Perf_1yr_{i,t} + b_{21} Low_Flowvol_{i,t} \times Perf_3yr_{i,t} \\
&+ b_{22} Mid_Flowvol_{i,t} \times Perf_3yr_{i,t} + b_{23} High_Flowvol_{i,t} \times Perf_3yr_{i,t} \\
&+ b_{31} Low_Flowvol_{i,t} \times Perf_5yr_{i,t} + b_{32} Mid_Flowvol_{i,t} \times Perf_5yr_{i,t} \\
&+ b_{33} High_Flowvol_{i,t} \times Perf_5yr_{i,t} + c_1 Low_Flowvol_{i,t} \times Min_Rank_{i,t} \\
&+ c_2 Mid_Flowvol_{i,t} \times Min_Rank_{i,t} + c_3 High_Flowvol_{i,t} \times Min_Rank_{i,t} \\
&+ Flow_Vol_{i,t} + Controls + \epsilon_{i,t}, \qquad (1.21)
\end{aligned}$$

where $Low\_Flowvol_{i,t}$ is a dummy variable that equals one if the fund i's flow volatility falls into the bottom tercile in quarter t, $Mid\_Flowvol_{i,t}$ is a dummy variable that equals one if it belongs to the medium tercile, and $High\_Flowvol_{i,t}$ is a dummy variable that equals one if the flow volatility is ranked in the top tercile. Each quarter we conduct a cross-sectional regression of flows on the interaction of the three dummy variables with $Perf\_1yr$, $Perf\_3yr$, $Perf\_5yr$ and Min_Rank, respectively, together with the set of control variables and report the time-series mean and Newey-West t-statistics of the coefficients. The control variables are the same as those in Equation (1.17). The coefficients on the interaction terms capture the differential flow sensitivity to a certain performance horizon or to the minimum performance for funds with low, medium or high flow volatility. For example, the coefficient of $High\_Flowvol \times Min\_Rank$, i.e., c_3, captures the additional sensitivity to the Min_Rank for funds with high flow volatility. Given Hypothesis 4, we expect a large and positive coefficient for $High\_Flowvol \times Min\_Rank$, and a small coefficient for $Low\_Flowvol \times Min\_Rank$.

Table 1.4 presents the flow-performance results based on flow volatility as a proxy for fund ambiguity level. Columns 1 and 2 report results using raw return and Carhart (1997) alpha as performance measures, respectively. As expected, in Column 1, the coefficients for the three interaction terms $Low\_Flowvol \times Min\_Rank$, $Mid\_Flowvol \times Min\_Rank$ and $High\_Flowvol \times Min\_Rank$ increase monotonically from -0.002 (statistically insignificant) to 0.110 (significant at the 1% level). The F-test statistic rejects the null that the coefficients of $Low\_Flowvol \times Min\_Rank$, $Mid\_Flowvol \times Min\_Rank$ and $High\_Flowvol \times Min\_Rank$ are equal. In Column

2 of Table 1.4, where fund performance is measured using the Carhart (1997) alpha, we observe a similar increase in the value of the coefficients across the three flow volatility terciles.

In conclusion, the results in Table 1.4 are consistent with Hypothesis 4 that funds with more volatile flow appear to be more ambiguous to investors.

1.5.4 Family Size as a Proxy for Fund Ambiguity

Hypothesis 5 states that family size can also be used as a proxy for a fund's ambiguity level. In this subsection, we use the sum of total net assets for each fund within a fund family as a measure of family size. We then test whether a larger family size is associated with a reduction in the marginal flow sensitivity to a fund's minimum performance measure.

To test this hypothesis, we estimate the following model:

$$\begin{aligned} Flow_{i,t} =& a + b_{11} Low\_Famsize_{i,t} \times Perf\_1yr_{i,t} + b_{12} Mid\_Famsize_{i,t} \times Perf\_1yr_{i,t} \\ &+ b_{13} High\_Famsize_{i,t} \times Perf\_1yr_{i,t} + b_{21} Low\_Famsize_{i,t} \times Perf\_3yr_{i,t} \\ &+ b_{22} Mid\_Famsize_{i,t} \times Perf\_3yr_{i,t} + b_{23} High\_Famsize_{i,t} \times Perf\_3yr_{i,t} \\ &+ b_{31} Low\_Famsize_{i,t} \times Perf\_5yr_{i,t} + b_{32} Mid\_Famsize_{i,t} \times Perf\_5yr_{i,t} \\ &+ b_{33} High\_Famsize_{i,t} \times Perf\_5yr_{i,t} + c_1 Low\_Famsize_{i,t} \times Min\_Rank_{i,t} \\ &+ c_2 Mid\_Famsize_{i,t} \times Min\_Rank_{i,t} + c_3 High\_Famsize_{i,t} \times Min\_Rank_{i,t} \\ &+ D\_FamilySize_{i,t} + Controls + \epsilon_{i,t}, \end{aligned} \tag{1.22}$$

where $Low\_Famsize_{i,t}$ is a dummy variable that equals one if fund i belongs to a fund family whose size falls into the bottom tercile in quarter t, $Mid\_Famsize_{i,t}$ is a dummy variable that equals one if family size belongs to the medium tercile and $High\_Famsize_{i,t}$ is a dummy variable that equals one if the family size is in the top tercile. The variable $D\_Familysize_{i,t}$ is a dummy variable that equals 1 if the fund family size is above the median value for that quarter. We regress quarterly flows on the interaction of the three dummy variables with $Perf\_1yr$, $Perf\_3yr$, $Perf\_5yr$ and Min_Rank, respectively. The coefficients on the interaction terms represent the differential flow sensitivity to a certain performance horizon or to the minimum performance rank for funds that belong to a small-sized family, medium-sized family and large-sized family, respectively. For example, the coefficient c_1 for the variable $Low\_Famsize \times Min\_Rank$ in Equation (1.22) captures the additional flow sensitivity to the minimum performance rank for funds in a small family. Given Hypothesis 5, we expect a large and positive coefficient for $Low\_Famsize \times Min\_Rank$, but a small coefficient for $High\_Famsize \times Min\_Rank$.

Table 1.5 presents results using fund family size as a proxy for the fund's ambiguity level. Columns 1 and 2 report results using raw returns and the Carhart (1997) alpha as performance measures, respectively. Consistent with our expectation, in Column 1, the coefficients for the three interaction terms $Low\_Famsize \times Min\_Rank$, $Mid\_Famsize \times Min\_Rank$ and $High\_Famsize \times Min\_Rank$ decrease monotonically from 0.071 (significant at the 1% level) to 0.025 (significant at the 1% level). This means that the flow sensitivity to minimum performance rank for funds in a large family is nearly three times that for funds belonging to a small family. The F-test statistic rejects the null that the coefficients of $Low\_Famsize \times Min\_Rank$,

$Mid\_Famsize \times Min\_Rank$ and $High\_Famsize \times Min\_Rank$ are equal. In Column 2, when performance is measure in terms of the Carhart (1997) alpha, the coefficient for $Low\_Famsize \times Min\_Rank$ (0.103) is more than twice the coefficient for $High\_Famsize \times Min\_Rank$ (0.043), both significant at the 1% level.

As seen from the estimated coefficients for the variable, $D\_FamilySize$, family size has a positive and significant impact on fund flows, consistent with the findings of Sirri and Tufano (1998). This suggests that funds belong to bigger families experience a faster growth in assets. As noted by Gallaher et al. (2005), strategic decisions regarding advertising, and distribution channels are made at the fund family level. Funds that belong to a large family have more resources in terms of both management and reputation, allowing them to grow at a faster rate. In conclusion, the results in Table 1.5 are consistent with Hypothesis 5 that funds belonging to smaller families appear more ambiguous to investors.

1.5.5 The Role of Advertising

In previous subsections we examined the impact of proxies for a fund's ambiguity level on the behavior of investors. The results suggest that investors' marginal sensitivity to a fund's historical minimum performance is increasing in the perceived ambiguity of a fund. Of course, the additional flow-performance sensitivity can be costly from a fund's standpoint. We now focus on the possible ways a fund may be able to reduce its ambiguity level from the perspective of fund investors. According to Hypothesis 6 a fund's marketing effort can help reduce investors' ambiguity towards the fund when making decisions. In our empirical test of this hypothesis, we measure marketing effort using the amount of 12b-1 fees borne by a fund.

In order to test the above hypothesis, we estimate the following regression model in the subsample of retail funds:

$$\begin{aligned} Flow_{i,t} =& a + b_{11} Low\_Exp_{i,t} \times Perf\_1yr_{i,t} + b_{12} Mid\_Exp_{i,t} \times Perf\_1yr_{i,t} \\ &+ b_{13} High\_Exp_{i,t} \times Perf\_1yr_{i,t} + b_{21} Low\_Exp_{i,t} \times Perf\_3yr_{i,t} \\ &+ b_{22} Mid\_Exp_{i,t} \times Perf\_3yr_{i,t} + b_{23} High\_Exp_{i,t} \times Perf\_3yr_{i,t} \\ &+ b_{31} Low\_Exp_{i,t} \times Perf\_5yr_{i,t} + b_{32} Mid\_Exp_{i,t} \times Perf\_5yr_{i,t} \\ &+ b_{33} High\_Exp_{i,t} \times Perf\_5yr_{i,t} + c_1 Low\_Exp_{i,t} \times Min\_Rank_{i,t} \\ &+ c_2 Mid\_Exp_{i,t} \times Min\_Rank_{i,t} + c_3 High\_Exp_{i,t} \times Min\_Rank_{i,t} \\ &+ Expense_{i,t} + Controls + \epsilon_{i,t}. \end{aligned} \tag{1.23}$$

To emphasize the particular role of advertising in reducing fund ambiguity, we also study the effect of the expense ratio and non-12b-1 expenses, defined as expense ratio minus 12b-1 fees, using the same model. In Equation (1.23), $Low\_Exp_{i,t}$ is a dummy variable that equals one if fund i's corresponding type of fees (12b-1, non-12b-1 or expense ratio) falls in the bottom tercile in quarter t, $Mid\_Exp_{i,t}$ is a dummy variable that equals one if it belongs to the medium tercile, and $High\_Exp_{i,t}$ is a dummy variable that equals one if the fund is in the top tercile in terms of the expenses. The variable $Expense_{i,t}$ is the 12b-1, non-12b-1 or expense ratio depending on which type of fees is under investigation.

We regress quarterly flows on the interaction of the three dummy variables with $Perf\_1yr$, $Perf\_3yr$, $Perf\_5yr$ and Min_Rank, respectively, for all three types of fees. For example, when we study the effect of 12b-1 fees, the coefficients on the interaction terms represent the marginal flow sensitivity to a certain performance horizon or to the minimum performance rank for funds with the low, medium and high 12b-1 fees, respectively. According to Hypothesis 6, when focusing on 12b-1 fees, we expect a large and positive coefficient for $Low\_Exp \times Min\_Rank$, which captures the additional flow sensitivity to the minimum performance rank for funds with low 12b-1 fees.

Table 1.6 reports results based on the above test. Columns 1 and 2 of the table report results of the effect of 12b-1 fees on the flow-performance relationship when performance is measured using raw returns and the Carhart (1997) alpha, respectively. In Column 1, the coefficient for $Low\_Exp \times Min\_Rank$ is 0.058, and is significant at the 1% level. The coefficient for $High\_Exp \times Min\_Rank$, however, is only 0.021 and it is insignificant. This suggests that moving from the bottom tercile of 12b-1 expenditures funds to the top tercile, investors' flow sensitivity to the minimum performance rank is reduced by 64%. The F-test statistic rejects the null that the coefficients of $Low\_Exp \times Min\_Rank$, $Mid\_Exp \times Min\_Rank$ and $High\_Exp \times Min\_Rank$ are equal. Column 2 reports qualitatively similar flow-performance sensitivity results when performance is measured using the Carhart (1997) alphas.

Columns 3 and 4 present results for the non-12b-1 expenditures for both measures of performance. In Columns 3 and 4, the three coefficients of interest display, surprisingly, an increasing pattern, suggesting that the higher the non-12b-1 fees charged by funds, the more ambiguous they appear to their investors. Columns 5 and 6 present results for the expense ratio and we observe similar patterns as for the non-12b1 fees. We conjecture that this set of results may be attributed to the fact that high expense ratio funds attract relatively less sophisticated investors, since the sophisticated investors would presumably avoid high expense funds. Thus, we observe stronger ambiguity aversion among the investors in funds with high expense ratios or non-12b-1 fees.

In conclusion, the evidence presented in this subsection confirms that 12b-1 fees that are spent on marketing and advertising do help reduce the ambiguity of funds from the investors' perspective.

1.6 Contrast between the Response of Ambiguity Averse Investors and Bayesian Investors

Finally, we develop a test to distinguish ambiguity averse investor behavior from the Bayesian learning benchmark. Following earlier discussion, we argue that a fund's performance volatility is another proxy for the fund's ambiguity level in addition to the flow volatility and family size. The more volatile the fund's past performance, the harder it is for investors to learn about the fund's future performance. Thus, a fund with a higher degree of performance volatility is more ambiguous from an investor's perspective. Accordingly, we expect to observe greater marginal flow sensitivity to the minimum performance rank for such funds.

In a recent study, Huang et al. (2012) hypothesize that the volatility of funds' past performance should have a dampening effect on flow-performance sensitivity if investors update their beliefs in a Bayesian manner. As we note below, the two seemingly contradictory hypotheses can in fact be reconciled.

We want to first document the two effects by performing two separate tests. We apply the following model to test our baseline hypothesis of ambiguity aversion:

$$\begin{aligned} Flow_{i,t} =& a + b_1 Min\_Ind_{i,t} \times Perf\_3yr_{i,t} + e_1 Low\_3yr_{i,t} + e_2 Mid\_3yr_{i,t} + e_3 High\_3yr \\ &+ Controls + \epsilon_{i,t}. \end{aligned} \tag{1.24}$$

For the purpose of comparison, we adopt similar measures and time periods as in Huang et al. (2012). For example, in this section, we expand our sample to include both retail and institutional funds. Also, we focus on a fund's previous 3 years performance ($Perf\_3yr$), defined as a fund's performance measured by its raw return rankings within its objective category over the past 36 months. We define the following fractional performance rankings over the low, medium and high performance ranges[12]. The fractional rank for funds in the bottom performance quintile (Low_3yr) is $Min(Perf\_3yr, 0.2)$, in the three medium quintiles (Mid_3yr) is $Min(0.6, Perf\_3yr - Low\_3yr)$, and in the top quintiles ($High\_3yr$) is $Perf\_3yr - Mid\_3yr - Low\_3yr$. We also include the identical set of control variables as in our baseline model specified in Equation (1.17).

In the above specification, the variable $Min\_Ind_{i,t}$ is a dummy variable that equals one if fund i's 3-year performance happens to be the worst among 1-year, 3-year and 5-year performance measures in quarter t. The coefficient b_1 for the interaction term $Min\_Ind_{i,t} \times Perf\_3yr$ captures the additional sensitivity to the 3-year

[12]See, for example, Huang et al. (2007, 2012) and Sirri and Tufano (1998).

performance, if it happens to be the worst among the three performance measures. Thus, the coefficient b_1 captures the ambiguity aversion effect. We expect b_1 to be positive and significant. The results are presented in Column 1 of Table 1.7. The coefficient b_1 is estimated to be 0.053 which is highly significant in both statistical and economic terms. It is worth noting that we do observe the convexity in the flow-performance relationship reflected in the respective coefficients for the performance ranges (Low_3yr, Mid_3yr and $High\_3yr$). In particular, the coefficient for the high performance range (0.407) is nearly 6 times the coefficient for the low performance range (0.070) suggesting a convex flow-performance relationship.

Next, we replicate the Huang et al. (2012) results using the following test:

$$\begin{aligned} Flow_{i,t} =& a + cVol_{i,t} \times Perf\_3yr_{i,t} + e_1 Low\_3yr_{i,t} + e_2 Mid\_3yr_{i,t} + e_3 High\_3yr_{i,t} \\ &+ Controls + \epsilon_{i,t}. \end{aligned} \tag{1.25}$$

The variable $Vol_{i,t}$ is the fund i's previous 36 months' raw return volatility in quarter t. If the dampening effect of performance volatility on the flow-performance relationship exists, we expect to see a significant negative coefficient for the interaction term, $Vol \times Perf\_3yr$. As shown in Column 2 of Table 1.7, this coefficient is estimated to be -0.777, which is significant at the 1% level. All of the other coefficients are also qualitatively similar to the values reported by Huang et al. (2012).

Finally, it is of interest to show that ambiguity aversion and the dampening effect of performance volatility could co-exist. We distinguish our ambiguity aversion phenomenon using the following model:

$$
\begin{aligned}
Flow_{i,t} =& a + b_2 Min\_Ind_{i,t} \times Vol_{i,t} \times Perf\_3yr_{i,t} + cVol_{i,t} \times Perf_3yr_{i,t} + Low\_3yr_{i,t} \\
&+ e_2 Mid\_3yr_{i,t} + e_3 High\_3yr_{i,t} + fVol_{i,t} + Controls + \epsilon_{i,t}. \qquad (1.26)
\end{aligned}
$$

Here b_2 captures the effect of performance volatility on 3-year performance if it happens to be the minimum performance rank and c captures the impact of performance volatility on the 3-year performance, on average. Under ambiguity aversion, we expect b_2 to be positive and significant, since past performance volatility is expected to increase the fund's ambiguity level. However, if investors are Bayesian learners as modeled in Huang et al. (2012), we also expect to observe a negative value for the coefficient c, since high signal noise should dampen flow sensitivity, in general.

The results from estimating Equation (1.26) are reported in Column 3 of Table 1.7. We find that the coefficient b_2 has a value equal to 1.143 while the coefficient c equals -0.906, and both coefficients are significant at the 1% level. The coefficients imply, in economic terms, a 1 unit increase in performance volatility will **decrease** sensitivity to performance by $0.906 \times 1 = 0.906$, on average. However, it will **increase** sensitivity to the minimum performance rank by $(1.143 - 0.906) \times 1 = 0.237$. Here we observe clearly that the presence of ambiguity aversion leads to a net increase in the overall flow-performance sensitivity despite the dampening effect of performance volatility.

These findings show that the dampening effect and ambiguity aversion are not contradictory to each other and may actually co-exist. This is intuitive, since it is reasonable to conjecture that there are two distinct types of fund investors.

The first type is the sophisticated investors who have a better understanding of the mutual fund industry. Upon observing past fund performance, these investors face less ambiguity and are able to update their beliefs in a manner similar to Bayesian rules. The other type of investors is less sophisticated. They have access to the same past performance information, but cannot update their beliefs about the fund manager's skill in a Bayesian fashion. They display aversion to ambiguity, as shown by the positive and significant coefficient on the minimum performance rank variable.

We further examine the above intuition by including a dummy variable $Inst_i$, which equals 1 if a fund is only open to institutional investors. We develop the following test to separate the two types of fund investors with different level of sophistication:

$$\begin{aligned} Flow_{i,t} =& a + b_2 Min\_Ind_{i,t} \times Vol_{i,t} \times Perf\_3yr_{i,t} + d_1 Inst_i \times Vol_{i,t} \times Min\_Ind_{i,t} \\ & \times Perf\_3yr_{i,t} + cVol_{i,t} \times Perf\_3yr_{i,t} + d_2 Inst_i \times Vol_{i,t} \times Perf\_3yr_{i,t} \\ & + d_3 Inst_i + e_1 Low\_3yr_{i,t} + e_2 Mid\_3yr_{i,t} + e_3 High\_3yr_{i,t} + Controls + \epsilon_{i,t}. \end{aligned}$$

We expect to observe a stronger volatility dampening effect in institutional fund flows and a weaker ambiguity aversion effect. Similar to Equation (1.26), the coefficient, b_2, captures ambiguity aversion and the coefficient, c captures the dampening effect. The coefficient d_1 captures the marginal ambiguity aversion for the institutional investors and d_2 captures the marginal volatility dampening effect for institutional investors. We expect d_1 to be negative and d_2 to be positive. The results of this test are presented in Column 4 of Table 1.7. As expected, the coefficient d_1 is estimated to be -0.237

and d_2 equals -0.405. This implies that a 1 unit increase in volatility will **decrease** flow sensitivity to performance by $0.885 + 0.405 = 1.29$ for institutional investors. However, it will **decrease** sensitivity by only $1.29 - (1.181 - 0.237) = 0.346$ if the performance measure is the minimum among 1-year, 3-year and 5-year performance ranks. Thus, for institutional funds, regardless of whether the 3-year performance is the minimum performance or not, there is a volatility dampening effect which dominates the impact of ambiguity aversion. This is consistent with our intuition that for sophisticated investors, we should see a stronger dampening effect of performance volatility on flow-performance sensitivity.

In order to more closely match the results of Huang et al. (2012) we repeat our tests using an identical sample period: 1983-2006. We find qualitatively similar results and report them in Columns 5-8 in Table 1.7. In conclusion, we find that in a population with a larger fraction of naive investors, ambiguity aversion behavior dominates the dampening effect of performance volatility on the flow-performance relation. Conversely, in a population with a greater fraction of sophisticated investors, the volatility dampening effect dominates ambiguity aversion.

1.7 Robustness

In previous tests we investigated the asymmetric sensitivity of fund flows to performance measured over 1-, 3-, and 5-year horizons caused by investor ambiguity aversion. In particular, we find that the more ambiguous a fund appears to its investors, the greater the flow-performance sensitivity to the minimum performance over the three time horizons. Of course, investors have access to additional information beyond the above three measures of performance. Other information available to investors includes the performance of the fund since inception and the Morningstar fund ratings. In fact, previous studies show that investor flows do respond to Morningstar ratings.[13] A natural question is whether the additional sensitivity to minimum performance documented in our tests is simply due to the omission of long term performance measures and/or Morningstar ratings. To address this concern we reexamine the baseline model in Equation (1.17) while controlling for the funds' performance since inception and the Morningstar fund ratings.

We measure a fund's performance since its inception in the same manner as its 1, 3 and 5 years performance. We rank funds based on their average monthly raw returns since inception within their objective category. The resulting ranking is a number between 0 and 1. In terms of the Morningstar fund ratings, according to Sharpe (1997), Morningstar ranks funds within four asset classes before 1996 and within smaller categories thereafter. We replicate the latter rating scheme because it covers most of our sample period. Since we do not have access to the criteria that Morningstar uses to classify funds into style categories, we continue to employ the classifications used for our primary tests. Morningstar rates all funds based on their 3-, 5- and 10-year return and risk, respectively and then a weighted overall rating is determined. We follow the procedure described in Nanda et al. (2004b) to replicate the funds' Morningstar fund ratings.

---

[13]See, e.g., Del Guercio and Tkac (2002).

Our robustness tests are based on the following specification:

$$\begin{aligned} Flow_{i,t} =& a + b_1 Perf\_1yr_{i,t} + b_2 Perf\_3yr_{i,t} + b_3 Perf\_5yr_{i,t} + b_4 Perf\_Incep_{i,t} + \\ & b_5 MorningstarRating_{i,t} + cMin\_Rank_{i,t} + Controls + \epsilon_{i,t}, \end{aligned} \tag{1.27}$$

The *Morningstar Rating* represents fund i's rating based on the Morningstar fund rating scheme in quarter t. The variable $Perf\_Incep_{i,t}$ captures a fund i's performance since inception in quarter t.

Table 1.8 presents results of the robustness tests. We note that the coefficient of Min_Rank is consistently positive and significant for the sample of retail funds when we control for performance since inception and/or Morningstar rating, while it is insignificant for the institutional sample. Moreover, the coefficient of $Perf\_Incep$ is negative and insignificant in all columns. However, the coefficient for *Morningstar Rating* is consistently positive and significant. For example, based on results in Column 5, if a retail fund's Morningstar rating increases by 1 star, there is a corresponding increase of 1.7% in terms of fund flows. By contrast, for an institutional fund, a 1 star increase in the Morningstar rating is associated with a fund flow increase of only 0.8%, based on the results in Column 6. These results suggest that the Morningstar rating is an important signal for retail investors when making investment decisions. On the other hand, institutional investors seem to rely to a much lesser extent on this rating.

One of the most well-known findings in mutual fund literature is the convex relationship between investor flows and past fund performance. Previous studies (e.g. Sirri and Tufano (1998), Chevalier and Ellison (1997)) find that investor flows respond strongly to good past performance while being less sensitive to bad performance. As we previously note (see footnote 5), in the context of our model, the flow-performance convexity is potentially complementary to the effect of investor ambiguity aversion. We now show that our baseline result is robust after controlling for the convex flow-performance relationship. Table 1.9 presents results of this test. As in the baseline model in Equation (1.17), we regress quarterly flow on past performance measured over 1-, 3-, and 5-year. We adopt fractional ranks for each of the 1-, 3-, and 5-year performance (same as Low_3yr, Mid_3yr and $High\_3yr$ defined in Equation (1.24) to capture the convexity in flow-performance relationship. The results presented in Table 1.9 confirm that the additional sensitivity to the minimum performance measure remains positive and significant in the retail fund sample even after controlling for the convexity of the flow-performance relationship.

Another possible concern is that some time variant factors, such as a change in the fund manager, may cause differential flow sensitivity to performance measured over different time horizons. To deal with this issue, in unreported tests, we consider various performance measures calculated over a fixed time horizon (1-, 3- or 5-year). The various performance measures considered include the CAPM alpha, Fama-French 3-factor alpha, Carhart 4-factor alpha, raw return, and the Morningstar fund rating. The various performance measures calculated over the identical time horizon may be viewed as multiple performance signals. We find that our results are qualitatively unchanged. In particular, we find that for each of the three time horizons considered, retail fund flows display additional sensitivity to the minimum performance measure.

In conclusion, the results of the robustness tests confirm that mutual fund flows display an additional sensitivity to the minimum performance rank even after controlling for performance since inception, the funds' Morningstar star ratings, and convexity in the flow-performance relation.

1.8 Concluding Remarks

This paper presents a model of an ambiguity averse mutual fund investor who faces multiple signals about the performance of the fund. The model implies that an ambiguity averse investor, in attempting to learn about manager skill, always puts additional weight on the worst signal. We empirically test the key implications of the model by using a fund's past performance measured over multiple time horizons, as multiple signals about manager skill observed by investors.

Our study provides novel evidence on the role of ambiguity aversion in determining the response of mutual fund investors to historical fund performance information. We find that consistent with our model, fund flows display heightened sensitivity to the minimum performance measure. Further, we observe this ambiguity averse behavior only among retail fund investors. By contrast, fund flows in institutional funds appear to be more consistent with standard Bayesian learning behavior. We also find that funds with more frequent/aggressive strategy changes, funds that have more volatile past flows, and funds belonging to smaller fund families appear more ambiguous to investors while advertising expenditure appears to reduce the degree of ambiguity perceived by investors. Furthermore, we distinguish between the effect of increased performance volatility on ambiguity averse investors and on investors whose behavior is more consistent with Bayesian learning. We find that fund volatility increases ambiguity averse investors' response to the minimum performance measure while it dampens the Bayesian investors' sensitivity to past performance in general. Taken together, these results suggest that fund investor behavior is best characterized as reflecting both Bayesian learning and ambiguity aversion.

Table 1.1: Summary Statistics of the Equity Mutual Fund Sample

	Share Classes		
	All Funds	Retail Shares	Institutional Shares
Total Fund Number	7,020	5,781	1,239
Average Fund Number	2,847	2,359	466
Flow (in % per quarter)	1.51	1.32	2.33
Age (in years)	11.60	12.07	8.67
TNA (in millions)	678.06	739.36	330.49
Expense Ratio (in %)	1.41	1.49	0.96
Total Expense (in %)	1.61	1.73	0.99
12b-1 Fee (in %)	0.56	0.57	0.20
Raw Return (in % per month)	0.73	0.72	0.79
Vol. of Raw Return (in % per month)	4.60	4.59	4.66

Note: This table reports the time-series averages of quarterly cross-sectional averages of fund characteristics for the period 1993-2011. TNA is the total net assets. Flow is the percentage change in TNA. Expense Ratio is the total quarterly management and administrative expenses divided by average TNA. Total Expense is estimated as expense ratio plus 1/7 of maximum front-end load. 12-b1 Fee are fees paid by the fund out of fund assets to cover marketing expenses, distribution expenses and sometimes shareholder service expenses. The 12-b1 fees are only available since 1992. Raw Return is the average monthly raw return during the prior 12 months and Vol. of Raw Return is the corresponding standard deviation. The statistics are reported for all funds (i.e., share classes), funds open only to retail investors, and funds available only to institutional investors.

