

Essays on Market Frictions

A thesis submitted

by

Jiali Gao

to

the Discipline of Finance

The University of Sydney Business School

in fulfilment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Finance

The University of Sydney

November, 2025

© 2025 Jiali Gao
All rights reserved.

Declaration of Authorship

I declare that this thesis titled, “Essays on Market Frictions” and the work presented in it are my own. I confirm that:

- This thesis has not been submitted for any degree or other purposes.
- The intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Jiali Gao

November 2025

Authorship Attribution Statement

This is to certify that I have not included any works (published or unpublished) that I have only made minor contribution to in the main body of this thesis.

I design the study, analyze the data, and write the original draft of the manuscript. The co-authors review and edit the writing.

Jiali Gao

November 2025

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Juan Yao

November 2025

Artificial Intelligence Statement

Following the University of Sydney's requirements for thesis submission, I confirm that:

- All conceptual work, analysis, and interpretation are my own
- The thesis structure, arguments, and conclusions are original
- All AI-generated content was critically evaluated and verified
- The use of AI tools was limited to assistance and did not replace my intellectual contribution

Jiali Gao

November 2025

Acknowledgements

I would like to express my sincere gratitude to my advisors, **Juan Yao** and **Stephen Satchell**, for their invaluable guidance, insightful discussions, and constant support throughout the course of my PhD. Their mentorship has not only shaped the direction of this thesis, but also deepened my understanding of the world and of life. I am particularly thankful for the intellectual freedom they afforded me, which allowed me to pursue the topics I was genuinely passionate about, while still keeping the research aligned with a coherent and rigorous academic direction.

I would like to thank Peter Pham and Jing Yu for their service to the PhD program in Finance at the University of Sydney, and for their constant encouragement and support throughout my journey. I am also grateful to Bei Chen, Wei Cui, Quan Gan, He Huang, Guanglian Hu, Oh Kang Kwon, Andre Lot, Yingdong Mao, Yancheng Qiu, Joakim Westerholm, and Huaixin Wang for their advice and help along the way (names listed in alphabetical order by surname).

Jiali Gao

The University of Sydney

November, 2025

Essays on Market Frictions

Jiali Gao

Abstract

This thesis investigates market frictions arising from technologies and tools designed to enhance efficiency and convenience in financial markets. In the first chapter, entitled “Visual Deception in Financial Markets”, I examine how the widely used auto-scaled digital charts distort investors’ risk evaluations, thereby affecting risk premium in the cross-section of stock returns. In the second chapter, entitled “Rating Driven Risk Shifting by Mutual Funds”, we demonstrate how the discrete nature of star ratings leads to deliberate risk shifting. Due to the widespread use of star ratings, rating-driven risk shifting accounts for 21% of the market-wide risk shifting, which has previously been shown to hurt performance. In the third chapter, entitled “Luck and Skill in the World of Diseconomies of Scale”, we find that most of the variation in mutual funds’ realized alphas is driven by luck. We show that lucky funds tend to underperform, raising concerns about performance-based, backward-looking metrics such as star ratings. The results indicate that forward-looking qualitative metrics provide better insights and are neutral to luck.

Table of Contents

1	Introduction	11
2	Chapter Two: Visual Deception in Financial Markets	15
2.1	Introduction	15
2.2	Data and Variables	22
2.2.1	Visual Scaling (VS) and Visual Scaling Exposure (VSE)	22
2.2.2	Data Statistics	25
2.3	Evidence from Experiments	29
2.4	Deceptive Visualizations and Asset Prices	34
2.4.1	Testable Propositions	34
2.4.2	Deceptive Visualizations in the post 1980s	36
2.4.3	Retail and Institutional Investors	38
2.4.4	Absence of Deceptive Visualizations in the Early Era (1920s - 1980s)	40
2.4.5	Portfolio Performances and Time-series Analysis	42
2.5	Further Analysis	48
2.5.1	Evidence from Large Brokerage Account	48
2.5.2	Visual Scaling Exposure and Beta	53
2.5.3	Visual Scaling Exposure and Lottery Preference	54
2.6	Conclusion	59
3	Chapter Two: Rating Driven Risk Shifting by Mutual Funds, co-author with Stephen Satchell and Juan Yao	61
3.1	Introduction	61
3.2	Star Rating and Methodology Reform	66
3.3	Data and Key Variables	69
3.3.1	Data	69

3.3.2	Key Variables	69
3.4	Rating-driven Risk-shifting under New Methodology	74
3.4.1	Rating-driven Risk-shifting and Methodology Reform	74
3.4.2	Rating-driven and Market-wide Risk-shifting	77
3.4.3	Cross-sectional Variation and 3Y, 5Y, 10Y Rating	79
3.4.4	Market Volatility and Unemployment Risk	84
3.4.5	Systematic Risk and Idiosyncratic Risk	85
3.4.6	Analysis of Fund Holdings	87
3.4.7	Analysis of Trading Activities	89
3.5	Rating-driven Style-shifting under Old Methodology	90
3.6	Further Analysis	93
3.6.1	Rating Change and Reference-Dependent Risk Preference	93
3.6.2	Agency Issue and Incentive	96
3.7	Conclusion	98

4	Chapter Three: Luck and Skill in the World of Diseconomies of Scale, co-author with Juan Yao	100
4.1	Introduction	100
4.2	Data	106
4.3	Realized Alpha and Luck	109
4.3.1	Linear Model	109
4.3.2	Nonparametric Model	112
4.4	Theoretical Framework	115
4.5	Chasing Performance and Future Expected Alpha	117
4.5.1	Luck, Flow and Size	117
4.5.2	Luck and Future Expected Alpha	122
4.6	Capital Misallocation and Skill	128

4.6.1	Luck and Capital Misallocation	128
4.6.2	Value Added as Skill	130
4.7	Luck and Ratings	135
4.7.1	Rating Statistics	135
4.7.2	Star Rating, Funding Status and Luck	138
4.7.3	Quantitative versus Qualitative	141
4.8	Conclusion	147

“The curious task of economics is to demonstrate to men how little they really know about what they imagine they can design.” - Friedrich Hayek

1 Introduction

We live within a complex system shaped by individuals who are often naïve, short-sighted, and driven by greed (Marx (2018)). History has repeatedly shown, in various forms, that underestimating the dynamics of such complexity often causes well-intentioned actions to result in unintended consequences.¹ This thesis focuses on technology, which originally developed to enhance human life, but often brings about unintended impacts to society (Harari (2014)). Technologies tend to diffuse gradually through society due to their convenience and efficiency. As society becomes increasingly dependent on a particular technology, it becomes difficult to recognize its long-term consequences, since the technology shapes the very way we think.² Because of this, the unintended impacts of technology are often irreversible and structural in nature.

As Albert Einstein once said, “We cannot solve our problems with the same thinking we used when we created them.” This thesis draws on Einstein’s insight to examine the frictions introduced by widely accepted “technologies” (or “innovations”) in financial markets. It asks whether these technologies—originally designed to enhance efficiency—have, in fact, generated unintended and far-reaching consequences that we are only beginning to recognize. In particular, this thesis focuses on digital charting in stock market and rating system in mutual fund industry.

In the first chapter, entitled “**Visual Deception in Financial Markets**”, I examine the unintended consequences of a legacy financial technology: auto-scaled digital charts.

¹The U.S. Prohibition and China’s Great Leap Forward, for instance, were launched with good intentions but ended in disaster. Nevertheless, policy failures—while sometimes catastrophic—are generally reversible.

²As argued by *Sapiens: A Brief History of Humankind* (Harari (2014)), the Agricultural Revolution was history’s biggest fraud—it seemed to promise abundance, but in reality, it trapped humans in a lifestyle marked by overwork, poor nutrition, social inequality, and increased suffering. Nevertheless, returning to foraging was no longer feasible, as humans had become dependent on farming and were already caught in the so-called “luxury trap”.

With the rise of graphical user interfaces since the 1980s, digital charts have become widely used in financial markets. These charts typically employ a default auto-scaling function that adjusts the price path to fit the screen. Although intended to enhance readability, this auto-scaling feature can distort the visual representation of risk. Despite substantial evidence from behavioral science showing that deceptive visualizations influence human behavior (Lauer and O'Brien (2020); Pandey et al. (2015)), their implications have received little attention in finance and economics. I provide experimental showing that auto-scaled charts mislead investors' risk perception. In particular, because auto-scaling tends to compress (low visual scales) the price paths of high-volatility stocks and stretch (high visual scales) those of low-volatility stocks, investors' perception of risk is less sensitive to actual volatility in auto-scaled charts.

A fundamental principle in finance is that risk should be compensated with higher expected returns, yet this relationship is empirically difficult to observe in the cross-section of stocks in recent decades (Baker et al. (2011)). I find that stocks' visual scaling exposure can explain the diminishing trend in the positive risk-return relationship (Ang et al. (2009)), which was evident before the 1980s but became less apparent afterward. This chapter highlights that fintech innovations, although designed to support investors in specific dimensions, may unintentionally alter investor behavior along other dimensions.

In the second chapter, entitled "**Rating Driven Risk Shifting by Mutual Funds**", we demonstrate the unintended consequences of the Morningstar funds rating system, which was originally designed to improve investors' evaluation of fund performance. Today, the Morningstar rating is widely used by investors and has even become a key determinant of their investment decisions (Ben-David et al. (2022a)). Given the non-linear payoff structure induced by the discrete (1 to 5 star) rating system, we hypothesize that managers strategically alter portfolio risk based on their relative position to a potential rating upgrade or downgrade. We find that managers tend to increase risk when approaching a rating upgrade and reduce risk when facing a potential downgrade. Prior to the 2002 rating methodol-

ogy reform —when ratings did not account for investment styles —managers strategically adjusted their styles to influence their ratings.

Huang et al. (2011) show that time-varying risk shifting, either increase risk or decrease risk, is at cost of hurting performance. We further document that rating-driven risk shifting accounts for 21% of the time-varying risk shifting observed in mutual fund industry. Although the generally convex shape of the flow-performance relationship may also incentivize managers to take on more risk (Chevalier and Ellison (1997)) - potentially contributing significantly to market-wide risk-shifting, Spiegel and Zhang (2013) challenge this view by documenting that the relationship is largely linear. This chapter highlights that although rating systems are intended to evaluate performance, those being evaluated often have incentives to manipulate the system. Therefore, the unintended consequences of introducing such systems should be carefully considered.

In the third chapter, entitled “**Luck and Skill in the World of Diseconomies of Scale**”, we extract “luck” from mutual funds’ past performance and demonstrate the downside of using backward-looking performance metrics to guide future investment decisions. We find that most of the variation in mutual funds’ realized alphas is driven by luck. We revisit Berk and Green (2004) by introducing luck into a world with performance-chasing investors and non-optimizing managers operating under diseconomies of scale. By extracting “luck” from realized alpha, we explain why skilled managers (Berk and Van Binsbergen (2015), Zhu (2018)) often lack alpha persistence (Jensen (1968), Fama and French (2010)) and why naive performance-chasing investors (Sirri and Tufano (1998), Ben-David et al. (2022b)) exhibit a degree of sophistication (Berk and Van Binsbergen (2015), Zhu (2018)), as documented in the literature.

We demonstrate that fund luck negatively predicts future performance, due to naive performance-chasing behavior and the decreasing returns to scale in mutual funds. Furthermore, we find that backward-looking quantitative performance metrics, such as Morningstar rating, primarily capture luck, leading to poor long-term predictability. In contrast, we show

that qualitative metrics, such as Morningstar analyst ratings, are forward-looking therefore neutral to funds luck, leading to better performance predictability. Given the widespread use of backward-looking performance evaluation systems in the portfolio management industry and investors' reliance on them, this chapter highlights the important role of luck when using such system as a signal of managerial skill.

2 Chapter Two: Visual Deception in Financial Markets

“We shape our tools, and thereafter our tools shape us.” - Marshall McLuhan

2.1 Introduction

Computer technology has revolutionarily altered how investors gather financial market information. In the past, stock quotes were reported in newspapers, and technical analysis required manually drawing prices on paper with a pencil. Today, investors can easily access real-time data on online platforms and their brokerage accounts. Digital price charts are arguably the most crucial feature offered by these platforms and have become an indispensable tool for investors. For instance, [Glaser et al. \(2019\)](#) find that over 90% of financial information providers offer prices charts. Furthermore, [Laarits and Wurgler \(2025\)](#) use browser data to show that 93% of individual investors view charts, spending the most of their pre-trade research time on chart viewing (76 out of 144 minutes) compared to other information such as analyst opinions or traditional risk statistics.

The digital price charts utilize candlesticks³ of varying lengths to represent returns, collectively forming a stock’s price path. To better visualize changes in a stock’s price path, the selected range of digital price charts, and most charting tools, is typically determined by the historical high and low prices. The stock’s price path is then scaled to fit within this range, a feature referred to as auto-scaling. However, while intended to enhance readability, the visual scaling differs across stocks due to the unique range of each price path. Humans rely heavily on visuals when processing information ([Stenberg et al. \(1995\)](#), [Paivio et al. \(1968\)](#)), a series of behavioral science experiments have shown that visual scaling can be deceptive and mislead participants.⁴ As noted by [Lauer and O’Brien \(2020, p. 337\)](#), “*The*

³This paper uses candlestick charts to illustrate the effect of visual scaling, but such effect also exist in different charts such as line charts, bar charts, or pie charts ([Lauer and O’Brien \(2020\)](#)).

⁴The discussions of deceptive visualizations are not new. In 1954, Darrell Huff, an active freelance writer, published a slim volume titled *How to Lie with Statistics*, in which he illustrated how deceptive visuals can be used to mislead readers. More recently, a growing body of experimental research has documented both the prevalence of deceptive visualizations and proposed potential solutions (see, for example, [Lauer and O’Brien \(2020\)](#), [Pandey et al. \(2015\)](#), [Yang et al. \(2021\)](#), [O’Brien and Lauer \(2018\)](#), [Camba et al. \(2022\)](#), [Lauer and](#)

programs’ defaults adapt the visual range of a graph’s axes to the data source. This fact becomes especially problematic when a user wants to compare several graphs... As a result, the data shown across the several graphs are visually misleading when they are compared with each other.” Despite the widespread use of digital charts in financial markets, the effect of the digital charts on investors’ perceptions and behavior is relatively understudied in the literature. This paper investigates how the auto-scaling, a key function of digital charting, distorts investors’ perception of risk, and how such misperception influences asset prices.

Figure 1 (a), illustrates how the trading platform TradingView visually presents the 3-month daily candlestick chart for Tesla. Auto-scaling adjusts the price-axis range to the high and low prices within a 3-month window. Users can manually adjust the time-axis, as shown in Figure 1 (b), but the price-axis range is determined by the highest and lowest prices over the selected time window. As the window is expanded to 6 months shown in Figure 1 (b), the price path is scaled down due to the inclusion of a new high price at the end of December 2023. I define this scaling effect experienced by chart users as Visual Scaling (VS), which is determined by the average price, along with the historical high and low. In the cross-section of stocks displayed in auto-scaled charts, a stock with a wider price range relative to its average price, despite exhibiting higher volatility, will appear visually compressed, resulting in a lower VS value. In contrast, a stock with a narrower price range relative to its average price, despite having lower volatility, will be visually expanded, leading to a higher VS value.

Sanchez (2023), Fan et al. (2022)). These studies are discussed in greater detail later in the introduction.

Figure 1: Tesla in TradingView



(a) 3-month chart



(b) 6-month chart

I first use experiment to examine whether auto-scaled charts can distort participants' risk perception. I present randomly generated price paths in auto-scaled charts to one group of participants (Auto), and the same price paths in fixed-scaled charts to the other group (Fixed). I find participants' risk perceptions to price paths is consistent with actual volatility measured by standard deviation in the Fixed group, yet this pattern largely disappears in the Auto group. Regression analysis shows that VS significantly explains risk evaluation in the Auto group but not in the Fixed group. First, the experiment results suggest that auto-scaled charts distort participants' risk perception, as their perceptions rely on depicted volatility rather than actual volatility. Second, the experiment validates VS as a measure of visual scaling in auto-scaled charts, capturing the scaling effect investors are exposed to rather than merely reflecting the historical price range.

Next, I investigate how auto-scaled charts affect asset prices. I estimate the Visual Scaling Exposure (VSE) of each stock by averaging its VS values over multiple window lengths (e.g., 3-month, 6-month and 12-month periods). The selection of these windows is supported [Laarits and Wurgler \(2025\)](#), which document that the most frequently utilized time windows by individual investors are 3-month and 1-year. Stocks with high (low) VSE are more likely to be assigned with high(low) VS values. I find that VSE significantly and positively impacts volatility premium, suggesting that depicted volatility carries positive risk premium. Once this effect is accounted for, positive volatility premium, which is difficult to empirically identify,⁵ becomes evident. Additionally, I find that depicted volatility premium is weaker for stocks with a higher proportion of institutional investors, as they are more sophisticated and less influenced by visual scaling. Moreover, in the pre-1980s period — prior to the introduction of digital charting technologies —VSE had no predictive power for the volatility premium. This is because, during that period, VSE, constructed based

⁵[Wang et al. \(2017\)](#) finds a positive relationship between realized volatility and expected return after accounting for reference-dependent preferences, which becomes insignificant once risk-adjusted returns are used. [Liu \(2022\)](#) reports a positive relationship (17 basis points per month with a t-statistic of 2.40) between filtered short-run volatility and expected return, based on 55 years of data.

on price, historical highs, and lows, merely captured characteristics of the price path rather than influencing investors through visual scaling.

Since high-volatility stocks tend to have wider price paths, resulting in lower VSE values, in the digital charting era, high-volatility stocks do not appear as highly volatile. This visual distortion predicts a weaker volatility-return relationship in the era of digital charting. I conduct a time series portfolio analysis to clearly document how the rise of deceptive visualization affects the diminishing of the volatility premium. I show that a long-short portfolio based on volatility earned a positive premium before the 1980s but experienced a significant drop thereafter. Consistent with this finding, early literature documents a positive volatility-return relationship (Lintner (1965), Lehmann (1990), Tinic and West (1986)), whereas more recent studies do not find such a relationship (Ang et al. (2006), Bali et al. (2017)). Conversely, a volatility-neutral portfolio constructed based on VSE showed no premium before the 1980s but began to earn a premium post-1980s. The key implication is that it is essential to control for VSE when attempting to document the volatility premium, as investors rely on the visual appearance of the price path to assess a stock's volatility. As a result, I find that the long-short portfolio on volatility constructed within VSE quintiles continues to earn a positive premium across the full sample period.

In further analysis, I investigate whether investors' trading behavior is influenced by auto-scaled digital charts. I utilize a large discount brokerage (LDB) dataset from Barber and Odean (2001), which includes trading data from 78,000 households between 1991 and 1996.⁶ I find that among phone-based trades, high-risk trades exhibit a monotonic increase in participation by males, younger individuals, and singles, relative to low-risk trades — consistent with expected risk-taking norms. However, this demographic pattern weakens substantially, or even disappears, in computer-based trades. I acknowledge that PC-based

⁶Empirically studying the effect of deceptive visualizations in real trading data is challenging, as it is nearly impossible to establish an individual investor's normative behavior. Motivated by Barber and Odean (2001), I use demographic-based risk preferences as a proxy for normative behavior.

traders and phone-based traders differ endogenously in many respects, which could potentially contribute to the violation of risk-taking norms. Supported by the finding in [Laarits and Wurgler \(2025\)](#) that price charts are the most utilized function by individual investors prior to executing a trade, it is reasonable to attribute this observed violation of risk-taking norms to digital price charts and their misleading visual scaling. Through regression analysis, I find that a stock's VSE significantly explains demographic preferences for risk in computer-based trades, whereas VSE has no effect on phone-based trades. The results suggest that, compared to phone-based traders, computer-based traders are more likely to be exposed to digital charting and thus more susceptible to the influence of auto-scaling.⁷

Finally, I test whether the visual scaling of digital charts affects the pricing of systematic risk and skewness preference, and find no significant effect. This outcome is consistent with the premise that beta (a covariance measure) and skewness (the third moment of a return distribution) are not visually discernible from the price paths displayed on the charts. Therefore, the visual scaling of the charts should not be expected to influence participants' perception of these characteristics.

This paper makes several contributions to the literature. First, this paper contributes to the growing body of experimental studies on visual elements in finance. Prior research has shown that the format in which investment information is presented ([Weber et al. \(2005\)](#)), the use of color in charts ([Bazley et al. \(2021\)](#)), and the visual salience of past returns ([Bose et al. \(2022\)](#)) can influence individual investment decisions. However, these studies do not have implications for asset prices. My paper is the first to demonstrate how visuals influence asset prices.

Second, it contributes to the existing literature on how visual presentation influences human behavior in psychology and behavioral science. Experimental studies ([Lauer and](#)

⁷This website shows the interfaces of the trading platform in 1990s, where users can easily access charts. <https://web.archive.org/web/19970405061629/http://www.schwab.com/SchwabNOW/SNLibrary/SNLib001/SN001.html>

O’Brien (2020), Pandey et al. (2015)) have shown that participants’ evaluations of numerical data can be misled by modifications in visual scaling. Other studies, including Yang et al. (2021) and O’Brien and Lauer (2018), provide further evidence that such visual effect persists across diverse demographic groups in real-world contexts.⁸ Recent research (Camba et al. (2022), Lauer and Sanchez (2023), Fan et al. (2022)) has explored ways to help individuals overcome the influence of deceptive visualizations.⁹ My paper is the first to provide empirical evidence of the deceptive impact of visuals on investor behavior and financial markets.

Additionally, this paper contributes to the asset pricing literature. Empirical evidence in recent decades finds a weak or negative relationship between risk and return (Ang et al. (2006), Hou and Loh (2016), Bali et al. (2017)), contradicting the fundamental finance principle that risk is compensated with higher expected returns (Baker et al. (2011)). This paper provides an explanation for this puzzle. After accounting for visual scaling exposure of stocks, I document a strong firm-level positive relationship between realized volatility and expected return.

Given individual investors’ strong reliance on charts (Laarits and Wurgler (2025)), the findings of this study also contain practical implications for trading and charting platforms, and their users. I emphasize the need for financial software providers to clearly communicate the use of auto-scaling to users, which can help users interpret chart-based information more accurately. More broadly, this study revisits a legacy technology from the 1980s —digital charts —to highlight the unintended consequences that can arise from financial technologies. As fintech continues to reshape the investor experience, these insights underscore the importance of anticipating and mitigating potential downsides of emerging technologies that

⁸Yang et al. (2021) find that approximately 85% of people are misled by deceptive visualizations, and this effect persists even among individuals with extensive experience working with data and statistics (e.g., PhD students in quantitative fields). O’Brien and Lauer (2018) provide evidence of its existence even when non-deceptive text is included.

⁹Camba et al. (2022) and Lauer and Sanchez (2023), using experiments and eye-tracking methods, suggest that only explicit and active educational interventions can help overcome deceptive visualizations. Fan et al. (2022) develop a tool to identify and annotate deceptive visualizations, demonstrating its effectiveness in mitigating such effects.

are originally intended to assist investors.

The rest of the paper proceeds as follows: Section 2 describes the data and VSE construction. Section 3 presents the evidence that auto-scaled charts are deceptive using experiments. Section 4 investigates how deceptive visualizations affect asset prices. Section 5 provides further analysis. Section 6 concludes.