Table 1.2: Ambiguity Aversion of Retail and Institutional Investors

Performance Measured by	Raw Return		4-Factor Alpha	
	Retail	Institutional	Retail	Institutional
Perf_1yr	0.050***	0.055***	0.036***	0.045***
	(8.33)	(3.55)	(7.11)	(5.23)
Perf_3yr	0.039***	0.035**	0.049***	0.067***
	(6.78)	(2.41)	(10.55)	(5.78)
Perf_5yr	0.011***	0.042***	-0.015**	0.012
	(2.64)	(3.16)	(-2.50)	(1.12)
Min_Rank	0.048***	0.017	0.057***	0.010
	(7.02)	(0.89)	(7.74)	(0.85)
Age	-0.017***	-0.027***	-0.010***	-0.018***
	(-9.79)	(-6.94)	(-10.41)	(-4.80)
Previous Quarter Flow	0.177***	0.137***	0.209***	0.167***
	(7.52)	(5.71)	(8.25)	(5.08)
Size	-0.004***	-0.011***	-0.004***	-0.010***
	(-6.77)	(-7.42)	(-8.45)	(-6.49)
Volatility	-0.394***	-0.622***	0.322**	0.293***
	(-3.98)	(-2.89)	(2.08)	(1.31)
Total Expense	-0.434***	-1.912***	-0.278**	-1.947***
	(-2.72)	(-3.42)	(-2.10)	(-2.85)
Category Flow	0.288***	0.060	0.208***	0.028
	(3.46)	(0.38)	(3.52)	(0.20)
Intercept	0.020***	0.089***	-0.011	0.040***
	(3.06)	(5.69)	(-1.50)	(3.49)

Note: This table examines the ambiguity aversion behavior of both retail and institutional investors during period 1993-2011. The sample is divided into retail shares and institutional shares. Each quarter, funds are assigned ranks between zero and one according to their performance during the past 12 months (Perf_1yr), 36 months (Perf_3yr) and 60 months (Perf_5yr) respectively. Performance is measured by the average monthly raw returns or the Carhart (1997) 4-factor alphas. Funds are ranked based on raw returns within the funds' objective category while ranks based on alphas are computed across all funds in the sample. Min_Rank is defined as the minimum performance rank among three periods. A linear regression model is estimated by regressing quarterly flows on funds' three performance ranks (Perf_1yr, Perf_3yr and Perf_5yr) and the minimum rank (Min_Rank). The control variables include fund age (Age), defined as log (1+age), quarterly flow in previous quarter (Previous Quarter Flow), logarithm of lagged fund TNA (Size), volatility of monthly raw return in prior 12 months (Volatility), lagged total expense (Total Expense), and aggregate flow to the fund' objective category (Category Flow). Time-series averages of coefficients and the Newey-West t-statistics (in parentheses) are reported. The symbols *, **, and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 1.3: Strategy Changes as a Proxy for Fund Ambiguity Level

Panel A: Strategy Changes Proxied by Change in Factor Loadings						
Performance Measured by	Raw Return			4-Factor Alpha		
	Low	Mid	High	Low	Mid	High
Perf_1yr	0.053***	0.046***	0.045***	0.038***	0.031***	0.035***
	(5.91)	(6.89)	(7.27)	(5.01)	(4.43)	(5.92)
Perf_3yr	0.038***	0.030***	0.036***	0.053***	0.046***	0.039***
	(6.50)	(4.55)	(4.58)	(5.79)	(7.30)	(4.25)
Perf_5yr	0.032***	0.015**	0.008	-0.005	-0.020**	-0.012**
	(6.31)	(2.55)	(1.33)	(-0.73)	(-2.06)	(-2.33)
Min_Rank	0.011	0.044***	0.061***	0.033***	0.053***	0.061***
	(1.13)	(4.03)	(6.12)	(3.04)	(4.56)	(4.67)
Age	-0.014***	-0.016***	-0.015***	-0.007***	-0.009***	-0.011***
	(-6.88)	(-8.44)	(-9.84)	(-6.86)	(-5.89)	(-6.19)
Previous Quarter Flow	0.189***	0.260***	0.253***	0.217***	0.291***	0.279***
	(7.64)	(9.19)	(9.70)	(8.31)	(8.62)	(9.03)
Size	-0.004***	-0.005***	-0.006***	-0.004***	-0.004***	-0.005***
	(-6.14)	(-8.05)	(-4.99)	(-5.28)	(-5.80)	(-4.78)
Volatility	-0.378***	-0.368***	-0.370***	0.628**	0.253	0.140
	(-3.07)	(-2.73)	(-3.45)	(2.49)	(1.02)	(1.08)
Total Expense	-1.085***	-0.869***	-0.092	-0.462	-0.576**	-0.174
	(-3.36)	(-3.49)	(-0.58)	(-1.61)	(-2.44)	(-1.34)
Category Flow	0.364***	0.130	0.225	0.131	0.036	0.323**
	(3.71)	(1.41)	(1.60)	(1.58)	(0.43)	(2.35)
Intercept	0.023**	0.027***	0.016*	-0.026*	-0.000	0.002
	(2.13)	(3.16)	(1.87)	(-1.91)	(-0.03)	(0.20)

Table 3 Continued

Panel B: Panel B: Strategy Changes Proxied by Fund R-Squared Values

Performance Measured by	Raw Return			4-Factor Alpha		
	Low	Mid	High	Low	Mid	High
Perf_1yr	0.046***	0.058***	0.046***	0.032***	0.049***	0.020**
	(10.39)	(6.53)	(6.37)	(5.54)	(6.08)	(2.29)
Perf_3yr	0.031***	0.037***	0.034***	0.034***	0.051***	0.056***
	(3.76)	(6.28)	(6.76)	(4.19)	(8.46)	(4.79)
Perf_5yr	0.005	0.020***	0.026***	-0.014**	-0.009	-0.021**
	(0.96)	(3.61)	(5.01)	(-2.51)	(-1.53)	(-2.37)
Min_Rank	0.056***	0.028***	0.021	0.063***	0.027**	0.054***
	(6.01)	(3.04)	(1.54)	(5.52)	(2.62)	(3.97)
Age	-0.013***	-0.019***	-0.014***	-0.007***	-0.013***	-0.007***
	(-6.93)	(-11.25)	(-5.63)	(-5.07)	(-7.76)	(-5.22)
Previous Quarter Flow	0.274***	0.228***	0.199***	0.298***	0.286***	0.201***
	(11.00)	(8.30)	(7.75)	(11.83)	(8.47)	(9.72)
Size	-0.005***	-0.004***	-0.003***	-0.005***	-0.004***	-0.004***
	(-5.17)	(-6.73)	(-4.70)	(-5.85)	(-6.03)	(-5.73)
Volatility	-0.409***	-0.388***	-0.398***	0.188	0.446**	0.674*
	(-3.64)	(-3.41)	(-2.93)	(1.44)	(2.18)	(1.93)
Total Expense	-0.169	-0.900***	-0.676**	-0.151	-0.493*	-0.568*
	(-1.31)	(-3.15)	(-2.62)	(-1.30)	(-1.73)	(-1.85)
Category Flow	0.291**	0.150*	0.398***	0.207*	0.083	0.160
	(2.22)	(1.89)	(5.28)	(1.67)	(1.11)	(1.35)
Intercept	0.013*	0.034***	0.024**	-0.009	-0.003	-0.012
	(1.86)	(3.51)	(2.18)	(-0.97)	(-0.33)	(-0.70)

Note: This table presents results of tests that use a measure of a fund's strategy changes as a proxy for the fund's ambiguity level. The sample includes only the retail funds. Each quarter, the sample is divided into three groups, Low, Mid, and High based on the value of the strategy change measure. Panel A reports results when strategy changes are measured by the average absolute change in the fund's (Carhart (1997) model) factor loadings between the prior 1-30 and 31-60 months. In Panel B, fund strategy changes are measured by a fund's R-squared value over its lifetime based on the Carhart (1997) 4-factor model. The table presents coefficient estimates obtained by regressing quarterly fund flows on the funds' 1-year, 3-year, and 5-year performance ranks (Perf_1yr, Perf_3yr and Perf_5yr) and the minimum rank (Min_Rank). Performance is measured by the average monthly raw returns or the Carhart (1997) 4-factor alphas. The control variables are the same as the ones in Table 2. Time-series averages of coefficients and the Newey-West t-statistics (in parentheses) are reported. The symbols *, **, and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 1.4: Flow Volatility as a Proxy for Fund Ambiguity Level

Performance Measured by	Raw Return	4-Factor Alpha
Low_Flowvol×Perf_1yr	0.046***	0.033***
	(9.94)	(6.84)
Mid_Flowvol×Perf_1yr	0.050***	0.044***
	(7.35)	(6.35)
High_Flowvol×Perf_1yr	0.072***	0.045***
	(6.78)	(4.50)
Low_Flowvol×Perf_3yr	0.026***	0.028***
	(5.63)	(5.60)
Mid_Flowvol×Perf_3yr	0.038***	0.052***
	(6.04)	(8.53)
High_Flowvol×Perf_3yr	0.049***	0.067***
	(4.70)	(6.36)
Low_Flowvol×Perf_5yr	0.036***	0.016***
	(9.65)	(2.56)
Mid_Flowvol×Perf_5yr	0.008	-0.022***
	(1.59)	(-2.90)
High_Flowvol×Perf_5yr	-0.035***	-0.064***
	(-4.06)	(-7.07)
Low_Flowvol×Min_Rank	-0.002	0.007
	(-0.54)	(1.08)
Mid_Flowvol×Min_Rank	0.041***	0.043***
	(5.10)	(4.31)
High_Flowvol×Min_Rank	0.110***	0.122***
	(8.12)	(7.41)
Flow_Vol	0.060***	0.034***
	(6.36)	(3.11)
Age	-0.008***	-0.004***
	(-6.20)	(-3.43)
Previous Quarter Flow	0.158***	0.192***
	(7.11)	(7.74)
Size	-0.003***	-0.003***
	(-5.08)	(-6.44)
Volatility	-0.527***	0.135
	(-5.08)	(0.92)
Total Expense	-0.292**	-0.187
	(-2.05)	(-1.58)
Category Flow	0.281***	0.201***
	(3.89)	(3.39)
Intercept	-0.006	-0.026***
	(-1.12)	(-3.48)

Note: This table presents results of tests that use flow volatility as a proxy for a fund's ambiguity level. The sample includes only the retail funds. Flow_Vol is the standard deviation of quarterly flows over prior 12 quarters. Each quarter, a dummy variable Low_Flowvol equals one if the flow volatility falls in the bottom tercile, a dummy variable Mid_Flowvol equals one if the flow volatility belongs to the medium tercile and a dummy variable High_Flowvol equals one if the flow volatility is in the top tercile. A linear regression is performed A linear regression is performed as in Table 1.3. The control variables are the same as the ones in Table 1.2. Time-series averages of coefficients and the Newey-West t-statistics (in parentheses) are reported. The symbols *, **, and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 1.5: Fund Family Size as a Proxy for Fund Ambiguity Level

Performance Measured by	Raw Return	4-Factor Alpha
Low_Famsize×Perf_1yr	0.050***	0.019**
	(7.59)	(2.01)
Mid_Famsize×Perf_1yr	0.045***	0.042***
	(6.00)	(5.54)
High_Famsize×Perf_1yr	0.049***	0.032***
	(6.12)	(7.12)
Low_Famsize×Perf_3yr	0.028***	0.026
	(3.12)	(1.41)
Mid_Famsize×Perf_3yr	0.030***	0.061***
	(3.48)	(7.67)
High_Famsize×Perf_3yr	0.043***	0.040***
	(6.03)	(7.73)
Low_Famsize×Perf_5yr	0.002	-0.028***
	(0.29)	(-3.24)
Mid_Famsize×Perf_5yr	0.018**	-0.016***
	(2.18)	(-3.28)
High_Famsize×Perf_5yr	0.027***	0.010
	(4.93)	(1.44)
Low_Famsize×Min_Rank	0.071***	0.103***
	(5.14)	(4.37)
Mid_Famsize×Min_Rank	0.056***	0.039***
	(5.94)	(3.36)
High_Famsize×Min_Rank	0.025***	0.043***
	(3.25)	(5.34)
D_FamilySize	0.005***	0.004
	(2.78)	(1.13)
Age	-0.017***	-0.010***
	(-8.56)	(-9.95)
Previous Quarter Flow	0.175***	0.207***
	(7.52)	(8.25)
Size	-0.006***	-0.006***
	(-5.78)	(-7.97)
Volatility	-0.409***	0.310**
	(-4.11)	(2.02)
Total Expense	-0.486***	-0.331**
	(-3.21)	(-2.65)
Category Flow	0.305***	0.225***
	(3.56)	(4.01)
Intercept	0.025***	-0.005
	(3.86)	(-0.58)

Note: This table presents results of tests using family size as a proxy for a fund's ambiguity level. The sample includes only retail funds. Family size is total net asset of all fund share classes that belong to the same fund family. Each quarter, a dummy variable Low_FamSize equals one if fund i belongs to a family whose size falls into the bottom tercile, a dummy variable Mid_FamSize equals one if family size belongs to the medium tercile, and a dummy variable High_FamSize equals one if the family size is in the top tercile. D_FamliySize is a dummy variable that equals one if family size is above the median value for that quarter. A linear regression is performed as in Table 1.3. The control variables are the same as the ones in Table 1.2. Time-series averages of coefficients and the Newey-West t-statistics (in parentheses) are reported. The symbols *, **, and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 1.6: Impact of Fund Advertising on Funds' Ambiguity

	12b-1		Non-12b-1		Expense Ratio	
Performance Measure by	Raw Return	Alpha	Raw Return	Alpha	Raw Return	Alpha
Low_Exp×Perf_1yr	0.045***	0.020**	0.063***	0.042***	0.067***	0.051***
	(4.97)	(2.20)	(6.03)	(6.82)	(7.53)	(8.52)
Mid_Exp×Perf_1yr	0.038***	0.028***	0.048***	0.036***	0.044***	0.030***
	(4.70)	(4.64)	(6.32)	(6.08)	(7.81)	(5.33)
High_Exp×Perf_1yr	0.064***	0.049***	0.043***	0.030***	0.040***	0.031***
	(5.94)	(5.52)	(8.25)	(4.03)	(5.78)	(3.57)
Low_Exp×Perf_3yr	0.030***	0.048***	0.047***	0.045***	0.034***	0.041***
	(3.78)	(7.42)	(6.50)	(7.27)	(4.81)	(6.15)
Mid_Exp×Perf_3yr	0.052***	0.032***	0.027***	0.045***	0.042***	0.049***
	(5.87)	(4.80)	(4.12)	(6.97)	(5.82)	(8.22)
High_Exp×Perf_3yr	0.056***	0.076***	0.042***	0.052***	0.039***	0.051***
	(4.46)	(6.03)	(4.92)	(7.77)	(4.59)	(6.37)
Low_Exp×Perf_5yr	0.012	-0.005	0.032***	0.013*	0.043***	0.018**
	(1.55)	(-0.55)	(6.17)	(1.93)	(8.45)	(2.44)
Mid_Exp×Perf_5yr	0.006	-0.014*	0.015***	-0.017**	0.009	-0.024***
	(0.71)	(-1.83)	(2.84)	(-2.23)	(1.35)	(-2.99)
High_Exp×Perf_5yr	0.003	-0.023*	-0.009	-0.034***	-0.020***	-0.036***
	(0.57)	(-1.95)	(-1.21)	(-5.14)	(-3.37)	(-4.32)
Low_Exp×Min_Rank	0.058***	0.069***	0.005	0.020***	0.007	0.013*
	(4.67)	(4.56)	(0.49)	(3.15)	(0.83)	(1.80)
Mid_Exp×Min_Rank	0.048***	0.081***	0.060***	0.058***	0.065***	0.078***
	(3.27)	(5.96)	(6.93)	(6.02)	(7.10)	(7.23)
High_Exp×Min_Rank	0.021	0.024	0.084***	0.092***	0.078***	0.074***
	(1.40)	(1.22)	(6.41)	(5.64)	(6.27)	(4.20)
Expense	-3.587***	-2.599***	0.144	-0.202	-0.258*	-0.445***
	(-6.75)	(-4.74)	(0.98)	(-1.23)	(-1.84)	(-2.72)
Age	-0.023***	-0.014***	-0.018***	-0.010***	-0.018***	-0.012***
	(-8.73)	(-8.04)	(-8.98)	(-10.74)	(-9.10)	(-10.28)
Previous Quarter Flow	0.209***	0.215***	0.176***	0.209***	0.174***	0.207***
	(6.91)	(7.19)	(7.58)	(8.26)	(7.42)	(8.14)
Size	-0.004***	-0.004***	-0.004***	-0.004***	-0.006***	-0.005***
	(-6.67)	(-5.30)	(-5.88)	(-9.89)	(-11.92)	(-10.60)
Volatility	-0.542***	0.129	-0.437***	0.303*	-0.354***	0.355**
	(-5.23)	(0.79)	(-4.05)	(1.92)	(-3.71)	(2.25)
Category Flow	0.216**	0.163***	0.297***	0.207***	0.270***	0.186***
	(2.57)	(3.07)	(3.66)	(3.13)	(3.10)	(2.96)

Table 1.6 Continued

	12b-1		Non-12b-1		Expense Ratio	
Performance Measure by	Raw Return	Alpha	Raw Return	Alpha	Raw Return	Alpha
Intercept	0.051***	0.012	0.013*	-0.014*	0.021***	-0.005
	(6.03)	(1.30)	(1.88)	(-1.71)	(3.26)	(-0.60)

Note: This table examines the impact of various types of fund expenses on funds' ambiguity. The sample includes only retail funds during 1993-2011. Expense Ratio is the total quarterly management and administrative expenses divided by average TNA. 12b-1 fees are the part of total management and administrative expenses that cover distribution expenses and sometimes shareholder services. The non-12b-1 fees are the expense ratio less 12b-1 fees. The 12b-1 fees are only available in CRSP database after 1992. Each quarter, a dummy variable Low_Exp equals one if the fund i's corresponding type of fees falls into the bottom tercile, a dummy variable Mid_Exp equals one if it belongs to the medium tercile, and a dummy variable High_Exp equals one if it is in the top tercile. Expense is 12b-1 fees, non-12b-1 fees or Expense Ratio. A linear regression is performed by regressing quarterly flows on the funds' three performance ranks (Perf_1yr, Perf_3yr and Perf_5yr) and the minimum rank (Min_Rank) interacted with the three dummy variables for each of the three expense types. Performance is measured by the average monthly raw returns or the Carhart (1997) four-factor alphas (Alpha). The control variables are the same as the ones in Table 1.2. Time-series average coefficients and the Newey-West t-statistics (in parentheses) are reported. The symbols *, **, and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 1.7: Contrast between the Response of Ambiguity Averse Investors and Bayesian Investors

	Panel A: 1993-2011				Panel B: 1983-2006			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Min_Ind×Perf_3yr	0.053***	.	.	.	0.052***	.	.	.
	(13.50)	.	.	.	(9.05)	.	.	.
Min_Ind×Vol×Perf_3yr	.	.	1.143***	1.181***	.	.	1.110***	1.151***
	.	.	(8.91)	(8.67)	.	.	(7.48)	(7.61)
Vol×Perf_3yr	.	-0.777***	-0.906***	-0.885***	.	-0.575**	-0.686***	-0.683**
	.	(-2.32)	(-2.98)	(-2.87)	.	(-2.06)	(-2.65)	(-2.60)
Inst×Min_Ind×Vol×Perf_3yr	.	.	.	-0.237	.	.	.	-0.441**
	.	.	.	(-1.60)	.	.	.	(-2.30)
Inst×Vol×Perf_3yr	.	.	.	-0.405**	.	.	.	-0.312
	.	.	.	(-2.26)	.	.	.	(-1.65)
Inst	.	.	.	0.005	.	.	.	-0.004
	.	.	.	(1.50)	.	.	.	(-0.90)
Low_3yr	0.070***	0.147***	0.128***	0.126***	0.057***	0.120***	0.098***	0.097***
	(3.61)	(6.60)	(5.70)	(5.63)	(2.36)	(4.30)	(3.50)	(3.37)
Mid_3yr	0.091***	0.134***	0.137***	0.139***	0.075***	0.104***	0.107***	0.109***
	(18.28)	(8.25)	(9.06)	(9.39)	(11.55)	(8.64)	(9.58)	(9.83)
High_3yr	0.407***	0.428***	0.467***	0.469***	0.358***	0.364***	0.398***	0.402***
	(17.29)	(12.44)	(15.06)	(15.55)	(11.76)	(11.49)	(12.89)	(12.82)
Age	-0.020***	-0.022***	-0.020***	-0.021***	-0.014***	-0.016***	-0.015***	-0.015***
	(-10.01)	(-10.74)	(-10.28)	(-11.11)	(-6.94)	(-7.47)	(-7.21)	(-7.46)
Previous Quarter Flow	0.127***	0.129***	0.127***	0.126***	0.170***	0.174***	0.170***	0.169***
	(6.14)	(6.07)	(6.11)	(6.12)	(6.38)	(6.53)	(6.31)	(6.26)
Size	-0.005***	-0.005***	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***
	(-8.43)	(-8.28)	(-8.67)	(-8.39)	(-11.93)	(-11.55)	(-12.09)	(-12.03)
Vol	-0.668***	-0.204	-0.272	-0.274*	-0.450***	-0.187	-0.232*	-0.206
	(-4.31)	(-1.24)	(-1.65)	(-1.72)	(-3.30)	(-1.41)	(-1.72)	(-1.56)
Total Expense	-0.619***	-0.661***	-0.643***	-0.694***	-0.287**	-0.320**	-0.303**	-0.436***
	(-2.89)	(-3.13)	(-3.02)	(-3.81)	(-2.29)	(-2.55)	(-2.43)	(-3.49)
Category Flow	0.197**	0.189**	0.203**	0.216**	0.424***	0.407***	0.415***	0.418***
	(2.33)	(2.20)	(2.41)	(2.61)	(9.91)	(9.33)	(9.88)	(10.04)
Intercept	0.064***	0.042***	0.042***	0.044***	0.051***	0.042***	0.040***	0.044***
	(5.96)	(5.15)	(5.32)	(5.20)	(4.17)	(3.38)	(3.33)	(3.52)

Note: This table examines the differences in the response of ambiguity averse investors and Bayesian investors during the periods: 1983-2006 and 1993-2011. The sample includes all retail and institutional funds. Perf_3yr is a fractional rank ranging from 0 to 1 reflecting a fund's ranking within an objective category based on the average monthly raw returns

Table 1.7–continued

over prior 36 months. The fractional rank for funds in the bottom performance quintile (Low_3yr) is Min(Perf_3yr,0.2), in the three medium quintiles (Mid_3yr) is Min(0.6, Perf_3yr-Low_3yr), and in the top quintiles (High_3yr) is Perf_3yr-Mid_3yr-Low_3yr. Vol is the standard deviation of raw returns over prior 36 months. The dummy variable Min_Ind equals one if the 3-year performance rank is the minimum among the performance ranks during the 1-year, 3-year and 5-year ranking periods. A dummy variable Inst equals one if the fund is open only to institutional investor. Each quarter, a linear regression is performed by regressing quarterly flows on funds' Perf_3yr and its interaction with Min, Vol, and Inst. The control variables are the same as the ones in Table 1.2. Time-series averages of coefficients and the Newey-West t-statistics (in parentheses) are reported. The symbols *, **, and *** denote significance at the 10%, 5% and 1% level, respectively.

Table 1.8: Robustness Test–Morningstar and Peformance Since Inception

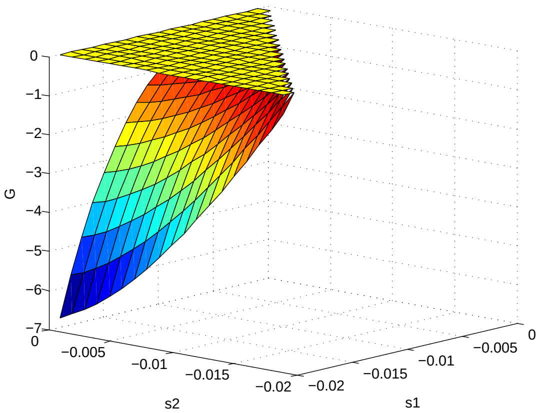
	Retail	Institutional	Retail	Institutional	Retail	Institutional
Perf_1yr	0.050***	0.063***	0.052***	0.068***	0.052***	0.068***
	(10.13)	(3.48)	(9.68)	(3.47)	(9.82)	(3.44)
Perf_3yr	0.032***	0.023	0.014*	0.014	0.014**	0.012
	(4.68)	(1.16)	(1.95)	(0.65)	(2.06)	(0.51)
Perf_5yr	0.017***	0.049**	-0.01	0.026	-0.009	0.036**
	(3.05)	(2.47)	(-1.53)	(1.66)	(-1.45)	(2.04)
Perf_Incep	-0.001	-0.013	.	.	-0.007	-0.016
	(-0.16)	(-1.24)	.	.	(-1.29)	(-1.54)
Morningstar Rating	.	.	0.016***	0.007*	0.017***	0.008**
	.	.	(9.05)	(1.87)	(8.92)	(2.02)
Min_Rank	0.054***	0.019	0.045***	0.012	0.045***	0.012
	(7.68)	(0.93)	(6.38)	(0.50)	(6.37)	(0.52)
Age	-0.007***	-0.021***	-0.008***	-0.020***	-0.006***	-0.021***
	(-6.67)	(-4.31)	(-9.01)	(-5.11)	(-5.37)	(-4.34)
Previous Quarter Flow	0.199***	0.161***	0.197***	0.161***	0.196***	0.161***
	(8.35)	(5.09)	(8.48)	(5.03)	(8.45)	(5.03)
Size	-0.006***	-0.011***	-0.007***	-0.011***	-0.007***	-0.011***
	(-8.37)	(-6.68)	(-9.56)	(-6.72)	(-8.77)	(-6.70)
Volatility	-0.300***	-0.431*	0.001	-0.369*	0.055	-0.350
	(-4.06)	(-1.97)	(0.01)	(-1.69)	(0.75)	(-1.53)
Total Expense	-0.230*	-2.001***	-0.186	-1.955***	-0.195	-1.970***
	(-1.72)	(-3.03)	(-1.44)	(-2.99)	(-1.49)	(-3.01)
Category Flow	0.411***	0.318**	0.389***	0.279*	0.412***	0.313**
	(6.90)	(2.17)	(6.05)	(1.77)	(6.30)	(2.00)
Intercept	-0.006	0.066***	-0.034***	0.050***	-0.042***	0.053***
	(-1.35)	(4.11)	(-6.92)	(2.96)	(-7.30)	(3.14)

Note: This table reexamines the baseline model that studies ambiguity aversion behavior of both retail and institutional investors when controlling for performance since inception and/or the Morningstar star rating during the period 1993-2011. The sample includes all retail and institutional shares. Morningstar overall rating (Morningstar Rating) is a weighted average of a fund's 3, 5 and 10 years category star rating ranging from 1 to 5 (See Nanda et al. (2004b) for details). A linear regression is performed by regressing quarterly flows on funds' three performance ranks and the minimum rank. The control variables are the same as the ones in Table 1.2. Time-series averages of coefficients and the Newey-West t-statistics (in parentheses) are reported.

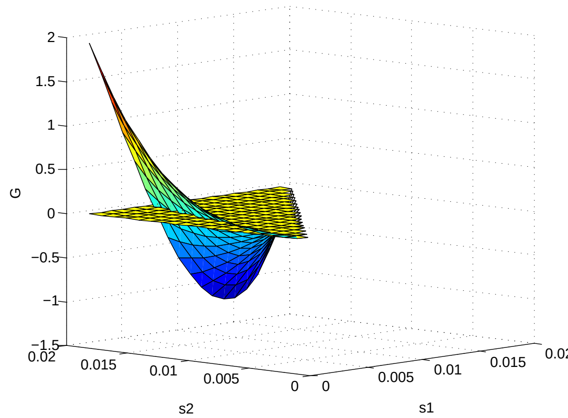
Table 1.9: Robustness Test–Flow-Performance Convex Relationship

Performance Measured by	Raw Return		4-Factor Alpha	
	Retail	Institutional	Retail	Institutional
Low_1yr	0.070***	0.091**	0.076***	0.038
	(4.27)	(2.57)	(4.04)	(1.03)
Mid_1yr	0.029***	0.052***	0.018***	0.043***
	(4.81)	(2.98)	(3.96)	(2.97)
High_1yr	0.241***	0.101**	0.200***	0.098**
	(10.47)	(2.50)	(10.36)	(2.20)
Low_3yr	0.042***	0.028	0.029**	0.147**
	(2.80)	(0.49)	(2.23)	(2.07)
Mid_3yr	0.026***	0.027*	0.036***	0.056***
	(4.26)	(1.77)	(7.35)	(4.89)
High_3yr	0.184***	0.139**	0.207***	0.215***
	(6.17)	(2.56)	(7.72)	(2.87)
Low_5yr	-0.033*	-0.190***	-0.015	-0.094
	(-1.82)	(-3.62)	(-0.81)	(-1.34)
Mid_5yr	0.013***	0.052***	-0.015*	0.011
	(2.87)	(3.38)	(-1.92)	(0.81)
High_5yr	0.056**	0.118***	0.009	0.059
	(2.34)	(2.69)	(0.33)	(1.01)
Min_Rank	0.036***	0.005	0.041***	-0.006
	(4.85)	(0.24)	(6.20)	(-0.34)
Age	-0.018***	-0.027***	-0.010***	-0.017***
	(-9.59)	(-7.32)	(-10.30)	(-4.16)
Previous Quarter Flow	0.172***	0.137***	0.203***	0.162***
	(7.57)	(5.94)	(8.38)	(5.06)
Size	-0.004***	-0.010***	-0.004***	-0.009***
	(-6.92)	(-7.40)	(-8.24)	(-5.98)
Volatility	-0.573***	-0.693***	0.106	0.012
	(-5.47)	(-3.06)	(0.71)	(0.05)
Total Expense	-0.509***	-2.031***	-0.351**	-2.172***
	(-3.14)	(-3.79)	(-2.54)	(-2.93)
Category Flow	0.295***	0.130	0.219***	0.083
	(3.45)	(0.77)	(3.17)	(0.54)
Intercept	0.038***	0.125***	0.001	0.058***
	(5.18)	(7.68)	(0.15)	(4.36)

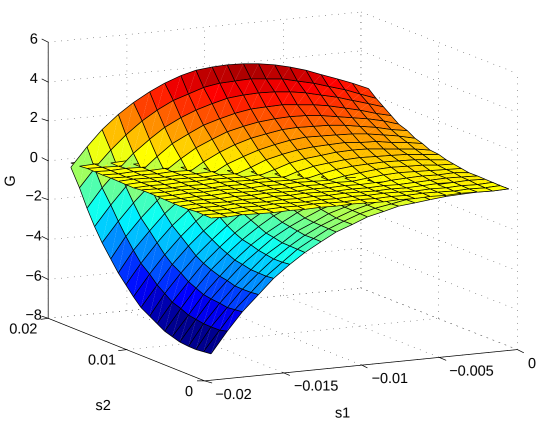
Note: This table reexamines the baseline model that studies ambiguity aversion behavior of both retail and institutional investors when controlling for the convexity in flow-performance relationship during the period 1993-2011. The sample includes all retail and institutional shares. Each quarter, we adopt fractional ranks for funds' performance measured over past 1-, 3-, and 5-year horizon. For example, the 1-year performance of a fund in the bottom quintile (Low_1yr) is defined as Min(Perf_1yr, 0.2), in the three medium quintiles (Mid_1yr) is defined as Min(0.6, Perf_1yr-Low_1yr), and in the top quintiles (High_1yr) is defined as Perf_1yr-Low_1yr-Mid_1yr. Performance is measured by the ranking within category of the average monthly raw returns. Min_Rank is defined as the minimum performance rank among 1-year, 3-year and 5-year performance ranks. A linear regression is performed by regressing quarterly flows on funds' fractional performance ranks and the minimum rank. The control variables are the same as the ones in Table 1.2. Time-series averages of coefficients and the Newey-West t-statistics (in parentheses) are reported. The symbols *, **, and *** denote significance at the 10%, 5% and 1% level, respectively.



Graph A: $s_1 < s_2 < 0$



Graph B: $0 < s1 < s2$



Graph C: $s_1 < 0 < s_2$

Figure 1.1: Difference in Flow-Performance Sensitivity to Signal $\mathbf{s_1}$ Relative to $\mathbf{s_2}$ across Difference Signal Realizations

The figure shows the difference in flow-performance sensitivity to signal s_1 relative to s_2. In particular, it graphs the value of the term G in Equation (2) as we vary s_1 and s_2. Graph A, B and C shows the value of G when $s_1 < s_2 < 0$, $0 < s_1 < s_2$, and $s_1 < 0 < s_2$, respectively.

CHAPTER 2
A BLESSING OR A CURSE? THE IMPACT OF HIGH FREQUENCY TRADING ON INSTITUTIONAL INVESTORS

2.1 Introduction

In recent years, financial markets have undergone tremendous changes with the adoption of new technology. Trades are now mostly placed and executed electronically, while there are over a dozen for-profit exchanges as well as alternative trading venues competing for volume and liquidity. Equally prominently, computer-based high frequency trading (HFT) has grown from being virtually non-existent, to becoming a dominant force in the market. By some statistics, HFT firms account for 70% of the U.S. stock trading volume in 2009.[1] The rapid growth of HFT has led to considerable media attention and policy interest in the issue of the impact of HFT on market quality and on the welfare of other market participants. Despite this interest, there is so far scant evidence on the question of how the recent explosion in HFT has affected a particularly important class of market participants, namely, institutional investors. The goal of this study is to provide evidence regarding the impact of HFT activity on the trading costs of institutional investors.