2.2 Data and Variables

2.2.1 Visual Scaling (VS) and Visual Scaling Exposure (VSE)

Consider a stock with price P and a return r on day t , the candle for this stock on day t represents a dollar return of $r \times P$. This candle is presented in an auto-scaled chart with the upper bound at the highest price (H) and the lower bound at the lowest price (L), so the candle vertically occupies $(r \times P)/(H - L)$ of the chart window. Since the actual return is r , I model the scaling effect as the $P/(H - L)$, a ratio referred to in this paper as the visual scaling (VS). Thus, the depicted return r' on day t is expressed as:

$$r'_t = r_t \times \frac{P}{H - L} = r_t \times VS. \quad (1)$$

Under the assumption that the standard deviation of returns captures the actual volatility, the key argument for the actual volatility σ and depicted volatility σ'_t on day t in auto-scaled charts are formulated as:

$$\left\{ \begin{array}{l} \text{Actual volatility: } \sigma_t = \sqrt{\frac{1}{N} \sum_{i=t-N+1}^t (r_i - \bar{r})^2} \\ \text{Depicted volatility: } \sigma'_t = \sqrt{\frac{1}{N} \sum_{i=t-N+1}^t (VS \cdot r_i - VS \cdot \bar{r})^2} = \sigma_t \times VS. \end{array} \right. \quad (2)$$

For the stock i presented in an auto-scaled chart at time t , with a charting window of w days, the VS applied to the stock is estimated as:

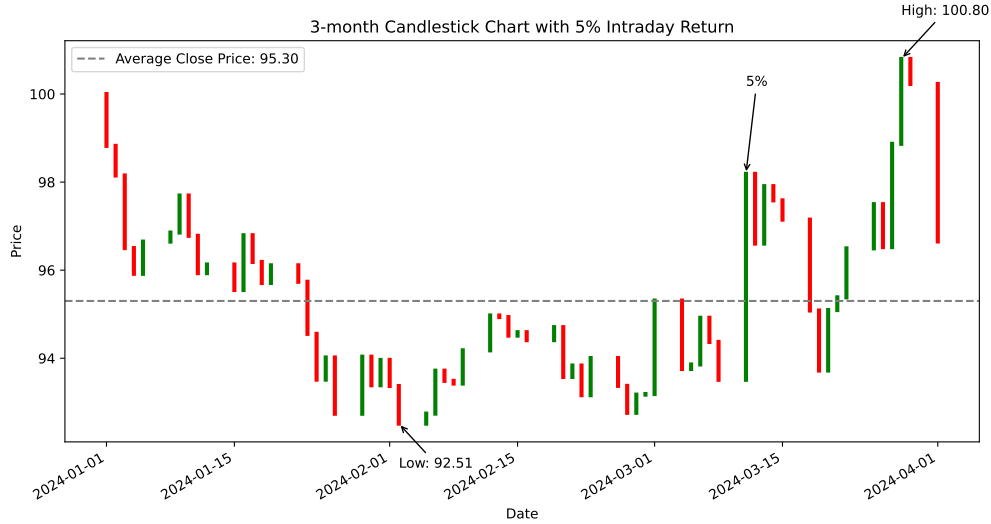
$$VS_{i,t,w} = \frac{\overline{Price}_{i,[t-w,t]}}{H_{i,[t-w,t]} - L_{i,[t-w,t]}}, \quad (3)$$

where $\overline{Price}_{i,[t,t-w]}$, $H_{i,[t-w,t]}$ and $L_{i,[t-w,t]}$ denote the average price,¹⁰ the highest price and the lowest price of stock i over the window from day $t - w$ to day t , respectively. Figure 2 (a) illustrates a candlestick chart over a 3-month window, which has a high VS of 11.48. The stock has a narrow price range, with a past high price of 100.80 and a past low price of 92.51, relative to its average price of 95.30. The auto-scaling function sets the past high and low as the upper and lower bounds, respectively, and “scales up” each candlestick to fit the price path into the fixed-size chart. In contrast, Figure 2 (b) presents a candlestick chart over a 3-month window with a low VS of 4.05. This stock has a wider price range, with a past high price of 111.30 and a past low price of 86.50, compared to an average price of 100.48. The auto-scaling function sets the past high and low as the upper and lower bounds, and “scales down” each candlestick to fit the price path into the fixed-size chart. Both stocks in Figure 2 (a) and Figure 2 (b) exhibit a return of 5% over the selected window. However, due to auto-scaling, the visual representation of this 5% return appears significantly larger in Figure 2 (a) than in Figure 2 (b).

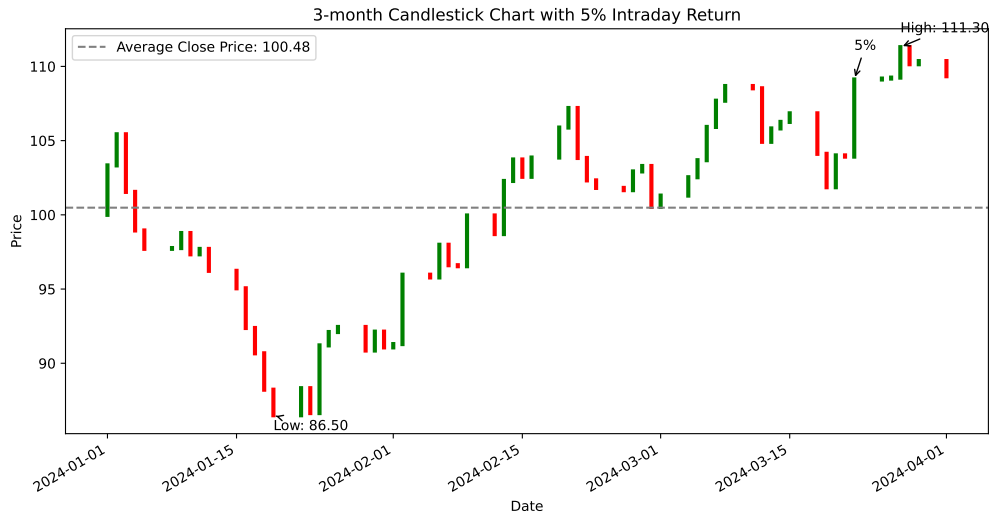
¹⁰I use the average price over the window, rather than the latest price at time t , to construct the VS for two reasons. First, the latest price does not affect the chart’s visual scaling unless it coincides with the historical high or low. Second, using the latest price would introduce some degree of momentum exposure into the VS measure. That said, all empirical results remain highly consistent when the VS is alternatively computed using the latest price.

Figure 2: Candlesticks and Visual Scaling (VS)

Figure (a) and (b) depict two simulated 3-month candlestick charts with auto-scaling in the price-axis. For simplicity and ease of illustration, these figures assume 0 overnight returns. VS is calculated as average price divided by the difference between high price and low price. Figure (a) is an example of candlestick chart with high visual scaling ($VS = 11.48 \approx 95.30/(100.80-92.51)$), and Figure (b) is an example of candlestick chart with low visual scaling ($VS = 4.05 \approx 100.48/(111.30-86.50)$).



(a) High Visual Scaling ($VS_{3m} = 11.48$)



(b) Low Visual Scaling ($VS_{3m} = 4.05$)

In real world, investors have varying preferences for time window lengths, and softwares

have different default settings —both of which are unobservable at the individual level. For simplicity, I assume that investors of stock i have homogeneous visual scaling exposure. I estimate the visual scaling exposure (VSE) for investors of stock i in time t as the mean value of the VS over different time windows, and it is formulated as:

$$VSE_{i,t} = \frac{1}{N} \sum_{w \in W} VS_{i,t,w}, \quad (4)$$

where W represents the set of selected time windows (e.g., 3 months, 6 months and 12 months), and N is the number of windows used in the calculation.

2.2.2 Data Statistics

The demographic summary statistics presented in Table 1 compare the Fixed-Scaled (Control) group and the Auto-Scaled (Treatment) group across seven variables. Overall, the mean values demonstrate that the demographic characteristics are highly similar between the two groups. In both groups, Male participants constitute approximately half, with a mean of 0.53 in the Fixed-Scaled group and 0.51 in the Auto-Scaled group. The average Age group is between 5.36 (Fixed-Scaled) and 5.53 (Auto-Scaled), which generally corresponds to the 36 – 45 age range. The mean Education level is around 3.88 to 3.87, falling between "Some college" and "Undergraduate degree." Income group averages are also very close, at 3.26 and 3.24, corresponding to the \$40,001 to \$80,000 range. Finally, the remaining characteristics—Self-stated risk aversion, Stock market participation, White, and US —also exhibit similarly close means, further supporting the comparability of the two groups. The statistics reported is generally consistent with [Bose et al. \(2022\)](#) and [Bazley et al. \(2021\)](#).

Table 1: Demographic Summary Statistics

This table reports the summary statistics of key demographic variables for the control group (Fixed-Scaled) and the treatment group (Auto-Scaled). The variables are defined and coded as follows: Male is an indicator variable equal to 1 for male participants and 0 for female participants. Age is a categorical variable with 11 levels: 1. 18–20; 2. 21–25; 3. 26–30; 4. 31–35; 5. 36–40; 6. 41–45; 7. 46–50; 8. 51–55; 9. 56–60; 10. 61–65; and 11. Above 65 years old. Education is a categorical variable with seven levels: 1. Some high school; 2. High school graduate; 3. Some college; 4. Undergraduate degree; 5. Professional degree; 6. Master’s degree; and 7. Doctoral degree. Income Group is a categorical variable representing annual income ranges: 1 if \$0–25,000; 2 if \$25,001–40,000; 3 if \$40,001–60,000; 4 if \$60,001–80,000; 5 if \$80,001–100,000; 6 if \$100,000–150,000; and 7 if above \$150,000. Self-stated Risk Aversion is a categorical variable ranging from 1 to 7. Stock Market Participation is an indicator variable that is 1 if the experiment participant reports investing in the stock market, and 0 otherwise. White is an indicator variable equal to 1 if the participant identifies as white, and 0 otherwise. US is an indicator variable that is 1 if the experiment participant resides in the United States, and 0 otherwise. The p -values of the means in two-tail test is reported.

	Fixed-Scaled			Auto-Scaled			p -value
	Mean	SD	N	Mean	SD	N	
Male	0.53	0.50	48	0.51	0.50	59	0.83
Age group	5.36	2.86	48	5.53	2.73	59	0.76
Education	3.88	1.26	48	3.87	1.33	59	0.97
Income group	3.26	1.68	48	3.24	1.74	59	0.95
Self-stated risk aversion	3.72	1.65	48	3.88	1.70	59	0.62
Stock market participation	0.40	0.49	48	0.38	0.49	59	0.83
White	0.72	0.45	48	0.71	0.45	59	0.91
US	0.89	0.31	48	0.90	0.30	59	0.87

The stock return data are from the Center for Research in Security Price (CRSP), while accounting data are from Compustat. Institutional ownership data are obtained from the Thomson Financial 13F file. Two risk proxies, return volatility (RVol) and idiosyncratic volatility (IVol), are estimated using monthly stock return from past 12 months. The control variables are size, momentum, price, turnover, distance to 52-week high, distance to 52-week low. Variable definitions are provided in Appendix 1A. Each month, I construct VSE,

risk proxies and controls. I use the average of 3 months, 6 months and 12 months VS to estimate VSE according to Equation (4). The selection of these windows is supported Laarits and Wurgler (2025), which document that the most frequently utilized time windows by individual investors are 3-month and 1-year. The descriptive statistics are presented in Table 2.

Table 2: Summary Statistics for Stock Month Data

Panel A reports the distribution of visual scaling value of 3-month chart (VS_{3m}), 6-month chart (VS_{6m}), 12-month chart (VS_{12m}), and the VSE which equals the mean of VS_{3m} , VS_{6m} and VS_{12m} . Panel B reports the average size (Size(\$M)), institutional ownership (IO), log of (1+number of analysts) (AC), and market to book (M/B), distance to 52-week High ($High_{52w}$), distance to 52-week Low (Low_{52w}), return volatility (RVol), idiosyncratic volatility (IVol), momentum (Mom), Price, maximum return (MAX), and turnover (TO) by the VSE quintiles. Standard errors of coefficients are adjusted for heteroskedasticity and clustering by firm and quarter. *, **, *** indicate significance at the 10%, 5%, and 1% level, respectively.

Panel A: Summary Statistics of VS and VSE							
	Mean	Std Dev	Percentiles				
			Min	25th	Median	75th	Max
VSE	3.75	2.95	0.43	1.86	2.94	4.65	132.17
VS_{3m}	5.44	4.64	0.43	2.54	4.16	6.68	184.32
VS_{6m}	3.52	2.81	0.52	1.69	2.73	4.39	15.69
VS_{12m}	2.23	1.67	0.35	1.11	1.76	2.79	9.48

Panel B: Firm Characteristics by VSE						
VSE Rank	Size(\$M)	IO	AC	M/B	$High_{52w}$	Low_{52w}
1	843.00	0.40	1.59	3.50	-0.47	0.23
2	1755.85	0.50	1.73	3.15	-0.21	0.16
3	2778.74	0.52	1.80	2.93	-0.13	0.12
4	3644.23	0.50	1.84	2.77	-0.09	0.09
5	2798.52	0.46	1.82	2.53	-0.05	0.05
5-1	1955.52*** (36.40)	0.06*** (8.80)	0.24*** (24.90)	-0.97*** (-19.18)	0.42*** (32.58)	0.18*** (56.44)

VSE Rank	RVol	IVol	Mom	Price	MAX	TO
1	0.71	0.18	0.22	10.32	0.12	1.95
2	0.50	0.12	0.14	18.41	0.07	1.75
3	0.39	0.10	0.11	26.03	0.05	1.52
4	0.30	0.07	0.09	33.70	0.04	1.24
5	0.21	0.05	0.06	30.22	0.02	0.82
5-1	-0.50*** (-68.94)	-0.12*** (-59.08)	-0.16*** (-10.47)	19.91*** (50.05)	-0.09*** (-52.76)	-1.12*** (-53.87)

The trading data is used in this study from a large discount brokerage firm (LDB) and were first used in Barber and Odean (2000).¹¹ The data contains over 3 million trades from 78,000 households from February 1991 through January 1997. The trading data document how each trade was initiated (e.g., by phone or by personal computer), which is then merged with the demographic information provided by the brokerage house. Appendix 1B reports the summary statistics of the trading data. This study focuses on buy trades of common stocks only, as buy decisions better reflect investors’ preferences for specific stocks, whereas sell decisions may be driven by various factors such as liquidity needs, capital gains realization, or changes in investment preferences. To ensure a cleaner test of demographic-based risk preferences, I exclude households in the top 1% of trade volume. Although these observations account for approximately 20% of trades, they contribute only 1% to demographic variation, introducing potential noise. I also remove observations with missing demographic information (gender, age, or marital status). After applying these filters, the final sample consists of 179,137 trades from 22,613 households.

Given that the average estimated holding period for computer-based trades in the LDB data is approximately one year (Barber and Odean (2002)), I deliberately choose 6-month and 12-month windows to estimate the VSE. This is because I assume that the longer an investor expects to hold a stock, the longer the window of historical chart they are likely to track.¹² For instance, an investor planning to hold a stock for 12 months may refer to the past six months or the past year of digital charts, making them more influenced by the VS over these periods rather than, for example, the past three months.¹³ The VSE as well as the volatility (return volatility and idiosyncratic volatility) are estimated one day before the

¹¹I thank Terrance Odean for sharing the data. The brokerage dataset has been widely used in both earlier studies (e.g., Barber and Odean (2000), Barber and Odean (2001), Barber and Odean (2002), Barber et al. (2008), Kumar (2009)) and more recent research (e.g., Shue and Townsend (2021), An et al. (2024)).

¹²The same argument is made by Laarits and Wurgler (2025)

¹³Empirically, VS estimates from different window lengths are highly correlated (over 0.9 correlation) due to overlapping periods. As a result, the arbitrary selection of VS windows for VSE estimation does not significantly impact the main results.

buy trade. The mean VS for the 6-month window is 2.63, while the mean for the 12-month window is 1.74. The VSE measure, calculated following Equation (4), has a mean of 2.18.

2.3 Evidence from Experiments

I first present experimental evidence in this subsection to examine the effect of auto-scaling function on risk perception. I randomly generate price paths and present them to two groups of investors: one group is presented with fixed-scaled charts, and the other is presented with auto-scaled charts. The price paths are generated as follows. First, I randomly draw 66 daily returns over 3 months, assuming 22 trading days per month, from a normal distribution with specified volatility levels, and construct price paths starting from an initial price of \$100. Then, I generate 20 candidate paths for each volatility level for the filtering process. Two filtering criteria are applied. First, to prevent final price levels or trends from influencing participants' perception of risk, I exclude any path with an absolute return greater than 30%. This restriction ensures that perceived risk is driven by price fluctuations rather than directional movement. Second, within each volatility level, I select the two paths that exhibit the highest and lowest VS values, in order to maximize the variation of visual scaling. I consider five levels of annualized volatility: 20%, 40%, 60%, 80%, and 100%. Overall, I obtain two price paths (one with high VS and one with low VS) for each of the five volatility levels, resulting in a total of ten price paths. As the treatment, I present the ten price paths using auto-scaled charts to one group (Auto). As the control, I show the same ten price paths using fixed-scale charts with a y-axis range from \$0 to \$200 to the other group (Fixed). Participants of the experiment are then asked to rate the perceived risk of the displayed stock on a scale from 1 (low risk) to 5 (high risk), responding to a question: "How risky do you think the stock is?"

I run the experiment on Amazon's Mechanical Turk (MTurk), which provides a relatively low-cost way to access online experiment participants. The MTurk platform allows "Requesters" to post "Human Intelligence Tasks" (HITs) for "Workers", who are registered MTurk participants (Mason and Suri (2012)). As discussed in Bazley et al. (2021), MTurk

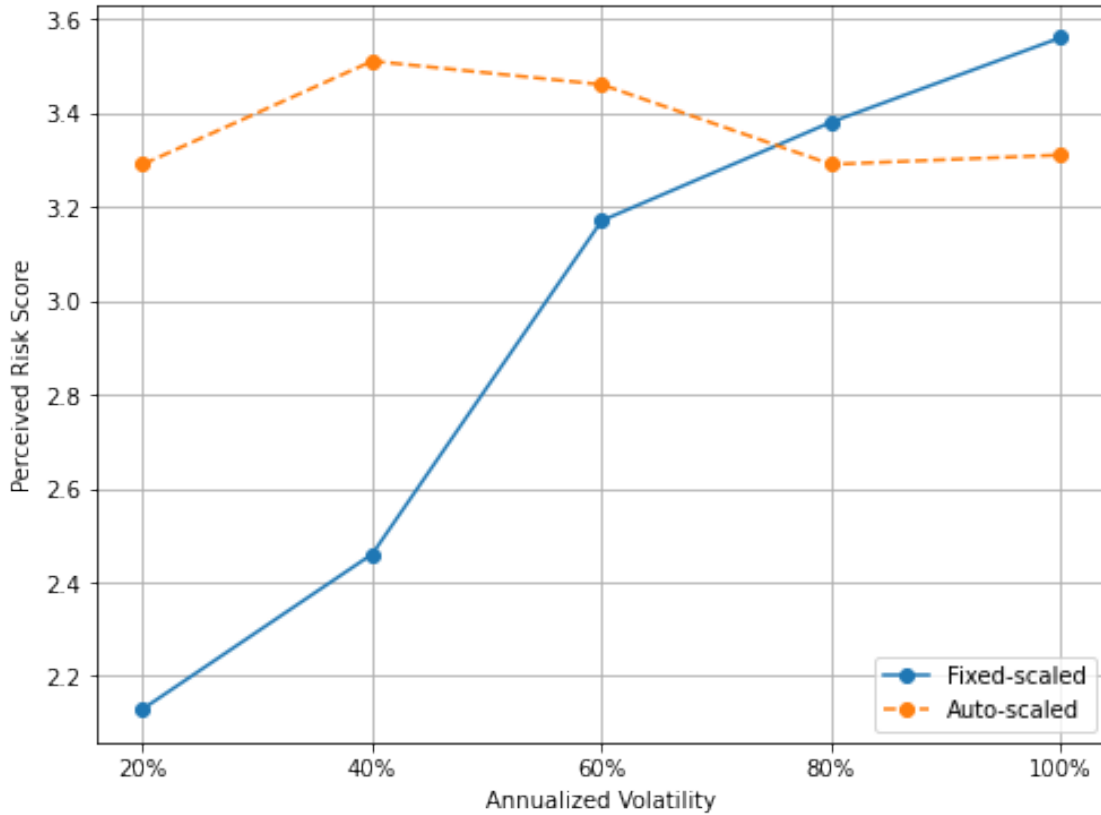
participants exhibit a greater heterogeneity in individual characteristics compared to lab samples. Furthermore, MTurk is a double-blind platform that allows for randomized assignment across both control and treatment groups. For these reasons, MTurk has been widely used in many economics research (e.g., [Montiel Olea and Strzalecki \(2014\)](#), [Kuziemko et al. \(2015\)](#)) and, more recently, in finance research (e.g., [Duarte et al. \(2012\)](#) and [Bazley et al. \(2021\)](#)). As a standard practice in MTurk-based research, I do not disclose the specific nature or objectives of the experiments to participants. Workers complete HITs in exchange for compensation from the Requester.¹⁴

There are 48 participants in the Auto group and 59 in the Fixed group, thus the total observations are 480 from the Fixed group and 590 from the Auto group. Before investigating how visual scaling affects risk evaluation, I first examine if the risk evaluation by participants differs among the Fixed group and the Auto group. For each volatility levels, I calculate the average score for perceived risk from both groups. As shown in [Figure 3](#), participants in the Fixed group appear to accurately assess volatility, as stocks with higher volatility consistently receive higher perceived risk scores (Blue). However, this pattern does not exist in the Auto group (Orange).

¹⁴In the experiments, I provide participants with compensation at a typical market-based rate on the MTurk platform—approximately 10 cents per minute—consistent with [Bazley et al. \(2021\)](#).

Figure 3: Perceived Risk and Actual Risk

This figure plots the average score for perceived risk of price paths across risk levels (20%, 40%, 60%, 80%, 100%) for participants exposed to auto-scaled charts (Orange dashed line) and participants exposed to fixed-scaled charts (Blue solid line). The score for perceived risk ranges from 1 to 5, indicated by participants. Risk levels are annualized volatility.



Then I investigate whether visual scaling can explain the lack of risk evaluation in the Auto group using an OLS regression. The dependent variable is the perceived risk score ranging from 1 to 5. Table 3 presents the results. The coefficient estimates for Volatility are consistently positive and significant in the Fixed group (columns (1) and (2)), indicating that participants' risk perception is consistent with actual volatility when presented with fixed-scaled charts. In contrast, the coefficient estimates are insignificant in the Auto group (columns (4) and (5)), suggesting that participants fail to detect risk in a way consistent

with actual volatility when viewing auto-scaled charts. The coefficient for Volatility \times VS is negative and significant in the Fixed group (column (2)). Since VS reflects the range of the price path in fixed-scaled charts, this negative coefficient suggests that participants' risk perception increase when the stock exhibits a wider price range (i.e., lower VS). On the other hand, the coefficient for Volatility \times VS is positive and significant in the Auto group (column (5)). This positive coefficient can be interpreted in two ways. First, it indicates that participants' perceived risk increases when they see a chart with high visual scaling, after controlling for actual volatility. Second, since the term Volatility \times VS represents the depicted volatility (as shown in Equation 2), the positive coefficient of the interaction term suggests that participants' risk perceptions align more closely with depicted volatility (0.46, t-statistic = 3.42) than with actual volatility (-0.18, t-statistic = -0.46). The inclusion of demographic control in the regressions (columns (3) and (6)) and the incorporation of participant fixed effects (Panel B) do not alter either the magnitude or the statistical significance of the main findings. Table 3 suggests that visual scaling affects participants' risk perception when they are exposed to auto-scaled charts, which is consistent with the experimental evidence of deceptive visualizations documented in behavioral science studies (Lauer and O'Brien (2020), Yang et al. (2021)). Furthermore, the experiment confirms that the VS measure in auto-scaled charts influences investors' risk perception through its visual scaling effect.