Traditional institutional investors such as mutual funds, pensions, insurance firms, and hedge funds account for over 50% of the public equity ownership in the U.S. (French (2008)). They play a critical role in price discovery by trading based on new information or in response to price deviations from fundamentals. Moreover,

[1] See, e.g., "High-frequency trading under scrutiny," Financial Times, July 28, 2009.

they generate a huge volume of trading and trading costs are a critical determinant of their performance. Hence, institutional trading costs are often viewed as an important yardstick for measuring the quality and liquidity of the financial markets. For this reason, facilitating efficient execution of institutional trades has been a key objective of the securities markets design and regulation. Whether HFT is good news or bad news for traditional institutional investors has been extensively discussed and debated in public media. Some institutional investors have expressed serious concerns that high frequency (HF) traders may negatively impact their trading profits (e.g., Arnuk and Saluzzi (2008)). Such concerns are apparently heard by regulators, as noted in the 2010 speech by Mary Schapiro, the former SEC Chairperson, "Institutional investors also have expressed serious reservations about the current equity market structure. Institutional investors questioned whether our market structure meets their need to trade efficiently and fairly, in large size." In fact, asset managers' concerns regarding HFT have led to the growing popularity of off-exchange trading venues, e.g., "dark pools".[2]

Interestingly, the widespread concerns about the negative impact of HFT on institutional trading costs are in sharp contrast to the findings of a few recent academic studies. Academic evidence so far seems to suggest that, predominantly, HFT is associated with improved market liquidity, reduced volatility, and increased price efficiency/discovery; see, for example, Chaboud et al. (2009), Brogaard (2010), Hen-

[2]The trading volume in dark pools has grown by almost one-half between the years 2009-2012; see "U.S. 'dark pool' trades up 50%," Financial Times, November 19, 2012.

dershott et al. (2011), Boehmer et al. (2012), Menkveld (2010), Hasbrouck and Saar (2013), Brogaard et al. (2013), and Malinova et al. (2013). The evidence produced by these studies is consistent with the view that HFT firms are the modern day version of market makers with enhanced technology. If technology expedites the execution of trades and/or improves the efficiency of market making, HFT should benefit market participants, including institutional investors.[3]

However, some researchers have raised the concern that the liquidity provided by HF traders may be illusory. Since HF traders do not have an affirmative obligation to provide liquidity, their trading is opportunistic in nature, and the liquidity they create may disappear quickly when it is most needed on the market. Kirilenko et al. (2011) and Easley et al. (2011a) both note that during the Flash Crash of May 6, 2010, many HF traders withdrew from the market while others turned into liquidity demanders. In the context of institutional trading, an open question is whether HFT is a reliable source of liquidity when liquidity is most demanded by institutional investors.

The illusory nature of liquidity created by HFT may also be understood in the context of specific HFT strategies. Two particular types of directional HFT strategies appear to directly take advantage of the large trades made by institutional investors –

[3]A study perhaps most related to mine is Brogaard et al. (2012). Using UK data, they find no clear evidence that increases in HFT activities due to speed changes at London Stock Exchange affect institutional trading costs. However, to my knowledge, so far there is no study on the impact of HFT on institutional trading cost in the context of the U.S. market.

order anticipation (front running) and momentum ignition.[4] An HF trader following an order anticipation strategy detects large orders from institutional investors and trades in front of them. For example, a HF trader who buys in front of a large buy order will subsequently attempt to sell to the large buyer at a higher price or to hold on to the position in case of a permanent price increase. The institutional investor who submits the large buy order is adversely impacted in either case. With momentum ignition, HF traders may ignite rapid price movement along one direction through a series of submissions and cancelations of orders, and profit by establishing an early position. Such strategies may increase intraday price volatility and drive up the trading costs of institutional investors.

In this study, I combine two sources of data to examine the relationship between HFT and institutional trading costs. Data on institutional trading costs are from Ancerno (formerly Abel/Noser). The main measure of trading cost is execution shortfall, defined as the percentage difference between the execution price and a benchmark price that is prevailing in the market when the ticket is placed with the broker. The execution shortfall captures the bid-ask spread, the market impact, and the drift in price while the ticket is executed. Data on HFT is provided by NASDAQ. This dataset contains all trades on NASDAQ for a randomly selected sample of 120 stocks during 2008 and 2009, with identification of trades executed by HFT firms.

I assess the relation between HFT and institutional trading costs using both

[4]Several popular types of HFT strategies are discussed in the Concept Release on Equity Market Structure by Commission (2010). In addition to directional trading strategies, three other broad types of strategies include passive market making, arbitrage, and structural trading.

sorted portfolios and multivariate regressions. Using sorted portfolios, I show that while HFT is positively associated with stock liquidity and the latter is negatively associated with institutional trading costs, the relation between HFT and institutional trading costs is positive. The multivariate panel regressions confirm this relation by controlling for various stock characteristics and institutional trading characteristics. The regression coefficient suggests that one standard deviation increase in HFT activity is associated with an increase in average execution shortfall by one third. Considering that an average institution in the sample has a daily trading volume of \$20.5 million for the sample stocks, one third increase in execution shortfall cost implies an additional transaction cost of more than \$10,000 per day. I also find that the impact of HFT on institutional trading costs is stronger for both small-cap and large-cap stocks, relative to mid-cap stocks.[5]

I consider alternative explanations for the positive relation between HFT and institutional trading cost. These include the possibility of omitted variables causing both HFT activity and institutional trading costs to increase at the same time. Alternatively, it could be that HF traders find it more attractive to trade on stocks that have high trading costs. I seek to rule out the alternative interpretations through several approaches.

[5]The main measure of trading cost in this study is execution shortfall, which captures the bid-ask spread as well as the price impact (e.g., Anand et al. (2012)). I have also examined the timing delay component of trading cost to test a hypothesis that HFT reduces delays in trade execution. However I do not find evidence in favor of this hypothesis. In addition, the main regressions performed in the study are based on stock-day observations. I have also obtained similar results using regressions at individual trade level that control for heterogeneity in institutional trading skills.

First, the sorted portfolio analysis indicates that HF traders are most active in liquid stocks, rather than illiquid stocks which have high trading costs. Second, I include firm- and time-fixed effects in the multivariate regression specification, which helps to ensure that unobserved slow-moving stock characteristics and time-invariant factors do not cause the positive relationship between HFT activity and trading costs. Third, since days with news releases may also affect both HFT and trading costs, I control for earnings announcements and mergers and acquisitions events in the sample and the results still hold. Fourth, I study the short selling ban on financial stocks instituted on September 19, 2008, which is an exogenous shock to execution shortfall. I find that, as expected, the execution shortfall increases sharply on that day due to the ban. If HF traders choose to be more active when the execution shortfall is high, we would expect an increase in HFT after the implementation of the ban. However, I find that the HFT activity drops sharply subsequent to the ban being implemented. This evidence also suggests that when liquidity is low, HF traders withdraw from the market. Fifth, Granger causality tests provide further evidence that intensive HFT activity contributes to an increase in trading costs, but not vice versa.

Finally, I perform two sets of analysis to understand the specific mechanisms through which HFT may increase the costs of traditional institutional investors. First, I examine whether HF traders profit from providing liquidity when institutional investors exhibit large buy-sell imbalance, i.e., when institutional investors on the net are either large buyers or sellers of a stock. I find that on days with large institutional buy-sell imbalance on a given stock, HFT activities are more intense, but at market close HF traders manage to keep virtually no open positions on the stock. Further, the impact of HFT on institutional trading costs is more pronounced when institutions exhibit large imbalance on the buy side. Therefore, if anything, HFT represents an ephemeral and expensive source of liquidity provision to institutional investors.

Second, I use the non-randomness of HF trades to test whether directional trading, electronic marketing making, and other types of HFT strategies have different impact on institutional trading costs. In the case of directional strategies such as momentum ignition and front running, one would observe long sequences of HF trades in the same direction.[6] As for electronic market making, HF traders have to buy and sell the same stocks very fast so that one should observe rapid reversals of HF trade directions. I use the runs test to detect non-randomness in HF trade directions on each stock on a given day. The runs tests detect the pervasive use of directional trading and market making strategies by HF traders. More importantly, the impact of HFT on institutional trading costs is most pronounced when HF traders engage in directional trading strategies. This lends support to the anecdotal observations made by institutional investors that their trades are front-run by HF traders.

The rest of the paper is organized as follows: Section 2.2 discusses the literature related to HFT. Section 2.3 describes the data. Section 2.4 presents the baseline results and analyses on causality between HFT and institutional trading costs. Section 2.5 provides further analysis on how and when HFT affects institutional trading

[6] Front-running trades by HF traders are more likely in the form of a sequence of small trades in the same direction than a few large trades, because in recent years both institutions and HF traders split large orders into small sizes for execution.

costs as well as the robustness of the results. Section 2.6 concludes.

2.2 Related Literature

This paper fits in the growing literature on algorithmic trading and HFT. Theoretical models in this area focus primarily on the interaction between HF traders and traditional investors. Such studies generally predict undesirable impacts of HFT and a wealth transfer from slow traders to HF traders. Hoffmann (2013) finds that algorithmic traders suffer less from adverse selection because of their speed advantage and that they decrease the profits of human traders. Cartea and Penalva (2005) present a model with a liquidity trader, a market maker and a HF trader. Their model predicts an increase in volatility and price impact of the liquidity trader. In the model built by McInish and Upson (2011), HF traders use their speed advantage to learn quote updates quicker than slow traders, which allows the former to profit from trading at stale prices with the latter. Jarrow and Protter (2011) find that HF traders create temporary mispricing and profit from it. Biais et al. (2011) document that multiple equilibriums can arise for a given level of algorithmic trading and some of them are associated with a sharp increase in the price impact of trades. Jovanovic and Menkveld (2012) model HF traders as middlemen between the buyers and sellers. Their model suggests that HF traders can exert positive or negative effects depending on their informational advantage stemming from their speed.

In contrast to the overall negative predictions of theoretical models, most empirical studies document a positive impact of HFT. Using the same dataset as in this study, Brogaard et al. (2013) provide evidence that HF traders facilitate price efficiency by placing marketable orders in the direction of permanent price changes and in the opposite direction of transitory pricing errors on average days and the days with highest volatility. Their limit orders are adversely selected but are compensated by liquidity rebates. With the same dataset, Brogaard (2010) finds no evidence that HF traders withdrawing from markets in bad times or that they front run large non-HFT trades. Using message counts as a proxy for algorithmic trading (AT), Hendershott et al. (2011) find that AT improves liquidity and brings about more efficient price discovery. With the same proxy, Boehmer et al. (2012) document that on average AT improves liquidity and informational efficiency. Another study by Chaboud et al. (2009) also documents that algorithmic traders increase their supply of liquidity over the hour following macroeconomic data releases, even though they restrict activity in the minute following each release. Also, Hasbrouck and Saar (2013) find improved spreads, depth and volatility associated with HFT. Menkveld (2010) finds that the bid-ask spreads of a new market for Dutch stocks, Chi-X, were reduced by about 30% within a year with the entry of a new HF trader on the market.

There are also some empirical studies that document negative effects of HFT. The major concerns are the quality of the liquidity provided by HF traders and whether they increase volatility. Kirilenko et al. (2011) find evidence that instead of supplying liquidity, some HF traders withdrew from the market and some demanded liquidity during the Flash Crash on May 6, 2010. Hasbrouck and Saar (2013) document the "fleeting" nature of many limit orders in electronic markets and point out the liquidity provided by HF traders is short-lived. Similarly, Egginton et al. (2014)

question the degraded quality of liquidity and elevated volatility caused by HFT. Easley et al. (2011a) find that extraordinary flow toxicity, i.e., market makers being adversely selected without knowing, in the hours leading up to the Flash Crash causes HF traders to withdraw from the market.

Overall, even though theoretical models predict the shift in wealth from slow traders to HF traders, there is limited empirical evidence along this direction. In fact, most empirical evidence suggests an improvement in market quality with the occurrence of HFT. However, this improvement does not immediately lead to more efficient trading for traditional investors. A related study by Malinova et al. (2013) examines the impact of HFT on retail investors. They find that a reduction of HFT causes a decline in market liquidity and trading profits of retail traders. In a recent study, Brogaard et al. (2013) use data from the London Stock Exchange and find no clear evidence of change in trading costs caused by increases in HFT activities due to speed changes at the exchange. However, in the U.S. market, there is so far no direct analysis on whether HFT increases institutional investors' trading costs. This paper fills the gap.

2.3 Data and Descriptive Statistics

2.3.1 Measuring HFT

The HFT dataset is provided by NASDAQ under a non-disclosure agreement. The dataset contains trading data from 2008 and 2009 for a sample of 120 randomly selected stocks listed on NASDAQ or the New York Stock Exchange (NYSE). The timestamp for trades in the dataset is to the millisecond. For each trade in the dataset, a variable named "Type" identifies the liquidity demander and supplier as a high-frequency (HF) trader or non-high-frequency (nHF) trader based on NASDAQ's knowledge of its customers and analysis of the firm's trading, such as how often its net trading in a day crosses zero, its order duration, and its order to trade ratio.

NASDAQ identifies a total of 26 HFT firms in the data. However, HFT firms that route their orders through large integrated firms such as Goldman Sachs and Morgan Stanley cannot be identified and thus are excluded. As noted in Brogaard et al. (2013), even though the 26 HFT firms represent a significant amount of HFT activity, it is not possible to completely identify all HF trades. Despite this limitation, this dataset is by far the most suitable for this study. Previous academic studies that use this dataset include Brogaard (2010), Brogaard et al. (2013), and Carrion (2013).

The dataset categorizes 120 stocks into three market capitalization groups: large, medium and small. Each size group contains 40 stocks, with 20 stocks listed on NYSE and the other 20 listed on NASDAQ. The top 40 stocks are from the largest market capitalization stocks. The medium-size category consists of stocks around the 1000th largest stocks in the Russell 3000, and the small-size category contains stocks around the 2000th largest stock in the Russell 3000. For each stock, the dataset contains the following fields: Ticker Symbol, Date, Time (in milliseconds), Shares, Price, Buy/Sell Indicator, and Type (HH, HN, NH, NN). The Type variable identifies whether the two participants in a trade are HFT firms (H) or not (N). For example, "HN" means that an HF firm demands liquidity and an nHF (non-HF) firm supplies liquidity in the trade. See Brogaard et al. (2013) for additional details on this dataset.

In this paper, I focus on the total HFT activity on a stock. To construct the measure of HFT activity, I first calculate the trading volume of each trade in the dataset by multiplying Price and Shares traded. Each day, the aggregate trading volume of all trades that HFT firms participate in (with Type of HH, HN or NH) for a particular stock captures the total HFT volume on that stock. The measure of HFT daily activity on stock i, denoted as HFT Intensity$_{it}$, is defined as the aggregate HFT volume for stock i on day t divided by the stock's average daily trading volume in the past 30 days .

2.3.2 Measuring institutional trading cost

The NASDAQ dataset is merged with a proprietary database of institutional investors' equity transactions compiled by Ancerno Ltd., from which I construct the measure of institutional trading cost. There are 204 institutions in the Ancerno dataset that are involved in trading the 120 sample stocks during 2008 and 2009, with an average trading volume of $20.5 million per institution per day. Previous academic studies that use Ancerno's data include Anand et al. (2010, 2012), Goldstein et al. (2009), Chemmanur et al. (2009), Goldstein et al. (2010), and Puckett and Yan (2011).

A typical order from a buy-side institution is large in size and usually has high information content. To reduce market impact, the trading desk of the buy-side institution splits the large order to several brokers. The allocation to each broker is defined as a ticket and each ticket may result in several distinct trades or executions. For each execution, the database reports identity codes for the institution, the CUSIP and ticker for the stock, the stock price at placement time, date of execution, execution price, number of shares executed, whether the execution is a buy or sell, and the commissions paid. See Anand et al. (2012) for additional details on this dataset.

Following Anand et al. (2012), the cost of each trade (referred to as "ticket" in the Ancerno data) is defined in terms of execution shortfall:

$$\text{Execution Shortfall} = \frac{P_1 - P_0}{P_0} \times D, \tag{2.1}$$

where P_1 measures the value-weighted execution price of the ticket, P_0 is the price at the time when the broker receives the ticket, and D is a dummy variable that equals 1 for a buy trade and -1 for a sell trade. I calculate the volume-weighted average of the execution shortfall of all trading tickets for stock i on day t and denote it as Execution Shortfall$_{it}$.

In this study, I conduct most of the tests at the stock level using the daily measures of HFT Intensity$_{it}$ and Execution Shortfall$_{it}$. As a robustness test, I also examine the relationship between HFT activity and execution shortfall at the trading ticket level.

Another aspect of institutional trading costs is the execution timing delay cost incurred between the initial trading decision point (market open) and the price at the time the order is placed with the broker:

$$\text{Timing Delay} = \frac{P_0 - \text{Open Price}}{\text{Open Price}} \times D, \tag{2.2}$$

where Open Price is the opening price on the execution day. This timing delay cost can be thought of as the cost of seeking liquidity (e.g, Group (2009)). This measure is constructed for each trading ticket in the sample. I calculate the volume-weighted average of the timing delay of all trading tickets for stock i on day t and denote it as Timing Delay_{it}. The main focus of this paper is to examine the impact of HFT on execution shortfall which is a major component of institutional investors' trading costs. However, it is also of interest to examine if HFT helps to reduce the timing delay costs.

2.3.3 Sample descriptive statistics

I obtain data on institutional trading and HFT from 2008 to 2009 on a sample of 120 stocks. To minimize observations with errors I impose several data screens. I delete tickets with execution shortfall greater than an absolute value of 10%. Also, I delete tickets with ticket volume larger than the stock's total trading volume on the execution date. I obtain data on stock daily trading volume, daily returns, close price, and total shares outstanding from *CRSP*. In addition, I identify earnings announcement dates from *I/B/E/S* and *COMPUSTAT*. I obtain information on mergers and acquisitions from SDC Platinum.

Table 1 reports the summary statistics of HFT and the institutional trading. These numbers reveal some notable patterns in HFT. The HF traders are most active in large stocks. The average daily HFT volume on large stocks, medium stocks and small stocks is \$158.23, \$3.65 and \$0.38 million, respectively. This pattern raises a natural question about the role of HF traders. If, as the proponents of HFT typically advocate, HF traders play a role in providing liquidity, they should be more active in small stocks where liquidity is scarce. The average Execution Shortfall for large, medium and small stocks is 0.15%, 0.16%, and 0.20%, respectively. The results indicate that trading tickets placed on small stocks are more difficult to execute, as shown by the larger execution shortfall. This observation is consistent with the findings of Anand et al. (2010). The size of an average trading ticket placed on large stocks is $487,871 and it takes more than three executions to implement the ticket. The average ticket size on small stocks is only $63,943 and it takes about 1.8 executions to implement the ticket.

2.3.4 Determinants of HFT

Before an examination on the relation between HFT and institutional trading cost, it is useful to understand the firm characteristics that may be associated with the intensity of HFT. These characteristics may also be related to trading costs and serve as control variables in my main analysis.

I consider the following characteristics. 1) firm size (Log Market Cap), the logarithm of a stock's daily market capitalization; 2) Book-to-Market Ratio, measured using information available at the beginning of each calendar quarter; 2) Event Dummy, a dummy variable that equals one for a stock on a given day if there is a corporate event (earnings announcement or merger and acquisition announcement), and equals zero otherwise; 3) Daily Return Volatility, which is a stock's range-based estimate of daily volatility (annualized), following Parkinson (1980); 4) Prior 1-day Return, Prior 1-month Return, and Prior 12-month Return, which are a stock's lagged daily return, lagged monthly return, and lagged 12 months return, respectively; 5) stock illiquidity as measured by the Amihud Illiquidity Ratio, i.e., the daily absolute return divided by the dollar trading volume on that day; 6) Daily Dollar Turnover, a stock's daily dollar trading volume scaled by the stock's total shares outstanding; 7) Average Institutional Order Size, the average dollar volume of all tickets placed on a stock, scaled by the average trading volume of that stock in prior 30 days; 8) Absolute Institutional Imbalance, the absolute value of the daily total dollar volume of all institutional buy tickets minus that of all sell tickets on a stock, scaled by the average trading volume of that stock in the past 30 days; 9) Average Trades Per Order, defined as the average number of trades to complete a trading ticket on a stock; 10) Prior 1-month Market Volatility, annualized daily return volatility of the CRSP value-weighted index in prior month; 11) Prior 1-day Market Return, the return of the CRSP value-weighted index during the previous day.

A panel regression model is estimated by regressing daily stock HFT Intensity on these firm characteristics. The estimated coefficients and two-way clustered t-statistics are reported in Table 2. The results suggest that HFT intensity is positively related to firm size, return volatility, and negatively related to illiquidity. HF trading is also more active in stocks with high daily dollar turnover and high absolute institutional trading imbalance, stocks with large number of institutional trades per order, and on days with event announcements.

2.4 Impact of HFT on Institutional Trading Costs

2.4.1 HFT, liquidity, and trading costs: sorted portfolios

I begin with a sorted portfolio analysis to present an intuitive picture on the relations among HFT activity, liquidity, and trading costs of institutional investors.

First, I look at the relation between HFT and the conventional measure of stock liquidity, the Amihud Illiquidity Ratio. Since that the 120 stocks are in three distinctive size categories, I first sort all stocks into three groups based on size. Within each size group stocks are further divided into three groups based on the Amihud Illiquidity Ratio on each day. I calculate the average HFT Intensity of all stock-days in each of the nine (3×3) groups. Figure 2.1 plots the average HFT Intensity against the Amihund Illiquidity Ratio across the nine groups; it shows clearly a positive relation between HFT and liquidity, within each size group. This finding complements those reported by the existing literature. However, we cannot infer the direction of the causality from such a simple statistical association. It may be the case that HF traders choose to trade more in liquid stocks, given their reliance on rapid-fire trading strategies.

Next, I look at the relation between stock liquidity and institutional trading costs measured by Execution Shortfall. I continue to rely on the nine groups of stocks sorted on size and Amihud Illiquidity Ratio. Figure 2.2 plots the average Execution Shortfall across the nine groups; it shows a clear negative relation between execution shortfall and liquidity within each size group. That is, trading costs are lower for liquid stocks.

Combining the patterns from the first two panels of Figure 2.1 and 2.2, one may expect a negative relation between HFT Intensity and Execution Shortfall. However, Figure 2.3 shows that the opposite holds. In this plot, I sort stocks into terciles based on HFT Intensity within each size group to form nine portfolios and compute the average Execution Shortfall within each portfolio. The plot shows that within each size group, when HFT is more active, the average Execution Shortfall for institutional investors is also higher. In other words, the HFT activity is positively correlated with institutional trading costs.

Figure 2.1-2.3 present rather intriguing relationship among HFT activity, liquidity, and institutional execution shortfall. If HFT activity could improve liquidity, as documented in the extant literature, why does execution shortfall increase when HFT activity is more intensive? Considering the distinctive features of institutional trading, HFT may indeed bring more harm than good to institutional investors. First of all, the liquidity provided by HFT may be illusory and may disappear when institutional investors most need it. Moreover, the large order sizes and potentially high information content make institutional trades most vulnerable to HFT strategies such as front running (see Hirschey (2011)). Such strategies can dramatically increase the price drifts and market impact during the execution of a large order.

2.4.2 Multivariate analysis

In order to control for other relevant factors that may affect trading costs, I move on to conduct the following tests in a multivariate panel regression setting with controls of various firm characteristics. Specifically, I estimate a panel regression model of the form:

$$\text{Execution Shortfall}_{it} = \alpha_i + y_t + a \times \text{HFT Intensity}_{it} + b \times X_{it} + \epsilon_{it}, \tag{2.3}$$

where α_i and y_t represent firm-fixed effects and time(day)-fixed effects, respectively. HFT Intensity$_{it}$ is the measure of daily HFT activity on stock i. Execution Shortfall$_{it}$ is volume-weighted average execution shortfall of all trading tickets on stock i at day t. X_{it} represents a set of firm characteristics that have been considered in Table 2 when I examine the determinants of HFT activity. These include firm size, book-to-market ratio, stock returns during prior one day, one month, and 12 months, the Amihud illiquidity ratio, a range-based daily stock volatility measure, daily trading turnover, average institutional order size, absolute institutional trade imbalance, and average number of trades per order. For inference I use standard errors that are robust to cross-sectional and time-series heteroskedasticity and within-group autocorrelation based on Petersen (2009).

Table 3 presents estimates of coefficients and the two-way clustered t-statistics. The first two columns report the estimates of the model without controlling for day- and firm-fixed effects. However, to control for market conditions I additionally include the prior 1-day market return and prior 1-month market volatility as control variables. In the last two columns, the linear regression model in Equation (2.3) is estimated with both day dummies and firm-fixed effects, but without the two market-condition variables.

In both sets of tests, the coefficient on HFT Intensity is positive and significant at the 1% level. This positive coefficient suggests that after controlling for other economic determinants of trading costs, HFT activity has an *increasing* effect on execution shortfall of institutional investors. In particular, the coefficient from the fixed-effects regression indicates that a one standard deviation increase in HFT activity leads to a 5bp increase in execution shortfall. Considering that an average institution in my sample generates a daily trading volume of \$20.5 million, a 5bp increase in execution shortfall means an additional cost of more than \$10,000 per day on the sample stocks.

To better evaluate the effects of control variables on execution shortfall, I focus on the estimation results of the model without day- and firm-fixed effects, as shown in the first two columns of Table 3. The coefficients for the control variables are of expected signs. The coefficient of the illiquidity measure is positive and significant since a higher illiquidity measure means lower liquidity which leads to a higher execution shortfall. The coefficient of the absolute value of institutional buy-sell imbalance is positive and significant at the 1% level. This is because the higher imbalance leads to more competition for liquidity in one direction, thus execution shortfall is higher. Similar to prior studies, I find that execution shortfall increases with stock volatility.

In sum, the results from the multivariate panel regression indicate that when HFT activity is more intense, institutional investors' execution shortfall is higher. More importantly, this positive relationship holds when I control for various firm characteristics as well as the time- and firm-fixed effects.

2.4.3 Impact of HFT across firm size

I further examine the differential effects of HFT on execution shortfall for stocks with different sizes. To do this, I estimate the baseline model in Equation (2.3) within each size group. I expect the impact of HFT on execution shortfall to be stronger for small stocks. This is because it is more costly for HF traders to participate in small stocks and they will charge a higher premium to do so. In fact, in order to make profit, HFT strategies require such traders to be able to buy and sell in a timely manner, yet this is harder to accomplish in the case of small stocks (e.g.,Arnuk and Saluzzi (2008)).

Table 4 reports the estimates of coefficients and the two-way clustered t-statistics. The regression model is estimated with both day dummies and firm-fixed effects. From left to right, the table reports the estimation results in the subsamples of large, mid, and small stocks. The coefficient of HFT Intensity suggests that, as expected, the increasing effect of HFT activity on execution shortfall is strongest on small stocks. Thus, HF traders charge a high premium when they trade small stocks. It is further noted that the coefficient for HFT Intensity is also significantly positive for large-cap stocks, suggesting an important impact by HFT on the trading costs of such stocks. Finally, the coefficient for HFT Intensity is insignificantly positive in the subsample of midcap stocks.

2.4.4 Direction of causality

There are two alternative explanations for the multivariate test results. This includes the possibility of some omitted variables that cause both HFT activity and execution shortfall to increase at the same time. Alternatively, it could be that it is precisely when execution shortfall is high that it is more profitable for HF traders to trade actively.

In fact, the tests conducted in the previous subsections have already help to rule out the alternative interpretations to certain degree. First, the sorted portfolio analysis indicates that HF traders are most active in liquid stocks, rather than illiquid stocks featured with high trading costs. Second, I include firm- and time-fixed effects in the multivariate regression specification, which helps ensure that unobserved slow-moving stock characteristics and time-invariant factors do not cause the positive relationship between HFT activity and execution shortfall.

In this subsection, I conduct further analysis on this issue.

The above results establish the increasing effect of HFT activity on execution shortfall for institutional investors after controlling for time- and firm-fixed effects. However, there may be certain special events that cause an increase in both HFT activity and execution shortfall. To rule out this possibility, I control for two types of important corporate events: earnings announcements and mergers and acquisitions (M&A). I identify earnings announcement days from $COMPUSTAT$ (and augmented with $I/B/E/S$ data in the case of missing earnings announcement dates in $COMPUSTAT$). The M&A dates are identified from SDC. In total, during the two year period, there are 960 quarterly earnings announcements and 323 M&A announcements where the 120 firms in my sample are either acquirers or targets.

In order to observe the different impact of HFT on execution shortfall on event days and non-event days, I create a dummy variable Event Dummy that equals one for a stock-day observation falling within a 5-day window of a corporate event for that stock. It is zero otherwise. No-Event Dummy is a dummy variable that equals one for a stock-day not in any 5-day corporate event window for that firm. I then interact HFT Intensity with Event Dummy and No-Event Dummy, respectively, and use the interaction terms in place of HFT Intensity in the panel regression analysis. Other variables in the regression remain the same as those reported in Table 3.

Table 5 presents estimates of the coefficients and the two-way clustered t-statistics. The coefficient of the interaction between HFT Intensity and Event Dummy is positive but not significant. However, the interaction between HFT Intensity and No Event Dummy is positive and significant at the 1% level. The results indicate that the increasing effect of HFT activity on execution shortfall mainly occurs on days without corporate events. This is inconsistent with the hypothesis that certain corporate events drive both HFT Intensity and Execution Shortfall higher.

In the previous subsection, I find that when HF traders participate more, institutional investors encounter a higher execution shortfall. Alternatively, it could also be that HF traders choose to be more active when execution shortfall is high. In this subsection, I will rule out this possibility through analysis of an exogenous event - the short selling ban.

I study the behavior of HF traders and the pattern of execution shortfall around the short selling ban from September 19, 2008 to October 8, 2008. On September 19, 2008, the SEC released an emergency order prohibiting short selling in a group of 799 financial stocks. The initial list of securities covers 13 stocks in my sample. On September 22, the list expanded to cover 16 stocks in my sample, and one more stock was added to the banned list on September 23.[7] This short selling ban was instituted

[7]The trading symbols of the sample stocks in the initial short-selling ban list are: AINV,

immediately without any advance notice, and thus can be viewed as an exogenous event. The prohibition on short selling has an immediate impact on institutional investors' execution shortfall cost in the banned stocks. This ban, however, does not by itself impact HF traders directly.

Figure 2.4 presents the time-series pattern of the average Execution Shortfall of the banned and unbanned stocks around the short selling ban. As expected, the execution shortfall of banned stocks increases sharply when the ban is imposed on September 19. Figure 2.5 plots the time-series of the average HFT Intensity for the banned and unbanned stocks around the same period. On September 19, when execution shortfall reaches its highest level in the picture, I observe a sharp decrease in HFT activity. If the increasing effect of HFT activity on execution shortfall is because that the HF traders choose to participate more when trading costs are high, one should observe an increase in HFT activity instead. This pattern also raises a question on the HF trader' role in providing liquidity. Clearly when liquidity is most needed, they appear to withdraw from the market altogether (e.g., Carrion (2013)).

In conclusion, through observations of institutional trading costs and the behavior of HF traders during the shore selling ban, I further rule out the alternative explanation that the positive relation between HFT and trading cost is due to a selection effect, i.e. HF traders choose to be more active when trading cost is high.

I use the Granger causality test to further establish the direction of causality.

BXS, CB, CRVL, DCOM, EWBC, FFIC, FMER, FULT, MIG, PNC, PTP, SF. The list is expanded to cover GE, AXP, and CSE on 9/22/2008 and ARCC on 9/23/2008.

The Grander causality test enables one to infer, in a statistical sense, whether a lagged variable (e.g., lagged HFT Intensity) bears a causal effect on another variable (e.g., Execution Shortfall). Specifically, for a given stock, the Granger causality test is performed under the following VAR(1) framework:

$$\begin{pmatrix} ES_{i,t} \\ HFT_{i,t} \end{pmatrix} = \begin{pmatrix} a_{1,i} \\ a_{2,i} \end{pmatrix} + \begin{pmatrix} b_{11,i} & b_{12,i} \\ b_{21,i} & b_{22,i} \end{pmatrix} \begin{pmatrix} ES_{i,t-1} \\ HFT_{i,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,i,t} \\ \epsilon_{2,i,t} \end{pmatrix}, \tag{2.4}$$

where $ES_{i,t}$ and $HFT_{i,t}$ are the Execution Shortfall and HFT Intensity for stock i on day t, respectively. $a_{1,i}$, $a_{2,i}$, $b_{11,i}$, $b_{12,i}$, $b_{21,i}$, $b_{22,i}$ are parameters. $\epsilon_{1,i,t}$ and $\epsilon_{2,i,t}$ are innovation terms.

I examine the following two null hypotheses: (1) HFT Intensity does not Granger cause Execution Shortfall; (2) Execution Shortfall does not Granger cause HFT Intensity. If $b_{12,i} \neq 0$ then null hypothesis (1) is rejected, indicating that HFT Intensity Granger causes Execution Shortfall. On the other hand, if $b_{21,i} \neq 0$ then null hypothesis (2) is rejected, which means that Execution Shortfall Granger causes HFT Intensity.

A statistical issue here is that inference has to be made jointly on 120 stocks. Take the inference on the first hypothesis (i.e., HFT Intensity does not Granger cause Execution Shortfall) for example. Even when the true values of $b_{12,i}$s are all zero across the 120 stocks, by statistical randomness the sample estimates of some of the $b_{12,i}$s will be significantly different from zero. Therefore, in the presence of a relatively large cross-section of stocks, inference in a stock-by-stock fashion is likely problematic. Instead, I focus on the distribution of the estimated coefficients (i.e., $b_{12,i}$ and $b_{21,i}$)

across the 120 stocks, and assess whether the sample distribution of the coefficients is different from what one would observe under the null hypothesis of no causality. To do so, a further complication to take into account is that the variables of interest, $b_{12,i}$s or $b_{21,i}$s, are correlated across stocks.[8]

I take a bootstrap approach to perform statistical inference jointly on the 120 stocks, in a way similar to the bootstraps performed by Kosowski et al. (2006) and Jiang et al. (2007) in their studies of mutual fund performance. In the context of this study, the bootstrap procedure generates randomized observations of $\text{ES}_{i,t}$ and $\text{HFT}_{i,t}$ under the null of no causality (i.e., $b_{12,i}$=0 and $b_{21,i}$=0 for all i), while at the same time keep the time-series persistence parameters of $\text{ES}_{i,t}$ and $\text{HFT}_{i,t}$ per se, the correlation between $\epsilon_{1,i,t}$ and $\epsilon_{2,i,t}$ for any given stock, as well as the correlations among $\epsilon_{1,i,t}$ and $\epsilon_{2,i,t}$ across 120 stocks.[9] For each bootstrap, I estimate the cross-sectional statistics including the mean, median, 1st and 3rd qunitiles of the t-statistics for the

[8] In addition to inference based on the cross-sectional distribution of the coefficients, one can also use more conventional Wald-type test on the hypothesis that the coefficients $b_{12,i}$s (or 120 $b_{21,i}$s) are jointly zero across all 120 stocks. However, in the presence of a large cross-section relative to the length of the time series, the power and size of the conventional test are likely an issue.

[9] Specifically, the procedure involves the following steps. Across the 120 stocks, I compute the cross-sectional distribution statistics such as mean, median, 1st and 3rd quintiles of the t-statistics. First, I estimate the VAR(1) model described in (2.4) using the sample data, and obtain the coefficients, corresponding t-statistics, and the estimated residuals for all stocks. Second, I bootstrap (i.e., resampling with replacements) the residuals to reconstruct the bootstrapped time series of $ES_{i,t}$ and $HFT_{i,t}$, using the bootstrapped residuals and the estimated parameters from the model (2.4) but restricting $b_{12,t}$ and $b_{21,i}$ to be zero. Third, I estimate the model (2.4) using the bootstrapped $ES_{i,t}$ and $HFT_{i,t}$, and obtain a new set of coefficients and the corresponding t-statistics. Across 120 stocks, I obtain the cross-sectional distribution statistics of the bootstrapped t-statistics. Step 2 and 3 are repeated for 2,000 times to obtain 2,000 bootstrapped observations of the cross-sectional statistics (i.e., mean, median, 1st and 3rd quintiles of the t-statistics). Note that I bootstrap t-statistics rather than the coefficients per se, because the t-statistics are pivotal statistics that have a better convergence property.

estimated coefficients. The bootstraps are performed 2,000 times, and the sample cross-sectional statistics (e.g., the mean of the t-statistics) are compared with the the corresponding bootstrapped statistics to assess statistical significance. Specifically, the bootstrapped p-value is computed as the percentage of bootstrapped statistics that exceed the sample statistics. A bootstrapped p-value close to 1 indicating that the sample statistic is abnormally low relative to the distribution under the null hypothesis of no causality; and a bootstrapped p-value of 0 indicating that the sample statistic is abnormally high relative to what one would expect under the null of no causality.

Table 6 presents the results of the Granger causality test. As shown in Panel A, across the 120 stocks, $b_{12,i}$, the coefficient related to the causality of HFT on ES, has a positive mean of 0.317, and its corresponding t-statistic has a positive mean of 0.311. The bootstrapped p-value is 0.002, indicating that the mean of the sample t-statistic is abnormally high relative to what is expected under the null of no causality. Note that the p-values for other cross-sectional statistics, i.e., median, 1st and 3rd quintiles, are all very low. Therefore, we infer that across the 120 stocks, there is a pervasive pattern that the intensity of HFT Granger-causes institutional trading cost.

On the other hand, as shown in Panel B of the table, the coefficient related to the causality of ES on HFT, $b_{21,i}$, has a small mean of 0.001; and the corresponding t-statistic has a small mean of 0.039, with a bootstrapped p-value of 0.341. This suggests that the mean of the sample t-statistic is within the normal range of what one would expect under the null of no causality. In addition, the p-values for the median and 1st and 3rd quintiles are in the range of 0.14 to 0.70. Overall, this suggests that there is no pervasive support to the hypothesis that institutional trading cost Granger-causes HFT.

In sum, the Granger causality tests provide further confirmation that more intensive HFT activities lead to an increase in institutional trading costs, but not vice versa.

2.5 Further Analysis of HFT activities

The analysis in this section consists two parts. The first part includes two sets of robustness results, based on the timing delay component of trading costs and on trade-level regression analysis. The second part includes two sets of results on the specific mechanisms through which HFT impacts institutional trading costs.

2.5.1 Robustness: Timing delay costs and trade-level regressions

I have provided evidence that intensive HFT activities lead to an increase in institutional investors' execution shortfall. This finding suggests that even though HFT improves the overall market quality, as documented in current literature, it causes additional trading costs for institutional investors. A natural question to ask is whether improved market quality may benefit institutional investors in some other ways, and to some extent offset the increase in trading costs. Considering the large amount of quotes sent by HF traders, one possible benefit to institutional investors may be that the costs incurred while waiting for liquidity may go down. Here, I perform analysis to address this possibility.

The cost incurred while seeking liquidity is known as timing delay in the literature. The specific measure of the timing delay cost is defined in Equation (2.2). To study the impact of HFT on timing delay, I estimate the following panel regression model:

$$\text{Timing Delay}_{it} = \alpha_i + y_t + a \times \text{HFT Intensity}_{it} + b \times X_{it} + \epsilon_{it} \tag{2.5}$$

where α_i are the firm-fixed effects, the y_t are day dummies, and HFT Intensity$_{it}$ is the measure of daily HFT activity on stock i as describe in subsection 2.3.1, Timing Delay$_{it}$ is the volume-weighted average timing delay of all institutional trades on stock i at day t, and X_{it} represents the same set of control variables as in Equation (2.3).

Table 7 presents the estimates of coefficients, with t-statistics computed using the two-way (by stock and by day) clustered standard errors. The regression model is estimated with both day dummies and firm-fixed effects. The coefficient of HFT Intensity is insignificant, which suggests that after controlling for other economic determinants of trading costs, HFT activity has no effect on the timing delay costs of institutional investors. Thus, while HFT activity increases institutional investors' execution shortfall, it does not provide the benefit of reduced timing delay costs.

So far, I conduct all the multivariate panel regression analyses at the stock-day level, where execution shortfall costs are aggregated for each stock on each trading day. The aggregation at stock-day level provides a strong indication that HFT increases institutional trading costs. However, one factor may be missing in the analysis of the data at the stock-day level, which is the difference in the trading skills of institutional investors. As pointed out by Anand et al. (2012), some institutions consistently execute trades with lower execution shortfalls than the others. If trades are executed by different institutions at different days on different stocks, the heterogeneity of institutional trading skills likely influences the aggregated measure of trading costs at stock-day level. To control for this factor, I estimate the following regression model based on trade-level observations:

$$\text{Execution Shortfall}_{i,j,t} = \alpha_j + \gamma_m + a \times \text{HFT Intensity}_{it} + b \times X_{it} + \epsilon_{it} \qquad (2.6)$$

where Execution Shortfall$_{i,j,t}$ is the execution shortfall of each trade (referred to as a “ticket” in the Ancerno data) for stock i on day t by institution j. α_j represents the institution-fixed effects, and γ_m represents the time(month)-fixed effects. X_{it} represents the same set of control variables as in Equation (2.3).

Table 8 presents the estimates of coefficients, with the t-statistics computed using the two-way clustered standard errors. The coefficient of HFT Intensity is positive and significant at the 1% level. This suggests that after controlling for heterogenous institutional trading skills, HFT increases execution shortfall at the trade level, consistent with the conclusion drawn from stock-day level analysis.

2.5.2 When and how does HFT impact institutional trading costs

In this subsection, I investigate two specific conjectures related to the mechanisms via which HFT affects institutional trading costs. The first is that HFT may profit from providing liquidity to institutions when the latter have large buy-sell imbalance among themselves. The second is that HF traders front run institutional investors' large trades.

I first investigate the possibility that HFT profits from providing liquidity to traditional institutional investors when the latter have large trade imbalances. If this notion of liquidity provision turns out to be true in the data, then the profits made by HF traders in a way resemble the profits made by traditional market makers. After all, electronic market making is an important form of HF strategies. However, even in this case, it is important to question whether the liquidity provision by HFT comes with extra costs to institutional investors, given the same level of trade imbalances among the institutions.

To begin with, I compare the daily buy-sell imbalance of the two types of investors–institutional investors and HF traders. I define the daily institutional imbalance on each stock as the buy dollar volume minus sell dollar volume of all institutions (HF traders) normalized by the stock's average daily trading volume over the prior 30 days. Panel A of Table 9 presents the distribution of such buy-sell imbalances for the sample stocks from 2008 to 2009. The table shows that while the daily imbalance by traditional institutional investors exhibits large variations, the daily imbalance for HF traders is mostly very close to zero. This contrast is consistent with the notion that institutional investors trade on information or mispricing that may pay off over a relatively long horizon, while HF traders profit mostly from price swings at very short horizons.Both anecdotal evidence and academic researchers have suggested that holding overnight positions can be very costly for HF traders (e.g., Menkveld (2010)).

Next, I use sorted portfolios to examine the relation of institutional buy-sell imbalance with both HFT activity and HFT buy-sell imbalance. Specifically, within each of the three size group, I sort stocks into terciles based on institutional buy-sell imbalance, and examine the average HFT Intensity and average HF buy-sell imbalance across the nine groups.

Panel B and C of Table 9 report the average institutional buy-sell imbalance and HFT buy-sell imbalance in each of the nine groups, respectively. The numbers suggest that despite the large swings of institutional imbalances, the imbalances of HF traders tend to be very small. This is consistent with the statistics reported in Panel A on HF trade imbalances. Finally, Panel D shows that when institutions exhibit buy-sell imbalance on either the buy or sell side, HFT Intensity becomes higher relative to the case when institutional trades are balanced.

Combining results from all panels of Table 9, one can make the following inferences. First, HFT becomes more active when institutions encounter large trade imbalances; presumably this is consistent with a liquidity provision role played by HF traders. However, the results in Panel C suggest that HF traders have minimum trade imbalances at the end of a trading day. Thus, if they provide liquidity to institutions, such liquidity provision is quite ephemeral – within a day, literally. Therefore, a more accurate description of the liquidity provision role of HF traders is that they serve as intra-day intermediaries and quickly pass the imbalances from institutions to other market participants.

We then investigate another important question regarding the liquidity provision role of HF traders. Our analysis in Table 3 shows that institutional trading costs are higher when institutions face large trade imbalances. If the presence of HFT reduces institutional trading costs on such occasions, then liquidity provision by HFT has a socially beneficial element. On the other hand, if the presence of HFT increases trading costs on such occasions, it is likely that HF traders are successful in taking advantage of institutional investors when the latter face large trade imbalances.

To address this question, I examine the differential impact of HFT on execution shortfall when institutions are net sellers, net buyers, or trading with relative balance. Specifically, I divide all stock-days into three groups based on institutional buy-sell imbalance, and then estimate the panel regression model specified in Equation (2.3) within each group. The results are reported in Table 10. The first two columns of the table report results when institutions are net selling. The coefficient of HFT Intensity is negative but not significant at the 5% level, suggesting that HFT activity does not hurt institutional investors significantly when the latter are net selling. The middle two columns report results when institutional trading is relatively balanced. The coefficient of HFT Intensity is 0.524 and significant at the 5% level, suggesting that HFT activity significantly increases institutional investors' trading costs when their trading is balanced. The most striking results are reported in the last two columns, for the case when institutional investors are net buyers. The coefficient of HFT Intensity is 0.612 and significant at the 1% level, which suggests that the impact of HFT activity on execution shortfall is most pronounced when institutional investors are net buyers. Overall, there is no evidence that HFT helps reduce trading costs when institutional investors have large trade imbalances; rather, HF traders appear to have successfully taken advantage of institutions when the latter are net buyers on a stock, making their trades extra costly.

In sum, the evidence presented in this part of the analysis suggests that HFT serves as a sort of intraday liquidity providers to institutions when the latter have large buy-sell imbalance among themselves; however such liquidity provision is extra costly to institutions, especially when they are net buyer of a stock.

I now turn to the second conjecture, that is, HF traders use certain strategies (e.g., front-running) to take advantage of institutional investors and increase the latter's trading costs. Here, I rely on the non-randomness, or sequences and reversals, of HF trade directions to detect the presence of HF strategies. For example, if HF traders engage in electronic market making, a type of HFT strategy considered to provide liquidity to the market, they have to buy and sell the same stocks very fast so that one should observe rapid reversals of trade directions. In contrast, directional trading strategies such as momentum ignition and front-running large institutional orders typically involve long sequences of trades in the same direction.

The non-randomness of HF trading is tested using the runs test on all trades made by HF traders on a stock on a given day. The runs test has been used in early studies on the random walk properties of stock prices (e.g., Fama (1965) and Campbell et al. (1970)). In the context of this study, I create a trading direction variable that equals 1 if an HF trader is on the buy side of a trade and -1 otherwise. I then use the runs statistic to test the null hypothesis of randomness in the sequence of HF trade directions at the stock-day level.[10] A negative and significant runs test statistic indicates frequent reversals in trade directions, an indication of market making strategies in play. A positive and significant test statistic means the popularity of sequential trades in the same direction, an indication directional trading strategies in use.

Based on the one-way critical value at the 2.5% level (i.e., -1.96 and 1.96), I identify 18506 cases at the stock-day level where the runs statistics are significantly positive, 18195 cases where the runs statistics are significantly negative, and 18262 cases of insignificant runs statistics. This translates into approximately one-third of stock-day cases where directional HF strategies are detected, and approximately one-third of cases where market making strategies are detected. Such high frequencies are striking; if HF trades are random, one would expect the significant cases to be only 2.5% in each direction. Therefore, both market making and directional trading are important strategies employed by HF traders.

The important question is what these strategies mean to the trading costs of institutional investors. To address this question, I perform panel regressions following the model specified in Equation (2.3), but separately for the cases where the runs tests at stock-day level are significantly positive, significantly negative, and insignificant. The results are presented in Table 11. First, as shown in the first two columns of

[10]Runs test is also known as the Wald-Wolfowitz test and is used to test the hypothesis that a series of numbers is random. A run is a series of numbers below or above the benchmark. The test statistic is: $Z = (R - E(R))/\sqrt{(V(R))}$, where R is the number or runs, $E(R)$ and $V(R)$ are expectation and variance of R. The test statistic is asymptotically normally distributed; see Wald and Wolfowitz (1940).

the table, when HF trades exhibit directional sequences (i.e., when the runs statistics are significantly positive), the coefficient of HFT Intensity is 0.409, significant at the 1% level. This result indicates that HF traders' use of directional trading strategies significantly increases the execution shortfall of institutional investors. Second, as shown in the middle two columns of the table, when HF trades exhibit frequent reversals, the coefficient of HFT Intensity is 0.291, significant at the 5% level. This suggests that the electronic market making strategies employed by HF traders also increases institutional trading costs, although at a smaller magnitude relative to the case when HF traders engage in direction trading. Finally, the results reported in last two columns of the table show that when neither directional trading nor market making strategies are detected (i.e., when the runs statistics are insignificant), HFT Intensity does not have a significant impact on institutional trading costs (with a coefficient of 0.196 and a t-statistic of 1.64).

2.6 Conclusions

This paper fills a gap in the literature by directly examining the impact of HFT on the trading costs of institutional investors in the U.S. market. To establish the relation, I first construct daily measures of trading costs and HFT activity during 2008 and 2009 from two datasets. I obtain daily measures of HFT activity from a dataset of 120 stocks, representing a subset of HFT activity, which NASDAQ makes available to academics. To measure trading costs I use a proprietary database of institutional investors' equity transactions compiled by Ancerno.

Using direct measures of institutional trading costs and daily HFT activity on each of 120 sample stocks, I conduct a sorted portfolio test and a panel regression with control for various firm characteristics. I find strong evidence that an increase in HFT is associated with an increase in the trading costs of institutional investors. The regression result suggests that a one standard deviation increase of HFT activity leads to an additional trading cost of more than $10,000 per day for an average institution in the dataset. I also find that this incremental effect of HFT on execution shortfall is stronger on smaller stocks.

I adopt a variety of approaches to rule out the alternative interpretation that it is precisely when execution shortfall is high that it is more profitable for HF traders to trade more aggressively. First, the sorted portfolio analysis indicates that HF traders are most active in liquid stocks, rather than in illiquid stocks which tend to have high trading costs. Second, I include firm- and time-fixed effects in the multivariate regression specification, which helps ensure that unobserved slow-moving stock characteristics and time-invariant factors do not cause the positive relationship between HFT activity and execution shortfall. Third, I control for corporate events such as earnings announcements and M&A announcements and the results still holds. Fourth, I use the short selling ban imposed on financial stocks on September 19, 2008 as an exogenous shock to execution shortfall. I find that for the stocks in my sample that are subject to the short selling ban, HF traders' market participation rate declined while institutional trading costs rose sharply. Fifth, I apply the Granger causality test to establish the direction of causality between HFT activity and execution shortfall. The results provide further evidence that intensive HFT activity contributes to an increase in trading costs, but not vice versa.

I perform further analysis to understand the mechanisms via which HFT affects institutional trading costs. My analysis shows that HFT provides liquidity to the market when institutions have large trade imbalances. However, the liquidity provision by HFT is short-lived as HF traders maintain zero open positions at market close. And such liquidity provision proves particularly expensive for institutions in terms of their trading costs. My analysis also shows the prevalence of both directional strategies and market making strategies used by HF traders. The presence of either type of strategies results in increased institutional trading costs; but the impact is most pronounced when the directional trading strategies are in use. This lends support to the anecdotal observations among institutional investors that their trades have been front-run by HF traders.

In sum, the evidence provided in this paper suggests a significant impact of HFT on traditional institutional investors. An increase in HF traders' participation rate is associated with higher trading costs for institutional investors. This finding underscores the need for further investigation into the broader impact of the rapid growth in high frequency trading, particularly in terms of its implications for long term investors.

Table 2.1: Summary statistics

	All	Large Cap	Mid Cap	Small Cap
Average Market Capital ($billion)	17.500	46.780	1.590	0.400
Average HFT Total Trading Volume (million)	54.570	158.230	3.650	0.380
Average Execution Shortfall (%)	0.167	0.146	0.163	0.196
Amihud Illiquidity Ratio	0.006	7.6E-05	0.002	0.019
Average Institutional Order Size	244,286	487,871	154,823	63,943
Average Trades Per Order	2.303	3.126	1.861	1.850

Note: this table reports the averages of stock characteristics, HFT activity, and execution shortfall of all stock-days, as well as the averages by market capital, during the periods of 2008 and 2009. All the variables are measured on a daily basis. Market Capitalization is a stock's market value. HFT Total Trading Volume is the daily total trading volume of HFT on a stock. Average Execution Shortfall is the volume-weighted average execution shortfall of all institutional trades on a stock. Amihud Illiquidity Ratio is the daily absolute return divided by the dollar trading volume on that day. Average Institutional Order Size is the average dollar volume of all institutional trades placed on a stock. Average Trades Per Order is the average number of trades to complete an order ("ticket") on a stock.

Table 2.2: Determinants of HFT

Dependent Variable	HFT Intensity	
	Coefficient	t-value
Intercept	-0.179	(-3.65)
Log Market Cap	0.022	(6.75)
Book-to-Market Ratio	-3.080	(-1.92)
Event Dummy	0.058	(10.89)
Daily Return Volatility	0.098	(1.98)
Prior 1-day Return	0.192	(6.83)
Prior 1-month Return	-0.003	(-0.36)
Prior 12-month Return	-0.010	(-2.69)
Amihud Illiquidity Ratio	-0.570	(-3.20)
Daily Dollar Turnover	0.036	(3.18)
Average Institutional Order Size	-0.161	(-1.69)
Absolute Institutional Imbalance	0.132	(3.80)
Average Trades Per Order	0.000	(2.06)
Prior 1-month Market Volatility	-0.003	(-0.24)
Prior 1-day Market Return	-0.376	(-3.98)
Day-fixed Effects	No	
Stock-fixed Effects	No	
Two-way Clustered Standard Deviations	Yes	
Adjusted R-squared (%)	29.2	
Number of Observations	52809	

Note: this table reports the determinants of HFT intensity based on panel regressions. The dependent variable is HFT Intensity. The explanatory variables include the following. Log Market Cap is the logarithm of a stock's daily market capitalization. Book-to-Market Ratio is the quarterly book-to-market ratio. Event Dummy is a dummy variable that equals one for a stock within a 5-day window of corporate events (earnings announcement or M&A announcement), and zero otherwise. Daily Return Volatility is a stock's annualized range based daily volatility. Prior 1-day Return is a stock's lagged daily return. Prior 1-month Return is a stock's lagged monthly return. Prior 12-month Return is a stock's lagged 12 months return. Amihud Illiquidity Ratio is the ratio of the daily absolute return to the dollar trading volume on a trading day. Daily Dollar Turnover is a stock's daily dollar trading volume scaled by the stock's total shares outstanding. Average Institutional Order Size is the average dollar volume of all tickets placed on a stock on a trading day, scaled by the average trading volume of that stock in prior 30 days. Absolute Institutional Imbalance is the absolute value of the daily total dollar volume of all institutional buy trades minus that of all sell trades on a stock on a trading day, scaled by the average trading volume of that stock in the past 30 days. Average Trades Per Order is the average number of trades to complete a trading ticket on a stock for a trading day. Prior 1-month Market Volatility is the market's annualized monthly return volatility in prior month. Prior 1-day Market Return is the market return in prior day. The t-statistics are computed using two-way (by stock and by day) clustered standard errors.

Table 2.3: HFT's impact on Execution Shortfall

Dependent Variable	Execution Shortfall		Execution Shortfall	
	Coefficient	t-value	Coefficient	t-value
Intercept	0.025	(0.24)	-1.144	(-1.77)
HFT Intensity	0.336	(4.48)	0.309	(3.37)
Log Market Cap	-0.004	(-0.66)	0.043	(1.08)
Book-to-Market Ratio	-5.978	(-0.95)	6.303	(1.23)
Prior 1-day Return	-0.072	(-0.24)	-0.178	(-0.64)
Prior 1-month Return	0.017	(0.25)	-0.037	(-0.69)
Prior 12-month Return	0.013	(0.92)	-0.004	(-0.26)
Amihud Illiquidity Ratio	3.955	(3.14)	4.687	(3.36)
Daily Return Volatility	0.324	(1.42)	0.046	(0.30)
Daily Dollar Turnover	-0.007	(-1.66)	-0.001	(-0.19)
Average Institutional Order Size	0.743	(1.37)	0.735	(1.42)
Absolute Institutional Imbalance	0.271	(2.56)	0.281	(2.67)
Average Trades Per Order	0.000	(0.16)	0.000	(-0.44)
Prior 1-month Market Volatility	0.285	(3.24)		
Prior 1-day Market Return	-0.031	(-0.05)		
Day-fixed Effects	No		Yes	
Stock-fixed Effects	No		Yes	
Two-way Clustered Standard Deviations	Yes		Yes	
Adjusted R-squared (%)	0.69		3.47	
(Number of Observations)	54963		54963	

Note: this table reports the results of panel regressions that examine the impact of HFT intensity on the execution shortfall costs of institutional investors. The dependent variable is Execution Shortfall, the volume-weighted average execution shortfall of all institutional trades on a stock for a trading day. The main explanatory variable, HFT Intensity, is the total daily trading volume of HFT on a stock for a trading day scaled by the average trading volume of that stock in the prior 30 days. The control variables include the following. Log Market Cap is the logarithm of a stock's daily market capitalization. Book-to-Market Ratio is the quarterly book-to-market ratio. Stock Volatility is a stock's annualized range based daily volatility. Prior 1-day Return is a stock's lagged daily return. Prior 1-month Return is a stock's lagged monthly return. Prior 12-month Return is a stock's lagged 12 months return. Amihud Illiquidity Ratio is the daily absolute return to the dollar trading volume on that day. Dollar Turnover is a stock's daily dollar trading volume scaled by the stock's total shares outstanding. Average Institutional Order Size is the average dollar volume of all tickets placed on a stock, scaled by the average trading volume of that stock in prior 30 days. Absolute Institutional Imbalance is the absolute value of the daily total dollar volume of all institutional buy tickets minus that of all sell tickets on a stock, scaled by the average trading volume of that stock in the past 30 days. The first two columns report the panel regression results with only day-fixed effects but no stock-fixed effects. The last two columns report the panel regression results with both day- and stock-fixed effects. The t-statistics are computed using two-way (by stock and by day) clustered standard errors.

Table 2.4: HFT's impact on execution shortfall across stock size

Dependent Variable	Execution Shortfall					
	Large Stocks		Mid Stocks		Small Stocks	
	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value
Intercept	-1.078	(-1.97)	-1.101	(-1.32)	-2.125	(-2.16)
HFT Intensity	0.188	(1.99)	0.152	(1.56)	0.622	(2.49)
Log Market Cap	0.063	(1.90)	0.059	(0.92)	0.145	(1.76)
Book-to-Market Ratio	114.861	(1.33)	-46.562	(-0.97)	9.820	(1.75)
Prior 1-day Return	-0.098	(-0.30)	-0.402	(-1.23)	-0.060	(-0.10)
Prior 1-month Return	0.083	(1.51)	-0.105	(-1.33)	-0.086	(-0.79)
Prior 12-month Return	-0.012	(-0.41)	0.028	(1.06)	-0.018	(-0.68)
Amihud Illiquidity Ratio	648.124	(4.25)	24.085	(4.19)	4.771	(3.32)
Daily Return Volatility	-0.055	(-2.81)	0.203	(0.40)	0.157	(0.30)
Daily Dollar Turnover	0.004	(0.56)	-0.002	(-0.07)	-0.016	(-0.17)
Average Institutional Order Size	14.595	(2.26)	2.848	(1.86)	0.909	(1.72)
Absolute Institutional Imbalance	1.064	(5.26)	0.126	(0.81)	0.163	(1.05)
Average Trades Per Order	-0.005	(-1.87)	0.007	(1.87)	-0.007	(-1.39)
Day-fixed Effects	Yes		Yes		Yes	
Stock-fixed Effects	Yes		Yes		Yes	
Two-way Clustered Std.	Yes		Yes		Yes	
Adjusted R-squared (%)	4.54		4		5.79	
Number of Observations	20119		18981		15863	

Note: this table report the results of panel regressions that examine the differential impact of HFT activity on execution shortfall for different stock size groups. The 120 stocks are divided into three groups based on their market capitalizations. The baseline regression model (as described in Table 3) is estimate within the three size groups, respectively. The regression model is estimated with both day- and stock-fixed effects. The t-statistics are computed using two-way (by stock and by day) clustered standard errors.

Table 2.5: HFT's impact on execution shortfall on event days and no-event days

Dependent Variable	Execution Shortfall	
	Coefficient	t-value
Intercept	-1.129	-(1.74)
HFT Intensity × Event Dummy	0.155	(1.29)
HFT Intensity × No-Event Dummy	0.375	(3.88)
Event Dummy	0.058	(1.39)
Log Market Cap	0.041	(1.03)
Book-to-Market Ratio	6.284	(1.23)
Prior 1-day Return	-0.181	-(0.65)
Prior 1-month Return	-0.037	-(0.70)
Prior 12-month Return	-0.005	-(0.31)
Amihud Illiquidity Ratio	4.711	(3.37)
Daily Return Volatility	0.039	(0.26)
Daily Dollar Turnover	0.002	(0.24)
Average Institutional Order Size	0.725	(1.40)
Absolute Institutional Imbalance	0.285	(2.69)
Average Trades Per Order	0.000	-(0.49)
Day-fixed Effects	Yes	
Stock-fixed Effects	Yes	
Two-way Clustered Standard Deviations	Yes	
Adjusted R-squared (%)	3.49	
Number of Observations	54963	

Note: this table reports the results of panel regressions that examine the differential impact of HFT activity on the execution shortfall on days with and without corporate events. Event Dummy is a dummy variable that equals one for a stock within a 5-day corporate event window (earnings announcement or M&A announcement), and zero otherwise. No-Event Dummy is a dummy variable that equals zero for a stock not within a corporate event window, and zero otherwise. All other variables are defined in Table 3. The regression model is estimated with both day- and stock-fixed effects. The t-statistics are computed using two-way (by stock and by day) clustered standard errors.