Table 3: Visual Scaling and Risk Evaluation

This table reports the coefficient estimates from a OLS regression. The dependent variable is the score for perceived risk ranging from 1 to 5. Volatility is annualized standard deviation of returns (20%, 40%, 60%, 80%, 100%) of the 10 price paths used in the experiment. VS is the calculated using the 3-month window, consistent with charts displayed in the experiment. Controls include Male, Age group, Education, Income group, Self-stated risk aversion, Self-stated risk aversion, White, and US. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: OLS						
Sample =	Fixed-scaled (N=48)			Auto-scaled (N=59)		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.80*** (15.29)	2.96*** (9.07)	2.84*** (8.37)	3.43*** (29.80)	2.92*** (8.92)	2.32*** (7.62)
Volatility	1.90*** (10.68)	1.56*** (4.03)	1.91*** (4.77)	-0.08 (-0.49)	-0.18 (-0.46)	-0.20 (-0.55)
Volatility \times VS _{3m}		-0.59*** (-3.94)	-0.55*** (-3.41)		0.46*** (3.12)	0.55*** (3.38)
VS _{3m}		-0.06 (-0.99)	-0.08 (-1.32)		-0.02 (-0.25)	-0.00 (0.00)
Controls	NO	NO	YES	NO	NO	YES
Adj R ²	19.33%	25.02%	28.53%	-0.01%	2.04%	3.21%
Panel B: Participants Fixed Effect						
Sample =	Fixed-scaled (N=48)			Auto-scaled (N=59)		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.80*** (16.95)	2.96*** (10.32)	2.84*** (8.63)	3.43*** (33.30)	2.91*** (10.02)	2.32*** (8.79)
Volatility	1.90*** (11.84)	1.56*** (4.43)	1.91*** (4.92)	-0.08 (-0.55)	-0.18 (-0.52)	-0.20 (-0.60)
Volatility \times VS _{3m}		-0.59*** (-4.43)	-0.55*** (-4.80)		0.46*** (3.51)	0.55*** (3.53)
VS _{3m}		-0.06 (-1.10)	-0.08 (-1.55)		-0.02 (-0.29)	-0.00 (0.03)
Controls	NO	NO	YES	NO	NO	YES
FE	YES	YES	YES	YES	YES	YES
Adj R ²	19.27%	25.05%	28.58%	0.00%	2.53%	3.82%

2.4 Deceptive Visualizations and Asset Prices

After documenting evidence that visual scaling influences investors’ risk perception through experiment, I further investigate its asset pricing implications. In financial markets, many investors remain undiversified. As documented in the *Capital Markets Fact Book*, 90% of investors in the 1950s were individual investors, and even as of 2023, this figure remains at 40.7%. Most individual investors’ portfolios are highly undiversified (Goetzmann and Kumar (2008)). In an imperfect financial market where investors hold undiversified portfolios, Levy (1978), Merton et al. (1987) and Malkiel and Xu (2002) predict that idiosyncratic risk should be compensated with positive returns. However, the positive idiosyncratic risk compensation observed in early literature (Lintner (1965), Lehmann (1990), Tinic and West (1986)) has become increasingly difficult to empirically identify in recent decades.¹⁵ In fact, many studies find a negative risk-return relationship (Ang et al. (2006), Ang et al. (2009)), leading to what is now referred to as the “low risk anomalies”. Even though lotteries (Bali et al. (2017)) or skewness (Schneider et al. (2020)) have been proposed to explain such “low-risk anomalies”, these studies generally fail to find evidence of positive risk compensation. In this section, I examine how visual scaling affects the risk-return trade-off in a market with undiversified investors in the era of digital charting.

2.4.1 Testable Propositions

I introduce an equilibrium model with n poorly diversified investors, a fraction of whom evaluate risk based on depicted volatility, hereafter referred to as visual investors. The expected return of stock i in equilibrium is

$$\mathbb{E}(R_i) = r + \frac{\sigma_i^2 S_i}{\tau \left[n_{1,i} + \sum_{j=1}^{n_{2,i}} \frac{1}{VS_{ij}^2} \right]}, \quad (5)$$

¹⁵This phenomenon is also discussed in Ang et al. (2006) Page 261.

where R_i is the return of stock i , r is the risk-free rate, σ_i^2 is the variance of the stock i , S_i is the total supply of stock i , and $\tau > 0$ is the coefficient of risk aversion. $n_{1,i}$ is the number of investors who observe the actual variance σ_i^2 of stock i 's return, while $n_{2,i}$ is the number of visual investors who were presented with a visually-scaled variance $(VS_{ij}\sigma_i)^2$, where VS_{ij} represents investor j 's degree of visual scaling in assessing the risk of stock i . $n_i = n_{1,i} + n_{2,i}$. The derivation of Equation 5 is presented in Appendix 1C.

Three testable propositions can be developed from Equation 5, depending on the fraction of visual investors (n_2). If the visual investors dominate the market such that $n_{2,i} \rightarrow n_i$ and the visual scaling VS_{ij} are homogeneous and can be estimated as VSE_i , then the expected return is:

$$\mathbb{E}(R_i) \approx r + VSE_i^2 \frac{S_i}{\tau n_i} \sigma_i^2. \quad (6)$$

Proposition 1: When visual investors are present, they demand premium for the depicted volatility ($\sigma_i \times VSE_i$).¹⁶ Thus, there is a positive relationship between depicted volatility and expected returns.

If there are no visual investors for stock i , i.e., $n_{2,i} \rightarrow 0$, the expected return simplifies to:

$$\mathbb{E}(R_i) = r + \frac{S_i}{\tau n_i} \sigma_i^2. \quad (7)$$

Proposition 2: When visual investors are absent, they demand premium for the actual volatility (σ_i), resulting in a positive relationship between actual volatility and expected returns, but there is no premium for depicted volatility.

Proposition 3: When visual investors are present with asset pricing implication in Equation 6, and the volatility (σ_i) is negatively related to visual scaling (VSE_i), the risk premium associated with the actual volatility is smaller. A detailed proof is provided in Appendix 1C.

¹⁶Since both σ_i and VSE are non-negative, although $\sigma_i \times VSE_i$ and $\sigma_i^2 \times VSE_i^2$ are not mathematically interchangeable, their economic interpretation in this setting remains directionally consistent.

2.4.2 Deceptive Visualizations in the post 1980s

Since digital charting is based on GUI technology, which was developed in the 1980s, investors were influenced by auto-scaling since then. As predicted by *Proposition 1*, when undiversified visual investors are present in the market, depicted volatility ($\sigma_i \times VSE_i$) should be compensated with positive returns. One empirical challenge is that stocks with high risk are also likely to be lotteries (Bali et al. (2017), Bali et al. (2011)). Visual scaling not only visually magnifies (compresses) the risk, but also magnifies (compresses) the lottery characteristic associated with risky stocks. Therefore, it lacks logic to test the effect of visual scaling on the risk-return trade-off without considering lottery characteristics. To address this issue, I construct residual risk proxies (RVol \perp and IVol \perp) by regressing the risk proxies against lottery proxies, including maximum return (MAX) and idiosyncratic skewness (ISkew).

Table 4 reports the results of the Fama-MacBeth regression showing how risk compensation is influenced by VSE. In Panel A, the coefficients for Risk \times VSE are all positive and significant across multiple subsample periods and for both two risk proxies. This provides evidence of positive compensation for depicted volatility as predicted by *Proposition 1*. Including controls, as shown in Panel B, does not diminish the effect of VSE on risk premium. An alternative interpretation of the result is that investors demand greater premium for risk as visual scaling increases. Intuitively, when stocks are presented with low visual scaling, the risk is inadequately communicated to investors, leading to poor risk assessment and, consequently, lower risk premium for these stocks.

Table 4: Visual Scaling Exposure and Risk-Return Trade-Off: Post-1980s

Every month, I run a cross-sectional regression of excess returns on lagged variables. The time-series average of the regression coefficients is reported. $RVol_{\perp}$ is residual return volatility, and $IVol_{\perp}$ is residual idiosyncratic volatility. $RVol_{\perp}$ ($IVol_{\perp}$) is constructed by regressing $RVol$ ($IVol$) against the lottery proxies (MAX and ISkew). VSE is the mean of VS_{3m} , VS_{6m} and VS_{12m} . For the purpose of interpretation, VSE is standardized with mean of 0, standard deviation of 1. The control variables are size, momentum, price, turnover, distance to 52-week high, distance to 52-week low and their interaction with risk measures. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: No Control						
Risk =	$RVol_{\perp}$			$IVol_{\perp}$		
Sample Period	1980s- 1990s	2000s- present	1980s- present	1980s- 1990s	2000s- present	1980s- present
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.73** (1.81)	0.81** (2.13)	0.78*** (2.88)	0.72** (2.13)	0.81** (2.25)	0.77*** (3.07)
Risk	0.64*** (3.00)	0.40* (1.83)	0.50*** (3.21)	1.78** (2.45)	1.22** (2.00)	1.46** (3.06)
Risk \times VSE	0.90*** (5.96)	0.79*** (3.07)	0.84*** (5.25)	2.95*** (4.14)	3.04*** (4.49)	3.02*** (6.94)
VSE	0.24** (2.18)	-0.02 (-0.35)	0.09 (1.47)	0.21* (1.85)	-0.04 (-0.31)	0.07 (1.40)
Panel B: With Controls						
Risk =	$RVol_{\perp}$			$IVol_{\perp}$		
Sample Period	1980s- 1990s	2000s- present	1980s- present	1980s- 1990s	2000s- present	1980s- present
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.67** (1.69)	0.72*** (3.09)	0.70*** (3.58)	0.65* (1.92)	0.70** (2.04)	0.68*** (2.97)
Risk	0.47* (1.77)	1.01*** (3.03)	0.77*** (3.21)	0.95 (0.95)	3.18** (3.31)	2.20** (3.06)
Risk \times VSE	0.82*** (3.79)	0.60*** (2.61)	0.69*** (4.11)	2.65*** (3.54)	2.30*** (3.39)	2.46*** (4.54)
VSE	0.22** (2.48)	0.00 (0.01)	0.10 (1.32)	0.18** (2.31)	-0.02 (-0.19)	0.07 (0.98)
Controls	YES	YES	YES	YES	YES	YES
Controls Inter.	YES	YES	YES	YES	YES	YES

2.4.3 Retail and Institutional Investors

I investigate the cross sectional variation of visual scaling effect across retail and institutional investors in this section. Institutional investors are more sophisticated and thus less likely to be visual investors. As predicted by *Proposition 2*, in the absence of visual investors, the positive risk premium is associated with actual risk (σ_i) rather than depicted risk ($\sigma_i \times VSE_i$). I proxy institutional investor presence using institutional ownership, as well as firm size and stock price, which are most strongly correlated with institutional ownership among other firm characteristics (Gompers and Metrick (2001)). I compute the first principal component (PCA_{inst}) of the three proxies as a combined measure of institutional investor presence. Stocks are then sorted into quintiles based on PCA_{inst} , with quintile 1 (low PCA_{inst}) representing the retail subsample and quintile 5 (high PCA_{inst}) representing the institutional subsample. Table 5 presents the Fama-MacBeth regression results using retail and institutional subsamples. In Panel A of Table 5, I find that the coefficient on Risk \times VSE is positive and significant for stocks with higher retail ownership, but insignificant for those with greater institutional ownership. Moreover, the coefficient on Risk alone is positive and significant for stocks with more institutional investors, while it is insignificant for stocks dominated by retail investors. These findings hold for both residual return volatility and residual idiosyncratic volatility, and remain robust after the inclusion of control variables in Panel B. Overall, Table 5 provides evidence that stocks with higher retail ownership exhibit a premium on depicted volatility as visual scaling effect is more pronounced among retail investors. In contrast, stock with greater institutional ownership display a positive premium on actual volatility, consistent with institutions being less susceptible to visual scaling effect.

Table 5: Retail versus Institutional Investors

This table presents the Fama-MacBeth regression results of visual deception in Table 4, using retail and institutional subsamples. I use price, size, and institutional ownership as proxies for institutional investor presence and compute their first principal component (PCA_{inst}). Stocks are then sorted into quintiles based on PCA_{inst} , with quintile 1 (low PCA_{inst}) representing the retail subsample and quintile 5 (high PCA_{inst}) representing the institutional subsample. $RVol_{\perp}$ is residual return volatility, and $IVol_{\perp}$ is residual idiosyncratic volatility. $RVol_{\perp}$ ($IVol_{\perp}$) is constructed by regressing $RVol$ ($IVol$) against the lottery proxies (MAX and ISkew). VSE is the mean of VS_{3m} , VS_{6m} and VS_{12m} . For the purpose of interpretation, VSE is standardized with mean of 0, standard deviation of 1. The control variables are size, momentum, price, turnover, distance to 52-week high, distance to 52-week low and their interaction with risk measures. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: No Control				
Risk =	$IVol_{\perp}$		$RVol_{\perp}$	
	Retail (1)	Institutional (2)	Retail (3)	Institutional (4)
Intercept	0.50** (2.45)	0.94*** (3.69)	0.47** (1.99)	0.94*** (4.85)
Risk	0.23 (1.43)	0.73*** (3.28)	0.55 (0.93)	2.63*** (4.11)
Risk \times VSE	1.06*** (6.44)	0.06 (0.55)	3.78*** (6.26)	0.18 (0.29)
VSE	0.33*** (3.48)	-0.03 (-0.33)	0.28*** (2.99)	-0.05 (-0.54)
Panel B: With Controls				
Risk =	$IVol_{\perp}$		$RVol_{\perp}$	
	Retail (1)	Institutional (2)	Retail (3)	Institutional (4)
Intercept	0.86** (3.00)	0.99*** (4.45)	0.83** (2.99)	0.97*** (4.64)
Risk	0.26 (0.60)	2.04*** (5.03)	0.51 (0.33)	6.78*** (4.72)
Risk \times VSE	1.01*** (5.07)	0.24 (1.56)	3.62*** (5.26)	0.76 (1.46)
VSE	0.28*** (3.62)	0.06 (1.19)	0.25*** (3.36)	0.03 (0.58)
Controls	YES	YES	YES	YES
Controls Inter.	YES	YES	YES	YES

2.4.4 Absence of Deceptive Visualizations in the Early Era (1920s - 1980s)

The extent of stock price visualization in financial markets has evolved over time, driven by the development of computer technology. Before the 1960s, when computer graphics had not yet been developed,¹⁷ stock price charting involved manually drawing charts on graph paper with stock quotations sourced from newspapers, quotation boards, or ticker tapes.¹⁸ In the 1960s, computer technology had entered the financial sector,¹⁹ along with the development of programming tools like SPSS (1968) and SAS (1976), the text-based interfaces still made data visualization primarily accessible only to programmers. Since the 1980s, with the rise of graphical user interfaces (GUIs), tools like the Bloomberg Terminal (1981) and Reuters systems began to offer more sophisticated charting options. Traders could now view stock prices in real time through line charts, bar charts, and candlestick charts, making technical analysis more accessible, but also introducing the potential for visual scaling effect, as documented earlier.

In the absence of digital charting, visual investors would not exist. Therefore, as predicted by *Proposition 2*, depicted volatility ($\sigma_i \times VSE_i$) should not be associated with a risk premium in the pre-GUI period. Table 6 presents the impact of VSE on risk premium before the 1980s. In Panel A, the coefficients for Risk \times VSE are insignificant for both risk proxies across multiple subsample periods, providing evidence that depicted volatility was not priced, consistent with *Proposition 2*. The same results hold after including additional controls, as shown in Panel B.

¹⁷Computer Graphics (CG) was first created as a visualization tool for scientists and engineers in government and corporate research centers such as Bell Labs and Boeing in the 1950s.

¹⁸In the early days, trading relied on the open outcry system, where brokers vocally communicated buy and sell orders on the trading floor. The invention of the ticker tape in 1867 by Edward A. Calahan revolutionized this process by transmitting stock prices via telegraph, allowing investors to track prices in real-time without being physically present. Ticker tape remained in use for over a century.

¹⁹In the 1960s, the computer-based Quotron system introduced electronic price quotations, displaying stock prices on screens and eliminating the need for ticker tapes. In 1972, NASDAQ, the first electronic stock market, was founded.

Table 6: Visual Scaling Exposure and Risk-Return Trade-Off: Pre-1980s

Every month, I run a cross-sectional regression of excess returns on lagged variables. The time-series average of the regression coefficients is reported. $RVol_{\perp}$ is residual return volatility, and $IVol_{\perp}$ is residual idiosyncratic volatility. $RVol_{\perp}$ ($IVol_{\perp}$) is constructed by regressing $RVol$ ($IVol$) against the three lottery proxies (MAX and ISkew). VSE is the mean of VS_{3m} , VS_{6m} , and VS_{12m} . For the purpose of interpretation, VSE is standardized with mean of 0, standard deviation of 1. The control variables are size, momentum, price, turnover, distance to 52-week high, distance to 52-week low and their interaction with risk measures. The sample period begins in 1926, which is the earliest available stock price data in CRSP. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: No Control						
Risk =	$RVol_{\perp}$			$IVol_{\perp}$		
Sample Period	1920s- 1930s	1940s- 1950s	1960s- 1970s	1920s- 1930s	1940s- 1950s	1960s- 1970s
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.39** (2.31)	0.34*** (3.59)	0.53*** (3.67)	1.44** (2.29)	0.33*** (3.39)	0.49*** (3.42)
Risk	0.95 (1.13)	3.03*** (8.50)	1.84*** (6.54)	6.03 (1.61)	9.85*** (8.78)	5.38*** (5.12)
Risk \times VSE	-1.05 (-1.44)	0.91 (0.95)	0.50 (0.46)	-1.44 (-0.47)	3.26 (0.97)	0.94 (0.68)
VSE	-0.98 (-1.29)	0.11 (1.15)	0.10 (0.79)	-0.90 (-1.50)	0.04 (0.41)	0.02 (0.15)
Panel B: With Controls						
Risk =	$RVol_{\perp}$			$IVol_{\perp}$		
Sample Period	1920s- 1930s	1940s- 1950s	1960s- 1970s	1920s- 1930s	1940s- 1950s	1960s- 1970s
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	1.12* (1.91)	0.87*** (6.50)	1.83*** (8.68)	1.02* (1.95)	0.88*** (6.42)	1.81*** (8.52)
Risk	5.72*** (2.75)	4.37*** (6.21)	2.31*** (4.71)	26.77*** (2.95)	15.20*** (6.25)	5.52*** (2.77)
Risk \times VSE	-0.32 (-0.44)	0.75 (0.95)	-0.03 (-0.26)	-1.90 (-0.67)	2.39 (0.69)	-0.99 (-0.51)
VSE	0.03 (0.12)	0.20*** (3.35)	-0.02 (-0.27)	-0.03 (-0.08)	0.14** (2.35)	-0.10 (-1.26)
Controls	YES	YES	YES	YES	YES	YES
Controls Inter.	YES	YES	YES	YES	YES	YES

2.4.5 Portfolio Performances and Time-series Analysis

In this subsection, I construct portfolios to trade on the effect of deceptive visualization on risk premium. Since previous tests have shown that depicted volatility carries a premium, one might initially consider constructing a portfolio based on depicted volatility ($RVol\perp \times VSE$). However, because volatility and VSE are negatively correlated in auto-scaled charts,²⁰ the composite measure of depicted volatility, which is an interaction of the two, might suffer from collinearity. Specifically, the resulting portfolio on depicted volatility might simply consist of stocks with high VSE and low $RVol\perp$, or stocks with low VSE and high $RVol\perp$. This approach would fail to cleanly isolate the independent effects of VSE or $RVol\perp$. To address this issue and also potential challenge on multicollinearity in regression models, a double sort is employed.²¹

Table 7, Panel A, presents the impact of VSE on the risk premium using independent double sort. Each month, stocks are independently sorted into VSE quintiles (V1 to V5) and into volatility (R1 to R5) quintiles based on $RVol\perp$. Under this independent double-sort approach, all stocks within the same VSE quintile have a similar VSE regardless of their volatility quintiles. However, due to the negative correlation between VSE and volatility, the resulting 5×5 portfolios have an unbalanced number of constituent stocks, a feature I will address later. These 25 portfolios are formed at the beginning of each month and held for one month, and I report the equal-weighted returns for each portfolio along with the performance of the volatility spread ($R5-R1$) of each VSE quintile. The results show that the

²⁰Stocks with high volatility are more likely to exhibit wider price paths, leading to smaller value of visual scaling exposure. To illustrate this relationship, I sort stocks into quintiles based on their volatility and report the 25th percentile, 75th percentile, and mean of VSE for each quintile. The result is reported in Appendix 1D

²¹The negative relationship between VSE and volatility measures is economic or statistical in nature. However, this relationship is not strictly mathematically dependent. If VSE were mechanically dependent on Volatility, which could be expressed as $VSE = f(Volatility)$, then the construction of an independent double-sorted portfolio would be empirically infeasible. The extreme cells of the sort (e.g., the High VSE and High Volatility portfolios, $V5R5$) would contain none or an extremely small number of stock observations, making portfolio formation impossible.

volatility spread yields a positive and significant alpha across all VSE quintiles. This finding suggests that high-volatility stocks carry a premium as long as they are displayed to investors with the same visual scaling, thereby mitigating the effect of deceptive visualization.