Table 2.6: Granger causality

Panel A: Distribution of $b_{12,i}$				
	Q1	Mean	Median	Q3
Sample Coefficients	-0.215	0.317	0.117	0.486
Sample t-statistic	(-0.456)	(0.311)	(0.265)	(0.977)
Bootstraped p-value	[0.043]	[0.002]	[0.010]	[0.008]
Panel B: Distribution of $b_{21,i}$				
	Q1	Mean	Median	Q3
Sample Coefficients	-0.002	0.001	0.000	0.002
Sample t-statistic	(-0.725)	(0.039)	(-0.031)	(0.793)
Bootstraped p-value	[0.695]	[0.341]	[0.583]	[0.141]

Note: this table reports the result of the Granger-causality test on the relation between HFT Intensity and Execution Shortfall. The following VAR(1) model is estimated for each stock:

$$\begin{pmatrix} ES_{i,t} \\ HFT_{i,t} \end{pmatrix} = \begin{pmatrix} a_{1,i} \\ a_{2,i} \end{pmatrix} + \begin{pmatrix} b_{11,i} & b_{12,i} \\ b_{21,i} & b_{22,i} \end{pmatrix} \begin{pmatrix} ES_{i,t-1} \\ HFT_{i,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1,i,t} \\ \epsilon_{2,i,t} \end{pmatrix},$$

where $ES_{i,t}$ and $HFT_{i,t}$ are the Execution Shortfall and HFT Intensity for stock i on day t, respectively. The table reports the cross-sectional distribution (mean, median, the 1st and 3rd quartiles) of the coefficients $b_{12,i}$ and $b_{21,i}$ across 120 stocks, and the cross-sectional distribution of the t-statistics for these two coefficients. The p-values reported in the table are obtained via a bootstrapping procedure to assess the statistical significance of these cross-sectional statistics. The bootstraps are performed under the null of no causality (i.e., $b_{12,i} = b_{21,i} = 0$) but retain the time-series persistence of each variables in the sample, the correlations of the residuals $\epsilon_{1,i,t}$ and $\epsilon_{2,i,t}$ for a given stock, as well as the cross-stock correlations of these residuals. The bootstrapped p-values are calculated as the percentages of bootstrapped distributional statistics (e.g., mean, median, Q1 and Q3) of the t-statistics for the estimated coefficients exceed the corresponding sample distributional statistics.

Table 2.7: HFT's impact on timing delay costs

Dependent Variable	Execution Shortfall	
	Coefficient	t-value
Intercept	-0.127	(-0.96)
HFT Intensity	0.115	(2.90)
Log Market Cap	-0.005	(-1.35)
Book-to-Market Ratio	-0.617	(-0.08)
Prior 1-day Return	0.165	(0.55)
Prior 1-month Return	-0.020	(-0.28)
Prior 12-month Return	0.003	(0.27)
Amihud Illiquidity Ratio	2.543	(2.16)
Daily Return Volatility	-0.073	(-0.60)
Daily Dollar Turnover	-0.002	(-1.26)
Institutional Order Size	1.467	(6.58)
Absolute Institutional Imbalance	0.037	(0.57)
Trades Per Order	0.000	(0.01)
Month-fixed Effects	Yes	
Institution-fixed Effect	Yes	
Two-way Clustered Standard Deviations	Yes	
Adjusted R-squared (%)	1.13	
Number of Observations	1689919	

Note: this table reports the results of panel regressions that examine the impact of HFT activity on the timing delay costs of institutional investors. The dependent variable, Timing Delay Cost, is the volume-weighted average timing delay costs of all institutional trades on a stock for a trading day. All the other variables are defined in Table 3. The regression model is estimated with both day- and stock-fixed effects. The t-statistics are computed using two-way (by stock and by day) clustered standard errors.

Table 2.8: Trade-level analysis of HFT's impact on execution shortfall

Dependent Variable	Execution Shortfall	
	Coefficient	t-value
Intercept	-0.127	(-0.96)
HFT Intensity	0.115	(2.90)
Log Market Cap	-0.005	(-1.35)
Book-to-Market Ratio	-0.617	(-0.08)
Prior 1-day Return	0.165	(0.55)
Prior 1-month Return	-0.020	(-0.28)
Prior 12-month Return	0.003	(0.27)
Amihud Illiquidity Ratio	2.543	(2.16)
Daily Return Volatility	-0.073	(-0.60)
Daily Dollar Turnover	-0.002	(-1.26)
Institutional Order Size	1.467	(6.58)
Absolute Institutional Imbalance	0.037	(0.57)
Trades Per Order	0.000	(0.01)
Month-fixed Effects	Yes	
Institution-fixed Effect	Yes	
Two-way Clustered Standard Deviations	Yes	
Adjusted R-squared (%)	1.13	
Number of Observations	1689919	

Note: this table reports the results of trade-level panel regressions that examine the impact of HFT activity on institutional execution shortfall. The dependent variable, Execution Shortfall, is measured for each trade. Institutional Order Size is the dollar volume of an institutional trading ticket, scaled by the average trading volume of that stock in the past 30 days. Trades Per Order is number of executions used to complete a ticket. All the other variables are the same as described in Table 3. The linear regression model is estimated with both month- and institution-fixed effects. The t-statistics are computed using two-way clustered standard errors.

Table 2.9: HFT and institutional buy-sell imbalances

Panel A: Distribution of HFT and institution buy-sell imbalance				
	Q1	Mean	Median	Q3
HFT Buy-Sell Imbalance	-0.009	0.000	0.000	0.009
Institution Buy-Sell Imbalance	-0.022	0.003	0.001	0.024

Panel B: Institutional buy-sell imbalance			
	Institutions net selling	Institutions balanced	Institutions net buying
Large Stocks	-0.062	0.000	0.060
Mid Stocks	-0.104	0.002	0.106
Small Stocks	-0.116	0.002	0.138
Panel C: HFT Intensity			
	Institutions net selling	Institutions balanced	Institutions net buying
Large Stocks	0.246	0.226	0.255
Mid Stocks	0.171	0.151	0.166
Small Stocks	0.093	0.082	0.095
Panel D: HFT buy-sell imbalance			
	Institutions net selling	Institutions balanced	Institutions net buying
Large Stocks	0.001	0.000	-0.001
Mid Stocks	0.003	0.000	-0.002
Small Stocks	0.002	-0.001	-0.002

Note: this table reports the results of analysis on the relations among institutional trade imbalances, HFT intensity, and HFT trade imbalances. Institutional (HFT) trade imbalance is the buy volume minus sell volume of all institutions (HF traders) normalized by the stock's average daily trading volume over the prior 30 days. HFT Intensity, is the total daily trading volume of HFT on a stock for a trading day scaled by the average trading volume of that stock in the prior 30 days. Panel A reports the sample distribution of institutional trade imbalances and HFT trade imbalances. Panel B reports the institutional trade imbalances for nine groups of stocks classified by size and institutional trade imbalances. Panel C report the HFT Intensity for the same nine groups of stocks. Panel D reports the HFT trade imbalances for the same nine groups of stocks.

Table 2.10: HFT's impact on execution shortfall when institutional trading is imbalanced

Dependent Variable	Execution Shortfall					
	Institutions net selling		Institutions balanced		Institutions net buying	
	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value
Intercept	3.176	(2.82)	-1.525	(-1.37)	-2.473	(-2.86)
HFT Intensity	-0.178	(-1.77)	0.524	(2.24)	0.612	(4.78)
Log Market Cap	-0.198	(-2.79)	0.083	(1.18)	0.177	(2.36)
Book-to-Market Ratio	32.267	(2.72)	1.584	(0.34)	24.048	(2.41)
Prior 1-day Return	0.594	(1.50)	-0.603	(-0.99)	-0.438	(-1.01)
Prior 1-month Return	0.157	(1.40)	-0.041	(-0.45)	-0.172	(-1.22)
Prior 12-month Return	0.065	(1.76)	0.019	(0.61)	-0.054	(-1.31)
Amihud Illiquidity Ratio	3.236	(1.38)	2.657	(0.96)	-0.054	(-1.31)
Daily Return Volatility	0.312	(0.89)	-0.174	(-0.75)	0.043	(0.20)
Daily Dollar Turnover	0.024	(2.36)	-0.009	(-1.05)	-0.015	(-2.14)
Average Institutional Order Size	0.528	(1.04)	0.531	(0.30)	0.858	(1.11)
Absolute Institutional Imbalance	0.359	(3.12)	7.783	(2.02)	0.258	(2.09)
Average Trades Per Order	0.000	(0.12)	-0.003	(-1.28)	0.000	(-0.23)
Day-fixed Effects	Yes		Yes		Yes	
Stock-fixed Effects	Yes		Yes		Yes	
Two-way Clustered Std.	Yes		Yes		Yes	
Adjusted R-squared (%)	12.2		16.1		8.96	
Number of Observations	18362		18398		18203	

Note: this table reports the results of panel regressions that examine the differential impact of HFT on execution shortfall when institutions are net selling, net buying, or trading with balance. All stock-days are divided into three groups based on Institutional Buy-Sell Imbalance. The baseline regression model (as described in Table 3) is estimate within each group, respectively. The linear regression model is estimated with both day and firm-fixed effects. The t-statistics are computed using two-way (by stock and by day) clustered standard errors.

Table 2.11: Impact of HFT strategies on execution shortfall

Dependent Variable	Execution Shortfall					
	Directional		Market Making		Random Walk	
	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value
Intercept	0.143	(0.14)	0.217	(-0.18)	-1.371	(-1.44)
HFT Intensity	0.409	(2.60)	0.291	(1.94)	0.196	(1.64)
Log Market Cap	-0.019	(-0.30)	0.054	(0.69)	0.093	(1.63)
Book-to-Market Ratio	10.538	(2.49)	2.742	(0.42)	-2.678	(-0.15)
Prior 1-day Return	0.075	(0.21)	-0.339	(-0.65)	-0.316	(-0.66)
Prior 1-month Return	-0.019	(-0.20)	0.046	(0.38)	-0.130	(-1.48)
Prior 12-month Return	0.038	(1.38)	0.001	(0.03)	-0.026	(-0.85)
Amihud Illiquidity Ratio	9.170	(4.61)	5.798	(2.43)	2.208	(1.09)
Daily Return Volatility	-0.213	(-1.41)	0.172	(0.69)	0.223	(0.62)
Daily Dollar Turnover	-0.024	(-1.70)	0.004	(0.43)	0.010	(0.92)
Average Institutional Order Size	1.275	(0.95)	-0.903	(-1.45)	1.525	(2.95)
Absolute Institutional Imbalance	0.220	(1.18)	0.595	(3.83)	0.135	(0.88)
Average Trades Per Order	0.000	(-0.53)	0.000	(-0.20)	-0.001	(-0.39)
Day-fixed Effects	Yes		Yes		Yes	
Stock-fixed Effects	Yes		Yes		Yes	
Two-way Clustered Std.	Yes		Yes		Yes	
Adjusted R-squared (%)	3.45		4.02		3.98	
Number of Observations	18506		18195		18262	

Note: this table reports the results of panel regressions that examine the differential impact of HFT on execution shortfall when different types of HF strategies are in detected. Stock-day observations are divided into three groups based on the non-randomness of HF trades. The non-randomness of HF trades is measured by runs tests on all HF trades on a stock on a given day. The regression model (as described in Table 3) is estimate within each group, respectively, with both day- and stock-fixed effects. The t-statistics are computed using two-way (by stock and by day) clustered standard errors.

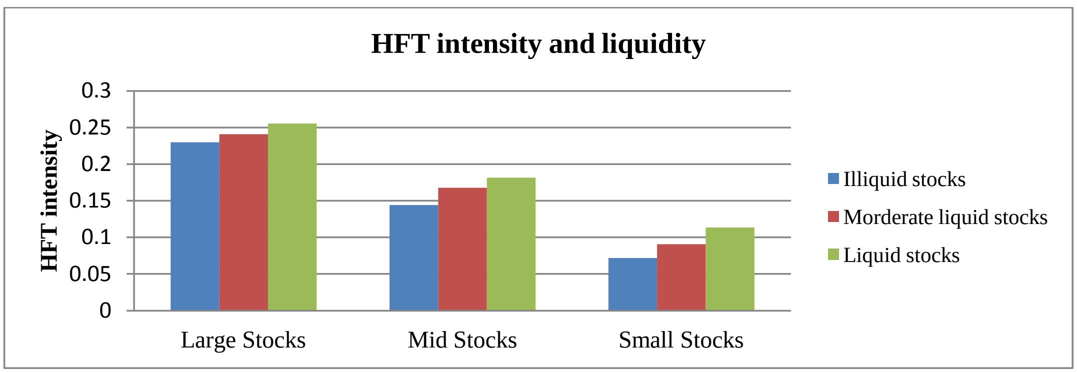


Figure 2.1: Relation between HFT intensity and liquidity.
Note: this figure plots the HFT Intensity for different levels of liquidity in each of the three size groups. Liquidity is measured by Amihud Illiquidity Ratio. HFT Intensity is the total daily trading volume that HFT involves on a stock scaled by the average trading volume of that stock in the prior 30 days. Each day, I sort all stocks into three portfolios based on their size. Then each portfolio is further divided into three groups based on Amihud Illiquidity Ratio. The columns in the figure represent the average HFT Intensity in each group.

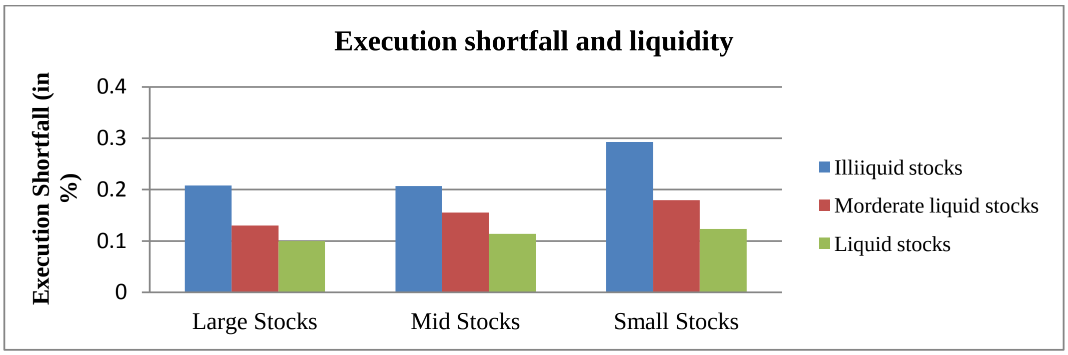


Figure 2.2: Relation between liquidity and execution shortfall
Note: this figure plots the Execution Shortfall for different levels of liquidity in each of the three size groups. Liquidity is measured by Amihud Illiquidity Ratio. Execution Shortfall is the volume-weighted average execution shortfall of all institutional trading tickets on a stock. Each day, I sort all stocks into three portfolios based on their size. Then each portfolio is further divided into three groups based on the Amihud Illiquidity Ratio. The columns in the figure represent the average Execution Shortfall in each group.

Figure 2.3: Relation between HFT intensity and execution shortfall

Note: this figure plots the Execution Shortfall for different levels of HFT Intensity in each of the three size groups. Execution Shortfall and HFT Intensity are defined in Figure 1 and 2. Each day, I sort all stocks into three portfolios based on their size. Then each portfolio is further divided into three groups based on HFT Intensity. The columns in the figure represent the average Execution Shortfall in each group.

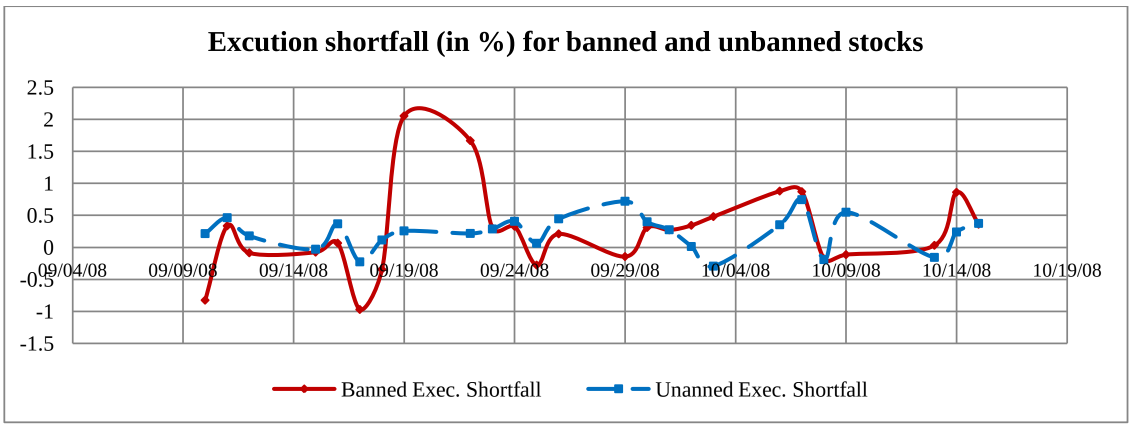


Figure 2.4: Execution shortfall around the Short-selling Ban of September 18, 2008

Note: this figure plots the time-series of the average Execution Shortfall for banned and unbanned stocks around the short selling ban period from September 18, 2008 to October 8, 2008. Execution Shortfall is the volume-weighted average execution shortfall of all institutional trading tickets on a stock. There are 13 stocks in my sample in the initial short selling ban list on 9/18/2008. On 9/22/2008, the list expanded to cover 16 stocks in the sample, and one more stock was added to the list on 9/23/2008.

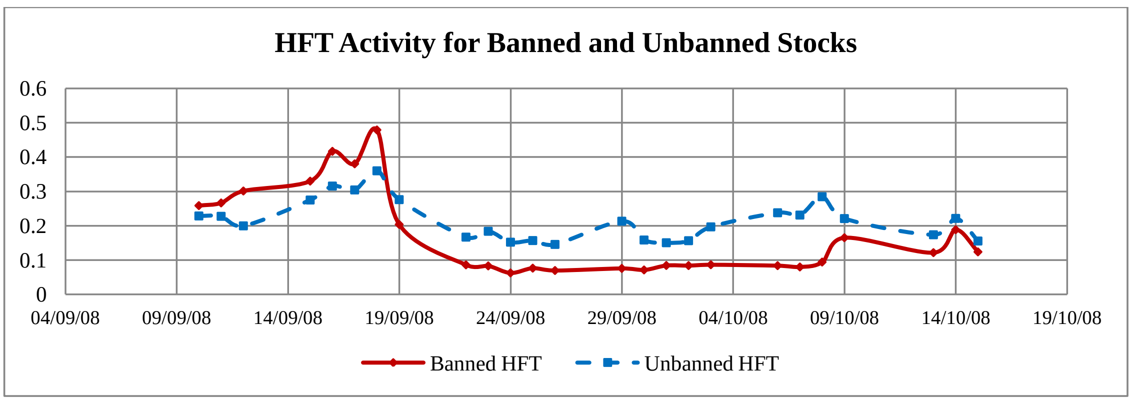


Figure 2.5: HFT activity around the Short-selling Ban of the September 18, 2008
Note: this figure plots the time-series of the average HFT Intensity for banned and unbanned stocks around the Short-selling Ban period from September 18, 2008 to October 8, 2008.

CHAPTER 3
MUTUAL FUND TOURNAMENTS AND FUND ACTIVE SHARE

3.1 Introduction

The tendency to chase past performance is one of the best known facts regarding the behavior of mutual fund investors.[1] Funds with superior relative performance within a category attract the lion's share of the cash inflows in the future. Consequently, economists have often viewed the competition in the fund industry as a tournament in which players are competing with each other for investor cash flows. A number of papers provide empirical evidence on the incentive effects of this tournament-like competition among funds. For example, Brown et al. (1996) find that funds that trail at the half-year mark tend to subsequently increase the volatility of their portfolios. Similarly, Chevalier and Ellison (1997) study the relationship between fund flows and past performance of mutual funds and conclude that the flow-performance relationship provides incentives for the funds to alter the risk of their portfolios at the interim stage.[2]

The more recent evidence on this issue, however, casts doubt on the earlier findings with respect to the adverse risk-taking incentives in mutual funds. For example, Busse (2001) analyzes daily returns for a sample of 230 equity funds over the period 1985–1995, but fails to find support for the hypothesis that fund managers

---

[1] A number of studies have documented a strong positive relation between a fund's past performance and its future fund inflows. See, for example, Ippolito (1992), Chevalier and Ellison (1997), Sirri and Tufano (1998), and Sapp and Tiwari (2004).

[2] See, also, Koski and Pontiff (1999).

actively alter the risk of their portfolios in response to past performance. However, he recognizes that "uncovering a more complex behavior pattern should be a fruitful area for future research" (p. 73). In this paper we re-examine the incentive effects created by the fund flow-performance relationship. We first analyze a simple model which suggests that the trailing funds have an incentive to strategically deviate from their relevant benchmark, while the leading funds have a similar incentive to closely track their respective benchmarks' allocation weights. The model suggests a direct measure of the extent of risk-taking by a fund, namely, a fund's Active Share.[3] Unlike the noisy measures of risk estimated from fund returns that are employed in earlier studies, the Active Share is a direct measure constructed from portfolio holdings, of the strategic (risk) choice of the fund manager. We find that consistent with our theoretical model, the empirical results confirm that the risk-taking incentives of fund managers are a function of (a) the fund cash flow sensitivity to past performance, i.e., the strength of the tournament effect, and (b) the magnitude of the interim performance gap separating the funds.

As a first step, we develop a framework to analyze the risk-taking incentives of fund managers. We consider a model in which two risk-neutral fund managers with exogenously determined unequal performances at an interim stage, compete for investor cash flows. Given the unequal performances of the managers at the interim stage, we may think of one manager as the interim winner, and the other as the

[3]The Active Share measure, proposed by Cremers and Petajisto (2009) is discussed in more detail in Section II.

interim loser. Managers' compensation is assumed to be a fixed proportion of the assets under management. Each manager has the choice of investing in a market index in addition to a security that represents pure idiosyncratic risk. Our focus is on the portfolio choices made by the two managers after they observe their interim performances.

We show that there exists an equilibrium in which it is optimal for the fund manager who is trailing behind at the interim stage (i.e., the interim loser) to increase the *activeness* of her portfolio by reducing the exposure to the market index. The intuition for the portfolio choices made by the two managers after their interim performance records become known is straightforward. The interim winner fund manager has an incentive to try to "lock in" her gains by maintaining an indexed position after it becomes known that she is ahead of her competitor. On the other hand, the optimal policy for the interim loser manager is to hold a portfolio that differs from that of the interim winner, as that is the only way she can hope to make up the initial performance shortfall. Importantly, our result regarding the interim loser manager's incentive to increase her portfolio's *activeness* differs from the previous empirical studies (e.g., Brown et al. (1996)) whose primary focus is on the interim loser's incentives to increase the *total* volatility of her portfolio.

Our theoretical analysis yields a number of testable predictions. In particular, it suggests that the increase in the *activeness* of the interim loser manager's portfolio is directly related to the magnitude of the performance gap at the interim stage, and to the strength of the investor (cash flow) response to the relative performance rankings of the funds (i.e., the strength of the tournament effect). The analysis also suggests that the incentive to increase active risk following poor performance is meaningful for only those funds that are within striking distance of the winning fund(s) at the interim evaluation stage, i.e., for funds for whom the performance deficit is not excessively large.

We use quarterly data for a sample of domestic stock funds for the period 1990-2008 obtained from the CRSP Survivor-Bias Free US Mutual Fund Database and the Active Share data [4] to test the key implications of our basic model. The funds in the sample are grouped into five categories, namely, Aggressive Growth (AGG), Growth (GRO), Growth and Income (GI), Income (I) and others. We use the change in the funds' Active Share measures during the last quarter of a particular year to capture their risk shifting incentives. The empirical evidence suggests that, consistent with the model, changes in the Active Share of the funds are negatively related to their relative return performance at the interim evaluation stage. The relationship is statistically significant for all funds and is more pronounced for funds with the greatest cash flow sensitivity to performance, namely, Aggressive Growth funds, as well as for the Growth funds. Our findings confirm that the incentives of fund managers to strategically change their portfolio risk profiles are related to the strength of the investor (cash flow) response to the relative performance rankings of the funds (i.e., the strength of the tournament effect). These results are broadly

[4]The Active Share data is constructed by Cremers and Petajisto (2009) and updated by Petajisto (2013). It can be downloaded at: http://www.petajisto.net/data.html.

consistent with the implications of our theoretical analysis. They also suggest that the failure to control for the performance gap at the interim evaluation stage, as well the use of noisy risk measures, may help explain the conflicting results in the earlier studies.

Our study contributes to a growing literature that examines managerial incentives and investor behavior in the mutual fund industry. On the theoretical front, a number of recent studies analyze the incentives facing mutual fund managers in a "tournament" framework (see, e.g., Bagnoli and Watts (2000), Goriaev et al. (2003), and Taylor (2003)). Our basic model is closest in spirit to that of Taylor (2003) who develops a model of a mutual fund tournament in which two fund managers with unequal midyear performances compete for new cash inflows. Taylor derives a mixed-strategy equilibrium for the case when both funds are actively managed. The pure strategies that are the basis for the mixed-strategies are those of the extreme form, i.e., they involve investing fully in risky assets or investing fully in the risk-free asset. In contrast, we allow the fund managers the option to take on idiosyncratic risk in addition to investing in the market portfolio and the risk free asset, and focus on the pure strategies equilibrium.

The use of tournaments involving relative performance evaluation as incentive devices has been studied by Lazear and Rosen (1981), Green and Stokey (1983), Nalebuff and Stiglitz (1983) and Meyer and Vickers (1997), among others.[5] Our study

[5]Lazear and Rosen (1981) examine the efficiency of three compensation schemes, namely, a linear piece rate, comparison against a fixed benchmark, and a tournament, in the context of labor contracts. In their modeling framework the output of each agent consists of the

complements this literature by focusing on a setting that involves *explicit* incentives for fund managers as well as the *implicit* incentives generated by the cash flow-performance relationship.

Other relevant papers in this area include Chen and Pennacchi (1996), Hu et al. (2011), and Basak et al. (2000). In the analytical framework of these papers, managers performance is measured against a benchmark. In this sense, the modeling framework of these papers does not involve a tournament structure within which competition among the players is a key feature and relative performance among peers is used in determining the reward. The model presented in this paper on the other hand, is based on a tournament structure of competition among fund managers.

The rest of the paper proceeds as follows. Section II discusses the Active Share measure employed in the study. Section III presents the theoretical model. Section IV presents empirical evidence on the key implications of our model while Section V offers concluding remarks.

3.2 Active Share

The Active Share measure, originally proposed by Cremers and Petajisto (2009), captures the degree of active stock selection by a fund manager. In essence, Active Share is the difference in the portfolio weights between a fund's portfolio holdings and its benchmark index. Thus, when a fund manager overweights (or underweights) a particular stock compared with the benchmark index, she is taking an

---

agent specific effort and an additive common shock. They show that the tournament scheme is the most efficient contracting arrangement when the variance of the common shock is relatively high.

active long (or short) position on the stock. Active Share is formally defined as:

$$\text{Active Share}_{i,t} = \frac{1}{2}\sum_{j=1}^{N}\left|\omega_{j,i,t} - \omega_{j,indexi,t}\right|, \tag{3.1}$$

where $\omega_{j,i,t}$ and $\omega_{j,indexi,t}$ are the portfolio weights of asset j in fund i and in the index of fund i in quarter t, and the sum is taken over all assets. A fund's benchmark index is defined as the index that produces the lowest Active Share. In total, there are 19 indices considered: the S&P 500, S&P 400, S&P 600, S&P 500/Barra Value, S&P 500/Barra Growth, Russell 1000, Russell 2000, Russell 3000, Russell Midcap, the value and growth components of the four Russell indexes (i.e., eight Russell style indexes), Wilshire 5000, and Wilshire 4500.

A fund's Tracking Error is traditionally used to capture the activeness of fund management. Unlike Active Share which measures active stock picking, Tracking Error captures the activeness in factor timing. It is defined as the volatility of the difference between a fund's return and its benchmark index return. For example, a portfolio that concentrates on a particular industry or sector is associated with high Tracking Error. Following Cremers and Petajisto (2009), Tracking Error is defined as the standard deviation of the residual, $\epsilon_{i,t}$, estimated from the following regression:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{indexi,t} - R_{f,t}) + \epsilon_{i,t}, \tag{3.2}$$

where $R_{i,t}$, $R_{f,t}$, and $R_{indexi,t}$ are daily fund, risk-free, and index returns, respectively, for fund i over a six-month period before each quarterly portfolio holdings are reported. Quarterly Tracking Error is the annualized standard deviation of the error term.

Active Share and Tracking Error measure different aspects of active management. These two measures capture two different types of deviation from the passive benchmark index. For example, a fund that actively picks individual stocks within each industry but also has a highly diversified portfolio across industries would have a high Active Share but a low Tracking Error. However, a fund that focuses on actively rotating across sectors or industries while holding diversified (or passive) positions within those sectors or industries would typically have a high Tracking Error but a low Active Share.

3.3 The Model

Consistent with theoretical studies in the area, we begin our analysis by considering a one-period model in which we explore the nature of managerial risk-taking incentives and the implications for portfolio performance when portfolio managers are faced with investors who chase recent performance. We further assume that there is no asymmetric information.

3.3.1 Model Structure

Consider two risk neutral managers who manage a pool of assets for a single period. Let the single period be divided into two (not necessarily equal) sub-periods (indexed by dates $t = 0$ to $t = 1$, and $t = 1$ to $t = 2$), with the end of the first sub-period marking an interim evaluation stage at which managers observe their performances to date, and make portfolio choices for the second sub-period. Figure 1 contains a timeline representing the sequence of events. Let m_g and m_b be the year-to-date cumulative return performances of the two managers at the end of the first sub-period (i.e., at the interim evaluation stage), with the assumption that $m_g > m_b$. Based on their relative performances in the first sub-period, we denote manager g (b) a "winning" ("losing") manager. Let the performance gap between manager g and manager b at the interim stage (date $t = l$) be denoted by m_δ, where $m_\delta = m_g - m_b$.

We assume that both managers begin with an asset base of size s at date $t = 0$. At date $t = 2$, the manager with the higher average return over the two sub-periods realizes additional fund inflows. We use notation λ, to denote the relative size of this additional fund inflow with respect to the original asset base, s. In other words, we assume that there is a group of investors, with aggregate wealth λ_s, who chase the winner fund. We further assume that managers are paid a fixed percentage k of the assets under their management at date $t = 2$.[6] Hence, each manager's objective is to enlarge the size of their asset base which can be achieved by earning a higher return over the single period and by attracting the performance chasing investors at the end of the period.

Each manager faces an identical structure of investment opportunity that consists of a risk-free asset with return $R_f > 0$, and a market index portfolio with return R_m, which represents the manager's passive strategy. Furthermore, we model the manager's active investment by assuming that she can invest in a fund-specific

[6]This assumption appears reasonable even though until the SEC rule changed in the sense that the terms of the employment contract of the fund portfolio managers were not required to be publicly disclosed. However, it is well known that the vast majority of Investment Advisory firm fee contracts are based solely on assets under management (see, for example, Deli (2002)). It is reasonable to expect that a fund manager's compensation will have a positive correlation with the profits of the Investment Advisory firm that employs her.

active-managed portfolio. Let R_g denote the random return realized by the active portfolio chosen by the winning manager and let R_b be the random return realized by the active portfolio chosen by the losing manager. We assume R_m, R_g, and R_b are mutually independent normal random variables, where $R_m \sim N(\mu, \sigma_m)$ with $\mu > R_f$, and $R_g, R_b \sim N(R_f, \sigma_i)$. The properties of the asset returns are public knowledge, so neither manager possesses superior knowledge.[7]

3.3.2 Analysis of the Model

We now focus on the portfolio choices made by each manager at the start of the second sub-period. Manager g's portfolio allocation decision is characterized by $\alpha^g = \begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix} \geq 0$, with $\alpha_1^g + \alpha_2^g \leq 1$, where α_1^g is the proportion of the total assets she decides to invest in the market index portfolio, α_2^g is the proportion invested in the actively managed portfolio, and $1 - (\alpha_1^g + \alpha_2^g)$ is the proportion invested in the risk-free asset. Similarly manager b's decision is characterized by $\alpha^b = \begin{bmatrix} \alpha_1^b \\ \alpha_2^b \end{bmatrix} \geq 0$, with $\alpha_1^b + \alpha_2^b \leq 1$.[8] Since the portfolio allocation to the actively managed portfolio represents the deviation from the passive strategy for a fund, it can be naturally interpreted as the fund's Active Share.