Table 7: Portfolio Performances

Panel A: At the beginning of every month, I independently sort stocks into five groups based on VSE (V1 to V5) and residual volatility (R1 to R5). The 5×5 portfolios are then held for one month. This table reports the monthly returns for each portfolios. Volatility spreads are calculated as R5 - R1 for each VSE quintiles. Return and 4-factor alpha (%) are reported for volatility spreads. VSE is the mean of VS_{3m} , VS_{6m} and VS_{12m} . Residual volatility (RVol \perp) is constructed by regressing RVol against the lottery proxies (MAX and ISkew). I only report the bottom-, middle-, and top-quintile volatility portfolios to save space. Panel B: Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: Independent Double Sort						
	VSE	V1	V2	V3	V4	V5
RVol \perp						
R1		-0.12	0.39	0.55	0.62	0.36
R3		0.55	0.72	0.85	0.89	0.82
R5		0.49	0.93	1.02	1.07	0.88
R5-R1 Return		0.61*** (4.28)	0.54*** (3.58)	0.47*** (3.09)	0.45*** (3.50)	0.51*** (4.05)
R5-R1 4-factor α		0.41*** (3.10)	0.39*** (2.77)	0.27** (2.17)	0.26** (2.02)	0.43*** (2.84)
Panel B: Volatility Portfolios						
		RVol \perp		RVol \perp Within VSE		
		(1)	(2)	(3)	(4)	
		EW	VW	EW	VW	
R1		0.40	0.57	0.45	0.54	
R3		0.81	0.85	0.76	0.74	
R5		0.72	0.99	0.83	0.95	
R5-R1 Return		0.32 (1.45)	0.42** (1.99)	0.38*** (3.06)	0.41*** (3.48)	
R5-R1 4-factor α		0.00 (0.00)	0.18 (1.30)	0.24*** (2.77)	0.34*** (3.61)	

As predicted by *Proposition 3*, the negative correlation between VSE and volatility can lead to a diminishing risk-return trade-off in the presence of visual investors. In Columns (1)

and (2) of Panel B, I report the performance of portfolios formed solely on $\text{RVol}\perp$. Consistent with both *Proposition 3* and existing literature, the volatility premium largely disappears, particularly for the equal-weighted portfolios. However, some evidence of a positive risk-return trade-off remains in Column (2) under the value-weighted portfolio. This suggests that the relationship might persist among large-cap stocks. Since investors in large-cap stocks are more likely to be institutional and thus less likely to be visual investors, their stock prices are less affected by visual scaling, as predicted by *Proposition 2*.

As discussed earlier, the independent sorting results in an unbalanced number of constituent stocks across portfolios. To address this, I employ a dependent sort, first sorting on VSE and then sorting on $\text{RVol}\perp$ within each VSE quintile. The performance of the portfolio formed on $\text{RVol}\perp$ is reported in Columns (3) and (4). Consistent with the results presented in Panel A, high volatility stocks continue to carry a positive and significant premium. However, the magnitude and statistical significance of this premium are generally lower than those reported in Panel A. This reduction is expected because the volatility spread, constructed conditional on VSE quintiles, exhibits less dispersion on $\text{RVol}\perp$ (i.e., less difference between the top and bottom $\text{RVol}\perp$ quintiles) compared to the unconditional spread.

Next, I extend the portfolio performance analysis to the full sample period from the 1920s to the present. I construct three distinct equal-weighted portfolios at the beginning of each month and monitor their returns for the corresponding month. The three portfolios are defined as follows:

- (1) **Sigma Portfolio:** This portfolio takes a long (short) position in stocks belonging to the top (bottom) quintiles of $\text{RVol}\perp$ based on an unconditional sort. This portfolio is identical to the one reported in Table 7, Panel B, Column (1).
- (2) **Sigma_within_VSE Portfolio:** This portfolio is based on volatility after controlling for VSE. It takes a long (short) position in stocks from the top (bottom) quintiles of $\text{RVol}\perp$ within each VSE quintile. This portfolio is identical to the one reported in

Table 7, Panel B, Column (3).

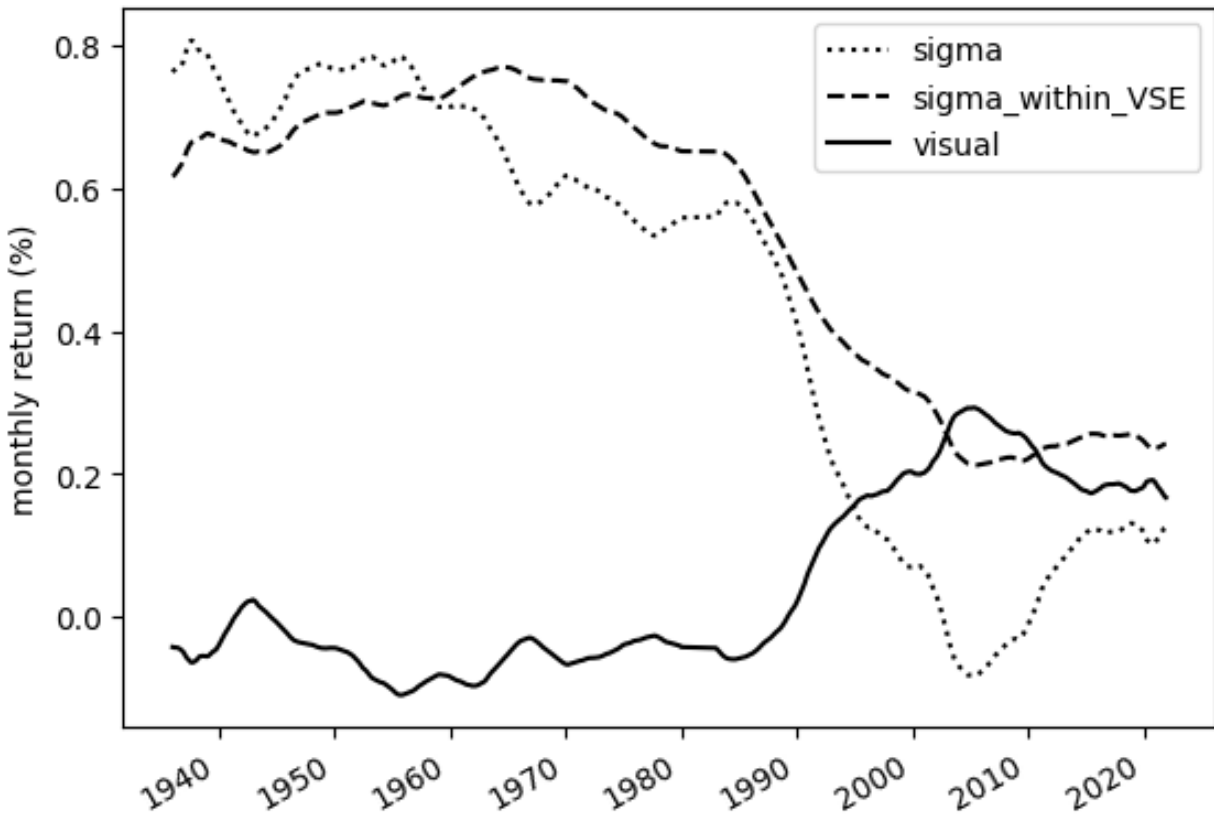
- (3) Visual Portfolio: This portfolio is a volatility-neutral portfolio on VSE, constructed using a 5×5 independent sort based on $RVol_{\perp}$ and VSE. This portfolio takes a long position in stocks that fall into both the top VSE quintile and the extreme $RVol_{\perp}$ quintiles (the top and bottom $RVol_{\perp}$ quintiles). Conversely, it takes a short position in stocks belonging to both the bottom VSE quintile and the extreme $RVol_{\perp}$ quintiles (the top and bottom $RVol_{\perp}$ quintiles). Equivalently, the portfolio's long positions are located in cells (R1, V5) and (R5, V5), and its short positions are in cells (R1, V1) and (R5, V1) of the Table 7, Panel A. The logic of this portfolio is that after controlling for volatility, the price charts with higher visual scaling are generally perceived as more risky.

The performances of the three portfolios are reported in Figure 4. *Proposition 3* predicts that the unconditional volatility premium should diminish when the market is affected by deceptive auto-scaled charts. Consistent with this prediction, the performance of the Sigma portfolio (the dotted line) is positive at approximately 0.6% per month before the 1980s but experiences a dramatic decline thereafter. While the decline of the Sigma portfolio post-1980s can be explained by both the introduction of digital charts and a general decline in the volatility premium due to increased diversification among market participants, the performance of the Sigma_within_VSE portfolio (the dashed line) provides better insight into the long-term trend of the volatility premium, as it controls for the effect of deceptive visualizations. The performance of the Sigma_within_VSE portfolio also shows a declining trend, reflecting the general diminishing of the volatility premium post-1980s, but at a smaller rate. Specifically, before the 1980s, the performance of the Sigma_within_VSE portfolio is approximately 0.6%, nearly identical to the Sigma portfolio. However, after the 1980s, the Sigma_within_VSE portfolio can still generate a monthly return of above 0.2% per month. Finally, the Visual portfolio continues to earn approximately zero return before

the 1980s but begins to earn a premium thereafter, reaching its peak at 0.3% – 0.4% per month between 2000 and 2010, and starts to decrease after 2010. This decline in the Visual portfolio’s performance after 2010 might indicate factors such as investor learning, increased sophistication of trading platforms, or the rise of algorithmic trading.

Figure 4: Portfolio Performances Over Time

This figure plots the monthly returns (in percent) of three key long-short portfolios: the “Sigma” portfolio, the “Sigma_within_VSE” portfolio, and the “Visual” portfolio. The “Sigma” portfolio is based on volatility, taking a long (short) position in stocks belonging to the top (bottom) quintiles of residual volatility (RVol \perp) based on an independent sort. The “Sigma_within_VSE” portfolio is based on volatility after controlling for VSE, taking a long (short) position in stocks from the top (bottom) quintiles of RVol \perp within each VSE quintile. The “Visual” portfolio is a volatility-neutral portfolio on VSE, constructed using a 5×5 independent sort based on RVol \perp and VSE. This portfolio takes a long position in stocks that fall into both the top VSE quintile and the extreme RVol \perp quintiles (the top and bottom RVol \perp quintiles). Conversely, it takes a short position in stocks belonging to both the bottom VSE quintile and the extreme RVol \perp quintiles (the top and bottom RVol \perp quintiles). Each portfolio is constructed monthly, and the equal-weighted return for the following month is recorded. The returns are smoothed using a 20-year moving average, requiring a minimum of 5 years of data.



2.5 Further Analysis

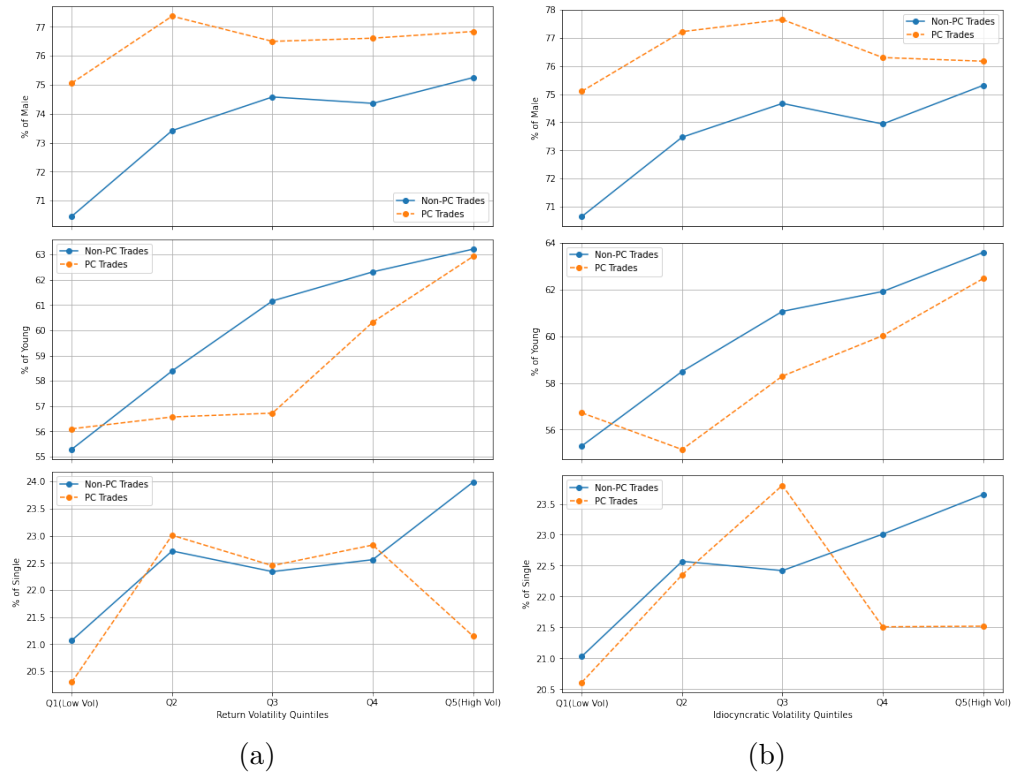
2.5.1 Evidence from Large Brokerage Account

Do auto-scaled charts influence trading behavior when investors assess stock risk? In this section, I present evidence of deceptive visualizations from trading records. Empirically studying the effect of deceptive visualizations in real trading data is challenging, as it is nearly impossible to establish an individual investor's normative behavior. Specifically, when an investor initiates a buy trade, unless they self-report their risk perception, it is unclear whether they were misled or genuinely intended to make the purchase. Extensive psychological studies (Byrnes et al. (1999), Deakin et al. (2004), Halek and Eisenhauer (2001)) have established that, differences in risk preferences across demographic groups are rooted in human nature. Barber and Odean (2001) show that males are more risk-seeking than females based on an analysis of their stock holdings. Following these studies, I use demographic-based risk preferences as a proxy for normative behavior to examine if visual scaling influences investors' behavior toward risk.

I first examine the demographic-based risk preferences among non-PC trades, then I study how PC trades deviate from this normative behavior. For non-PC buy trades, I sort trades into quintiles based on stocks' volatility estimated using historical monthly return over the past 12 months. For each quintile, I calculate the proportion of households that are male, young, and single. I then perform the same analysis for PC trades. Figure 5 (a) presents the results for return volatility, while Figure 5 (b) reports the results for idiosyncratic volatility.

Figure 5: Risk Preferences by Demographics

This figure plots the demographic characteristics of households across risk quintiles for both non-computer-based (Non-PC Trades) and computer-based trades (PC Trades). Specifically, for non-computer-based buy trades, I sort trades into quintiles based on return volatility (subfigure (a)) or idiosyncratic volatility (subfigure (b)). For each quintile, I plot the proportion of households who are male, young, and single, represented by the blue solid line. I perform the same analysis for computer-based trades, with results shown using the orange dashed line. Return volatility is estimated using monthly return from past 12 months. Idiosyncratic volatility is estimated using monthly return from past 12 months under market model.



The blue solid line in Figure 5 represents the demographic characteristics of non-PC trades. The blue lines exhibit an upward trend—indicating that higher risk is associated with a greater proportion of male, young, and single investors. This pattern aligns with findings from psychological studies (Byrnes et al. (1999), Deakin et al. (2004), Halek and Eisenhauer (2001)), reflecting human nature toward risk-taking.²² However, for PC trades, represented

²²I acknowledge that demographic risk preferences are context-dependent. However, the normative demographic risk preference in this study aligns with findings in the trading context (Kumar (2009), Barber and

by the orange dashed line, this human nature —or normative behavior —diminishes for age and even disappears for gender and marital status. This trading pattern aligns with the experimental findings shown in Figure 3. Both Figure 5 and Figure 3 demonstrate that when investors or participants evaluate risk based on numerical or textual information when they initiate trades on telephones, or viewing fixed-scaled charts in experiment, their risk perceptions are consistent with standard deviation-based volatility. However, when exposed to auto-scaled charts on PCs or in experimental settings, their risk perceptions become less aligned with standard deviation-based measures of volatility.

I acknowledge that PC trades and non-PC trades differ in many aspects. As discussed by Barber and Odean (2002), PC trades may involve more information, reflect greater confidence at the time of order placement, or be more speculative, these differences do not directly explain the observed deviation from demographic-based normative behavior toward risk. Since individuals investors spend most of their research time on viewing charts before place a trade, I argue that the key difference between PC trades and non-PC trades that drives the divergence in demographic-based risk-taking behavior might be due to the access to digital charting. PC traders are more likely to be influenced by the auto-scaling effect embedded in digital charting software. Empirically, high-risk stocks tend to have lower visual scaling (detailed in Figure 12), thus PC traders' risk perception may be distorted by auto-scaling, which ultimately leads to the departure from demographic-based normative behavior toward risk.

I further investigate whether visual scaling is a factor explaining the deviation from normative demographic-based behavior toward risk among PC trades using probit regressions. The dependent variables are indicator variables for male, young, and single. The first independent variable of interest is risk quintiles, based on either return volatility or idiosyncratic volatility, range from 1 to 5. The coefficient estimates for this variable are consistently positive and significant, indicating the presence of human normative behavior

Odean (2001)).

toward risk. The coefficients for the interaction term, risk quintile \times VSE, are insignificant, whereas the coefficients for risk quintile \times VSE \times PC are consistently positive and significant. This result suggests that VSE does not influence normative risk-taking behavior when households execute trades through non-PC channels. In contrast, when a trade is initiated via PC, households are more likely to be influenced by the auto-scaling embedded in the digital charting of the trading platform. Accordingly, I find that high-VSE stocks attract more buy trades from male, younger, and single households, suggesting that higher visual scaling amplifies perceived risk—consistent with the experimental findings reported in Table 3.

Table 8: Visual Scaling Exposure and Demographic-based Risk Preference

This table reports the coefficient estimates from a probit regression. The dependent variables are Male, Young, and Single. Male is a dummy variable indicating male investors. Young is a dummy variable indicating whether the household head is under 40. Single is a dummy variable for single investors. Risk quintiles are determined by sorting trades into quintiles based on return volatility or idiosyncratic volatility, where Quintile 1 represents the lowest risk and Quintile 5 the highest. Return volatility is estimated using monthly return from past 12 months. Idiosyncratic volatility is estimated using monthly return from past 12 months under market model. VSE is the mean of VS_{6m} and VS_{12m} . PC is a dummy variable that equals 1 if the trade is computer-based and 0 otherwise. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Dep. Var =	Male	Young	Single	Male	Young	Single
Risk =	RVol			IVol		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.69*** (39.04)	0.28*** (17.94)	-0.74*** (-43.04)	0.69*** (38.98)	0.26*** (16.47)	-0.75*** (-42.87)
Risk Quintiles	0.01** (2.20)	0.03*** (6.38)	0.01** (2.49)	0.01* (1.80)	0.03*** (6.45)	0.01** (1.99)
Risk Quintiles×VSE	0.00 (1.38)	-0.00 (-0.77)	-0.00 (-1.50)	0.00 (1.11)	-0.00 (-0.83)	-0.00 (-0.28)
Risk Quintiles×VSE×PC	0.03*** (3.65)	0.02*** (2.97)	0.03*** (3.03)	0.02** (2.17)	0.02** (2.56)	0.03*** (3.89)
VSE	-0.03*** (-5.93)	-0.04*** (-9.85)	-0.01** (-2.26)	-0.04*** (-6.07)	-0.04*** (-8.98)	-0.02** (-2.08)
PC	0.00 (0.07)	-0.03 (-0.95)	-0.12** (-2.24)	0.06 (1.31)	-0.06 (-1.30)	-0.09* (-1.85)
Risk Quintiles× PC	-0.03*** (-3.31)	-0.02** (-2.31)	-0.03*** (-3.01)	-0.04*** (-3.87)	-0.01 (-1.30)	-0.04*** (-3.61)
VSE× PC	0.04 (1.01)	-0.05 (-1.31)	0.03 (0.78)	-0.04 (-1.12)	0.02 (0.54)	-0.04 (-1.11)
Pseudo R ²	0.17%	0.35%	0.06%	0.17%	0.32%	0.06%

The coefficient for VSE is negative and significant, which is expected since VSE also captures the range-based fluctuations of the price path. This negative coefficient reflects the preference of male, young, and single households for stocks with a wider price ranges. The coefficient for PC is insignificant in most specifications. I also control for the interaction terms $PC \times$ risk quintiles as well as $PC \times$ VSE. Excluding these two interaction controls does not alter the main findings of this analysis.

2.5.2 Visual Scaling Exposure and Beta

In addition to idiosyncratic risk, which matters to undiversified investors, as predicted by CAPM, systematic risk should be priced by diversified investors. I examine whether visual scaling influences the premium of systematic risk, measured by beta. I first estimate beta using monthly returns over the past 12 months under the market model for each stock. To address the concern that beta is positively correlated with lottery proxies, I further regress beta on maximum return and return skewness following [Bali et al. \(2017\)](#). I run monthly Fama-MacBeth regressions with the dependent variable being the future abnormal return. The independent variables include residual beta, VSE, the interaction between residual beta and VSE, and a set of control variables. As shown in [Table 9](#), the positive and significant coefficient on residual beta (Beta_\perp) indicates that beta is compensated by mean-variance investors, a result consistent with [Bali et al. \(2017\)](#). The coefficient for $\text{VSE} \times \text{Beta}_\perp$ is insignificant for the full sample from 1980 to 2023, suggesting that visual scaling does not affect beta premium. This is because beta is not directly observable to investors in digital charts, meaning visual scaling does not influence their expected return on beta exposure.

Table 9: Visual Scaling Exposure and Beta

Every month, I run a cross-sectional regression of excess returns on lagged variables. The time-series average of the regression coefficients is reported. Residual beta (Beta_\perp) is constructed by regressing Beta against the lottery proxies (MAX and ISkew). VSE is the mean of VS_{3m} , VS_{6m} and VS_{12m} . For the purpose of interpretation, VSE is standardized with mean of 0, standard deviation of 1. The control variables are size, momentum, price, turnover, distance to 52-week high, distance to 52-week low and their interaction with Beta_\perp . Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Sample Period	1980s- 1990s	2000s- present	1980s- present	1980s- 1990s	2000s- present	1980s- present
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.68* (1.89)	0.75** (2.08)	0.72*** (2.82)	0.62* (1.74)	0.68* (1.96)	0.65*** (2.68)
Beta_\perp	0.19*** (2.99)	0.17* (1.81)	0.18*** (3.18)	0.22*** (2.72)	0.28*** (2.57)	0.26*** (3.69)
$\text{Beta}_\perp \times \text{VSE}$	0.14** (2.02)	0.05 (0.68)	0.09 (1.26)	0.17** (2.23)	0.04 (0.56)	0.09 (1.08)
VSE	0.18 (1.60)	-0.06 (-0.45)	0.05 (0.52)	0.19** (2.35)	-0.05 (-0.46)	0.06 (0.85)
Controls				YES	YES	YES
Controls Inter.				YES	YES	YES

2.5.3 Visual Scaling Exposure and Lottery Preference

I also attempt to investigate if visual scaling affect investors preference on lottery stocks. Investors with a lottery preference demand stocks that exhibit lottery-like characteristics in their past returns such as positive skewness, making stocks with lottery characteristics overpriced. High visual scaling in digital charts magnify these lottery characteristics, while low visual scaling compress them. I examine whether visual scaling can influence investors' skewness evaluation, thereby distorting their lottery preference.

Stocks' lottery characteristic is positively related to stock volatility (Bali et al. (2011)). To empirically test how visual scaling affects lottery preference without the influence of volatility, I consider maximum return (MAX), idiosyncratic skewness (ISkew) and stock salience (ST) as lottery proxies and construct residual lottery proxies (MAX_\perp , ISkew_\perp and ST_\perp) by regressing lottery proxies against volatility proxies (RVol and IVol). I run monthly

Fama-MacBeth regressions with the dependent variable being the future abnormal return. The independent variables include residual lottery proxies, VSE, the interaction between residual lottery proxies and VSE, and a set of control variables. The sample period covers the post-GUI era, from 1980 to 2023. As shown in Table 10, the negative significant coefficient of residual lottery proxies indicates that investors overprice lotteries. The coefficient estimates of the interaction between residual lottery proxies and VSE are all insignificant after including controls, implying that visual scaling does not affect investors' lottery preference. There are two potential explanations for why visual scaling does not affect lottery mispricing. First, the lottery characteristic, the skewness of stock returns, is not easily inferred from candlestick price charts, so investors' preference for lottery-like stocks may not be influenced by visual scaling of the price charts. Second, this result may suggest that lottery mispricing is not purely driven by investors' gambling preferences, but may also be confounded by other factors such as investor attention.

Table 10: Visual Scaling Exposure and Lottery Preference

Every month, I run a cross-sectional regression of excess returns on lagged variables. The time-series average of the regression coefficients is reported. $\text{MAX}\perp$ is residual maximum return, $\text{ISkew}\perp$ is residual idiosyncratic skewness, and $\text{ST}\perp$ is residual ST. $\text{MAX}\perp$ ($\text{ISkew}\perp$, $\text{ST}\perp$) is constructed by regressing MAX (ISkew , ST) against the three volatility proxies (RVol_{1m} and IVol_{1m}). VSE is the mean of VS_{3m} , VS_{6m} and VS_{12m} . For the purpose of interpretation, VSE is standardized with mean of 0, standard deviation of 1. The control variables are size, momentum, price, turnover, distance to 52-week high, distance to 52-week low and their interaction with the lottery proxies. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Lottery =	$\text{MAX}\perp$		$\text{ISkew}\perp$		$\text{ST}\perp$	
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.70*** (2.90)	0.65** (2.09)	0.70*** (2.91)	0.65** (2.10)	0.69*** (2.87)	0.64** (2.09)
Lottery	-2.24*** (-2.75)	-4.39*** (-3.15)	0.01 (0.63)	-0.01 (-0.16)	-3.67*** (-2.86)	-4.53** (-2.31)
Lottery \times VSE	1.13 (0.84)	-1.80 (-1.23)	0.05*** (3.77)	0.03 (1.10)	3.83** (2.55)	2.37 (1.40)
VSE	-0.02 (-0.24)	-0.12 (-1.37)	-0.02 (-0.23)	-0.13 (-1.44)	-0.03 (-0.31)	-0.13 (-1.40)
Controls		YES		YES		YES
Controls Inter.		YES		YES		YES

Further, I test the aggregate effect of visual scaling on lottery stocks. Lottery stocks are associated with high skewness and high volatility (Bali et al. (2011)). Section 2.4 provides evidence that visual scaling can shape investors' risk perceptions. Table 10 shows that visual scaling has no effect on investors' lottery preference. As a result, I hypothesize that lottery stocks are overvalued the most among stocks with low VSE. This is because investors are not visually exposed enough to the risk associated with lottery stocks, leading to overvaluation. On the other hand, lottery stocks are fairly valued or even undervalued among stocks with high VSE. This is because investors are visually exposed, or even overexposed, to the risk associated with lottery stocks, leading them to fully consider the risk, resulting in lottery stocks being fairly valued, or to overconsider the risk, resulting in lottery stocks being undervalued.

Table 11 presents the impact of visual scaling on lottery mispricing. Each month,

stocks are sorted by VSE into quintiles (V1 to V5). V1 represents stocks with low visual scaling exposure, while V5 represents those with high visual scaling exposure. Within each VSE quintile, stocks are further sorted based on lottery proxies into quintiles (R1 to R5). Portfolios are held for one month, and value-weighted returns are reported.

Table 11: Visual Scaling Exposure and Lottery Mispricing

At the beginning of every month, I first sort stocks into five groups (V1 to V5) based on VSE. Within each VSE groups, stocks are sorted into five groups (R1 to R5) based on lottery proxies. The portfolios are then held for one month. This table reports the monthly value-weighted raw returns for each portfolios. Lottery spreads are calculated as R5 - R1 within each VSE groups. Excess return and 4-factor alpha (%) are reported for lottery spreads. MAX is maximum daily return in the last month. ISkew is idiosyncratic volatility in the last month. ST is return salience in the past month. VSE is the mean of VS_{3m} , VS_{6m} , and VS_{12m} . I only report the bottom-, middle-, and top-quintile lottery portfolios to save space. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: MAX							
VSE \ MAX	V1	V2	V3	V4	V5	V5-V1	
R1	1.25	1.00	0.95	0.77	0.34		
R3	0.33	0.77	0.78	0.85	0.67		
R5	-0.54	0.34	0.72	0.82	0.86		
R5-R1 Return	-1.78*** (-12.53)	-0.66*** (-5.86)	-0.24** (-2.29)	0.05 (0.56)	0.52*** (4.45)	2.30*** (15.72)	
R5-R1 4-factor α	-1.63*** (-12.49)	-0.73*** (-7.51)	-0.32*** (-3.68)	-0.05 (-0.67)	0.23*** (2.93)	1.86*** (15.21)	
Panel B: ISkew							
VSE \ ISkew	V1	V2	V3	V4	V5	V5-V1	
R1	0.56	0.74	0.79	0.76	0.53		
R3	0.40	0.69	0.84	0.85	0.61		
R5	0.34	0.66	0.82	0.87	0.75		
R5-R1 Return	-0.22** (-2.34)	-0.08 (-1.10)	0.03 (0.59)	0.10* (1.79)	0.22*** (4.80)	0.44*** (6.34)	
R5-R1 4-factor α	-0.35*** (-3.50)	-0.19** (-2.45)	-0.02 (-0.38)	0.11** (2.19)	0.20*** (4.51)	0.54*** (7.45)	
Panel C: ST							
VSE \ ST	V1	V2	V3	V4	V5	V5-V1	
R1	0.64	0.90	0.95	0.93	0.67		
R3	0.65	0.77	0.77	0.78	0.53		
R5	-0.27	0.41	0.65	0.79	0.79		
R5-R1 Return	-0.91*** (-6.61)	-0.49*** (-4.90)	-0.30*** (-3.90)	-0.14* (-1.94)	0.12** (2.25)	1.03*** (7.57)	
R5-R1 4-factor α	-1.04*** (-7.58)	-0.67*** (-6.37)	-0.41*** (-4.96)	-0.13* (-1.85)	0.12** (2.30)	1.16*** (8.03)	

Panel A reports value-weighted returns and 4-factor alphas for lottery spreads (R5-R1) constructed using MAX within each VSE quintile. The lottery spread yields negative and significant alpha among stocks with low VSE. This suggests that lottery preferences are strongest due to investors' low visual exposure to risk. Conversely, the lottery spread generates positive and significant returns, along with risk-adjusted alpha, among stocks with high VSE. This indicates that investors in these stocks are well-aware of the risk due to heightened visual exposure, making lottery stocks even undervalued. Similar results is reported in Panel B and Panel C using ISkew and ST as lottery proxies.

2.6 Conclusion

Since the 1980s, the financial market has undoubtedly benefited from digital data visualization, enabling investors to process market information more quickly and intuitively. However, the impact of the charts' auto-scaling function has largely been overlooked by both investors and researchers. I present experimental evidence that investors' risk assessments are significantly influenced by auto-scaled digital charts. Specifically, investors' risk perceptions are distorted by the visual scaling imposed by this function. Consequently, in the cross-section of stocks, exposure to visual scaling of stocks significantly affects the volatility premium. This effect is more pronounced among stocks with a higher presence of retail investors compared to those dominated by institutional investors. I find that during the period from the 1920s to the 1980s, before graphical tools for stock market analysis had emerged, there was no evidence that visual scaling exposure affected the volatility premium. This deceptive visualization partially explains the diminishing risk-return trade-off, a relationship evident in earlier periods but increasingly difficult to observe in recent decades. I offer a simple potential explanation for this phenomenon: compared to the past, when investors relied primarily on numerical financial information, the era of digital charting makes risky stocks appear less risky to investors due to the inherent effect of the auto-scaling function.

This study yields four key implications. First, investors should be mindful of the auto-scaling effect when making investment decisions, as visual scaling can distort risk perception

and the interpretation of trading information. Second, financial service providers and data visualization developers should critically assess default charting settings and ensure that scaling choices are clearly communicated to users, thereby enhancing investor awareness. Third, findings in this paper underscore an important consideration for future research in behavioral asset pricing: the way return characteristics are visually presented can significantly shape investors' decision-making. Lastly, by studying a legacy fintech tool —digital charting, introduced in the 1980s —this study underscores the importance of anticipating the unintended consequences that may arise from the adoption of new technologies in financial markets, which are typically adopted for their convenience and efficiency.

3 Chapter Two: Rating Driven Risk Shifting by Mutual Funds, co-author with Stephen Satchell and Juan Yao

3.1 Introduction

Morningstar rates funds from 1 to 5 stars based on their past performance, providing investors with a quick and straightforward way to evaluate a fund's performance. As a result, Morningstar ratings have been identified as a key determinant of fund flows (Del Guercio and Tkac (2008), Reuter and Zitzewitz (2021), Ben-David et al. (2022a)). Mutual fund managers are incentivized by investor flows, thus star rating changes impose a non-linear payoff structure on managers.

To fix ideas, when managers are on the verge of a rating upgrade (e.g., The top performer among 4-star funds is poised for a potential upgrade to 5 stars in the near future.), consider the scenario where the manager performs well in the next few months and successfully achieves an upgrade in the near future. In this case, the fund would attract substantial inflows. However, if the manager underperforms in the next few months, the fund is unlikely to experience significant outflows, as a rating downgrade has not yet occurred. Therefore, for funds nearing a rating upgrade (hereafter top-tier funds), managers face a long call-like option position.

In contrast, when managers are close to a rating downgrade (e.g., The bottom performer among 4-star funds is poised for a potential downgrade to 3 stars in the near future.), consider the scenario where the manager performs well in the next few months, the fund may not attract inflows, as a rating upgrade has not occurred yet. However, if the manager underperforms, the fund is likely to experience a rating downgrade, resulting in substantial outflows. Therefore, for funds nearing a rating downgrade (hereafter bottom-tier funds), managers face a short put-like option position.

Under option pricing theory, we hypothesize that top-tier funds increase risk and

bottom-tier funds decrease risk²³ We find evidence that managers strategically alter portfolio risk based on their relative position to a potential rating upgrade or downgrade. After accounting for this rating-driven risk shifting, the market-wide risk shifting decreases by 21%. Building on [Huang et al. \(2011\)](#), which shows that time-varying risk shifting generally harms performance, this paper highlights the unintended agency issues arising from the widespread use of discrete performance metrics by financial service companies.

Following [Shu et al. \(2012\)](#) and [Qiu \(2003\)](#), we use total risk to test our hypotheses. Under the new methodology of Morningstar, we identify top-tier funds and bottom-tier funds based on past performance within each star rating for each Morningstar category, we find that top-tier funds increase total risk and bottom-tier funds decrease total risk. We use the methodology reform of Morningstar rating in June, 2002 as a natural experiment to identify the causality of uncovered rating-driven risk-shifting. Before the reform, Morningstar rated all equity funds together regardless of their investment style, which we refer to as the old methodology. Since then, Morningstar introduced the new methodology and rated equity funds within their peer groups (3×3 size-value Morningstar category). If the rating-driven risk-shifting observed among top- and bottom-tier funds in the post-reform period were driven by omitted factors, we would expect to find similar patterns in the pre-reform period. However, we do not find evidence for such patterns prior to the reform.

We also show that rating-driven risk-shifting behavior is more pronounced among small funds and funds with high expense ratio. Morningstar overall rating is calculated as the weighted average of 3-year, 5-year, and 10-year ratings. If the weighted average is a decimal number, the decimal rating is rounded up or down to the nearest integer. We show that the rating-driven risk-shifting behavior is less (more) pronounced for rounded-up (rounded-down) funds. We find managers also engage in risk-shifting to manage the 3-year and 5-year ratings, but with a much smaller magnitude compared to the overall rating, as expected,

²³Similar to our context, option pricing theory is also used to analyze performance-based fee contracts ([Grinblatt and Titman \(1989\)](#), [Massa and Patgiri \(2009\)](#)).

since the overall rating is much more visible to investors. Managers do not strategically engage in risk-shifting for 10-year rating as 10-year performance is less likely to be affected by short term risk-shifting.

We further document that investors rely more on systematic risk than on idiosyncratic risk to manage ratings. Top-tier funds increase systematic risk, while bottom-tier funds decrease systematic risk. In contrast, bottom-tier funds decrease idiosyncratic risk, but we do not find evidence that top-tier funds increase idiosyncratic risk. Managers' more common use of systematic risk than idiosyncratic risk to manage ratings might be attributed to the fact that systematic risk has less negative impact on fund performance compared to idiosyncratic risk (Huang et al. (2011)).

To further understand the specific risk-shifting behaviors contributing to the rating-driven risk-shifting, we examine the characteristics of fund holdings (Kacperczyk et al. (2005)), turnover (Grinblatt and Keloharju (2009)) and trading aggressiveness (Kacperczyk et al. (2008), Huang et al. (2011)). With respect to holding characteristics, we observe two main findings. First, top-tier funds tend to increase the holdings' beta, providing evidence of managers engaging in rating-driven risk-shifting through systematic risk. Second, bottom-tier funds decrease industry concentration, which is consistent with prior findings that bottom-tier funds decrease idiosyncratic risk.²⁴ Regarding to trading activities, top-tier (bottom-tier) funds tend to increase (decrease) turnover and interim trading aggressiveness.

How did managers manage ratings under the old rating methodology before the reform? We hypothesize that, if all equity funds are ranked together regardless of their investment styles, managers could either stick to their own investment styles and manage ratings by strategically altering risk, or they could strategically adjust investment style to manage their ratings (hereafter rating-driven style-shifting). We find evidence of the later. In particular, top-tier funds aggressively expose to styles other than the original style to gamble on potential

²⁴Huang et al. (2011) argue that exposure to idiosyncratic risk can be achieved by concentrating holdings on a few positions or industries.

upside. Besides, we also find that top-tier funds unintentionally decrease total risk due to style diversification.

Prospect theory ([Kahneman and Tversky \(2013\)](#)) offers a potential explanation for the uncovered rating-driven risk(style)-shifting. Managers may use the previous rating as a reference point and exhibit different risk preferences depending on whether the current rating is above or below that reference point. It is possible that top-tier funds are more likely to be funds that have recently experienced rating downgrades. Managers of these funds mentally anchor their previous high rating as a reference point, experiencing a sense of loss, and consequently increase risk (or shifting style). Conversely, bottom-tier funds are more likely to be funds that have recently received rating upgrades. Managers of these funds mentally anchor their previous low rating as a reference point, feeling a sense of gain, and consequently decrease risk (or concentrating style).

To confirm that rating-driven risk(style)-shifting is a strategic behavior by managers rather than a result of reference-dependent risk preference, we implement two approaches to control for rating changes, which may influence managers' perceptions of gain or loss. Both approaches suggest that the observed rating-driven risk(style)-shifting cannot be solely explained by prospect theory.

[Huang et al. \(2011\)](#) suggest agency problem is the main cause of the poor performance of risk-takers. After the reform, if the increasing risk by top-tier funds is solely driven by the agency problem, we would anticipate that top-tier funds earn negative abnormal returns as they engage in more risk-taking. If the option-like payoff faced by top-tier funds also serves as an incentive, motivating more managerial effort, we would expect top-tier funds to earn positive abnormal returns after controlling for the agency problem. Consistent with [Huang et al. \(2011\)](#), we find risk-takers earn negative abnormal return. However, after controlling for the poor performance of risk-takers, we find that top-tier funds earn positive abnormal returns, suggesting that the rating-induced option-like payoff can serve as an incentive to enhance managerial efforts. Before the reform, even though top-tier funds engage in more

style-shifting to other styles, they do not earn abnormal returns under the influence of rating-induced incentive, as their focus is on exposure to systematic risk premiums instead of seeking risk-adjusted alpha (Sha (2020)).

This paper contributes to the literature on star ratings. While Reuter and Zitzewitz (2021) and Ben-David et al. (2022a) document the influence of ratings on investor flows, other studies have examined the unintended side effects of ratings. For example, Ben-David et al. (2022b) show that rating-driven demand can lead to systematic price fluctuations, and Evans and Sun (2021) find that retail investors exhibit more sophisticated risk adjustments following improvements in Morningstar’s rating methodology. This paper adds to the discussion by highlighting a potential agency issue arising from the discrete nature of star ratings from manager side.²⁵

This paper is also related to the literature on mutual fund risk-shifting. Huang et al. (2011) provides strong evidence that time-varying risk-shifting harms fund performance. While a substantial body of work explores factors influencing risk-taking behavior²⁶, we argue that rating-driven risk-shifting has a broader market impact (21% of the market-wide risk-shifting) due to the widespread use of star ratings in the mutual fund industry. Although the generally convex shape of the flow-performance relationship may also incentivize managers to take on more risk (Chevalier and Ellison (1997)) - potentially contributing significantly to market-wide risk-shifting, Spiegel and Zhang (2013) challenge this view by documenting that the relationship is largely linear.

The remainder of the paper is organized as follows: Section 2 discusses the background

²⁵We are aware of two recent unpublished studies that also examine strategic rating manipulation. One is Cohen et al. (2024), which shows that managers strategically adjust their holdings to reclassify fund style and obtain higher ratings. The other is Kim (2022), which finds that managers inflate their month-end portfolios to manipulate ratings. Together with our study, these works highlight three independent approaches through which managers can engage in rating manipulation.

²⁶Brown et al. (1996) find that mid-year losing funds tend to increase portfolio volatility more than mid-year winners in the second half of the year. Ma and Tang (2019) show that managerial ownership can mitigate risk-taking, particularly among managers facing high agency frictions. Beyond agency problems, Massa and Patgiri (2009) demonstrate that managers with stronger incentives engage in more risk-taking and achieve higher risk-adjusted returns.

of the rating and the methodology reform. Section 3 describes the data and key variables. Section 4 presents the empirical results of rating-driven risk-shifting under the new methodology. Section 5 investigates the rating-driven style-shifting under the old methodology. Section 6 provides further analysis. Finally, Section 7 concludes.

3.2 Star Rating and Methodology Reform

The primary goal of Morningstar rating, introduced in 1985, is to help investors compare funds and choose one or a few among others. Morningstar rates each share class of a portfolio separately because each share class has different fees and return time periods available. Investments must have at least 36 continuous months of returns in order to receive a rating. At the end of each month, Morningstar assigns percentile rankings to mutual funds based on their Morningstar risk-adjusted return (MRAR) for different time periods (3, 5, and 10 years). Based on these percentile rankings, the top 10% of funds in terms of MRAR receive 5 stars; the next 22.5% receive 4 stars; the next 35% receive 3 stars; the next 22.5% receive 2 stars; and the bottom 10% receive 1 star. The overall rating for funds is then calculated as the weighted average of the 3-, 5-, and 10-year ratings, rounded to the nearest integer. For funds with only a 3-year rating, the overall rating is the 3-year rating. For funds with both 3-year and 5-year ratings, the overall rating is the sum of 60% of the 5-year rating and 40% of the 3-year rating. For funds with 3-, 5-, and 10-year ratings, the overall rating is the sum of 50% of the 10-year rating, 30% of the 5-year rating, and 20% of the 3-year rating.

On 2002 June 30, Morningstar introduced a methodology change (hereafter the reform) on the rating procedure. Before the reform, funds percentile ranking for 3-, 5-, and 10-year ratings were derived from ranking all US equity funds together based on their MRAR. After the reform, funds percentile ranking was derived from ranking US equity funds within their Morningstar category based on their MRAR, where Morningstar category is determined based on three years of portfolio statistics. Besides, the formula for MRAR is also modified to adjust for investor risk aversion under expected utility theory after the reform.²⁷

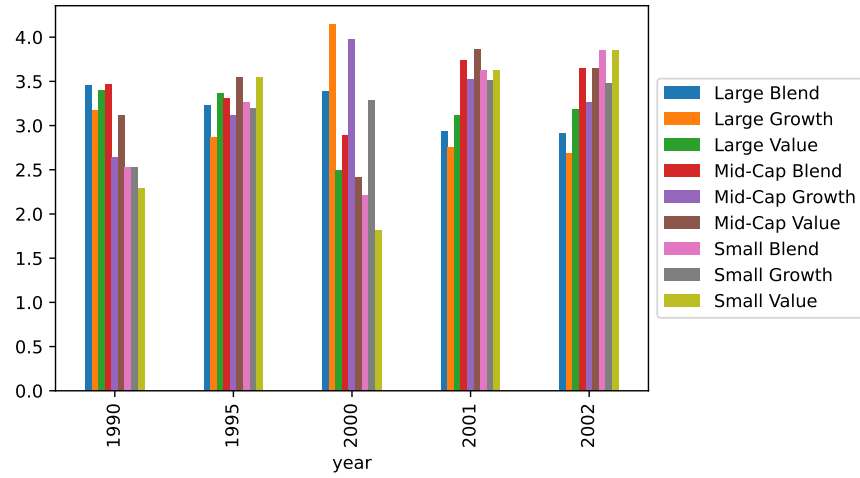
²⁷Before the reform, a fund was deemed to be riskless if the fund's return exceeded 90-day Treasury bill

Before the reform, ratings were determined without considering the investment style, resulting in a strong exposure to style. Consequently, funds with a particular Style A, tended to receive higher ratings than those with Style B, if Style A outperformed Style B. However, after the reform, when funds are rated within their investment style, this style-driven bias disappears, as illustrated in Figure 6. For example, in Panel A, mid-cap and value funds were systematically overrated in 2001, whereas funds in the large-cap and growth styles were underrated. In Panel B, the average rating for all styles is approximately the same after the reform.

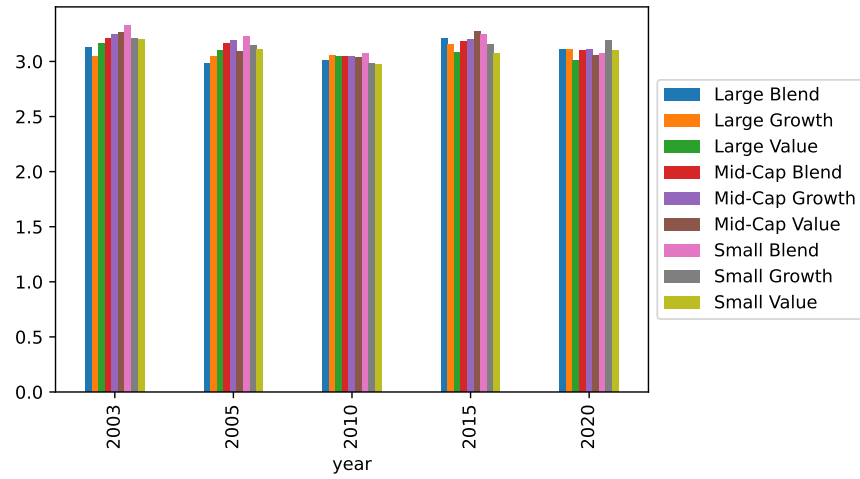
each month. See [Blume \(1998\)](#) for detailed calculation of MRAR before the reform.

Figure 6: Rating and Style

This figure illustrates, for specific years, the average Morningstar overall rating received by each of the 3x3 Morningstar categories. Figure (a) displays the years 1990, 1995, 2000, 2001, and 2002. Figure (b) displays the years 2003, 2005, 2010, 2015, and 2020.