At the end of the single period, one of the two funds will end up with a higher average return measured over the two sub-periods, and become the ultimate "winner"

---

[7]While the assumption that managers do not possess superior information, is certainly consistent with empirical evidence on the performance of actively managed funds (see, for example Jensen (1968), Gruber (1996), Grinblatt and Titman (1992), Carhart (1997), among others), it is not critical for our analysis. Our results hold as long as the two managers do not differ in their abilities.

[8]Consistent with the restrictions facing the typical mutual fund, we disallow short sales.

of the tournament. For the "winner" of the tournament, the end-of-period asset base under management will grow not only by the rate of the portfolio return, but also by an additional amount of λ_s. The "loser" fund will not get any additional cash inflow. We use an indicator variable, $\mathbb{1}[\alpha^g, \alpha^b]$, to indicate the event that manager g wins the tournament. That is, the indicator variable equals 1 when manager g ends up as the ultimate winner and it equals 0 when manager b ends up as the ultimate winner. Formally, we have

$$\mathbb{1}[\alpha^g, \alpha^b] = 1 \text{ iff } \frac{m_g + \alpha_1^g(R_m - R_f) + \alpha_2^g(R_g - R_f)}{2} > \frac{m_b + \alpha_1^b(R_m - R_f) + \alpha_2^b(R_b - R_f)}{2}$$

$\mathbb{1}[\alpha^g, \alpha^b] = 0$ otherwise.

The managers' end-of-period compensation is k times the end-of-period asset base:

$$C_g = ks\left((1+m_g)\left(1 + R_f + \alpha_1^g(R_m - R_f) + \alpha_2^g(R_g - R_f)\right) + \lambda\mathbb{1}[\alpha^g, \alpha^b]\right), \quad (3.3)$$

and

$$C_b = ks\left((1+m_b)\left(1 + R_f + \alpha_1^b(R_m - R_f) + \alpha_2^b(R_b - R_f)\right) + \lambda\left(1 - \mathbb{1}[\alpha^g, \alpha^b]\right)\right). \quad (3.4)$$

It is straightforward to note that

$$E[\mathbb{1}] = \Phi\left(\frac{m_\delta + (\alpha_1^g - \alpha_1^b)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + {\alpha_2^b}^2)\sigma_i^2 + (\alpha_1^g - \alpha_1^b)^2\sigma_m^2}}\right), \quad (3.5)$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal variable. We note that the above expression represents the probability of manager g

finishing the game as the ultimate winner. In the subsequent discussion we use the notation $\phi(\cdot)$ to denote the standard normal probability density function.

Since the constant k (proportional management fee charged by the managers) and s (the initial size of each fund) have no bearing on a manager's strategy, we will assume in this section, for brevity, that $ks = 1$. We now consider the expected utility of the two managers for a strategies pair $\{\alpha^g, \alpha^b\} = \left\{ \begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix}, \begin{bmatrix} \alpha_1^b \\ \alpha_2^b \end{bmatrix} \right\}$. The winning manager's expected utility is given by

$$U_g(\alpha^g, \alpha^b) = (1+m_g)\left(1 + R_f + \alpha_1^g(\mu - R_f)\right) + \lambda\Phi\left(\frac{m_\delta + (\alpha_1^g - \alpha_1^b)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + {\alpha_2^b}^2)\sigma_i^2 + (\alpha_1^g - \alpha_1^b)^2\sigma_m^2}}\right), \tag{3.6}$$

and the losing manager's expected utility is

$$U_b(\alpha^g, \alpha^b) = (1+m_b)(1+R_f+\alpha_1^b(\mu-R_f))+\lambda-\lambda\Phi\left(\frac{m_\delta + (\alpha_1^g - \alpha_1^b)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + {\alpha_2^b}^2)\sigma_i^2 + (\alpha_1^g - \alpha_1^b)^2\sigma_m^2}}\right). \tag{3.7}$$

Manager b chooses α^b to maximize her expected utility given by Equation (3.7). The following result characterizes the portfolio choices made by the two managers at date $t = 1$, in equilibrium.

Proposition 3. *In equilibrium, $\alpha_1^g > \alpha_1^b$, where the equality holds only when $\alpha_1^g = \alpha_1^b = 1$. Moreover, $\alpha_2^g = 0$ and $\alpha_1^b + \alpha_2^b = 1$.*

Proof. See Appendix B.

Proposition 1 establishes that, in equilibrium, (a) manager g (i.e., the winning manager at the interim stage) will invest a proportion of her portfolio in the market index that is at least as large as the corresponding proportion chosen by manager b, (b) manager g's portfolio will be completely passive, i.e., her portfolio's Active Share would be zero, and (c) manager b will choose to not invest in the risk free asset. We also note that the manager g will always desire a portfolio allocation to the market index that is strictly greater than the manager b's allocation to the market index when the short sale constraint is not binding. Intuitively, the manager who is ahead at the interim stage (date $t = 1$) would like to "lock in" her gains while the interim loser attempts to increase her portfolio's *activeness* in an attempt to take on some side bet by increasing the Active Share in her portfolio. Note that in our setting the interim loser fund has no incentive to increase her portfolio's systematic risk. The intuition for this result is that the only way for the loser fund to catch up to (and possibly surpass) the winner fund is to hold a portfolio that is different from that held by the winner fund. Given the winner fund manager's incentive to hold an indexed position, the loser fund manager attempts to take on risk that is fund-specific, and the natural way to do so is by increasing Active Share. Our result that the interim loser fund manager has an incentive to increase Active Shares is different from earlier studies (e.g., Taylor (2003)) that focus primarily on the losing fund manager's incentive to affect the total volatility of her portfolio. In our model, the total volatility of the loser fund may be either higher or lower than the winner fund.

The following lemma further describes the equilibrium in terms of the optimal portfolio allocations, α_g^*, and α_b^*.

Lemma 1. *In equilibrium either* $\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, *or* $\alpha^{b*} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, *or both.*

Proof. See Appendix B.

We focus on the equilibrium with $\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Later, we provide a sufficient condition for this to be the case. We make the following assumption that

$$(1+m_b)(\mu - R_f) \geq \lambda\phi\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}. \tag{3.8}$$

The left hand side of the above inequality is the expected rate of increase of manager b's assets if the manager chooses to increase her allocation to the market index portfolio, and the right hand side represents the additional reward the manager expects to get for having a larger chance of winning the tournament by lowering the allocation to the market index. Intuitively, the above assumption states that the tournament effect is not so overwhelming that manager b would want to go to the extreme of allocating 100% of her portfolio to the idiosyncratic risky asset in order to win the tournament. Inequality (8) is equivalent to

$$\begin{aligned}&\frac{\partial}{\partial \alpha_1^b} U_b\left(\alpha^g = \begin{bmatrix}1\\0\end{bmatrix}, \alpha^b = \begin{bmatrix}\alpha_1^b\\1-\alpha_1^b\end{bmatrix}\Bigg|_{\alpha_1^b=0^+}\right)\\ &= (1+m_l)(\mu - R_f) - \lambda\Phi\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}} > 0.\end{aligned} \tag{3.9}$$

Thus, the assumption represented by inequality (8) ensures that the interim loser fund has a non-zero allocation to the market index (i.e., $\alpha_1^{b^*} > 0$) in the equilibrium.

We next have the following statement regarding the existence of an equilibrium.

Proposition 4. *In the case that* $\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}} \geq C_1$, *the pair* $\left\{\alpha^{g*} = \begin{bmatrix}1\\0\end{bmatrix}, \alpha^{b*} = \begin{bmatrix}1\\0\end{bmatrix}\right\}$ *constitutes a Nash equilibrium where* $C_1 \propto \lambda$ *(see Appendix B for the definition of* C_1*).*

Proof. See Appendix B.

Proposition 2 states that if the interim performance gap, m_δ, is larger than $C_1\sqrt{\sigma_i^2+\sigma_m^2}$, it will be optimal for manager b to choose strategy $\alpha^b = \begin{bmatrix}1\\0\end{bmatrix}$. As a result, her chance of winning the tournament is exactly zero. Essentially this means that manager b withdraws from the tournament. The intuition for this outcome is as follows. When the performance gap is large, the chance for manager b to make up such a large gap before the end of the game is so slim that she simply gives up and decides instead to focus on the reward from the linear employment contract. Moreover, the bound on the size of the performance gap is proportional to λ, the reward from winning the tournament in the form of the end-of-period cash inflow. Intuitively, the higher the potential reward for winning the tournament, the stronger the desire for manager b to stay on and compete in the tournament. Given a large enough value of λ, she might still consider participating in the tournament even when she is lagging behind significantly at the interim stage.

The following proposition establishes a link between manager b's allocation to the idiosyncratic risky asset and the strength of the tournament effect captured by λ, and the interim performance gap, m_δ.

Proposition 5. *In case that* $\alpha_g{}^* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, *and* $\alpha_1^{b^*} < 1$, *we have* $\frac{\partial}{\partial \lambda}(1 - \alpha_1^{b^*}) > 0$, *and* $\frac{\partial}{\partial m_\delta}\Phi\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}\frac{1}{1-\alpha_1^{b^*}} + \frac{(\mu - R_f)}{\sqrt{\sigma_i^2+\sigma_m^2}}\right) > 0$.

Proof. See Appendix B.

Proposition 3 implies that when $\alpha^{b^*} = \begin{bmatrix} \alpha_1^{b^*} \\ 1 - \alpha_1^{b^*} \end{bmatrix}$ is manager b's optimal response with $\alpha_1^{b^*} < 1$, the higher the stakes in the tournament (i.e. the larger value of λ), the greater the Active Share in manager b's portfolio (i.e. higher α_2^b). Hence, the interim loser manager's propensity to increase Active Share is directly related to the strength of the tournament effect. The above proposition also points out that in the equilibrium, everything else equal, the larger the lead enjoyed by manager g at the interim stage of the game (date $t = 1$), the greater the probability of manager g

finishing the game as the ultimate winner, after taking into consideration the losing manager's incentive of increasing Active Share.

Thus far, our discussion of the equilibrium is based on the assumption that it exists. The following theorem establishes the existence of the equilibrium under the condition that the interim performance gap (i.e. m_δ) between the two managers is not too large.

Theorem 1. *There is a positive value C that depends on other parameters except m_δ, such that if $m_\delta < C\sqrt{\sigma_i^2 + \sigma_m^2}$, the pair $\left\{\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \alpha^{b*} = \begin{bmatrix} \alpha_1^{b^*} \\ 1 - \alpha_1^{b^*} \end{bmatrix}\right\}$ will be an equilibrium, where $\alpha_1^{b^*} = 1 - \frac{1}{x_1}$. Furthermore, $0 < \alpha_1^{b^*} < 1$, $\frac{\partial(1-\alpha_1^{b^*})}{\partial m_\delta} > 0$ and $\lim_{m_\delta \to 0^+}(1 - \alpha_1^{b^*}) = 0$. To be exact, we have that $1 - \alpha_1^{b^*} = O(\sqrt{m_\delta})$.*

Proof. See Appendix B for a more precise version of the theorem and the proof. Also notice that under the condition of the theorem, the claims made in Proposition 1 and 3 are true.

Corollary 2. *The Sharpe Ratio of fund b is lower than that of fund g, and it is decreasing with respect to m_δ and λ.*

Proof. See Appendix B.

In addition to assuring the existence of the equilibrium, this theorem also points out that manager b's Active Share is an increasing function of the interim performance gap, m_δ. That is, the larger the gap, the more the aggressive is manager b's Active Share. As the performance gap approaches zero, manager b's Active Share also approaches zero. Furthermore, the corollary to the theorem shows that the interim loser fund has a lower Sharpe ratio compared to the interim winner fund. That is, the tournament-driven Active Share is not beneficial to the fund investors in our model.

3.4 Empirical Evidence

Our analysis provides a number of testable predictions relating to the risk-taking incentives of fund managers. In particular, our results suggest that the increase in the Active Share of the interim loser manager's portfolio is directly related to the magnitude of the performance gap at the interim stage, and to the strength of the investor (cash flow) response to the relative performance rankings of the funds (i.e., the strength of the tournament effect). We now examine the empirical evidence on these issues.

3.4.1 Data and Sample Description

To test the key implications of our basic model, we use quarterly data for a sample of domestic stock funds for the period 1990-2008 constructed by Cremers and Petajisto (2009) and updated by Petajisto (2013), which includes information on the funds' Active Share, self-reported Active Share, Tracking Error, etc.[9] The dataset is available from 1980 to the third quarter of 2009. Since our tests require data to be available for an entire year, and the Active Share for all funds was very high in the 1980s (see Cremers and Petajisto (2009)), we restrict our sample period to be 1990-2008. We use data from the Center for Research in Security Prices (CRSP) Survivorship Bias Free Mutual Fund Database, which includes information on the funds' total net assets (TNA), returns and characteristics. During the 1990s, mutual funds started to offer different share classes that represent claims to the same underlying portfolios

[9]The dataset can be downloaded at: http://www.petajisto.net/data.html. See Petajisto (2013) for details of the data construction.

but with different fee structures. The Active Share database reports data for each fund. As noted in the CRSP mutual fund manual, however, CRSP treats each share class as a stand-alone fund and assigns it a separate fund identification number. To merge these two databases, we aggregate across all share classes for each fund: the net returns and expenses of the funds are calculated as TNA-weighted averages across all share classes; the TNA are calculated as the sum across all share classes; the age of the oldest share class of a fund represents the fund's age; the category of the largest share class is the category of the fund. We classify funds into five categories based on their objective codes:[10] aggressive growth (AGG), growth (GRO), growth and income, income and others. From this group of funds we exclude all stock index funds as the incentives of their fund managers and investors are quite different from that of the participants in active funds. Our final sample contains 2,202 fund-entities comprising 15,641 fund-years. Of them, 1,104 funds fall into the category of AGG and GRO.

Table 3.1 presents the characteristics of the funds in the sample. In an average quarter, the sample includes 813 funds with average total net assets (TNA) of $1037.46

[10]We categorize funds according to the following criteria. First, funds with Lipper objective codes G, LCGE, MCGE, MLGE, SCGE, with Wiesenberger objective codes G, G-S, S-G, GRO, LTG, SCG, or with Strategic Insight objective codes GRO, SCG are classified as growth funds. Second, funds with Wiesenberger objective code AGG or with Strategic Insight objective code AGG are classified as aggressive growth. Third, funds with Lipper objective code GI, with Wiesenberger objective codes G-I-S, G-S-I, I-G, I-G-S, GCI, G-I, I-S-G, S-G-I, S-I-G, GRI, or with Strategic Insight objective code GRI are classified as growth and income funds. Fourth, funds with Lipper objective codes EI, EIEI, I, with Wiesenberger objective codes I, I-S, IEQ, ING, or with Strategic Insight objective codes ING are classified as income funds. Fifth, all the other actively managed equity funds in our sample are classified as others.

million. Among them, there are 423 funds that fall into the category of AGG and GRO. The average Active Share for the entire sample is 79.7% and it is 82.1% for AGG and GRO funds.

3.4.2 Evidence of Performance Chasing Behavior by Investors

A number of previous studies have documented evidence of performance chasing behavior by fund investors. We begin our analysis by confirming this behavior in our sample using a cross-sectional regression framework. We examine quarterly cross-sectional regressions of the normalized cash flow experienced by a fund on several factors known to influence fund flows. The net cash flow to fund i during quarter t is measured as follows:

$$Flow_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1}(1 + R_{i,t})}{TNA_{i,t-1}}, \tag{3.10}$$

where $R_{i,t}$ denotes fund i's return during quarter t, and $TNA_{i,t}$ is the fund's total net asset value at the end of quarter t. Thus, our definition of flows reflects the percentage growth of the fund's assets in quarter t. The independent variables used in the regressions include the one quarter lagged flow, the fund's total return in the previous quarter (net of the average return for the fund's category), the logarithm of lagged total net assets (TNA), the logarithm of one plus age, previous quarter's total expense, and category flow. Total expense is estimated as expense ratio plus 1/7 of maximum front-end load. Category flow is the aggregate flow to the fund' objective category. These variables are suggested by previous studies of the determinants of fund flows (see, for example, Ippolito (1992), Chevalier and Ellison (1997), Sirri and Tufano (1998), and Sapp and Tiwari (2004)). Table 3.2 presents the estimated coefficients from these regressions. The reported coefficient estimates are time-series averages of 76 quarterly cross-sectional regression estimates over the period from January 1990 to December 2008. We present results for the entire sample as well as for funds that fall into the category of AGG and GRO.

It can be seen from the overall results that fund flows are significantly positively related to the previous quarter performance of the fund, confirming the results of previous studies. Additionally, fund flows are significantly negatively related to fund size reflecting the fact that smaller funds attract proportionately larger percentage flows compared to bigger funds. Comparing the results between the entire sample and the AGG and GRO funds, it is interesting to note that the fund flows experienced by the AGG and GRO funds have greater sensitivity to past relative performance, which presumably reflects the active nature of the investor clientele that is drawn to such funds. For example, the results suggest that in the case of AGG and GRO funds, outperforming the category average by one percent translates into an incremental flow equal to 0.41 percent of a fund's assets. The corresponding figures representing the incremental quarterly flow from beating the category average by one percent are 0.36 percent in the case of the entire sample. In this sense the implicit incentives created by the existence of the positive flow-performance relationship appear to be quite meaningful for the funds.

In summary, the results presented in Table 3.2 confirm that the recent relative performance of a fund has significant explanatory power for its future flows, even after controlling for a number of other fund characteristics. In particular, the "winner" funds attract significant future inflows. In the next sub-section we explore the risk-taking behavior of the funds in the context of the above incentives.

3.4.3 Do Funds Alter their Active Share in Response to Above Incentives?

We now examine the key question, namely, whether funds increase their Active Share following poor performance at an interim evaluation stage. To assess this issue, we divide each calendar year in the sample period into two sub-periods of three quarters and one quarter each. The quarterly Active Share is defined in Equation (3.1). It is the percentage of a fund's portfolio holdings that differ from the fund's benchmark index (the index that produces the lowest Active Share). We define the change in Active Share as the fund i's Active Share in the last quarter of year T subtract the average Active Share of fund i in the first three quarters of year T:

$$\Delta \text{Active Share}_{i,T} = \text{Active Share}_{i,4,T} - \frac{1}{3}\sum_{t=1}^{3} \text{Active Share}_{i,t,T} \tag{3.11}$$

Thus, for each fund we have a time series of the yearly changes in the Active Share from sub-period 1 to sub-period 2. We then conduct annual cross-sectional regressions of the change in Active Share of a fund on the return performance of the fund during the first sub-period of the year:

$$\Delta \text{Active Share}_{i,T} = a_T + b_T R_{i,1,T} + \epsilon_T, \tag{3.12}$$

In the above equation $R_{i,1,T}$ denotes the average of raw returns for fund i during the first sub-period (first three quarters) of calendar year T, net of the mean return earned by all funds in its peer group (category) for the sub-period. We next compute the time series averages of the yearly cross-sectional regression coefficients a_T and b_T.

Table 3.3 reports the estimated intercept and slope coefficients from Equation (3.12). The table reports coefficients for two groups of funds: funds with interim return (the average returns during the first three quarters of each year) above the median return of their categories and funds with interim return below the category median. Within each group, the table also reports coefficients for all funds in the group as well as aggressive growth (AGG) and growth (GRO) funds only.

We note that for funds with interim return above their category median, there is a negative relation between fund relative performance during the first three quarters of the year and the subsequent change in the fund's Active Share. Funds appear to increase their Active Share following poor performance at the interim stage. Equally striking is the fact that the relationship appears to be stronger, in both economic and statistical term, for funds in the aggressive growth (AGG) and growth (GRO) categories. Interestingly, it was these two categories for which the tournament effect (i.e., the sensitivity of flow to past relative performance) was found to be the most intense based on our earlier results in Table 3.2. These results are broadly consistent with our theoretical framework that suggests that the changes in Active Share are related to the performance gap at the interim stage, and to the strength of the tournament effect.

Recall that in our theoretical analysis, Theorem 1 places a restriction on the magnitude of the performance gap, m_δ , between the winning and losing managers (see also the discussion following Proposition 2). Intuitively, managers who trail too far behind at the interim stage and who do not have a realistic chance of making up the shortfall may exit the "tournament", i.e., they may choose to not increase their Active Share in the last quarter of the year. This suggests that the risk-taking incentives would be stronger for those fund managers who feel they are still "in the game" and have a chance of making up the performance deficit in the second half of the evaluation period. Consistent with this analysis, the last four columns of Table 3.3 show that when we restrict our sample to funds with interim return below their category median, the relationship between change in Active Share and the first sub-period return is insignificant.

3.4.4 Evidence Using Other Activeness Measures

Tracking Error is a traditional measure for active management. Unlike Active Share which measures active stock picking, Tracking Error captures the activeness in factor timing. However, under the tournament frame work, if the interim losing manager hopes to catch up in the last quarter, factor timing is a less efficient approach which involves shifting her portfolio towards a different industry or sector. Consistent with this analysis, we next provide evidence that there is no significant relationship between a fund's Tracking Error and its interim performance.

We follow Cremers and Petajisto (2009) and define Tracking Error as in Equation (3.2). Similar to the analysis using Active Share, we examine changes in fund Tracking Error across two sub-periods within each year. For each fund we have a time series of the yearly changes in the Tracking Error from sub-period 1 to sub-period 2. We then conduct annual cross-sectional regressions of the change in Tracking Error of a fund on the return performance of the fund during the first sub-period of the year (net of the mean return earned by all funds in its category). We next compute the time series averages of the yearly cross-sectional regression coefficients.

Table 3.4 reports the estimated intercept and slope coefficients for this analysis. The table reports coefficients for funds with interim return ($R_{i,1,T}$) above the median return of their categories only. It reports coefficients for all funds as well as aggressive growth (AGG) and growth (GRO) funds only. We note that for funds with interim return above their category median, there is a negative relation between fund relative performance during the first three quarters of the year and the subsequent change in the fund's Tracking Error. However, in contrast with the analysis with Active Share, the negative coefficient for Interim Return is not statistically significant. The results indicate that funds do not appear to increase their Tracking Error following poor performance at the interim stage.

We define a fund's Active Share as the deviation of a fund's portfolio holdings from the fund's benchmark index which is the index that produces the lowest Active Share. Alternatively, as proposed by Petajisto (2013), the benchmark index can also be chosen as the self-reported benchmark by a manager in the fund prospectus, rather than assigning the index that produces the lowest Active Share. We repeat our analysis with Active Share using this alternative measure. Table 3.5 reports the estimated intercept and slope coefficients for this analysis. The table reports coefficients for funds with interim return ($R_{i,1,T}$) above the median return of their categories only. It reports coefficients for all funds as well as aggressive growth (AGG) and growth (GRO) funds only. We note that for funds with interim return above their category median, there is a negative relation between fund relative performance during the first three quarters of the year and the subsequent change in the fund's self-reported Active Share. However, in contrast with the analysis with Active Share, the negative coefficient for Interim Return is not statistically significant. The results indicate that funds do not appear to increase their self-reported Active Share following poor performance at the interim stage.

3.5 Discussion and Conclusion

This paper provides a framework for the incentive effects created by the fund flow-performance relationship. As a first step we analyze a model in which two risk-neutral fund managers with unequal performances at an interim stage, compete for investor cash flows. Neither manager possesses superior information. Our focus is on the portfolio choices made by the two managers after they observe their interim performances. We show that there exists an equilibrium in which it is optimal for the fund manager whose performance lags behind at the interim stage (i.e., the interim loser) to increase the *activeness* of her portfolio. The model suggests a direct measure of the extent of risk-taking by a fund, namely, a fund's Active Share.

Our theoretical analysis affords a number of testable predictions. Specifically, our results suggest that the increase in the Active Share of the interim loser manager's portfolio is directly related to the magnitude of the performance gap at the interim stage, and to the strength of the investor (cash flow) response to the relative performance rankings of the funds (i.e., the strength of the tournament effect). Our empirical analysis of a sample of stock funds provides support for these implications.

Table 3.1: Summary Statistics of the Equity Mutual Fund Sample

	All funds	AGG and GRO funds
Number of funds	813	423
TNA (in millions)	1037.46	1076.09
Age (in years)	16.00	15.61
Quarterly Active Share (in %)	0.797	0.821
Quarterly self-reported Active Share (in %)	0.816	0.841
Quarterly raw return (in %)	2.190	2.213

Note: This table reports the time-series averages of quarterly cross-sectional averages of fund characteristics for the period 1990-2008. TNA is the total net assets. Age is the fund's age. Quarterly Active Share is defined following Cremers and Petajisto (2009) as the percentage of a fund's portfolio holdings that differ from the fund's benchmark index which is the index that produces the lowest Active Share:

$$\text{Active Share } = \frac{1}{2}\sum_{i=1}^{N} |\omega_{fund,i} - \omega_{index,i}|,$$

where $\omega_{fund,i}$ and $\omega_{index,i}$ are the portfolio weights of asset i in the fund and in the index, and the sum is taken over all assets. It is computed based on Spectrum mutual fund holdings data and index composition data for nineteen common benchmark indexes from S&P, Russell, and Wilshire. Quarterly self-reported Active Share is defined following Petajisto (2013) as the percentage of a fund's portfolio holdings that differ from the fund's self-reported benchmark index. Quarterly raw return is the fund's quarterly raw return. The statistics are reported for all funds and funds that fall into the categories of aggressive growth (AGG) and growth (GRO).

Table 3.2: Determinants of Quarterly Fund Cash Flows

	All Funds		AGG and GRO	
	Coefficient	t-value	Coefficient	t-value
Intercept	0.055	(5.290)	0.064	(4.750)
Previous quarter's flow	0.358	(14.210)	0.390	(11.520)
Previous quarter's return	0.357	(2.540)	0.413	(9.260)
Logarithm of total net assets	-0.003	(-4.380)	-0.004	(-4.130)
Logarithm of one plus age	-0.019	(-2.650)	-0.012	(-8.380)
Previous quarter's total expense	1.450	(1.280)	0.281	(1.320)
Category flow	0.784	(3.010)	-0.428	(-1.010)

Note: This table presents the coefficients from regressions of realized quarterly flow for a fund against the fund's lagged quarterly flow, the fund's total return in the previous quarter (net of the average return for the fund's category), the logarithm of lagged total net assets (TNA), the logarithm of one plus age, previous quarter's total expense, and category flow. The flow for a fund during a quarter is computed as the quarterly cash flow for the fund divided by the total net assets (TNA) at the beginning of the quarter. Total expense is estimated as expense ratio plus 1/7 of maximum front-end load. Category flow is the aggregate flow to the fund' objective category. The coefficients reported below are time-series averages of the quarterly cross-sectional regression coefficients from 1990 to 2008. The table reports coefficients for all funds as well as aggressive growth (AGG) and growth funds (GRO) only. Also shown are the t-statistics (in parentheses) based on the Fama Macbeth regression.

Table 3.3: Changes in Fund Active Share and Past Performance

	Rank above median				Rank below median			
	All Funds		AGG and GRO		All Funds		AGG and GRO	
	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value	Coefficient	t-value
Intercept	-0.003	(-1.443)	-0.003	(-1.342)	-0.003	(-0.984)	-0.004	(-1.726)
Interim Return	-0.076	(-2.409)	-0.089	(-3.121)	0.056	(1.417)	-0.005	(-0.103)

Note: We divide each year in the sample period into two three and one quarterly sub-periods. We compute the change in the Active Share of a fund from the first three quarters of the year to the last quarter of the year. The Quarterly Active Share is described in detail in Cremers and Petajisto (2009). It is the percentage of a fund's portfolio holdings that differ from the fund's benchmark index which is the index that produces the lowest Active Share:

$$\text{Active Share } = \frac{1}{2}\sum_{i=1}^{N}|\omega_{fund,i} - \omega_{index,i}|,$$

where $\omega_{fund,i}$ and $\omega_{index,i}$ are the portfolio weights of asset i in the fund and in the index, and the sum is taken over all assets. For each fund, the change in Active Share from the first three quarters of the year to the last quarter is regressed on the fund's Interim Return which is defined as the average quarterly raw return in the first three quarters adjusted for the mean return of all funds within the category during the first three quarters. The coefficients reported below are time-series averages of the 19 yearly cross-sectional regression coefficients from 1990 to 2008. The table reports coefficients for two groups of funds: funds with return (in the first three quarters of the year) above the median return of the category and funds with return below the median return. Within each group, it also reports coefficients for all funds in the group as well as aggressive growth (AGG) and growth funds (GRO) only. Also shown are the t-statistics (in parentheses) based on the Fama Macbeth regression.

Table 3.4: Changes in Fund Tracking Error and Past Performance

	All Funds		AGG and GRO	
	Coefficient	t-value	Coefficient	t-value
Intercept	0.002	(0.618)	0.002	(0.706)
Interim Return	0.023	(0.718)	0.009	(0.329)

Note: We divide each year in the sample period into two three and one quarterly sub-periods. We compute the change in the Tracking Error of a fund from the first three quarters of the year to the last quarter of the year. The Quarterly Tracking Error is described in detail in Cremers and Petajisto (2009). It is the annualized standard deviation of the error term when the excess return on a fund i is regressed on the excess return on its benchmark index (the index that produces the lowest Active Share). It is computed based on daily fund returns and daily index returns over a six-month period before the corresponding portfolio holdings are reported. For each fund, the change in average Quarterly Tracking Error from the first three quarters of the year to the last quarter is regressed on the fund's Interim Return which is defined as the average quarterly raw return in the first three quarters adjusted for the mean return of all funds within the category during the first three quarters. The coefficients reported below are time-series averages of the 19 yearly cross-sectional regression coefficients from 1990 to 2008. The table reports coefficients for funds with interim return (during the first three quarters of the year) above the median return of the category only. It reports coefficients for all funds as well as aggressive growth (AGG) and growth funds (GRO) only. Also shown are the t-statistics (in parentheses) based on the Fama Macbeth regression.

Table 3.5: Changes in Fund Self-reported Active Share and Past Performance

	All Funds		AGG and GRO	
	Coefficient	t-value	Coefficient	t-value
Intercept	-0.003	(-1.733)	-0.004	(-1.615)
Interim Return	-0.049	(-1.793)	-0.055	(-1.744)

Note: We divide each year in the sample period into two three and one quarterly sub-periods. We compute the change in the Self-reported Active Share of a fund from the first three quarters of the year to the last quarter of the year. The Quarterly Self-reported Active Share is defined in Petajisto (2013) as the percentage of a fund's portfolio holdings that differ from the fund's self-reported benchmark index. For each fund, the change in Self-reported Active Share from the first three quarters of the year to the last quarter is regressed on the fund's Interim Return which is defined as the average quarterly raw return in the first three quarters adjusted for the mean return of all funds within the category during the first three quarters. The coefficients reported below are time-series averages of the 19 yearly cross-sectional regression coefficients from 1990 to 2008. The table reports coefficients for funds with interim return (during the first three quarters of the year) above the median return of the category only. It reports coefficients for all funds as well as aggressive growth (AGG) and growth funds (GRO) only. Also shown are the t-statistics (in parentheses) based on the Fama Macbeth regression.

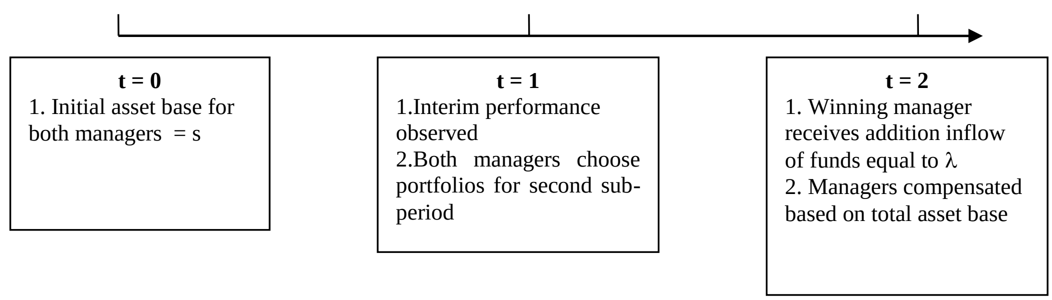


Figure 3.1: One Period Model

A. Numerical Analysis of Bayesian Benchmark Case with Signal Precision Uncertainty

In this appendix, we provide details for the numerical analysis for the Bayesian benchmark case with signal precision uncertainty discussed in Section B.2. We assume the investors are risk averse with the exponential utility function $U(c) = -e^{ac}$ and are uncertain about the precision of the realized signals, s_1 and s_2. Specifically, the precision of each signal may take on one of two values: h and l. This gives rise to the following four potential models: M_1: both s_1 and s_2 have high precision h, i.e., $H_1 = H_2 = h$; M_2: both s_1 and s_2 have low precision l, i.e., $H_1 = H_2 = l$; M_3: the worst signal s_1 has high precision and s_2 has low precision, i.e., $H_1 = h, H_2 = l$; M_4: the worst signal s_1 has low precision and s_2 has high precision, i.e., $H_1 = l, H_2 = h$. According to Bayes' rule, the posterior probabilities of the four models, denoted as $P_1 - P_4$, are as follows: $P_1 = \frac{A}{A+B+C+D}, P_2 = \frac{B}{A+B+C+D}, P_3 = \frac{C}{A+B+C+D}, P_4 = \frac{D}{A+B+C+D}$, where

$$
\begin{aligned}
A &= \frac{h}{\sqrt{1+2h}} \exp\left\{ -\frac{1}{2\sigma_\alpha^2}\left[h(s_1^2+s_2^2) - \frac{h^2}{1+2h}(s_1+s_2)^2\right]\right\},\\
B &= \frac{l}{\sqrt{1+2l}} \exp\left\{ -\frac{1}{2\sigma_\alpha^2}\left[l(s_1^2+s_2^2) - \frac{l^2}{1+2l}(s_1+s_2)^2\right]\right\},\\
C &= \frac{\sqrt{hl}}{\sqrt{1+h+l}} \exp\left\{ -\frac{1}{2\sigma_\alpha^2}\left[hs_1^2+ls_2^2 - \frac{1}{1+h+l}(hs_1+ls_2)^2\right]\right\},\\
D &= \frac{\sqrt{hl}}{\sqrt{1+h+l}} \exp\left\{ -\frac{1}{2\sigma_\alpha^2}\left[ls_1^2+hs_2^2 - \frac{1}{1+h+l}(ls_1+hs_2)^2\right]\right\}.
\end{aligned}
$$

The investor's expected utility is thus:

$$
\begin{aligned}
E_\pi(U(c)) &= P_1 * E(U(c)|M_1) + P_2 * E(U(c)|M_2) + P_3 * E(U(c)|M_3) + P_4 * E(U(c)|M_4)\\
&= P_1 * (-e^{Y_1}) + P_2 * (-e^{Y_2}) + P_3 * (-e^{Y_3}) + P_4 * (-e^{Y_4}),
\end{aligned}
$$

where

$$Y_1 = -a\left[\mu + \mu_\alpha + \frac{h}{1+2h}(s_1+s_2)\right] + \frac{1}{2}a^2\left(\sigma_\epsilon^2 + \frac{\sigma_\alpha^2}{1+2H}\right),$$
$$Y_2 = -a\left[\mu + \mu_\alpha + \frac{l}{1+2l}(s_1+s_2)\right] + \frac{1}{2}a^2\left(\sigma_\epsilon^2 + \frac{\sigma_\alpha^2}{1+2H}\right),$$
$$Y_3 = -a\left[\mu + \mu_\alpha + \frac{1}{1+h+l}(hs_1+ls_2)\right] + \frac{1}{2}a^2\left(\sigma_\epsilon^2 + \frac{\sigma_\alpha^2}{1+2H}\right),$$
$$Y_4 = -a\left[\mu + \mu_\alpha + \frac{1}{1+h+l}(ls_1+hs_2)\right] + \frac{1}{2}a^2\left(\sigma_\epsilon^2 + \frac{\sigma_\alpha^2}{1+2H}\right).$$

As in Equation (1.10), we have

$$k^* = \frac{1}{a}\left[\log(A+B+C+D) - \log(Ae^{Y_1}+Be^{Y_2}+Ce^{Y_3}+De^{Y_4})\right]. \tag{1}$$

The difference in the flow-performance sensitivity to the signal s_1 relative to signal s_2 is:

$$\begin{aligned}
&\frac{dF(k^\star)}{ds_1} - \frac{dF(k^\star)}{ds_2}\\
&= F' \times \left(\frac{dk^\star}{ds_1} - \frac{dk^\star}{ds_2}\right)\\
&= \frac{F'}{a} \times \left\{\frac{P_1}{\sigma_\alpha^2}(hs_2 - hs_1) + \frac{P_2}{\sigma_\alpha^2}(ls_2 - ls_1) + \frac{P_3}{\sigma_\alpha^2}\left[\frac{(h-l)(hs_1+ls_2)}{1+h+l} + ls_2 - hs_1\right]\right.\\
&+ \frac{P_4}{\sigma_\alpha^2}\left[\frac{(l-h)(ls_1+hs_2)}{1+h+l} + hs_2 - ls_1\right] + \frac{Q_1}{\sigma_\alpha^2}(hs_1 - hs_2) + \frac{Q_2}{\sigma_\alpha^2}(ls_1 - ls_2)\\
&+ \frac{Q_3}{\sigma_\alpha^2}\left[\frac{(l-h)(hs_1+ls_2)}{1+h+l} + \frac{a\sigma_\alpha^2(h-l)}{1+h+l} + hs_1 - ls_2\right] + \frac{Q_4}{\sigma_\alpha^2}\left[\frac{(h-l)(ls_1+hs_2)}{1+h+l}\right.\\
&\left.\left. + \frac{a\sigma_\alpha^2(l-h)}{1+h+l} + ls_1 - hs_2\right]\right\}
\end{aligned}$$

where $Q_1 = Ae^{Y_1}/(Ae^{Y_1}+Be^{Y_2}+Ce^{Y_3}+De^{Y_4}), Q_2 = Be^{Y_2}/(Ae^{Y_1}+Be^{Y_2}+Ce^{Y_3}+De^{Y_4}), Q_3 = Ce^{Y_3}/(Ae^{Y_1}+Be^{Y_2}+Ce^{Y_3}+De^{Y_4})$, and $Q_4 = De^{Y_4}/(Ae^{Y_1}+Be^{Y_2}+Ce^{Y_3}+De^{Y_4})$. We denote

$$\frac{dF(k^\star)}{ds_1} - \frac{dF(k^\star)}{ds_2} \equiv \frac{F'}{a} \times G. \tag{2}$$

Since both F' and the degree of risk aversion a are positive, the sign of $dF(k^\star)/ds_1 - dF(k^\star)/ds_2$ is determined by the sign of the term G.

We next calibrate the key parameters in the term G based on the sample of mutual funds described in detail in Section IV. In particular, the investor's prior regarding the average fund managers skill, μ_α, is set equal to zero. The prior variance, σ_α^2, is set equal to the cross-sectional variance of the Carhart 4-factor alphas across all funds. We use the cross-sectional average of the Carhart 4-factor model error variance across all funds as a measure of the noise in the performance-related signal.We compute this measure for two non-overlapping, equal sub-periods: 1991–2000 and 2001–2010. We set the high signal noise (σ_h^2) equal to the value of the measure computed over the period over the period 1991–2000. Similarly σ_l^2 is set equal to the value of the above measure obtained using data for 2001–2010. Thus, the high signal-to-noise ratio, $h = \sigma_\alpha^2/\sigma_l^2$, and the low signal-to-noise ratio, $l = \sigma_\alpha^2/\sigma_h^2$. Moreover, σ_ϵ^2 is computed as the cross-sectional average of the Carhart 4-factor model error variance across all funds from 1983 to 2011. The degree of risk aversion a is set equal to 2. The market risk premium, μ, is the average market access return from 1983 to 2011. Finally, the range of the signal realizations, i.e., s_1 and s_2, is set equal to one standard deviation, σ_ϵ, around zero. All parameters are chosen on a monthly basis. Based on the above discussion, we obtain the following values for the key parameters: $\sigma_\alpha^2 = 0.0007\%$, $\sigma_l^2 = 0.0282\%$, $\sigma_h^2 = 0.0845\%$, $\sigma_\epsilon^2 = 0.0467\%$, and $\mu = 0.5702\%$. Figure 1 shows the results of this numerical analysis.

B. Technical Details in Chapter 3

Proof of Proposition 1: We prove the proposition in 4 steps.

1. For any losing manager's feasible strategy, $\alpha^b = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$, the winning manager's strategy $\begin{bmatrix} \theta_1 \\ 0 \end{bmatrix}$ strictly dominates any one of her feasible strategies $\begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix}$ with $\alpha_1^g < \theta_1$, and any one of her feasible strategies $\begin{bmatrix} \theta_1 \\ \alpha_2^g \end{bmatrix}$ with $\alpha_2^g > 0$.

   For $\alpha_1^g \leq \theta_1$, we have

$$\frac{m_\delta + (\alpha_1^g - \theta_1)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + {\theta_2}^2)\sigma_i^2 + (\alpha_1^g - \theta_1)^2\sigma_m^2}} \leq \frac{m_\delta}{\sqrt{({\alpha_2^g}^2 + {\theta_2}^2)\sigma_i^2 + (\alpha_1^g - \theta_1)^2\sigma_m^2}} \leq \frac{m_\delta}{\theta_2\sigma_i}$$

   and therefore,

$$\Phi\left(\frac{m_\delta + (\alpha_1^g - \theta_1)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + {\theta_2}^2)\sigma_i^2 + (\alpha_1^g - \theta_1)^2\sigma_m^2}}\right) \leq \Phi\left(\frac{m_\delta}{\theta_2\sigma_i}\right),$$

   where the equality holds only if $\alpha_1^g = \theta_1$ and $\alpha_2^g = 0$. We thus have $\alpha_1^g \leq \theta_1$,

$$\begin{aligned}
&U_g\left(\alpha^g = \begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix}; \alpha^b = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}\right) \\
&= (1 + m_g)(1 + R_f + \alpha_1^g(\mu - R_f)) + \lambda\Phi\left(\frac{m_\delta + (\alpha_1^g - \theta_1)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + {\theta_2}^2)\sigma_i^2 + (\alpha_1^g - \theta_1)^2\sigma_m^2}}\right) \\
&\leq (1 + m_g)(1 + R_f + \alpha_1^g(\mu - R_f)) + \lambda\Phi\left(\frac{m_\delta}{\theta_2\sigma_i}\right) \\
&\leq (1 + m_g)(1 + R_f + \theta_1(\mu - R_f)) + \lambda\Phi\left(\frac{m_\delta}{\theta_2\sigma_i}\right) \\
&= U_g\left(\alpha^g = \begin{bmatrix} \theta_1 \\ 0 \end{bmatrix}; \alpha^b = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}\right)
\end{aligned}$$

   where the equality holds only if $\alpha_1^g = \theta_1$ and $\alpha_2^g = 0$.

2. For any losing manager's feasible strategy, $\alpha^b = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$, the winning manager's strategy $\begin{bmatrix} \alpha_1^g \\ 0 \end{bmatrix}$, if $\alpha_1^g > \theta_1$, strictly dominates any one of her feasible strategies $\begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix}$ with $\alpha_2^g > 0$.

For $\alpha_1^g > \theta_1$, we have:

$$\frac{m_\delta + (\alpha_1^g - \theta_1)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + {\theta_2}^2)\sigma_i^2 + (\alpha_1^g - \theta_1)^2\sigma_m^2}} \leq \frac{m_\delta + (\alpha_1^g - \theta_1)(\mu - R_f)}{\sqrt{\theta_2^2\sigma_i^2 + (\alpha_1^g - \theta_1)^2\sigma_m^2}}$$

and therefore,

$$\Phi\left(\frac{m_\delta + (\alpha_1^g - \theta_1)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + {\theta_2}^2)\sigma_i^2 + (\alpha_1^g - \theta_1)^2\sigma_m^2}}\right) \leq \Phi\left(\frac{m_\delta + (\alpha_1^g - \theta_1)(\mu - R_f)}{\sqrt{\theta_2^2\sigma_i^2 + (\alpha_1^g - \theta_1)^2\sigma_m^2}}\right),$$

where the equality holds only if $\alpha_2^g = 0$. We thus have $\alpha_1^g > \theta_1$,

$$U_g\left(\alpha^g = \begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix}; \alpha^l = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}\right) \leq U_g\left(\alpha^g = \begin{bmatrix} \alpha_1^g \\ 0 \end{bmatrix}; \alpha^b = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}\right),$$

where the equality holds only if $\alpha_2^g = 0$.

3. We have

$$\frac{\partial}{\partial \alpha_1^g} U_g\left(\alpha^g = \begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix}; \alpha^l = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}\right)\Bigg|_{\alpha^g = \begin{bmatrix} \theta_1 \\ 0 \end{bmatrix}} > 0$$

$$U_w\left(\alpha^w = \begin{bmatrix} \alpha_1^w \\ 0 \end{bmatrix}; \alpha^l = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}\right)$$
$$= (1 + m_w)(1 + R_f + \alpha_1^w(\mu - R_f)) + \lambda\Phi\left(\frac{m_\delta + (\alpha_1^w - \theta_1)(\mu - R_f)}{\sqrt{\theta_2^2\sigma_i^2 + (\alpha_1^w - \theta_1)^2\sigma_m^2}}\right).$$

We then have

$$\frac{\partial}{\partial \alpha_1^g} U_g\left(\alpha^g = \begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix}; \alpha^b = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}\right)\Bigg|_{\alpha^g = \begin{bmatrix} \theta_1 \\ 0 \end{bmatrix}}$$
$$= (1 + m_g)(\mu - R_f) + \lambda\frac{\mu - R_f}{\theta_2\sigma_i}\phi\left(\frac{m_\delta}{\theta_2\sigma_i}\right) > 0,$$

where ϕ. denotes the standard normal density function.

4. For any winning manager's feasible strategy, $\begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix}$, if $\alpha_1^g > \alpha_1^b - \frac{m_\delta}{\mu - R_f}$ (especially if $\alpha_1^g > \alpha_1^b$), then the losing manager's strategy $\begin{bmatrix} \alpha_1^b \\ 1 - \alpha_1^b \end{bmatrix}$ strictly dominates any one of her feasible strategies with $\begin{bmatrix} \alpha_1^b \\ \alpha_2^b \end{bmatrix}$ with $\alpha_1^b + \alpha_2^b < 1$.

If $\alpha_1^b + \alpha_2^b < 1$, in the presence of the short sale constraint, we have ${\alpha_2^b}^2 < (1-\alpha_1^b)^2$, and therefore

$$({\alpha_2^g}^2 + {\alpha_2^b}^2)\sigma_i^2 + (\alpha_1^g - \alpha_1^b)^2\sigma_m^2 < \sqrt{({\alpha_2^g}^2 + (1-\alpha_1^b)^2)\sigma_i^2 + (\alpha_1^g - \alpha_1^b)^2\sigma_m^2}$$

. In the case that $\alpha_1^g \geq \alpha_1^b - \frac{m_\delta}{\mu - R_f}$, we have $\Delta + (\alpha_1^g - \alpha_1^b)(\mu - R_f) > 0$, and therefore

$$\frac{m_\delta + (\alpha_1^g - \alpha_1^b)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + {\alpha_2^b}^2)\sigma_i^2 + (\alpha_1^g - \alpha_1^b)^2\sigma_m^2}} > \frac{m_\delta + (\alpha_1^g - \alpha_1^b)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + (1-\alpha_1^b)^2)\sigma_i^2 + (\alpha_1^g - \alpha_1^b)^2\sigma_m^2}}$$

Thus

$$U_b\left(\alpha^g = \begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix}; \alpha^b = \begin{bmatrix} \alpha_1^b \\ \alpha_2^b \end{bmatrix}\right)$$

$$= (1+m_b)(1+R_f+\alpha_1^b(\mu - R_f)) + \lambda - \lambda\Phi\left(\frac{m_\delta + (\alpha_1^g - \alpha_1^b)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + {\alpha_2^b}^2)\sigma_i^2 + (\alpha_1^g - \alpha_1^b)^2\sigma_m^2}}\right)$$

$$< (1+m_b)(1+R_f+\alpha_1^b(\mu - R_f)) + \lambda -$$

$$\lambda\Phi\left(\frac{m_\delta + (\alpha_1^g - \alpha_1^b)(\mu - R_f)}{\sqrt{({\alpha_2^g}^2 + (1-\alpha_1^b)^2)\sigma_i^2 + (\alpha_1^g - \alpha_1^b)^2\sigma_m^2}}\right)$$

$$= U_b\left(\alpha^g = \begin{bmatrix} \alpha_1^g \\ \alpha_2^g \end{bmatrix}; \alpha^b = \begin{bmatrix} \alpha_1^b \\ 1-\alpha_1^b \end{bmatrix}\right)$$

To finish the proof, we derive the claim of the proposition from the above four proved statements. From Statement 1, we have that in the equilibrium $\alpha_1^g \geq \alpha_1^b$. From Statement 2, we have $\alpha_2^g = 0$. Then from Statement 3, we have that $\alpha_1^g \geq \alpha_1^b$ with the equation holding only if $\alpha_1^g = \alpha_1^b = 1$. From Statement 4, we have that $\alpha_1^b + \alpha_2^b = 1$. Thus, the proposition is proved.

Proof of Lemma 1: Given the assumption that the equilibrium exists, we can denote the equilibrium to be $\left\{\begin{bmatrix} {\alpha_1^g}^* \\ 0 \end{bmatrix}, \begin{bmatrix} {\alpha_1^b}^* \\ 1-{\alpha_1^b}^* \end{bmatrix}\right\}$ because of the result in

Proposition 1. We have,

$$U_g\left(\alpha_g = \begin{bmatrix} \alpha_1^g \\ 0 \end{bmatrix}, \alpha^b = \begin{bmatrix} \alpha_1^b \\ 1-\alpha_1^b \end{bmatrix}\right)$$
$$= (1+m_g)(1+R_f+\alpha_1^g(\mu-R_f)) + \lambda\Phi\left(\frac{m_\delta + (\alpha_1^g-\alpha_1^b)(\mu-R_f)}{\sqrt{(1-\alpha_1^b)^2\sigma_i^2 + (\alpha_1^g-\alpha_1^b)^2\sigma_m^2}}\right)$$
$$U_b\left(\alpha_g = \begin{bmatrix} \alpha_1^g \\ 0 \end{bmatrix}, \alpha^b = \begin{bmatrix} \alpha_1^b \\ 1-\alpha_1^b \end{bmatrix}\right)$$
$$= (1+m_g)(1+R_f+\alpha_1^b(\mu-R_f)) + \lambda - \lambda\Phi\left(\frac{m_\delta + (\alpha_1^g-\alpha_1^b)(\mu-R_f)}{\sqrt{(1-\alpha_1^b)^2\sigma_i^2 + (\alpha_1^g-\alpha_1^b)^2\sigma_m^2}}\right)$$

Combining the derivatives of the above two functions, we get

$$\frac{\partial}{\partial\alpha_1^g}U_g\left(\alpha_g = \begin{bmatrix} \alpha_1^g \\ 0 \end{bmatrix}, \alpha^b = \begin{bmatrix} \alpha_1^b \\ 1-\alpha_1^b \end{bmatrix}\right) - \frac{\partial}{\partial\alpha_1^b}U_b\left(\alpha_g = \begin{bmatrix} \alpha_1^g \\ 0 \end{bmatrix}, \alpha^b = \begin{bmatrix} \alpha_1^b \\ 1-\alpha_1^b \end{bmatrix}\right)$$
$$= m_\delta(\mu-R_f) + \lambda\phi\left(\frac{m_\delta + (\alpha_1^g-\alpha_1^b)(\mu-R_f)}{\sqrt{(1-\alpha_1^b)^2\sigma_i^2 + (\alpha_1^g-\alpha_1^b)^2\sigma_m^2}}\right)$$
$$\cdot \frac{(m_\delta + (\alpha_1^g-\alpha_1^b)(\mu-R_f))(1-\alpha_1^b)\sigma_i^2}{((1-\alpha_1^b)^2\sigma_i^2 + (\alpha_1^g-\alpha_1^b)^2\sigma_m^2)^{3/2}}$$
$$> 0.$$

This shows that the first order optimality condition for the portfolio allocation of manager g and the first order condition for manager b will not be satisfied at the same time. As a consequence, in the equilibrium, the short sale constraint is binding for at least one of the two managers. If the short sale constraint is binding for manager g, we have, by Proposition 1, $\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. If the short sale constraint is binding for manager b, we have either $\alpha^{b^*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\alpha^{b^*} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. In the case that $\alpha^{b^*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, we have, again by Proposition 1, $\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Therefore, in equilibrium, either $\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, or $\alpha^{b^*} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, or both.

Discussion of the first order condition for the portfolio allocation of manager b: In general, the first order optimality condition for the portfolio allocation of manager b is

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \alpha_1^b} U_b \left(\alpha^g = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \alpha^b = \begin{bmatrix} \alpha_1^b \\ 1 - \alpha_1^b \end{bmatrix} \right) \\
&= (1 + m_l)(\mu - R_f) - \lambda\phi \left(\frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} \frac{1}{1 - \alpha_1^b} + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}} \right) \frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} \frac{1}{(1 - \alpha_1^b)^2}
\end{aligned}
$$

In the equilibrium, either the above first order condition is satisfied, or $\alpha^{b*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. The first order condition is equivalent to

$$
(1 + m_b)(\mu - R_f) = \lambda\phi \left(\frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} x + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}} \right) \frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} x^2,
$$

where $x = \frac{1}{1 - \alpha_1^b}$. Notice that both sides are positive. We can therefore take natural log of both sides, which yields the equivalent condition:

$$
\begin{aligned}
h(x) \equiv \log & \left(\frac{(1 + m_l)(\mu - R_f)\sqrt{\sigma_i^2 + \sigma_m^2}\sqrt{2\pi}}{\lambda m_\delta} \right) + \frac{1}{2} \left(\frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} x + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}} \right)^2 \\
& - 2 \log x = 0
\end{aligned}
$$

It is straightforward to see that $h(x)$ and $\frac{\partial}{\partial \alpha_1^b} U_b \left(\alpha^g = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \alpha^b = \begin{bmatrix} \alpha_1^b \\ 1 - \alpha_1^b \end{bmatrix} \right)$ have the same sign. Taking derivatives of $h(x)$, we have:

$$
\begin{aligned}
h'(x) &= \lambda\phi \left(\frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} x + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}} \right) \frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} - \frac{2}{x} \\
h''(x) &= \left(\frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} \right)^2 + \frac{2}{x^2} > 0.
\end{aligned}
$$

Therefore, $h(x)$ is a strictly convex function and thus has at most two roots. It is clear that $\lim_{x \to 0^+} h(x) = \lim_{x \to \infty} h(x) = +\infty$. In the case that $h(x)$ has only one root or no root, by the middle value theorem of continuous function, we have $h(x) \geq 0$ for $x \in [1, \infty)$. Therefore U_b will be a strictly increasing function with respect to the variable α_1^b. We can then conclude that, in this case, $\alpha^{b^*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ will be manager b's optimal response.

In the case that function $h(x)$ has two roots, we denote the roots by x_1 and x_2 with $1 < x_1 < x_2$. Because $h(x)$ is a convex function, we have $h(x) > 0$ for $x \in [1, x_1) \cup (x_2, \infty)$, and $h(x) < 0$ for $x \in [1 - 1/x_1, 1 - 1/x_2]$. Therefore,

$$\arg\max_{\alpha_1^b} U_b \left(\alpha^g = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \alpha^b = \begin{bmatrix} \alpha_1^b \\ 1 - \alpha_1^b \end{bmatrix} \right) = 1 \text{ or } 1 - \frac{1}{x_1}.$$

We derive some properties regarding the equation $h(x) = 0$, especially regarding x_1, and record the results in the following several lemmas. The existence of roots for $h(x)$ is given by the following lemma:

Lemma 2. *: $h(x)$ has two roots in $(0, +\infty)$ if and only if the following condition holds:*

$$\frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} < C_1, \tag{1}$$

where

$$C_1 = \frac{\lambda}{4(1 + m_l)(\mu - R_f)} \phi \left[\frac{1}{2} \left(\sqrt{\frac{(\mu - R_f)^2}{\sigma_i^2 + \sigma_m^2} + 8} + \frac{(\mu - R_f)}{\sqrt{\sigma_i^2 + \sigma_m^2}} \right) \right]$$
$$\cdot \left(\sqrt{\frac{(\mu - R_f)^2}{\sigma_i^2 + \sigma_m^2} + 8} - \frac{(\mu - R_f)}{\sqrt{\sigma_i^2 + \sigma_m^2}} \right)^2 .$$

In addition, $h(x)$ has one root if and only if $\frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} < C_1$.

Proof. The only critical point of $h(x)$, located by the first order condition $h'(x) = 0$, is at

$$x^* = \frac{1}{2} \left(\frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} \right)^{-1} \sqrt{\frac{(\mu - R_f)^2}{\sigma_i^2 + \sigma_m^2} + 8} - \frac{(\mu - R_f)}{\sqrt{\sigma_i^2 + \sigma_m^2}}. \tag{2}$$

Therefore, $h(x^*) = \min h(x)$. The ending behavior of $h(x)$ is given by:

$$\lim_{x \to \infty} h(x) \geq \lim_{x \to \infty} \frac{m_\delta^2}{2(\sigma_i^2 + \sigma_m^2)} x^2 - 2x = +\infty,$$
$$\lim_{x \to 0} h(x) \geq -2 \lim_{x \to \infty} \log x = +\infty.$$

Since $h(x)$ is convex, we then have that $h(x)$ has one root in $(0, +\infty)$ if and only if $h(x^*) = 0$, and it has two distinct roots in $(0, +\infty)$ if and only if $h(x^*) < 0$. By substituting x^* into $h(x)$, we get the proof the lemma.

The above Lemma 2 imposes a restriction on the performance gap between the winning manager and the losing manager. The gap, m_δ, can not be too large in order for $h(x)$ to have roots. If the condition of the lemma does not hold, $h(x)$ will always be positive. It will then lead to the conclusion that it is optimal for the losing manager to choose $\alpha^{l^*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ as her optimal strategy. Notice that the bound for the gap in the above condition is proportional to λ, the reward from winning the tournament.

Another way to express the bound in the above lemma is to say that there is at least one feasible portfolio holding for the losing manager at which this manager would prefer to increase her exposure to idiosyncratic risk. Increasing the exposure to idiosyncratic risk affects the losing manager's utility in two ways. First, increasing the portfolio allocation to idiosyncratic risk will increase the variance of the difference between her portfolio and the winning manager's portfolio. This, in turn, will introduce more noise in the outcome of the tournament, and therefore has the potential of increasing the chance for the losing manager to make up the gap and win the tournament. Second, an increase in the idiosyncratic risk exposure implies a reduction in the allocation to the market index and hence, a reduction in the expected return. The lower expected return has two consequences for the losing manager's utility. It will lower the losing manager's expected reward from the explicit linear contract, and it will lower her chance of winning the tournament. The losing manager has to take all these three factors into consideration when making the portfolio decision.

For the rest of the section, we will maintain the assumption that the condition of Lemma 2 holds. That is, we assume that $h(x)$ has two roots. It will be sufficient for that condition to hold if we assume that the tournament has some effect on at least one of the managers' portfolio decision. We will maintain the notation x_1 to denote the smaller one of the two roots of $h(x)$.

Proof of Proposition 2: It follows immediately from Lemma 2 that, under the conditions of the proposition, $\alpha^{b*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is optimal for manager b given $\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. The optimality of $\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ for manager g is assured by Proposition 1. Thus $\left\{ \alpha^{w*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \alpha^{b*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$ is a Nash equilibrium.

Lemma 3.

$$\frac{\partial x_1}{\partial \lambda} = \frac{1}{h'(x_1)\lambda} < 0.$$

Proof. Since x_1 is the smaller of the two roots of the convex function $h(x)$, therefore, $h(x_1) = 0$ and $h'(x_1) < 0$. From $h(x_1) = 0$, we have

$$\log\left((1+m_l)(\mu - R_f)\sqrt{\sigma_i^2 + \sigma_m^2}\sqrt{2\pi}\right) - \log\lambda - \log m_\delta$$
$$+ \frac{1}{2}\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}x_1 + \frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)^2 - 2\log x_1 = 0.$$

Differentiating the above equation, we get $\frac{\partial x_1}{\partial \lambda} = \frac{1}{h'(x_1)\lambda} < 0$. Thus the lemma is proved.

Lemma 4. $\frac{\partial [m_\delta x_1]}{\partial m_\delta} > 0$ *and* $\lim_{m_\delta \to 0^+} m_\delta x_1 = 0$. *Moreover, we have* $\lim_{m_\delta \to 0^+} x_1 = +\infty$ *, and to be more precise, we have* $x_1 = O(m_\delta^{-1/2})$.[11]

Proof. By $h'(x_1) < 0$, we have

$$\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}x_1 + \frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}} - \frac{2}{x_1} < 0.$$

Differentiate equation $h(x_1) = 0$, we get that

$$\frac{\partial x_1}{\partial m_\delta} = \frac{\left[\frac{1}{m_\delta} - \left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}x_1 + \frac{\mu-R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{x_1}{\sqrt{\sigma_i^2+\sigma_m^2}}\right]}{\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}x_1 + \frac{\mu-R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}} - \frac{2}{x_1}}.$$

[11] The notation O means that both x_1 and $m_\delta^{-1/2}$ approach infinity with the same order of speed when $m_\delta \to 0^+$.