(a) Before Reform



(b) After Reform

3.3 Data and Key Variables

3.3.1 Data

Our first source of data is Morningstar Direct, where we obtain fund ratings (3-, 5-, 10-year ratings and overall ratings), fund net-of-fee returns, gross expense ratios, turnover ratios, fund flows, total net assets (TNA), excess returns over Morningstar category index, and Morningstar risk-adjusted returns. Our sample consists of actively managed U.S. domestic equity funds within the 3×3 Morningstar category. The 3×3 Morningstar category includes Large Value, Large Blend, Large Growth, Mid-Cap Value, Mid-Cap Blend, Mid-Cap Growth, Small Value, Small Blend, and Small Growth. Following [Elton et al. \(2001\)](#), [Chen et al. \(2004\)](#), and [Pástor et al. \(2015\)](#), we exclude funds with less than \$15 million in TNA. Since multiple share classes are listed separately in the Morningstar Direct database, we aggregate the share class-level data to the fund portfolio level following [Ma and Tang \(2019\)](#) and [Barber et al. \(2016\)](#). We calculate the fund’s TNA as the sum of assets across all share classes and compute the value-weighted average of other fund characteristics across share classes. The fund-level rating is determined as the median rating of all share classes under the fund.

Second, we obtain quarterly equity holdings of U.S. open-end mutual funds from Thomson Reuters Mutual Fund Holdings (TRMFH) database. To merge the Morningstar Direct and TRMFH databases, we use fund tickers and fund CUSIP numbers whenever these are available following [Wermers \(2000\)](#). Our final sample consists of 2,609 unique funds from 1990 to 2022.

3.3.2 Key Variables

The key dependent variables are risk-shifting and style-shifting measures. Total risk has been used for many mutual fund risk-shifting literature ([Shu et al. \(2012\)](#), [Qiu \(2003\)](#)) and many previous papers²⁸ analyze risk shifting by comparing the standard deviations of the returns of mutual funds over two non-overlapping time periods. Therefore, we use the

²⁸See, for example [Brown et al. \(1996\)](#), [Koski and Pontiff \(1999\)](#), [Busse \(2001\)](#), [Elton et al. \(2003\)](#)

logarithmic change of total risk (ΔStd) estimated with previous and future 12-month monthly gross return as the proxy for managers' risk-shifting behavior. For style-shifting measures, we use the logarithmic change of R2 (ΔR2), which is estimated by regressing previous and future 12-month monthly gross returns against the Morningstar category index. A high R2 suggests that the fund's performance can be accurately explained by the style of its previous holdings, while a low R2 implies that the fund's performance is not explained by the style of its previous holdings. Therefore, a positive ΔR2 indicates style concentration, whereas a negative ΔR2 implies style shifting. We discuss the construction of all risk(style)-shifting measures in Appendix 2A.

The explanatory variables are indicator variables for top-tier funds (Top) and bottom-tier funds (Bottom). First, we construct the rank variable to identify the funds' relative position for receiving an upgrade or downgrade of a specific rating under a specific methodology. Under the new methodology, funds are rated within their Morningstar category peers based on their MRAR. The relative position for receiving a 3-year (5-year, 10-year) rating change, denoted as $\text{Rank}_{new,3}$ ($\text{Rank}_{new,5}$, $\text{Rank}_{new,10}$), is determined by sorting funds within each 3-year (5-year, 10-year) rating into quintiles based on their 3-year (5-year, 10-year) MRAR for each Morningstar categories. The overall rating is generated from the weighted average of the 3-, 5-, and 10-year ratings, instead of ranking based on past performance. Therefore, we can only use past performance to approximate the funds' relative position for receiving an overall rating change. We choose 3-year MRAR²⁹ as a proxy for past performance because the most recent 3-year period is included in all three rating periods, giving it the greatest impact.³⁰ The funds' relative position for receiving an overall rating change, labeled as Rank_{new} , is approximated by sorting funds within each overall rating into quintiles

²⁹Sharpe (1998) suggests that the resulting ranking is similar to that based on the Sharpe ratio. We do not use the Sharpe ratio due to endogeneity issues that arise when change in risk is independent variable. For instance, a top-tier fund with a high Sharpe ratio mechanically has low volatility, leading to a mechanical increase in risk in the subsequent period.

³⁰See page 15 of The Morningstar RatingTM for Funds | August 2021

based on their 3-year MRAR for each Morningstar category. Under the old methodology, funds are rated together based on their MRAR. Therefore, the funds' relative position for receiving an overall rating change, labeled as Rank_{old} , is approximated by sorting funds within each overall rating together into quintiles based on their 3-year MRAR. Appendix 2B provides a detailed explanation of the construction of the rank variables.

Secondly, we create the variable 'Top' for top-tier funds, which takes a value of 1 for funds with a rank equal to 5, and 0 otherwise. Similarly, we construct the variable 'Bottom' for bottom-tier funds, which takes a value of 1 for funds with a rank equal to 1, and 0 otherwise. For instance, Top_{new} identifies the funds closest to an overall rating upgrade under the new methodology, corresponding to funds with Rank_{new} equal to 5. Bottom_{old} represents the funds closest to an overall rating downgrade under the old methodology, denoted by funds with Rank_{old} equal to 1.

Table 12 presents the average characteristics of top-tier and bottom-tier funds. In the post-reform period, top-tier funds increase by an average of 0.079 stars in the following month, while bottom-tier funds decrease by 0.086 stars in the subsequent month. Moreover, 10.8% of top-tier funds receive a rating upgrade, with only 2.8% experiencing a downgrade in the following month. Conversely, 11.5% of bottom-tier funds undergo a rating downgrade, while merely 2.9% receive an upgrade in the subsequent month. These statistics underscore the effectiveness of the Top_{new} and Bottom_{new} variables in capturing funds' relative positions of receiving rating upgrades and downgrades.³¹ Top-tier funds exhibit better performance over the past 1 year and 3 years compared to bottom-tier funds. Top-tier funds receive higher flows than bottom-tier funds over the past 1 year. Top-tier funds are smaller in size compared to bottom-tier funds. We do not observe any significant differences in terms of

³¹Sharpe (1998) has suggested that the resulting ranking based on MRAR is akin to the ranking derived from the Sharpe ratio. In an untabulated test, we identified top-tier funds and bottom-tier funds based on the rank derived from the 3-year Sharpe ratio. The top-tier funds and bottom tier funds are also able to predict rating upgrades and downgrades, but with less predictive power compared to the top-tier and bottom-tier funds identified based on 3-year MRAR.

gross expense ratio and turnover between top-tier and bottom-tier funds. Both top-tier and bottom-tier funds have the same average overall rating, as they are identified within each rating. The statistics of top-tier funds and bottom-tier funds are similar in the pre-reform period.

Table 12: Summary Statistics for Top-tier and Bottom-tier Funds

This table presents the time-series average of characteristics for top-tier funds and bottom-tier funds. Top_{new} ($Bottom_{new}$) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $Rank_{new} = 5$ (=1). Top_{old} ($Bottom_{old}$) is an indicator variable for top-tier funds (bottom-tier funds) under the old methodology, which takes a value of 1 if $Rank_{old} = 5$ (=1). $\Delta rating_{t,t+1}$ is funds' rating change from month t to month $t+1$. Pos (Neg) $\Delta rating_{t,t+1}$ equals 1 if $\Delta rating_{t,t+1}$ is positive (negative) and 0 otherwise. Fund Size is total net asset (TNA) calculated as the sum of assets across all share classes. Fund $Flow_{t-11,t}$ is sum of monthly fund flow from past 12 month, where monthly fund flow is the aggregation of all share class flow then scaled by TNA. We also report funds gross return from past 1 year and past 3 years. Gross Expense ratio and Turnover ratio is annualized. The t-statistics in parentheses are Newey-West adjusted, with significance levels indicated by *, **, and ***, representing 10%, 5%, and 1%, respectively.

	Reform-2022			1990-Reform		
	$Bottom_{new}$	Top_{new}	Top_{new} - $Bottom_{new}$	$Bottom_{old}$	Top_{old}	Top_{old} - $Bottom_{old}$
$\Delta rating_{t,t+1}$	-0.086	0.079	0.165*** (18.817)	-0.09	0.091	0.181*** (11.200)
$Pos\Delta rating_{t,t+1}$	0.028	0.108	0.080*** (13.494)	0.029	0.129	0.100*** (21.232)
$Neg\Delta rating_{t,t+1}$	0.115	0.029	-0.085*** (-23.290)	0.119	0.038	-0.081*** (-6.565)
Gross Return $_{t-11,t}$ (%)	8.882	13.314	4.432*** (7.406)	10.688	22.024	11.198*** (4.990)
Gross Return $_{t-35,t}$ (%)	23.838	38.679	14.841*** (14.402)	41.753	65.111	24.469*** (16.953)
Fund Size (Bil)	2.165	1.627	-5.378*** (-3.609)	2.205	1.027	-1.188*** (-4.311)
Fund Flow $_{t-11,t}$ (%)	-0.147	0.021	0.168*** (13.045)	-0.021	0.184	0.206*** (7.435)
Gross Expense $_t$ (%)	1.154	1.208	0.055 (1.432)	1.175	1.247	0.008 (0.115)
Turnover $_t$ (%)	67.573	66.493	-1.080 (-0.391)	76.811	80.941	3.418 (1.455)
Overall Rating $_t$	3.186	3.189	0.003 (0.453)	3.22	3.226	0.006 (0.960)

3.4 Rating-driven Risk-shifting under New Methodology

3.4.1 Rating-driven Risk-shifting and Methodology Reform

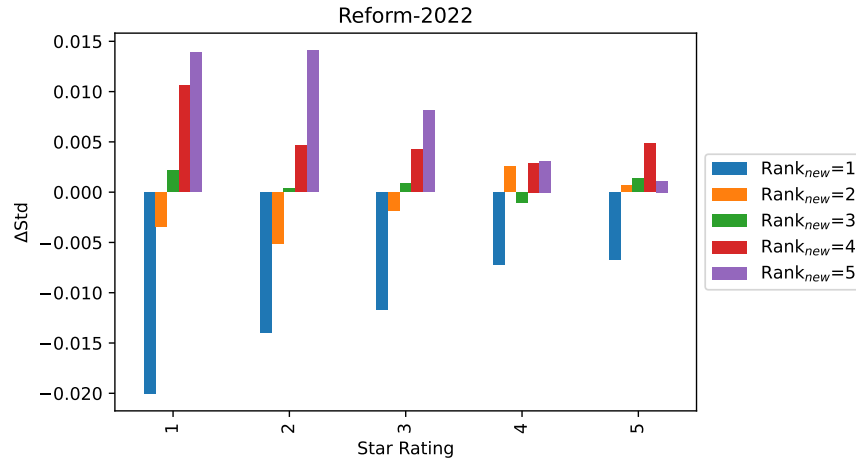
We investigate the relationship between risk-shifting behavior and a fund's relative position to a potential rating upgrade or downgrade using portfolio sorting. We calculate the time-series averages of ΔStd for funds with Rank_{new} ranging from 1 (bottom-tier) to 5 (top-tier), separately for each of the five Morningstar star ratings. Since the estimation of ΔStd involves two return windows, it may be systematically influenced by style³² and past performance.³³ To mitigate this, each month, we demean ΔStd within each Morningstar category and rating category. Figure 7 presents the result. In the post-reform period (Figure 7, Panel A), we observe that top-tier funds ($\text{Rank}_{new} = 5$) exhibit positive ΔStd for 1- to 4-star funds, consistent with our central hypothesis that these funds take on more risk when approaching a rating upgrade. Interestingly, within the 5-star group, the top-tier funds do not appear to increase risk, while the second-top-tier funds do. Given that [Chevalier and Ellison \(1997\)](#) and [Sirri and Tufano \(1998\)](#) find that a disproportionate amount of investor flow volume is directed toward the top-performing funds each year, this pattern is consistent with a tournament effect, as documented in the literature (See [Schwarz \(2012\)](#), [Kempf and Ruenzi \(2008\)](#)). In contrast, the ΔStd of bottom-tier funds ($\text{Rank}_{new} = 1$) is consistently negative across all rating levels, indicating that these funds reduce risk, likely in response to the fear of a downgrade.

³²The risk-shifting observed for a fund might be subject to the change of risk due to its style.

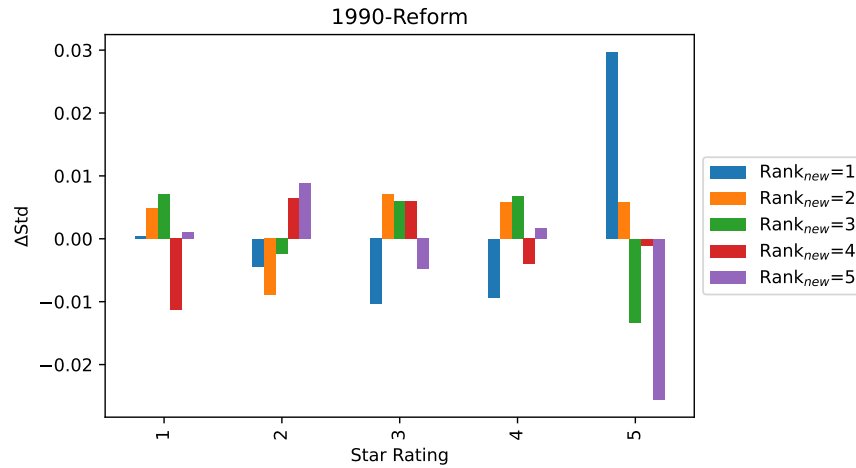
³³For example, previous studies expect or find a relationship between risk and return. (See [Markowitz \(1952\)](#) and [Sharpe \(1966\)](#).)

Figure 7: Rank and Risk-shifting

This figure illustrates the times-series average of funds' risk-shifting behaviour, as measured by ΔStd , across $Rank_{new}$ (1 to 5) within each rating (1 Star to 5 Star). The $Rank_{new}$ represents quintile sorting of equity funds based on their 3-year Morningstar risk-adjusted return (MRAR) from 1 (bottom-tier funds) to 5 (top-tier funds) within each rating for each 3x3 Morningstar category. ΔStd represents the log difference of the standard deviation of the next 12-month gross return and the past 12-month gross return. Every month, we demean ΔStd within each Morningstar category and rating. Panel (a) illustrates the sample after the methodology reform, while Panel (b) shows the sample before the methodology reform. Rating is the overall rating.



(a) After Reform



(b) Before Reform

However, Table 12 shows that top-tier funds have better past performance, experience

more inflows, and are smaller than bottom-tier funds under the new methodology. These factors individually, or interact in complex ways to influence risk-shifting behaviors.³⁴ If the uncovered rating-driven risk-shifting by top-tier funds and bottom-tier funds identified under the new methodology is driven by omitted factors, we would observe the top-tier funds increase risk and bottom-tier funds decrease risk in pre-reform period. In Panel B, for the period before the reform, the rating-driven risk-shifting is not found. Thus, we conclude that the uncovered rating-driven risk-shifting in the post-reform period is a strategic risk-shifting behavior to manage rating under the new methodology instead of a result of omitted factors.

Next, we use regression analyses to simultaneously control more factors and present the result in Table 13. The dependent variable is ΔStd , scaled by 100 for interpretation in percentage terms. Top_{new} (Bottom_{new}) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $\text{Rank}_{new} = 5$ ($=1$). Our control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense ratio, and overall rating with style fixed effects and month fixed effects. In column (1) and (2), the coefficients of Top_{new} are positive and significant at 1% level, while the coefficients of Bottom_{new} is negative and significant at 1% level, implying that top-tier funds increase risk by 0.372% to 0.878% and bottom-tier funds decrease risk by 1.029% to 1.269%. Column (3) and (4) further confirm the rating-driven risk-shifting primarily exists in post-reform period, implying such relationship is not caused by other omitted factors.

³⁴[Ammann and Verhofen \(2007\)](#) find prior performance positively affects risk-taking. [Brown et al. \(1996\)](#) find mid-year losers increase fund volatility than mid-year winners. [Chevalier and Ellison \(1997\)](#) find flow-performance relationship affect funds risk-taking. Small funds are subject more to agency issue which lead to more risk-taking. ([Chevalier and Ellison \(1997\)](#), [Ma and Tang \(2019\)](#), [Huang et al. \(2007\)](#))

Table 13: Rating-driven Risk-shifting

This table examines the rating-driven risk-shifting behavior by top-tier funds and bottom-tier funds in post-reform period. Top_{new} ($Bottom_{new}$) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $Rank_{new} = 5$ (=1). The $Rank_{new}$ represents quintile sorting of equity funds based on their 3-year Morningstar risk-adjusted return (MRAR) from 1 to 5 within each overall rating for each 3×3 Morningstar category. ‘Post’ takes the value of 1 if the observation is after the reform, 0 otherwise. Control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense, and overall rating with style fixed effects and month fixed effects. Standard errors are adjusted for heteroskedasticity and clustered by style and month. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Dep.Var(%) = ΔStd	Reform-2022		1990-2022	
	(1)	(2)	(3)	(4)
Intercept	0.077 (0.003)	-3.385*** (-2.612)	0.059 (0.003)	-2.806*** (-2.315)
Top_{new}	0.878*** (4.513)	0.372*** (2.675)	0.180 (0.516)	-0.572 (-1.218)
$Bottom_{new}$	-1.269*** (-5.063)	-1.029*** (-5.384)	0.471 (0.752)	0.367 (1.294)
$Top_{new} \times Post$			0.921** (1.994)	0.949** (2.055)
$Bottom_{new} \times Post$			-1.828** (-2.822)	-1.531** (-3.764)
Controls	No	YES	No	YES
Style Fixed Effect	YES	YES	YES	YES
Month Fixed Effect	YES	YES	YES	YES
adj.R squared	0.1%	1.4%	0.1%	1.3%

3.4.2 Rating-driven and Market-wide Risk-shifting

We investigate the extent to which rating-driven risk-shifting contributes to market-wide risk-shifting, which has been shown to harm performance and investor welfare (Huang et al. (2011)). We begin by sorting funds based on their change in volatility (ΔStd) within each Morningstar category and star rating. This within-group sorting helps control for systematic risk-shifting driven by fund style and past performance. Columns (1) to (3) of Table 14 report the time-series averages of ΔStd , as well as the percentages of funds facing rating upgrade (Top) and funds facing rating downgrade (Bottom). As shown in column (1), the average

risk-increasing fund raises its volatility by 0.192 (19.2%), while the average risk-reducing fund lowers its volatility by 0.191 (19.1%), implying a total spread in risk-shifting of 38.3%. Column (2) shows that risk-increasing funds contain 3.135% more managers positioned for a potential rating upgrade (Top) compared to risk-reducing funds. Conversely, column (3) shows that risk-reducing funds include 4.644% more managers facing a potential downgrade (Bottom) than their risk-increasing counterparts.

Table 14: Rating-driven and Market-wide Risk-shifting

This table presents the time-series averages of fund characteristics across risk-shifting quintiles (ΔStd quintiles) after the 2002 rating reform. The ΔStd quintiles sort funds into five groups based on ΔStd , ranging from 1 (funds that reduce risk) to 5 (funds that increase risk). Columns (1) to (3) report average characteristics of the ΔStd quintiles sorted within each Morningstar categories and star ratings, while columns (4) to (6) present the average characteristics of the ΔStd quintiles sorted within each Morningstar categories, star ratings, and Rank_{new} . The Rank_{new} represents quintile sorting of equity funds based on their 3-year Morningstar risk-adjusted return (MRAR) from 1 (bottom-tier funds) to 5 (top-tier funds) within each rating for each 3 \times 3 Morningstar category. Top_{new} (Bottom_{new}) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $\text{Rank}_{new} = 5$ (=1). The t-statistics in parentheses are Newey-West adjusted, with significance levels indicated by *, **, and ***, representing 10%, 5%, and 1%, respectively.

ΔStd quintiles	Within [Category, Ratings]			Within [Category, Ratings, Rank_{new}]		
	ΔStd (1)	Top_{new} (%) (2)	Bottom_{new} (%) (3)	ΔStd (4)	Top_{new} (%) (5)	Bottom_{new} (%) (6)
1	-0.191	21.254	25.804	-0.140	20.936	21.083
2	-0.023	19.004	20.159	-0.022	20.913	21.280
3	0.005	18.679	18.049	0.009	20.896	21.012
4	0.083	20.157	18.697	0.083	20.913	21.280
5	0.192	24.389	21.159	0.163	20.907	20.972
5-1	0.383*** (87.766)	3.135*** (4.982)	-4.644*** (-7.787)	0.303*** (78.608)	-0.029 (-0.774)	-0.111* (-1.624)

Next, we incorporate Rank_{new} into the within-group sort to ensure that the resulting risk-shifting quintiles are neutral with respect to the likelihood of a rating upgrade or downgrade. Columns (4) to (6) present the results of sorting funds into risk-shifting quintiles within Morningstar category, star rating, and Rank_{new} . Recall that Rank_{new} (ranging from 1 to 5) reflects a fund's relative position within its rating group under the new Morningstar

methodology. Funds with a Rank_{new} of 1 are those most likely to receive a rating downgrade, while those with a Rank_{new} of 5 are most likely to receive a rating upgrade. As shown in column (4), the average risk-increasing fund raises its volatility by 0.163 (16.3%), while the average risk-reducing fund lowers its volatility by 0.140 (14.0%), resulting in a total spread of 30.3%. This back-of-the-envelope calculation suggests that rating-driven risk-shifting leads to a 26.4% ($38.3\%/30.3\% - 1$) increase in market-wide risk-shifting,³⁵ or the rating-driven risk-shifting accounts for 20.9% ($30.3\%/38.3\% - 1$) of the market-wide risk-shifting observed in mutual fund industry. Moreover, after controlling for the rating effect, the distribution of funds facing potential upgrades (Top) and downgrades (Bottom) becomes balanced across ΔStd quintiles, as shown in columns (5) and (6), confirming that the risk-shifting quintiles are no longer driven by rating incentives.

3.4.3 Cross-sectional Variation and 3Y, 5Y, 10Y Rating

In this section, We investigate the cross-sectional variation of rating-driven risk-shifting. Spiegel and Zhang (2013) argue that the flow of smaller funds react more sensitively to performance, and funds with higher fees can charge a larger portion of incremental flow driven by rating upgrades. Therefore, we hypothesize that funds with smaller sizes and higher fees have more incentive to strategically managing ratings. We define funds with bottom quintile of TNA as small funds and use fund gross expense ratio as a proxy for fund fee³⁶ following Gil-Bazo and Ruiz-Verdú (2009) and define funds with top quintile of expense ratio as funds with high fee. The result is presented in Table 15. In column (1), the coefficient of $\text{Top}_{new} \times \text{Small}$ is positive and significant (1.252, t-statistic=2.441), suggesting that the top-tier small funds takes more risk than top-tier big funds. In column (2), the coefficient of $\text{Top}_{new} \times \text{High}$ is positive and significant (0.844, t-statistic=1.764), suggesting that the top-tier funds with high expense ratio takes more risk than top-tier funds with low

³⁵The increase in the spread is statistically significant at the 0.01% level.

³⁶Fund management fee, together with other operating costs, such as custodian, administration, accounting, registration, and transfer agent fees—comprise the fund’s expenses, which is typically computed as a fixed percentage of the value of assets under management.

expense ratio.