Therefore,

$$\begin{aligned}\frac{\partial[m_\delta x_1]}{\partial m_\delta} &= x_1 + m_\delta \frac{\partial x_1}{\partial m_\delta} \\ &= -\left[\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}x_1 + \frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}} - \frac{2}{x_1}\right]^{-1} \\ &= -\frac{1}{h'(x_1)} > 0\end{aligned}$$

We thus have that $m_\delta x_1$ is a increasing function of m_δ. Moreover, $m_\delta x_1 > 0$. Therefore, $(1+m_l)(\mu - R_f) = \lambda\phi\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}x_1 + \frac{\mu-R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}x_1^2$, $\lim_{m_\delta\to 0^+} m_\delta x_1$ exists, is bounded and non-negative. An equivalent equation of $h(x_1) = 0$ is:

$$(1+m_l)(\mu - R_f)m_\delta = \lambda\phi\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}x_1 + \frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{(m_\delta x_1)^2}{\sqrt{\sigma_i^2+\sigma_m^2}} \tag{3}$$

Take the limit of both sides of the above equality by letting $m_\delta \to 0^+$, we have:

$$0 = \lambda\phi\left(\frac{\lim_{m_\delta\to 0^+} m_\delta x_1}{\sqrt{\sigma_i^2+\sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{(\lim_{m_\delta\to 0^+} m_\delta x_1)^2}{\sqrt{\sigma_i^2+\sigma_m^2}},$$

which implies that $\lim_{m_\delta\to 0^+} m_\delta x_1 = 0$. From equation (3) we have

$$\begin{aligned}\lim_{m_\delta\to 0^+} m_\delta x_1^2 &= \frac{1}{\lambda}(1+m_l)(\mu - R_f)\sqrt{\sigma_i^2+\sigma_m^2}\left[\phi\left(\frac{\lim_{m_\delta\to 0^+} m_\delta x_1}{\sqrt{\sigma_i^2+\sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\right]^{-1} \\ &= \frac{1}{\lambda}(1+m_l)(\mu - R_f)\sqrt{\sigma_i^2+\sigma_m^2}\left[\phi\left(\frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\right]^{-1}\end{aligned}$$

The right hand side of the above identity is positive. Therefore, we have that $\lim_{m_\delta\to 0^+} x_1 = +\infty$, and to be more precise, $x_1 = O(m_\delta^{-1/2})$. Thus the lemma is proved.

Lemma 5. $\frac{\partial x_1}{\partial m_\delta} < 0$ *for sufficiently small* m_δ. *In fact, it is true for any*

$$m_\delta \in (0, \sqrt{\sigma_i^2+\sigma_m^2}\cdot \min(C_1, C_2)), \tag{4}$$

where C_1 *is given as in Lemma 2, and*

$$\begin{aligned}C_2 = &\frac{\lambda}{4(1+m_l)(\mu - R_f)}\cdot\phi\left[\frac{1}{2}\left(\sqrt{\frac{(\mu - R_f)^2}{\sigma_i^2+\sigma_m^2}+4} + \frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\right] \\ &\cdot\left(\sqrt{\frac{(\mu - R_f)^2}{\sigma_i^2+\sigma_m^2}+4} - \frac{\mu - R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)^2\end{aligned}$$

Proof. In the derivation of Lemma 3, we see

$$\frac{\partial x_1}{\partial m_\delta} = \frac{\left[\frac{1}{m_\delta} - \left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}x_1 + \frac{\mu-R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{x_1}{\sqrt{\sigma_i^2+\sigma_m^2}}\right]}{h'(x_1)}.$$

Since $h'(x_1) < 0$, we have $\frac{\partial x_1}{\partial m_\delta} < 0$ if and only if,

$$1 > \left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}x_1 + \frac{\mu-R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\frac{m_\delta x_1}{\sqrt{\sigma_i^2+\sigma_m^2}}.$$

The right hand side is a quadratic form of $\frac{m_\delta x_1}{\sqrt{\sigma_i^2+\sigma_m^2}}$. It is straightforward to solve the above inequality. Given $\frac{m_\delta x_1}{\sqrt{\sigma_i^2+\sigma_m^2}} > 0$, we have $\frac{\partial x_1}{\partial m_\delta} < 0$ if and only if

$$0 < \frac{m_\delta x_1}{\sqrt{\sigma_i^2+\sigma_m^2}} < \frac{1}{2}\left(\sqrt{\frac{(\mu-R_f)^2}{\sigma_i^2+\sigma_m^2}+4} - \frac{\mu-R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right). \tag{5}$$

By Lemma 3, we have $\lim_{m_\delta\to 0^+} m_\delta x_1 = 0$. Therefore, the above inequalities will be satisfied for sufficiently small m_δ. To answer the question of how small is sufficiently small, we evoke the equality $h(x_1) = 0$. With this equation, we can see that the necessary condition for

$$\frac{m_\delta x_1}{\sqrt{\sigma_i^2+\sigma_m^2}} = \frac{1}{2}\left(\sqrt{\frac{(\mu-R_f)^2}{\sigma_i^2+\sigma_m^2}+4} - \frac{\mu-R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)$$

is $\frac{m_\delta x_1}{\sqrt{\sigma_i^2+\sigma_m^2}} = C_2$. Therefore, for any $m_\delta \in (0, C_2\sqrt{\sigma_i^2+\sigma_m^2})$, we always have

$$\frac{m_\delta x_1}{\sqrt{\sigma_i^2+\sigma_m^2}} \neq \frac{1}{2}\left(\sqrt{\frac{(\mu-R_f)^2}{\sigma_i^2+\sigma_m^2}+4} - \frac{\mu-R_f}{\sqrt{\sigma_i^2+\sigma_m^2}}\right).$$

This together with the fact that inequality (6) is satisfied by sufficiently small m_δ, implies that inequality (6) is satisfied by any $m_\delta \in (0, C_2\sqrt{\sigma_i^2+\sigma_m^2})$, where we use the fact that $\frac{m_\delta x_1}{\sqrt{\sigma_i^2+\sigma_m^2}}$ is a continuous function of m_δ. Thus the lemma is proved.

Proof of Proposition 3: In case that $\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and $\alpha_1^{b*} < 1$, the condition in Lemma 2 is satisfied. Therefore, Lemma 3 and Lemma 4 imply that $\alpha_1^{b*} = 1 - 1/x_1$, and therefore, $\frac{\partial(1-\alpha_1^{b*})}{\partial\lambda} = \frac{\partial}{\partial\lambda}(1/x_1) = -\frac{1}{x_1^2}\frac{\partial x_1}{\partial\lambda} > 0$, because of Lemma 3. This proves the first part of the proposition. Taking the derivative and using Lemma 4, we have

$$\begin{aligned}&\frac{\partial}{\partial m_\delta}\Phi\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}\frac{1}{1-\alpha_1^{b*}} + \frac{(\mu-R_f)}{\sqrt{\sigma_i^2+\sigma_m^2}}\right)\\ &= \phi\left(\frac{m_\delta}{\sqrt{\sigma_i^2+\sigma_m^2}}\frac{1}{1-\alpha_1^{b*}} + \frac{(\mu-R_f)}{\sqrt{\sigma_i^2+\sigma_m^2}}\right) \times \frac{\partial(m_\delta x_1)}{m_\delta} > 0\end{aligned}$$

Thus the proposition is proved.

Lemma 6. *If* $\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$*, then for sufficiently small* m_δ *we have* $\alpha_1^{b^*} < 1$*. In fact, it is true for any* $m_\delta \in (0, \sqrt{\sigma_i^2 + \sigma_m^2} \cdot \min(C_1, C_2, C_3))$*, where* $C_3 = \tilde{m_\delta}/\sqrt{\sigma_i^2 + \sigma_m^2}$ *with* $\tilde{m_\delta}$ *being a root in* $(0, \sqrt{\sigma_i^2 + \sigma_m^2} \cdot \min(C_1, C_2))$*. Further,*

$$(1 + m_b)(\mu - R_f) = \lambda \left(1 - \Phi\left[\frac{m_\delta x_1}{\sqrt{\sigma_i^2 + \sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}}\right]\right) x_1 \tag{6}$$

if there is such a root, and $C_3 = +\infty$ *otherwise.*

Proof. Notice that

$$\begin{aligned}
&\frac{\partial}{\partial m_\delta}\left(1 - \Phi\left[\frac{m_\delta x_1}{\sqrt{\sigma_i^2 + \sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}}\right]\right) x_1 \\
&= -\phi\left[\frac{m_\delta x_1}{\sqrt{\sigma_i^2 + \sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}}\right] \frac{x_1^2}{\sqrt{\sigma_i^2 + \sigma_m^2}} \\
&+ \left(1 - \Phi\left[\frac{m_\delta x_1}{\sqrt{\sigma_i^2 + \sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}}\right]\right) \frac{\partial x_1}{\partial m_\delta} \\
&< 0
\end{aligned}$$

for m_δ in $(0, \sqrt{\sigma_i^2 + \sigma_m^2} \cdot \min(C_1, C_2))$, because the first term is clearly less than zero, and the second term is less than zero because of Lemma 4. Therefore, the right hand side expression of Equation (6) as a function of m_δ is decreasing in $(0, \sqrt{\sigma_i^2 + \sigma_m^2} \cdot \min(C_1, C_2))$. Thus equation (6) has at most one root in the interval. And if there is a root, any m_δ less than the root will satisfied the inequality with left hand side of (6) strictly less than the right hand side. Furthermore, we note the fact that

$$\begin{aligned}
&\lim_{m_\delta \to 0^+} \lambda \left(1 - \Phi\left[\frac{m_\delta x_1}{\sqrt{\sigma_i^2 + \sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}}\right]\right) x_1 \\
&= \lambda \left(1 - \Phi\left[\frac{\lim_{m_\delta \to 0^+} m_\delta x_1}{\sqrt{\sigma_i^2 + \sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}}\right]\right) \lim_{m_\delta \to 0^+} x_1 = +\infty
\end{aligned}$$

Therefore, in the case that Equation (6) has no root in the interval, we will always have the left hand side of (6) being strictly less than the right hand side. Here the continuity of the right hand side of (4) as a function of m_δ is used.

In summary, we have that the left hand side of Equation (6) is strictly less than the right hand side if $m_\delta \in (0, \sqrt{\sigma_i^2 + \sigma_m^2} \cdot \min(C_1, C_2, C_3))$. The earlier discussion about the first order condition of the losing manager's optimization problem leads us to conclude that $\arg\max_{\alpha_1^b} U_b \left(\alpha^g = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \alpha^b = \begin{bmatrix} \alpha_1^b \\ 1 - \alpha_1^b \end{bmatrix}\right) = 1$ or $1 - \frac{1}{x_1}$. By comparing the value of U_b at $\alpha_1^b = 1$ with the value of U_b at $\alpha_1^b = 1 - 1/x_1$, we have the necessary and sufficient condition for $\alpha_1^{b^*} < 1$ to be that the condition (1) in Lemma 2 holds and that at the same time the left hand side of Equation (6) is strictly less than the right hand side. For each $m_\delta \in (0, \sqrt{\sigma_i^2 + \sigma_m^2} \cdot \min(C_1, C_2, C_3))$, both the above conditions are satisfied. Thus the lemma is proved.

Based on the earlier lemmas, we derive the following sufficient condition for the existence of the equilibrium.

THEOREM 1: *Let* $\alpha_1^{b^*} = 1 - 1/x_1$. *The pair* $\left\{ \alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \alpha^{b^*} = \begin{bmatrix} \alpha_1^{b^*} \\ 1 - \alpha_1^{b^*} \end{bmatrix} \right\}$ *will be an equilibrium for any* $m_\delta \in (0, \sqrt{\sigma_i^2 + \sigma_m^2} \cdot \min(C_1, C_2, C_3, C_4))$, *where*

$$C_4 = \frac{\lambda(\mu - R_f)\sigma_i^2}{(1 + m_l)\sigma_m^2} \cdot n \left[\frac{(\mu - R_f)\sqrt{\sigma_i^2 + \sigma_m^2}}{\sigma_m^2} \right].$$

Furthermore, $\frac{\partial(1 - \alpha_1^{b^*})}{\partial m_\delta} > 0$ *and* $\lim_{m_\delta \to 0^+}(1 - \alpha_1^{b^*}) = 0$. *To be more precise, we have that* $1 - \alpha_1^{b^*} = O(\sqrt{m_\delta})$.

Proof. The optimality of the losing manager's strategy $\alpha^{b^*} = \begin{bmatrix} 1 - 1/x_1 \\ 1/x_1 \end{bmatrix}$ has already been established in Lemma 6. We now prove the optimality of the winning manager's strategy $\alpha^{g*} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ under the condition of the theorem. We evoke the equation $h(x_1) = 0$, which is equivalent to

$$\frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} = (1 + m_l)(\mu - R_f)\lambda\phi \left(\frac{m_\delta x_1}{\sqrt{\sigma_i^2 + \sigma_m^2}} + \frac{\mu - R_f}{\sqrt{\sigma_i^2 + \sigma_m^2}} \right) \left(\frac{m_\delta x_1}{\sqrt{\sigma_i^2 + \sigma_m^2}} \right)^2.$$

With this equation, we can see that the necessary condition for $(\mu - R_f)\sigma_i^2 - m_\delta x_1 \sigma_m^2 = 0$ is $\frac{m_\delta}{\sqrt{\sigma_i^2 + \sigma_m^2}} = C_4$. Therefore, there is no root for $(\mu - R_f)\sigma_i^2 - m_\delta x_1 \sigma_m^2 = 0$ with $m_\delta \in (0, C_4\sqrt{\sigma_i^2 + \sigma_m^2})$. This coupled with the fact that $\lim_{m_\delta \to 0^+}(m_\delta x_1) = 0$, implies that $(\mu - R_f)\sigma_i^2 - m_\delta x_1 \sigma_m^2 > 0$ for $m_\delta \in (0, C_4\sqrt{\sigma_i^2 + \sigma_m^2})$. We next compute the derivative of U_g with respect to α_1^g.

$$\frac{\partial}{\partial \alpha_1^g} U_g \left(\alpha^g = \left[\begin{array}{c} \alpha_1^g \\ 0 \end{array} \right]; \alpha^b = \left[\begin{array}{c} 1 - 1/x_1 \\ 1/x_1 \end{array} \right] \right)$$
$$= (1 + m_b + m_\delta)(\mu - R_f) + \lambda\phi \left(\frac{m_\delta x_1 + (x_1\alpha_1^g - x_1 + 1)(\mu - R_f)}{\sqrt{\sigma_i^2 + (x_1\alpha_1^g - x_1 + 1)^2 \sigma_m^2}} \right)$$
$$\frac{x_1(\mu - R_f)(\sigma_i^2 + (x_1\alpha_1^g - x_1 + 1)^2\sigma_m^2) - x_1(m_\delta x_1 + (x_1\alpha_1^g - x_1 + 1)(\mu - R_f))}{(\sigma_i^2 + (x_1\alpha_1^g - x_1 + 1)^2\sigma_m^2)^{3/2}}$$
$$\cdot \frac{(x_1\alpha_1^g - x_1 + 1)\sigma_m^2}{(\sigma_i^2 + (x_1\alpha_1^g - x_1 + 1)^2\sigma_m^2)^{3/2}}$$
$$= (1 + m_b + m_\delta)(\mu - R_f) + \lambda\phi \left(\frac{m_\delta x_1 + (x_1\alpha_1^g - x_1 + 1)(\mu - R_f)}{\sqrt{\sigma_i^2 + (x_1\alpha_1^g - x_1 + 1)^2 \sigma_m^2}} \right)$$
$$\frac{x_1[(\mu - R_f)\sigma_i^2 - m_\delta x_1^2(\alpha_1^g - 1 + 1/x_1)\sigma_m^2]}{(\sigma_i^2 + (x_1\alpha_1^g - x_1 + 1)^2\sigma_m^2)^{3/2})}$$
$$> 0$$

The last inequality is true for all $\alpha_1^g \in [\alpha_1^l, 1]$, and for $m_\delta \in (0, C_4\sqrt{\sigma_i^2 + \sigma_m^2})$. Therefore, $\alpha^{g*} = \left[\begin{array}{c} 1 \\ 0 \end{array} \right]$ is optimal for the winning manager. Furthermore, $\frac{\partial(1-\alpha_1^{b^*})}{\partial m_\delta} = -\frac{1}{x_1^2}\frac{\partial x_1}{\partial m_\delta} < 0$ by lemma 5, and $\lim_{m_\delta \to 0^+}(1 - \alpha_1^{b^*}) = \lim_{m_\delta \to 0^+} \frac{1}{x_1} = 0$ by lemma 4. Also by lemma 4, $1 - \alpha_1^{b^*} = O(\sqrt{m_\delta})$. Thus we have proved the theorem.

Proof of Corollary 1: The Sharpe Ratio of fund i, denoted by SR^i is:

$$SR^i = \frac{\alpha_1^i(\mu - R_f)}{\sqrt{(\alpha_1^i\sigma_m)^2 + ((1 - \alpha_1^i)\sigma_i)^2}}.$$

Taking the derivative of SR^i with respect to variable α_1^i and after simplification, we have

$$\frac{\partial}{\partial \alpha_1^i}(SR^i) = \frac{(\mu - R_f)\sigma_i^2}{(\sqrt{(\alpha_1^i\sigma_m)^2 + ((1 - \alpha_1^i)\sigma_i)^2})^3}(1 - \alpha_1^i)$$

Therefore, $\frac{\partial}{\partial \alpha_1^i}(SR^i) > 0$ for $\alpha_1^i < 1$. From Theorem 1, in the equilibrium, we have $\alpha_1^{g^*} = 1$ and $\alpha_1^{b^*} < 1$, and thus $SR^b < SR^g$. Moreover, we have that $\frac{\partial}{\partial m_\delta}(SR^b) = \frac{\partial}{\partial \alpha_1^b}(SR^b)\frac{\partial \alpha_1^{b^*}}{\partial m_\delta} < 0$ from Theorem 1, and $\frac{\partial}{\partial \lambda}(SR^b) = \frac{\partial}{\partial \alpha_1^b}(SR^b)\frac{\partial \alpha_1^{b^*}}{\partial \lambda} < 0$ from proposition.

REFERENCES

Ahn, David, Syngjoo Choi, Douglas Gale, and Shachar Karivl, 2011, Estimating ambiguity aversion in a portfolio choice experiment, Working paper, University of Chicago.

Anand, Amber, Paul Irvine, Andy Puckett, and Kumar Venkataraman, 2010, Market crashes and institutional trading, Working Paper, University of Georgia.

Anand, Amber, Paul Irvine, Andy Puckett, and Kumar Venkataraman, 2012, Performance of institutional trading desks: An analysis of persistence in trading costs, *Review of Financial Studies* 25, 557–598.

Anderson, E. W., E. Ghysels, and J.L. Juergens, 2009, The impact of risk and uncertainty on expected returns, *Journal of Financial Economics* 94, 233–263.

Arnuk, Sal L., and Joseph Saluzzi, 2008, Toxic equity trading order flow on wall street, White Paper, Themis Trading LLC.

Bagnoli, Mark, and Susan G. Watts, 2000, Chasing hot funds: The effects of relative performance on portfolio choice, *Financial Management* 29, 31–50.

Bailey, Warren, Alok Kumar, and David T. Ng, 2011, Behavioral biases of mutual fund investors, *Journal of Financial Economics* 102, 1–27.

Barber, B., T. Odean, and L. Zheng, 2005, Out of sight, out of mind: the effects of expenses on mutual fund flows, *Journal of Business* 78, 2095–2120.

Barber, Brad M., and Terrance Odean, 2000, Trading is hazardous to your wealth: The common stock investment performance of individual investors, *Journal of Finances* 55, 773–806.

Barber, Brad M., and Terrance Odean, 2001, Boys will be boys: Gender, overconfidence, and common stock investment, *Quarterly Journal of Economics* 116, 261–292.

Barber, Brad M., and Terrance Odean, 2002, Online investors: Do the slow die first?, *Review of Financial Studies* 15, 455–489.

Basak, Suleyman, Anna Pavlova, and Alex Shapiro, 2000, Offsetting the incentives: risk shifting and benefits of benchmarking in money management, London Business School.

Biais, B., T. Foucault, and S. Moinas, 2011, Equilibrium algorithmic trading, Working paper, IDEI Toulouse.

Boehmer, E., K. Fong, and J. Wu, 2012, International evidence on algorithmic trading, Working paper, University of Georgia.

Bossaerts, Peter, Paolo Ghirardato, Serena Guarnaschelli, and William Zame, 2010, Ambiguity in asset markets: Theory and experiment, *Review of Financial Studies* 23, 1325–1359.

Brogaard, J., 2010, High frequency trading and its impact on market quality, Working paper, University of Washington.

Brogaard, J., T.J. Hendershott, S. Hunt, T. Latza, L. Pedace, and C. Ysusi, 2012, High frequency trading and the execution costs of institutional investors, UK Government's Foresight Project: The future of computer trading in financial markets, Foresight Driver Review 21.

Brogaard, J., T.J. Hendershott, and R. Riordan, 2013, High frequency trading and price discovery, Working paper, University of Washington.

Brown, Keith, W. Harlow, and Laura Starks, 1996, Of tournaments and temptations: an analysis of managerial incentives in the mutual fund industry, *Journal of Finance* 51, 85–110.

Busse, Jeffrey, 2001, Another look at mutual fund tournaments, *Journal of Financial and Quantitative Analysis* 36, 53–73.

Campbell, J., A. Lo, and C. MacKinlay, 1970, *The econometrics of financial markets* (Princeton University Press, Princeton, NJ).

Carhart, Mark M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Carrion, A., 2013, Very fast money: High-frequency trading on nasdaq, *Journal of Financial Markets* forthcoming.

Cartea, A., and J. Penalva, 2005, Where is the value in high frequency trading?, Working paper, Universidad Carlos III de Madrid.

Chaboud, A., B. Chiquoine, E. Hjalmarsson, and C. Veg, 2009, Rise of the machines: Algorithmic trading in the foreign exchange market, Working paper, Federal Reserve Board.

Chemmanur, T., S. He, and G. Hu, 2009, The role of institutional investors in seasoned equity offerings, *Journal of Financial Economics* 94, 384–411.

Chen, Hsiu-lang, and George G. Pennacchi, 1996, Does prior performance affect a mutual fund's choice of risk? theory and further empirical evidence.

Chen, Z., and L. G. Epstein, 2002, Ambiguity, risk and asset returns in continuous time, *Econometrica* 70, 1403–1443.

Chevalier, Judith A., and Glenn Ellison, 1997, Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Commission, US Securities & Exchange, 2010, Concept release on equity market structure, Release No. 34-61458, File No. S7-02-10.

Condie, Scott, 2008, Living with ambiguity: prices and survival when investors have heterogeneous preferences for ambiguity, *Economic Theory* 36, 81–108.

Cremers, Martijn, and Antti Petajisto, 2009, How active is your fund manager? a new measure that predicts performance, *Review of Financial Studies* 22, 3329–3365.

Del Guercio, Diane, and Paula A. Tkac, 2002, The determinants of the flow of funds of managed portfolios: Mutual funds vs. pension funds, *Journal of Financial and Quantitative Analysis* 37, 523–557.

Deli, Daniel N., 2002, Mutual fund advisory contracts: An empirical investigation, *Journal of Finance* 57, 109–133.

Easley, D., M. Lopez de Prado, and M. O'Hara, 2011a, The microstructure of the 'flash crash': low toxicity, liquidity crashes and the probability of informed trading, *Journal of Portfolio Management* 37, 118–128.

Easley, David, and Maureen O'Hara, 2009, Ambiguity and nonparticipation: The role of regulation, *Review of Financial Studies* 22, 1818–1843.

Egginton, Jared F., Bonnie F. Van Ness, and Robert A. Van Ness, 2014, Quote stuffing, Working paper.

Ellsberg, D., 1961, Risk, ambiguity, and the savage axioms, *Quarterly Journal of Economics* 75, 643–669.

Epstein, Larry G., and Martin Schneider, 2003, Recursive multiple-priors, *Journal of Economic Theory* 113, 1–31.

Epstein, Larry G., and Martin Schneider, 2008, Ambiguity, information quality, and asset pricing, *Journal of Finance* 63, 197–228.

Epstein, Larry G., and Tan Wang, 1994, Intertemporal asset pricing under knightian uncertainty, *Econometrica* 62, 283–322.

Fama, E., 1965, The behavior of stock market prices, *Journal of Business* 38, 34–105.

Fama, Eugene F., and James D. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

French, Kenneth, 2008, The cost of active investing, Dartmoth College, NBER.

Gallaher, Steven, Ron Kaniel, and Laura T. Starks, 2005, Madison avenue meets wall street: Mutual fund families, competition and advertising, Working paper, University of Texas.

Gilboa, Itzhak, and David Schmeidler, 1989, Maxmin expected utility with non-unique prior, *Journal of Mathematical Economics* 18, 141–153.

Goldstein, M., P. Irvine, E. Kandel, and Z. Weiner, 2009, Brokerage commissions and institutional trading patterns, *Review of Financial Studies* 22, 5175–5212.

Goldstein, M., P. Irvine, and A. Puckett, 2010, Purchasing ipos with commissions, *Journal of Financial and Quantitative Analysis* 46, 1193–1225.

Goriaev, Alexei, Frederic Palomino, and Andrea Prati, 2003, Mutual fund tournament: risk taking incentives induced by ranking objectives.

Green, Jerry R., and Nancy L. Stokey, 1983, A comparison of tournaments and contracts, *Journal of Political Economy* 91, 349–364.

Grinblatt, Mark, and Sheridan Titman, 1992, The persistence of mutual fund performance, *Journal of Finance* 47, 1977–1984.

Group, ITG Investment Technology, 2009, Itg's global cost review, White Paper.

Gruber, Martin J., 1996, Another puzzle: The growth in actively managed mutual funds, *Journal of Finance* 51, 783–810.

Hasbrouck, J., and G. Saar, 2013, Low-latency trading, *Journal of Financial Markets* Forthcoming.

Hendershott, T., C. Jones, and A.J. Menkveld, 2011, Does algorithmic trading increase liquidity?, *Journal of Finance* 66, 1–33.

Hirschey, N., 2011, Do high-frequency traders anticipate buying and selling pressure?, Working paper, London Business School.

Hoffmann, P., 2013, A dynamic limit order market with fast and slow traders, Working Paper.

Hu, Ping, Jayant R. Kale, Marco Pagani, and Ajay Subramanian, 2011, Fund flows, performance, managerial career concerns, and risk-taking, *Management Science* 57, 628–646.

Huang, Jennifer, K. D. Wei, and H. Yan, 2007, Participation costs and the sensitivity of fund flows to past performance, *Journal of Finance* 62, 1273–1311.

Huang, Jennifer, K. D. Wei, and H. Yan, 2012, Investor learning and mutual fund flows, Working paper, University of Texas.

Illeditsch, P. K., 2011, Ambiguous information, portfolio inertia,and excess volatility, *Journal of Finance* 66, 2213–2247.

Ippolito, Richard A., 1992, Consumer reaction to measures of poor quality: Evidence from the mutual fund industry, *Journal of Forecasting* 35, 45–70.

Jain, Prem C., and Joanna S. Wu, 2000, Truth in mutual fund advertising: Evidence on future performance and fund flows, *Journal of Finance* 55, 937–958.

Jarrow, R., and P. Protter, 2011, A dysfunctional role of high frequency trading in electronic markets, Working paper, Cornell University.

Jensen, Michael, 1968, The performance of mutual funds in the period 1945-1964, *Journal of Finance* 23, 389–416.

Jiang, G., T. Yao, and T. Yu, 2007, Do mutual funds time the market? evidence from portfolio holdings, *Journal of Financial Economics* 86, 724–758.

Jovanovic, Boyan, and Albert J. Menkveld, 2012, Middlemen in limit-order markets, Working paper.

Ju, Nengjiu, and Jianjun Miao, 2012, Ambiguity, learning, and asset returns, *Econometrica* 80, 559–591.

Kahneman, Daniel, and Amos Tversky, 1979, Prospect theory: An analysis of decision under risk, *Econometrica* XLVII, 263–291.

Kirilenko, A., A. S. Kyle, M. Samadi, and T. Tuzun, 2011, The flash crash: The impact of high frequency trading on an electronic market, Working Paper, University of Maryland.

Klibanoff, Peter, Massimo Marinacci, and Sujoy Mukerji, 2005, A smooth model of decision making under ambiguity, *Econometrica* 73, 1849–1892.

Knight, Frank H., 1921, Risk, uncertainty, and profit, Houghton Mifflin, Boston.

Koski, Jennifer, and Jeffrey Pontiff, 1999, How are derivatives used? evidence from mutual fund industry, *Journal of Finance* 54, 791–816.

Kosowski, R., A. Timmermann, H. White, and R. Wermers, 2006, Can mutual fund stars really pick stocks? new evidence from a bootstrap analysis, *Journal of Finance* 51, 2551–2595.

Lazear, Edward P., and Sherwin Rosen, 1981, Rank-order tournaments as optimum labor contracts, *Journal of Political Economy* 89, 841–864.

Leippold, Markus, Fabio Trojani, and Paolo Vanini, 2008, Learning and asset prices under ambiguous information, *Review of Financial Studies* 21, 2565–2597.

Lynch, A., and D. Musto, 2003, How investors interpret past fund returns, *Journal of Finance* 58, 2033–2058.

Malinova, K., A. Park, and R. Riordan, 2013, Do retail traders suffer from high frequency traders?, Working paper, University of Toronto.

McInish, T., and J. Upson, 2011, Strategic liquidity supply in a market with fast and slow traders, Working paper, University of Memphis.

Menkveld, A., 2010, High frequency trading and the new-market makers, Working paper, Free University of Amsterdam.

Meyer, Margaret A., and John Vickers, 1997, Rank-order tournaments as optimum labor contracts, *Journal of Political Economy* 105, 547–581.

Nalebuff, Barry J., and Joseph E. Stiglitz, 1983, Prizes and incentives: Towards a general theory of compensation and competition, *Bell Journal of Economics* 14, 21–43.

Nanda, Vikram, Zhi Wang, and Lu Zheng, 2004b, Family values and the star phenomenon, *Review of Financial Studies* 17, 667–698.

Odean, Terrance, 1999, Do investors trade too much?, *American Economic Review* 89, 1279–1298.

Parkinson, M., 1980, The extreme value method for estimating the variance of the rate of return, *Journal of Business* 53, 61–65.

Petajisto, Antti, 2013, Active share and mutual fund performance, *Financial Analysts Journal* 69, 73–93.

Petersen, Mitchell A., 2009, Estimating standard errors in finance panel data sets: Comparing approaches, *Review of Financial Studies* 22, 435–480.

Puckett, A., and S. Yan, 2011, The interim trading skills of institutional investors, *Journal of Finance* 66, 601–633.

Sapp, Travis, and Ashish Tiwari, 2004, Does stock return momentum explain the smart money effect?, *Journal of Finance* 59, 2605–2622.

Sharpe, William F., 1997, Morningstar's performance measures, Stanford University.

Sirri, Erik R., and Peter Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Taylor, Jonathan, 2003, Risk-taking behavior in mutual fund tournaments, *Journal of Economic Behavior & Organization* 50, 373–383.

Wald, A., and J. Wolfowitz, 1940, On a test whether two samples are from the same population, *Annals of Mathematical Statist* 11, 147–162.