Table 15: Cross-sectional Variation

This table examines the cross-sectional variation of rating-driven risk-shifting from reform-2022. Top_{new} ($Bottom_{new}$) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $Rank_{new} = 5$ ($=1$). The $Rank_{new}$ represents quintile sorting of equity funds based on their 3-year Morningstar risk-adjusted return (MRAR) from 1 to 5 within each overall rating for each 3×3 Morningstar category. ‘Small’ (‘High’) is a dummy variable that takes the value of 1 if the fund is in the bottom (top) quintile of TNA (fund gross expense) each month and 0 otherwise. ‘RoundUp’ (‘RoundDown’) takes the value of 1 if the overall rating minus the precise rating calculated by the weighted average of the fund’s 3, 5, and 10 ratings is positive (negative), and 0 otherwise. Control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense, and overall rating with style fixed effects and month fixed effects. Standard errors are adjusted for heteroskedasticity and clustered by style and month. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Dep.Var(%)= Δ Std	(1)	(2)	(3)	(4)
Intercept	-4.215*** (-2.942)	-3.689*** (-3.006)	-3.342*** (-2.630)	-3.353 (-2.565)
Top_{new}	0.187* (1.931)	0.306** (2.470)	0.510*** (3.296)	0.400** (2.351)
$Bottom_{new}$	-1.054*** (-4.482)	-0.966*** (-4.833)	-1.137** (-2.304)	-1.121*** (-6.464)
$Top_{new} \times Small$	1.252** (2.441)			
$Bottom_{new} \times Small$	0.152 (0.214)			
$Top_{new} \times High$		0.844* (1.764)		
$Bottom_{new} \times High$		-1.039 (-0.968)		
$Top_{new} \times RoundUp$			-0.564** (-2.369)	
$Bottom_{new} \times RoundUp$			-0.004 (-0.007)	
$Top_{new} \times RoundDown$				0.652** (2.130)
$Bottom_{new} \times RoundDown$				0.595 (1.092)
Control Interaction	YES	YES	YES	YES
Controls	YES	YES	YES	YES
Style Fixed Effect	YES	YES	YES	YES
Month Fixed Effect	YES	YES	YES	YES
adj.R squared	1.4%	1.4%	1.4%	1.4%

Morningstar overall ratings are calculated as the weighted average of 3-year, 5-year, and 10-year ratings.³⁷ If the weighted average is a decimal number, the decimal rating is rounded up or down to the nearest integer to become the overall rating. We hypothesize that the overall ratings of rounded-up funds are susceptible to any downgrade in the 3, 5, or 10-year ratings, so managers of rounded-up funds need to be particularly mindful of the potential declines in short-term performance. In contrast, the rounded-down funds' overall rating are sensitive to any upgrade in the 3, 5, or 10-year rating, so managers of rounded-down funds have stronger exposure to potential increase in short-term performance than potential decrease. In column (3), the coefficient of $\text{Top}_{new} \times \text{RoundUp}$ is negative and significant (-0.564, t-statistic=-2.369), suggesting that the top-tier funds with rounded-up rating takes less risk than other top-tier funds. In column (4), the coefficient of $\text{Top}_{new} \times \text{RoundDown}$ is positive and significant (0.652, t-statistic=2.130), suggesting that the top-tier funds with rounded-down rating takes more risk than other top-tier funds.

We examine the rating-driven risk-shifting among the 3, 5, and 10-year ratings in Table 16. For each of the 3, 5, and 10-year ratings, similar to the overall rating, we first construct the rank variables ($\text{Rank}_{new,3}$, $\text{Rank}_{new,5}$, $\text{Rank}_{new,10}$) to determine the relative position of funds for receiving rating upgrades or downgrades. Subsequently, we identify top-tier funds ($\text{Top}_{new,3}$, $\text{Top}_{new,5}$, $\text{Top}_{new,10}$) and bottom-tier funds ($\text{Bottom}_{new,3}$, $\text{Bottom}_{new,5}$, $\text{Bottom}_{new,10}$).

³⁷For funds with only a 3-year rating, the overall rating is the 3-year rating. For funds with both 3-year and 5-year ratings, the overall rating is the sum of 60% of the 5-year rating and 40% of the 3-year rating. For funds with 3-, 5-, and 10-year ratings, the overall rating is the sum of 50% of the 10-year rating, 30% of the 5-year rating, and 20% of the 3-year rating.

Table 16: Risk-shifting in Response to 3, 5, and 10-Year Ratings

This table presents the risk-shifting behavior in response to the 3-year, 5-year, and 10-year ratings from reform-2002. $Top_{new,n}$ ($Bottom_{new,n}$) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology for n-year rating, which takes a value of 1 if $Rank_{new,n} = 5$ (=1). The $Rank_{new,n}$ represents quintile sorting of equity funds based on their n-year Morningstar risk-adjusted return (MRAR) from 1 to 5 within each n-year rating for each 3×3 Morningstar category. Control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense, and overall rating with style fixed effects and month fixed effects. Standard errors are adjusted for heteroskedasticity and clustered by style and month. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Dep.Var(%) = ΔStd	(1)	(2)	(3)
Intercept	-2.984** (-2.265)	-3.068** (-2.331)	-3.188** (-2.387)
$Top_{new,3}$	0.284** (2.121)		
$Bottom_{new,3}$	-0.504*** (-4.447)		
$Top_{new,5}$		0.270* (1.772)	
$Bottom_{new,5}$		-0.256** (-2.171)	
$Top_{new,10}$			0.087 (0.645)
$Bottom_{new,10}$			0.001 (0.010)
Controls	YES	YES	YES
Style Fixed Effect	YES	YES	YES
Month Fixed Effect	YES	YES	YES
adj.R squared	1.4%	1.4%	1.3%

In Table 16, by comparing the coefficients of $Top_{new,3}$ with $Top_{new,5}$ and $Top_{new,10}$, and the coefficients of $Bottom_{new,3}$ with $Bottom_{new,5}$ and $Bottom_{new,10}$, we observe that the rating-driven risk-shifting behavior by managers is more pronounced in response to the 3-year rating compared to the 5-year and 10-year ratings. We attribute this to the fact that the 5-year and 10-year ratings are less influenced by short-term performance than the 3-year rating. The risk-shifting driven by the 3-year rating is less prominent compared to the

risk-shifting driven by overall rating as reported in column (1) of Table 13. This could be attributed to the 3-year rating being less visible to investors than the overall rating, resulting in managers being less inclined to strategically manage the 3-year rating by adjusting risk levels.

3.4.4 Market Volatility and Unemployment Risk

We also investigate the impact of market volatility and unemployment risk on managers' rating-driven risk-shifting behavior in Appendix 2C. Each month, we estimate the standard deviation of market returns of past 12 months and around 24 months (from month -12 to month +12) as two proxies for market volatility. We define the period of low volatility (Low) as the month falling in the bottom quintile of market volatility across the full sample. In column (1) and (2), the coefficient of $\text{Top}_{new} \times \text{Low}_{past\ 12m}$ and $\text{Top}_{new} \times \text{Low}_{around\ 24m}$ is negative and significant. Hence, the rating-driven risk-shifting behavior, especially the risk-taking behavior by top-tier funds, is less pronounced or even disappears during periods of low market volatility. This suggests that it is difficult for fund managers to engage in risk-taking activities in stable market conditions.

In bear markets, according to Kempf et al. (2009), managers are more likely to face unemployment risk due to lower fund inflows and poorer overall fund performance (Chevalier and Ellison (1999), Zhao (2005)) implying a higher chance of losing their job with little chance of getting a promotion. Risk-shifting, especially risk-taking, hurts mutual fund performance (Huang et al. (2011)). Therefore, we predict that the rating-driven risk-shifting behavior is less pronounced during times of high unemployment risk. Following Shu et al. (2012), we consider year with negative annual returns and crisis years. We designate 2008, 2018, 2019, and 2020 as crisis years, encompassing the Global Financial Crisis (GFC) in 2008, as well as subsequent events such as the onset of the US-China trade war, the Quant Crisis,³⁸ and the COVID-19 pandemic. In column (3) and (4), the coefficient of $\text{Top}_{new} \times \text{Negative}$

³⁸Blitz (2021) finds that during 2018-2020, most factors failed to generate risk premiums except for expensive growth stocks, posing an exceptional challenge to quantitative managers.

Full-Year Return and $\text{Top}_{new} \times \text{Crisis Period}$ are negative and significant, suggesting the risk-taking behavior by top-tier funds is less pronounced or even disappears during periods of high unemployment risk.

3.4.5 Systematic Risk and Idiosyncratic Risk

Huang et al. (2011) find that the negative impact of risk-shifting on performance is significant primarily for funds that alter idiosyncratic risk exposure or their tracking error relative to their benchmarks, whereas the change in systematic risk leads to only modest reductions in fund performance. We are interested in understanding whether managers are aware of the performance consequences of risk-shifting through systematic risk and idiosyncratic risk. Table 17 reports managers' rating-driven risk-shifting through systematic risk and idiosyncratic risk. Systematic risk-shifting is calculated as the logarithmic difference of style beta, and idiosyncratic risk-shifting is calculated as the logarithmic difference of the standard deviation of tracking error with Morningstar category index. In column (1), we observe that top-tier funds increase systematic risk (1.066, t-statistic=4.645), while bottom-tier funds decrease risk (-1.108, t-statistic=2.969). In column (2), the negative and significant coefficient of $\text{Top}_{new} \times \text{Rating}_{overall}$ suggests that the top-tier funds' systematic risk-taking is concentrated in funds with low star ratings. However, the evidence of bottom-tier funds decreasing systematic risk loses significance (-0.794, t-statistic=-0.952) in column (2).

Table 17: Systematic risk and Idiosyncratic risk

This table examines funds' risk-shifting behavior through systematic risk and idiosyncratic risk, and their cross-sectional variation across ratings. Top_{new} ($Bottom_{new}$) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $Rank_{new} = 5$ (=1). The $Rank_{new}$ represents quintile sorting of equity funds based on their 3-year Morningstar risk-adjusted return (MRAR) from 1 to 5 within each overall rating for each 3×3 Morningstar category. $Beta_{MS}$ represent the betas of the 12 months fund gross returns against MorningStar Category Index, while $IVol_{MS}$ denotes the standard deviation of the 12 months excess returns of the fund gross returns on the MorningStar Category Index. The change of these risk measures (Δ) is computed as the logarithmic difference between the measures estimated using past 12-month data and future 12-month data. $Rating_{overall}$ takes value from 1 to 5. Control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense, and overall rating with style fixed effects and month fixed effects. Standard errors are adjusted for heteroskedasticity and clustered by style and month. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Dep.Var(%) =	Systematic Risk		Idiosyncratic Risk	
	$\Delta Beta_{MS}$		$\Delta IVol_{MS}$	
	(1)	(2)	(3)	(4)
Intercept	-0.397 (-0.467)	-0.752 (-1.083)	-1.776 (-1.304)	-1.210 (-0.742)
Top_{new}	1.066*** (4.645)	2.585*** (3.727)	-0.012 (-0.032)	2.334 (1.276)
$Bottom_{new}$	-1.018*** (-2.969)	-0.794 (-0.952)	-0.908*** (-3.077)	-4.491*** (-2.864)
Top_{new} × $Rating_{overall}$		-0.473** (-2.370)		-0.734 (-1.380)
$Bottom_{new}$ × $Rating_{overall}$		-0.130 (-0.618)		1.111** (2.331)
Controls	YES	YES	YES	YES
Style Fixed Effect	YES	YES	YES	YES
Month Fixed Effect	YES	YES	YES	YES
adj.R squared	0.5%	0.5%	0.1%	0.1%

In column (3), we observe that bottom-tier funds decrease idiosyncratic risk (-0.908, t-statistic=-3.077), while we do not find evidence that top-tier funds increase idiosyncratic risk. In column (4), the positive and significant coefficient of $Bottom_{new} \times Rating_{overall}$ suggests that the reduction in idiosyncratic risk among bottom-tier funds is concentrated in funds with low star ratings. Our findings in Table 17 indicate that rating-driven risk-shifting

predominantly occurs among funds with low performance, as captured by their star rating. Additionally, managers appear to understand the different performance consequences of risk-shifting through systematic risk and idiosyncratic risk, enabling them to strategically choose the type of risk that is most advantageous to them for managing ratings.

3.4.6 Analysis of Fund Holdings

We attempt to explain managers rating-driven risk-shifting using funds' holding information. Using funds' holdings, we first calculate two funds' holding betas measures: the weighted-average of holdings style beta and the style beta of the holdings portfolio return. Second, we also use holdings information to inspect portfolio concentration by estimating two variations of Herfindahl-Hirschman Index (HHI). In columns (1) and (2) of Table 18, the coefficients of Top_{new} are positive and significant, while the coefficients of $Top_{new} \times Rating_{overall}$ are negative and significant. This indicates that top-tier managers, particularly among low-rating funds, increase their holdings' style beta, consistent with the finding of increased systematic risk among top-tier managers uncovered in Table 17.

Table 18: Holdings Analysis

This table inspects the holdings of top-tier funds and bottom-tier funds from reform-2022. Top_{new} ($Bottom_{new}$) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $Rank_{new} = 5$ ($=1$). The $Rank_{new}$ represents quintile sorting of equity funds based on their 3-year Morningstar risk-adjusted return (MRAR) from 1 to 5 within each overall rating for each 3×3 Morningstar category. $Beta_{HB}$ represents the betas of the 12-month holding-based (HB) returns against the S&P 500, where the holding-based return is the value-weighted average return of the fund's most recent holding. $Beta_{Average}$ is the value-weighted average of stocks' beta for a fund's most recent holding. HHI, Herfindahl-Hirschman Index, is calculated as the sum of squared dollar weights (ranging from 0 to 1) of each industry (SIC 2-digit) within a fund. KSZ is the sum of squared differences between a fund's industry weights and the corresponding market's industry weights (Kacperczyk et al. (2005)). The change of these risk measures (Δ) is computed as the logarithmic difference between the measures estimated using past 12-month data and future 12-month data. Control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense, and overall rating with style fixed effects and month fixed effects. Standard errors are adjusted for heteroskedasticity and clustered by style and month. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Dep.Var(%) =	$\Delta Beta_{HB}$	$\Delta Beta_{Average}$	ΔHHI	ΔKSZ
	(1)	(2)	(3)	(4)
Intercept	-4.716*** (-3.511)	-0.843 (-1.113)	-4.549*** (-3.070)	-7.089*** (-6.230)
Top_{new}	3.150** (2.332)	1.572** (2.277)	-0.257 (-0.400)	0.638 (0.770)
$Bottom_{new}$	-1.415 (-1.116)	-0.557 (-1.222)	-1.679*** (-4.200)	-3.113** (-2.449)
$Top_{new} \times Rating_{overall}$	-1.003** (-2.099)	-0.558*** (-2.815)	-0.039 (-0.159)	-0.193 (-0.581)
$Bottom_{new} \times Rating_{overall}$	0.009 (0.027)	0.109 (0.884)	0.538*** (2.839)	0.797* (1.784)
Controls	YES	YES	YES	YES
Style Fixed Effect	YES	YES	YES	YES
Month Fixed Effect	YES	YES	YES	YES
adj.R squared	0.8%	0.4%	0.6%	0.4%

In columns (3) and (4) of Table 18, the coefficients of $Bottom_{new}$ are negative and significant, and the coefficients of $Bottom_{new} \times Rating_{overall}$ are positive and significant. This suggests that bottom-tier managers, especially among low-rating funds, decrease their holdings' industry concentration. Since higher industry concentration leads to greater exposure

to idiosyncratic risk (Huang et al. (2011)), the results in columns (3) and (4) support our previous finding that bottom-tier managers decrease idiosyncratic risk, as presented in Table 17.

3.4.7 Analysis of Trading Activities

We investigate managers rating-driven risk-shifting by inspecting funds' trading activities. Absolute return gap serves as a proxy for interim trading aggressiveness (Kacperczyk et al. (2008), Huang et al. (2011)), while turnover is a measure of sensation-seeking trading activity (Grinblatt and Keloharju (2009)). Table 19 indicates that for both trading proxies, we observe evidence suggesting that managers, particularly those with low ratings, tend to engage in more trading activities when they are in the top tier and reduce trading when they are in the bottom tier.

Table 19: Trading Analysis

This table inspects trading activities of top-tier funds and bottom-tier funds from reform-2022. Top_{new} ($Bottom_{new}$) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $Rank_{new} = 5$ ($=1$). The $Rank_{new}$ represents quintile sorting of equity funds based on their 3-year Morningstar risk-adjusted return (MRAR) from 1 to 5 within each overall rating for each 3×3 Morningstar category. Gap_{abs} represents the absolute gap of a fund, defined as the absolute value of the difference between the fund’s reported monthly return and the monthly buy-and-hold return of the fund’s portfolio disclosed in the previous period. Turnover is the fund’s monthly turnover ratio, calculated by dividing the annual turnover ratio by 12. The change of these risk measures (Δ) is computed as the logarithmic difference between the measures estimated using past 12-month data and future 12-month data. Control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense, and overall rating with style fixed effects and month fixed effects. Standard errors are adjusted for heteroskedasticity and clustered by style and month. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Dep.Var(%) =	ΔGap_{abs}	$\Delta Turnover$
	(1)	(2)
Intercept	2.048 (1.522)	-7.015** (-2.454)
Top_{new}	3.553*** (4.812)	0.969** (2.374)
$Bottom_{new}$	-1.271* (-1.710)	-0.805* (-1.956)
$Top_{new} \times Rating_{overall}$	-1.079*** (-3.878)	-0.262* (-1.850)
$Bottom_{new} \times Rating_{overall}$	0.115 (0.557)	0.283* (1.864)
Controls	YES	YES
Style Fixed Effect	YES	YES
Month Fixed Effect	YES	YES
adj.R squared	1.4%	0.3%

3.5 Rating-driven Style-shifting under Old Methodology

Before the reform, managers had the option to strategically shift their investment style to manage ratings, in addition to shifting risk. Our testing procedure is, first, to examine whether managers strategically shift their style in response to their relative position of rating change. If we do not find evidence of rating-driven style-shifting, we can then explore whether

rating-driven risk-shifting still exists in the pre-reform period. On the other hand, if we find evidence of managers strategically shifting style, we can further investigate the consequences of style-shifting on a portfolio's risk.

We use goodness-of-fit (R2) of a fund's gross return on the Morningstar category index to measure fund's style shifting from the current style. A decreasing R2 suggests a weakening explanatory power of fund's performance by the Morningstar category index, implying the fund is exposing to styles other than its Morningstar category. An increasing R2 indicates a strengthening explanatory power of fund's performance by the Morningstar category index, implying a shift back to its Morningstar category.

In columns (1) of Table 20 Panel A, the coefficient of Top_{old} is negative and significant (-1.170, t-statistic=-2.933), suggesting top-tier funds are more likely to expose themselves to other styles. Column (3) further shows this rating-driven style-shifting mainly exist in period before the reform, which indicates that the rating-driven style-shifting is not caused by some unknown omitted factors that is related to Top_{old} .

Table 20: Rating-driven Style-shifting before Reform

This table examines style-shifting behavior by top-tier funds and bottom-tier funds from 1990-reform. The main dependent variable in this analysis is the style shifting of funds, measured by $\Delta R2$. $R2$ represents goodness-of-fit of the 12 months fund gross returns against the Morningstar Category index. Top_{old} ($Bottom_{old}$) is an indicator variable for top-tier funds (bottom-tier funds) under the old methodology, which takes a value of 1 if $Rank_{old} = 5$ ($=1$). The $Rank_{old}$ represents quintile sorting within each rating with all equity funds based on their 3-year gross return from 1 to 5. The change of $R2$ ($\Delta R2$) is computed as the logarithmic difference between the measures estimated using past 12-month data and future 12-month data. ‘Pre’ takes the value of 1 if the observation is before the reform, 0 otherwise. Control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense, and overall rating with style fixed effects and month fixed effects. Standard errors are adjusted for heteroskedasticity and clustered by style and month. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Dep.Var(%) =	1990-Reform		1990-2022	
	$\Delta R2$ (1)	ΔStd (2)	$\Delta R2$ (3)	ΔStd (4)
Intercept	-3.223 (-1.036)	-5.140** (-2.128)	0.157 (0.206)	-2.838** (-2.337)
Top_{old}	-1.170*** (-2.933)	-1.416*** (-2.287)	0.038 (0.145)	0.621 (1.289)
$Bottom_{old}$	0.661 (0.877)	0.457 (0.503)	0.155 (0.588)	-0.704 (-0.977)
$Top_{old} \times Pre$			-1.119* (-1.838)	-1.142*** (-2.724)
$Bottom_{old} \times Pre$			0.593 (0.760)	0.361 (0.348)
Controls	YES	YES	YES	YES
Style Fixed Effect	YES	YES	YES	YES
Month Fixed Effect	YES	YES	YES	YES
adj.R squared	1.5%	5.0%	0.4%	1.4%

Does rating-driven style-shifting result in increased risk? Contrary to this expectation, column (2) presents an opposing view. The significant and negative coefficient of Top_{old} (-1.416, t-statistic=-2.287) suggests that top-tier funds engaging in strategic style-shifting reduce total risk. We attribute this finding to the unintentional diversification caused by style-shifting. Top-tier funds, by aggressively exposing themselves to styles other than their original style, inadvertently diversify their portfolios, leading to a decrease in total risk.

Column (4) further confirms this unintentional rating-driven diversification mainly exist in period before the reform.

3.6 Further Analysis

3.6.1 Rating Change and Reference-Dependent Risk Preference

Prospect theory by [Kahneman and Tversky \(2013\)](#) predicts that the feeling of gain can induce risk-aversion and feeling of loss can induce risk-taking. Managers tend to use the previous rating as a reference point, exhibiting risk-seeking behavior following a downgrade (when the current rating falls below the prior one), and risk-aversion following an upgrade (when the current rating is above the prior one). Therefore, this reference-dependent risk preference (RDRP) predicts that managers exhibit a preference to increase risk (shift style) after a rating downgrade and a tendency to reduce risk (concentrate style) following a rating upgrade in post-reform (pre-reform) period. ³⁹

Top-tier funds are also likely to be funds that have recently experienced rating downgrades (as shown in [Appendix 2D](#)), experiencing a sense of loss, and consequently increase risk or expose to other styles under the RDRP. Conversely, bottom-tier funds are more likely to be funds that have recently received rating upgrades (as shown in [Appendix 2D](#)), feeling a sense of gain, and consequently reduce risk or concentrate their style under the RDRP. Thus, RDRP could potentially explain the strategic rating-driven risk(style)-shifting.

³⁹We acknowledge that the predicted risk (style)-shifting in response to rating changes could also be interpreted as a strategic behavior under option pricing theory. Upgraded (downgraded) funds experience a shift from a long call (short put) to a short put (long call) position. Consequently, the strategy is to decrease (increase) risk or concentrate (shift) style. Since we cannot confirm whether the mentioned relation is a strategic behavior or a result of loss aversion, we will rule out this alternative from our analysis.

Table 21: Risk-shifting, Style-shifting and Reference-Dependent Risk Preference

This table controls the potential effect of reference-dependent risk preference on rating-driven risk(style)-shifting in post-reform (pre-reform) period in Panel A (B). Top_{new} ($Bottom_{new}$) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $Rank_{new} = 5$ (=1). Top_{old} ($Bottom_{old}$) is an indicator variable for top-tier funds (bottom-tier funds) under the old methodology, which takes a value of 1 if $Rank_{old} = 5$ (=1). $\Delta Rating_{t-n,t}$ equals 1 if the fund has experienced a rating upgrade compared to the rating n month ago, -1 if the fund has experienced a rating downgrade compared to the rating n month ago, and 0 if the rating remains the same as the rating n month ago. Columns (1) and (2) controls the effect of rating change by adding it in independent variables, while columns (3) and (4) controls the effect of rating change by focusing on subsample with no rating change in the past. Control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense, and overall rating with style fixed effects and month fixed effects. Standard errors are adjusted for heteroskedasticity and clustered by style and month. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: Reform-2022				
Dep.Var(%) = ΔStd	(1)	(2)	$\Delta Rating_{t-1,t}=0$ (3)	$\Delta Rating_{t-6,t}=0$ (4)
Intercept	-3.381*** (-2.597)	-3.251** (-2.449)	-3.124** (-2.491)	-3.292** (-2.378)
Top_{new}	0.352* (1.946)	0.380* (1.660)	0.421** (1.972)	0.445** (2.014)
$Bottom_{new}$	-1.010*** (-5.229)	-1.024*** (-5.504)	-0.986*** (-4.791)	-0.851*** (-4.865)
$\Delta Rating_{t-1,t}$	-0.428** (-2.298)			
$\Delta Rating_{t-6,t}$		-0.439 (-1.305)		
Controls	YES	YES	YES	YES
Style Fixed Effect	YES	YES	YES	YES
Month Fixed Effect	YES	YES	YES	YES
adj.R squared	1.4%	1.4%	1.3%	0.9%
% of Full Sample	100.0%	100.0%	74.0%	56.4%

Table 21 : Continue

Panel B: 1990-Reform				
Dep.Var(%) = $\Delta R2$	(1)	(2)	$\Delta Rating_{t-1,t=0}$	$\Delta Rating_{t-6,t=0}$
			(3)	(4)
Intercept	-3.423 (-1.113)	-3.852 (-1.439)	-4.465* (-1.678)	-5.283** (-2.145)
Top _{old}	-1.729*** (-3.607)	-1.983** (-4.444)	-1.743*** (-4.046)	-2.467*** (-4.507)
Bottom _{old}	0.623 (0.802)	0.539 (0.628)	0.428 (0.491)	-0.465 (-0.467)
$\Delta Rating_{t-1,t}$	0.579* (1.675)			
$\Delta Rating_{t-6,t}$		1.199*** (3.849)		
Controls	YES	YES	YES	YES
Style Fixed Effect	YES	YES	YES	YES
Month Fixed Effect	YES	YES	YES	YES
adj.R squared	1.5%	1.6%	1.4%	0.8%
% of Full Sample	100.0%	100.0%	56.6%	39.1%

We consider rating-driven risk-shifting under new methodology and RDRP in Panel A of Table 21. In columns (1) and (2), we present the regression of ΔStd on Top_{new} and $Bottom_{new}$ controlling the rating change with 2 windows (1-month and 6-month). We also present the regression in subsample focusing on funds with no rating change in the past in columns (3) and (4), which directly remove funds that potentially suffer loss aversion bias. The coefficient estimates of Top_{new} are positive and significant and the coefficient estimates of $Bottom_{new}$ are negative and significant from column (1) to (4), suggesting the strategic rating-driven risk-shifting under new methodology is not explained by RDRP.

Next, we consider rating-driven style-shifting under old methodology and RDRP in Panel B of Table 21. In columns (1) and (2) of Panel B, the coefficient estimates of Top_{old} are all negative and significant after controlling the two rating change proxies. Besides, the coefficient of Top_{old} remains negative and significant among funds with no rating change in

the past. As a result, the strategic rating-driven style-shifting under old methodology is not explained by RDRP either.

3.6.2 Agency Issue and Incentive

Huang et al. (2011) finds that funds that change risk, especially those that increase risk, tend to subsequently perform worse than others due to agency issues. Therefore, if the risk-taking by top-tier funds is merely an agency issue, top-tier funds should perform worse than other funds. However, the potential rating upgrade faced by top-tier funds can also serve as an incentive for managers to improve managerial effort (e.g., by increasing the collection of information, encouraging better techniques for constructing portfolios, or improving technological infrastructure), and incentives can increase performance (Massa and Patgiri (2009)). In terms of the net effect, top-tier funds should perform better than other funds after controlling for the poor performance caused by agency issues.

Table 22 examine the factor-adjusted return (Carhart (1997)) of top-tier funds. We implement dynamic fund risk adjustment proposed by Huang et al. (2011) to calculate Carhart factor-adjusted return to address the concern that funds that change risk also change risk exposure.⁴⁰ The Carhart factor-adjusted return is computed using the betas of the most recent holdings of the fund, obtained by regressing the hypothetical returns of the most recent portfolio holdings over the last 36 months on the Carhart factors. This adjustment allows the risk exposures of funds to change over time.

⁴⁰One commonly used methodology to adjust for risk and style is to first estimate the factor loadings for each fund over a rolling window using prior data and then compute abnormal returns in the subsequent period as the difference between the actual fund return and the expected fund return based on the estimated factor loadings. Huang et al. (2011) suggest that this traditional method is not appropriate since the factor loadings estimated over prior windows might not be accurate for funds that shift risk levels.

Table 22: Performance Analysis

This table presents the impact of risk-taking (style-shifting) by top-tier funds on funds future abnormal return. $\text{Alpha}_{\text{carhart},t+1,t+12}$ is factor-adjusted return (Carhart (1997)), but computed using the betas of the most recent holdings of the fund, obtained by regressing the hypothetical returns of the most recent portfolio holdings over the last 36 months on the Carhart factors. Top_{new} is an indicator variable for top-tier funds under the new methodology, which takes a value of 1 if $\text{Rank}_{\text{new}} = 5$. Top_{old} is an indicator variable for top-tier funds under the old methodology, which takes a value of 1 if $\text{Rank}_{\text{old}} = 5$. ‘Risk-Taker’ (‘Style-Shifter’) is a dummy variable that takes the value of 1 if the fund is in the top (bottom) quintile of $\Delta\text{Std}(\Delta R2)$ each month and 0 otherwise. Control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense, and overall rating with style fixed effects and month fixed effects. Standard errors are adjusted for heteroskedasticity and clustered by style and month. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: Reform-2022			
Dep.Var(%) = $\text{Alpha}_{\text{carhart},t+1,t+12}$	(1)	(2)	(3)
Intercept	13.039***	13.176**	13.193**
	(7.551)	(7.588)	(7.567)
Top_{new}	0.280*	0.293*	0.204
	(1.854)	(1.900)	(1.123)
Risk-Taker		-0.512*	-0.605*
		(-1.849)	(-1.938)
$\text{Top}_{\text{new}} \times \text{Risk-Taker}$			0.419*
			(1.717)
Controls	YES	YES	YES
Fund Fixed Effect	YES	YES	YES
Month Fixed Effect	YES	YES	YES
adj.R squared	1.6%	1.6%	1.6%
Panel B: 1990-Reform			
Dep.Var(%) = $\text{Alpha}_{\text{carhart},t+1,t+12}$	(1)	(2)	(3)
Intercept	21.551***	21.539**	13.193**
	(5.390)	(5.403)	(7.567)
Top_{old}	-0.170	-0.171	-0.108
	(-0.202)	(-0.203)	(-0.144)
Style-Shifter		0.031	0.097
		(0.137)	(0.326)
$\text{Top}_{\text{old}} \times \text{Style-Shifter}$			-0.309
			(-0.367)
Controls	YES	YES	YES
Style Fixed Effect	YES	YES	YES
Month Fixed Effect	YES	YES	YES
adj.R squared	3.3%	3.3%	3.3%

Panel A reports the performance of top-tier funds under new methodology. In column (1), the coefficient of Top_{new} is positive and significant (0.280, t-statistic=1.854), indicating that top-tier funds can outperform other funds by 0.28% in the following 12 months. In column (2), we include risk-takers in the regression to control for the effect of agency issues. Risk-takers are defined as funds within the top quintile of ΔStd each month. We find that risk-takers significantly perform worse (-0.512, t-statistic=-1.849) than others, which is consistent with [Huang et al. \(2011\)](#). After controlling for the poor performance of risk-takers, the coefficient of Top_{new} remains positive and significant, implying that the potential rating upgrade can serve as an incentive for top-tier funds to improve managerial efforts. In column (3), the coefficient of Top_{new} loses its significance after controlling $\text{Top}_{new} \times \text{Risk-Taker}$ and coefficient of $\text{Top}_{new} \times \text{Risk-Taker}$ is positive and significant (0.419, t-statistic=1.717) suggesting that the risk-takers in top-tier funds are incentivized the most by potential rating upgrades.

Panel B reports the performance of top-tier funds under old methodology. We do not find any significant relationship between Top_{old} and future factor-adjusted return from column (1) to (3). This suggests, even though top-tier funds engage in more style-shifting to other styles before the reform, they do not earn abnormal returns, as their focus is on exposure to systematic risk premiums instead of seeking risk-adjusted alpha. [Sha \(2020\)](#) also finds style-shifting does not lead to abnormal factor-adjusted returns.

3.7 Conclusion

This paper highlights the issue of strategic risk-shifting in the mutual fund industry arising from the discrete nature of star ratings. Because investors heavily rely on these ratings when selecting funds, changes in ratings lead to discrete shifts in investor flows ([Reuter and Zitzewitz \(2021\)](#)). This creates a non-linear payoff structure: funds nearing a rating upgrade face a long call-like incentive, encouraging them to increase risk in hopes of capturing rating upgrade; those approaching a downgrade face a short put-like incentive, leading them to reduce risk to avoid rating downgrade.

We find that rating-driven risk-shifting accounts for 21% of market-wide risk-shifting. Given the negative impact of risk-shifting behavior on fund performance, this paper raises concerns about the unintended consequences of discrete performance metrics. We further support our findings by analyzing changes in both systematic and idiosyncratic risk, as well as fund holdings and trading behavior. Before the Morningstar methodology reform, ratings were assigned without considering investment style, which gave managers more flexibility to shift across styles. We find evidence that, under this old system, managers strategically altered their investment styles to influence ratings. In further analysis, we first rule out reference-dependent risk preference as the primary driver of risk and style shifting. Second, we find that ratings can also serve as an incentive, encouraging risk-taking managers to exert greater effort.

4 Chapter Three: Luck and Skill in the World of Diseconomies of Scale, co-author with Juan Yao

“谋事在人，成事在天”

– 《三国演义》

“*Man proposes, but God disposes*”

– 《*Romance of the Three Kingdoms*》

4.1 Introduction

The asset management industry has been experiencing a global shift from active to passive investing in recent decades (Anadu et al. (2020)). This phenomenon appears consistent with the literature documenting that most funds lack skill (Kosowski et al. (2006), Barras et al. (2010), Fama and French (2010), Harvey and Liu (2022)). These studies typically treat alpha as a measure of skill and find that most funds cannot persistently generate positive alpha over their lifetimes. However, such findings are puzzling, as argued by Berk and Green (2004) (BG): If most funds indeed lack skill, why would investors entrust managers with their capital, and why would the active investment industry exist in the first place?

Berk and Green (2004) argues that the lack of alpha in the fund industry is an equilibrium outcome of rational investors chasing performance and market competition. The BG model contends that funds capable of generating alpha represent scarce investment opportunities and face diseconomies of scale. As expressed in Eq. (8), the expected alpha of a fund (α_t , or $E[r_t]$ where r_t is realized alpha) depends on the manager’s alpha-generating ability on the first dollar invested (a), the diseconomies of scale (b), and the assets under management (q). Thus, a positive alpha does not indicate managerial skill but rather reflects that the fund is underfunded with capital (hereafter referred to as “funding status”). Consequently, investors’ alpha-chasing behavior can be viewed as a rational response to learning about a fund’s funding status. This rational dynamic leads to a lack of alpha persistence in equilib-

rium, as capital inflows inflate the size of funds with positive alpha, making it increasingly difficult to outperform in the presence of decreasing returns to scale.

$$E[r_t] = \alpha_t = a - b \times q_{t-1}, \quad (8)$$

Skill, on the other hand, is a more complex concept. A skilled manager not only needs to generate alpha on the first dollar invested (a), but also to manage the fund size in accordance with the scalability (b) of their strategy, thereby actively choosing their size (q) at which to operate to maximize value extraction from financial markets. Thus, investors cannot fully infer managers' skill solely from realized alpha. Building on this insight, [Berk and Van Binsbergen \(2015\)](#) and [Barras et al. \(2022\)](#) use value added ($r \times q$)—the product of realized alpha and fund size—as a measure of skill. [Berk and Van Binsbergen \(2015\)](#) show that value added is strongly persistent for up to ten years, consistent with the notion that a genuine measure of skill should persist, and that it is positive at the industry level, supporting the existence of skill in the mutual fund industry.

However, the key implications of the BG model are questioned by other studies. [Fama and French \(2010\)](#) dismiss its equilibrium prediction, stating that “*Their model (i.e. BG) is an attractive theory, but our results reject most of its predictions about mutual fund returns.*” Similarly, [Zhu \(2018\)](#) and [Barras et al. \(2022\)](#) find that most funds do not operate at their optimal size to extract the maximum value from the market, suggesting that the equilibrium implications of the BG model have not yet been realized. Moreover, the rational assumptions by [Berk and Green \(2004\)](#) and [Berk and Van Binsbergen \(2015\)](#) that managers optimize and that investors learn rationally have been challenged empirically. For example, studies reveal that fund investors are often naive, heavily chasing past performance ([Chevalier and Ellison \(1997\)](#), [Sirri and Tufano \(1998\)](#)), and relying on simple performance metrics such as ratings ([Ben-David et al. \(2022a\)](#), [Del Guercio and Tkac \(2008\)](#), [Reuter and Zitzewitz \(2021\)](#)). [Zhu \(2018\)](#) further shows that managers often fail to optimize value added because they are not

fully aware of their alpha-generating ability (a) and scalability (b).

If realized fund alpha were well explained by expected alpha in Eq.(8), investors' performance-chasing behavior could be considered as rational learning about funding status (Berk and Green (2004)). However, in this study, we find that realized fund alpha contains a substantial idiosyncratic component that cannot be explained by the expected alpha. Using both linear and nonparametric models, we find that the idiosyncratic component accounts for at least 55% of realized alpha variation. Building on prior studies of luck in private equity funds (Korteweg and Sorensen (2017), Rossi (2019)), we interpret this component as luck, largely driven by the idiosyncratic returns of fund holdings beyond managers' control. Given that luck constitutes a substantial share of fund alpha, naive performance-chasing reflects not only rational learning about funding status as in the BG model, but also the chasing of luck.

Follow the spirit of the BG model, we re-examine the debate on skill versus luck in the presence of diseconomies of scale, in a setting where investors naively chase performance and managers fail to optimize fund size. In this semi-rational world, the supply side of the fund industry is suboptimal. Managers do not maximize value added, but instead actively manage all assets allocated by investors, at the cost of diseconomies of scale. Moreover, while managers' expected alpha suffers from diseconomies of scale as in Eq.(8), realized alpha is noisy, containing a "luck" that cannot be explained by expected alpha. On the demand side, investors naively chase past performance, thereby allocating capital both to productive underfunded funds and to funds experienced periods of good luck.

We model "luck" in this semi-rational world and predict a negative relation between luck and future expected alpha. We show that funds experiencing a lucky period tend to grow larger, both through the appreciation of their holdings and through capital inflows. In the presence of diseconomies of scale, such funds are more likely to underperform in the future due to their oversized scale. This mechanism is confirmed numerically using bootstrapping simulations. Motivated by this insight, we further test the consequence of chasing perfor-

mance. In the cross-section of funds, chasing alpha can sometimes help investors identify underfunded funds that subsequently outperform. However, because variation in size of a fund overtime is much smaller than variation in its realized alpha, once we control for fund fixed effects, historical outperformance is more likely dominated by luck, leading to future underperformance.

Our framework also provides insights into the relation between luck and funding status, thereby explaining potential capital misallocation by investors. We show that, due to naive performance-chasing, underfunded funds continue to experience outflows after periods of bad luck, whereas overfunded funds continue to attract inflows after periods of good luck. We further use simulation to show that capital misallocation induced by performance chasing destroys value at the industry level.

Next, we examine the validity of using value added as a measure of skill. Intuitively, non-optimizing managers who are unaware of their limits actively manages all of the capital provided by investors. However, because investors adjust their capital based on realized alpha only gradually, underfunded(overfunded) funds are more likely to remain underfunded(overfunded) in subsequent periods, leading to persistence in value added. This persistence does not necessarily indicate that value-creating managers are skilled; it simply reflects that underfunded managers have not yet received sufficient capital. In our simulations, when investors' performance chasing speed (measured by the coefficient of flow-alpha relationship) increases, the persistence of value added disappears. We therefore caution against using value added to identify skilled managers. Instead, we suggest that mutual fund companies should use value added as an incentive mechanism to encourage managers to extract more value from financial markets, thereby fostering a more rational equilibrium.

Our results have practical implications for fund performance evaluation. Morningstar assigns funds a 1 to 5-star rating based on past performance,⁴¹ providing investors with

⁴¹Morningstar assigns ratings to mutual funds based on their Morningstar risk-adjusted return (MRAR) for different time periods (3, 5, and 10 years). The overall rating for funds is then calculated as the weighted

a simple metric for evaluating fund quality. Such performance-based ratings are shown to be key drivers of investor capital allocation (Ben-David et al. (2022a), Del Guercio and Tkac (2008), Reuter and Zitzewitz (2021)). Do these star ratings primarily capture a fund’s funding status or do they merely reflect luck? We show that while star ratings capture funding status to some extent, they are more heavily influenced by luck. This is concerning, as our model predicts that lucky funds tend to underperform in the future. Consistent with this prediction, we find that 5-star funds significantly underperform 1-star funds in the following period, a result generally in line with findings from Blake and Morey (2000) and Morey (2003).

Using full fund-life observations, we find that much of the past performance is attributed to luck, which cannot persist and may even result in future underperformance. As Benjamin Graham aptly noted, “*In the financial markets, hindsight is forever 20/20, but foresight is legally blind.*” For investors, it is inherently difficult to distinguish whether a fund’s past outperformance was due to luck or underfunding. This creates a paradox: evaluating funds based solely on past performance appears fundamentally flawed. How, then, can we evaluate funds in a forward-looking manner? We argue that a qualitative approach is essential. We investigate the “analyst rating” introduced by Morningstar in 2011. We show that the “analyst rating” can partially capture a fund’s funding status and, is far less affected by historical luck compared to the traditional backward-looking star ratings. Consistent with our model’s predictions, funds with a “gold medal” rating from analysts tend to outperform in the future, aligning with the findings in Armstrong et al. (2019). We also examine Morningstar’s “quantitative rating”, which uses machine learning model and quantitative fund characteristics to mimic analyst ratings. Although Morningstar claims it is as forward-looking as the analyst rating, we find that the quantitative rating remains heavily influenced by historical luck, making it functionally similar to the star ratings. While machine learning can capture qualitative information in principle, we argue that the “quantitative rating” fails because its

average of the 3-, 5-, and 10-year ratings.

inputs (e.g. quantitative fund characteristics) are inherently backward-looking.

By relaxing the assumptions of the influential BG model, this paper contributes to several strands of the literature. First, prior studies (Kosowski et al. (2006), Barras et al. (2010), Fama and French (2010), Harvey and Liu (2022)) often treat luck as the opposite of skill and attempt to distinguish skilled from lucky funds based on past performance. We instead view luck as a time-varying idiosyncratic component of performance, that plays a crucial role in attracting capital flows. While the role of luck in performance measurement has been discussed in private equity (Rossi (2019), Korteweg and Sorensen (2017)), corporate finance (Bebchuk et al. (2010), Bertrand and Mullainathan (2001)), and management (Parnell et al. (2012), Liu and De Rond (2016)), this study is, to the best of our knowledge, the first to systematically investigate the magnitude and implications of luck in mutual fund performance.

Second, we contribute to the literature on investor sophistication by explaining why naive performance-chasing investors (Sirri and Tufano (1998), Ben-David et al. (2022a)) may appear rational and sophisticated (Berk and Van Binsbergen (2015), Zhu (2018), Barras et al. (2022)). Literature on investor learning (Schwarz and Sun (2023), Brown and Wu (2016), Choi et al. (2016), Barras et al. (2022)) typically assumes rationality and, in some cases, documents incomplete or slow learning. We show that naive performance-chasing can simultaneously generate rational capital allocation consistent with BG model's prediction and capital misallocation due to the presence of luck.

Third, we contribute to the literature on managerial skill, particularly studies using value added as a proxy for skill (Zhu (2018), Berk and Van Binsbergen (2015), Barras et al. (2022)). We highlight the limitations of interpreting value added as an indicator of managerial skill in a semi-rational world. We propose value added to be employed as an incentive mechanism to promote a more rational managerial behavior.

Finally, while existing studies on mutual fund ratings (e.g., Blake and Morey (2000), Armstrong et al. (2019)) primarily focus on their predictive power, we provide a theoretical

explanation, rationalizing the success and failure of different rating methods through their varying exposure to luck, thus providing practical guidance for the rating systems design.

The rest of the paper is organized as follows: Section 2 describes the data, Section 3 documents that a large portion of realized alpha is attributed to luck, Section 4 models luck and provides testable implications, Section 5 documents the negative relationship between luck and expected alpha, Section 6 discusses how luck leads to capital misallocation and the challenges of using value added as a proxy for skill, Section 7 addresses the practical implications on ratings, and Section 8 concludes.

4.2 Data

The mutual fund data used in this study come from Morningstar, covering the period from January 1990 to December 2023. To mitigate survivorship bias, we include both live and dead funds. Following standard practice in the mutual fund literature, we limit the sample to actively managed, domestic, equity-only funds in the U.S. market. Specifically, we include funds classified into one of the following nine Morningstar categories: large blend (LB), large growth (LG), large value (LV), mid-cap blend (MB), mid-cap growth (MG), mid-cap value (MV), small blend (SB), small growth (SG), and small value (SV). We further exclude index funds by removing funds with names containing “index”. In this study, we aggregate all share classes of the same fund. A fund’s assets under management (AUM) are calculated as the total AUM across all its share classes. When aggregating metrics such as returns, turnover, and expense ratios across share classes, we compute the AUM-weighted average for all non-missing values.

Following [Pástor et al. \(2015\)](#), we inflate fund AUM at the end of the each month to December 2023 dollars using the ratio of the total market value of all CRSP stocks in December 2023 to their value at the end of the prior month. This inflator ensures that AUM reflects the size of the fund relative to the investment universe of stocks—a reasonable way to capture the constraints imposed by fund size. A mutual fund enters the sample once its combined AUM across all share classes exceeds \$20 million in December 2023 dollars.

Additionally, we exclude funds with fewer than two years of data. After applying these criteria, the final sample consists of 3,662 mutual funds.

We construct three performance measures: the benchmark-adjusted gross alpha, the four-factor-adjusted gross alpha, and the benchmark-adjusted net alpha. The benchmark-adjusted gross alpha is calculated as the difference between a fund's gross return and its benchmark portfolio return, capturing the fund manager's ability to outperform the benchmark. The four-factor-adjusted gross alpha is an alternative measure of benchmark-adjusted performance that reflects the manager's ability to generate returns beyond common risk factors (Market, SMB, HML, UMD) following [Carhart \(1997\)](#). Finally, the benchmark-adjusted net alpha is calculated as the difference between a fund's net return and its benchmark return, reflecting the returns investors receive after accounting for fees and expenses.

Table [23](#) lists the descriptive statistics of the sample. As shown in Panel A, on average, funds outperform by 4 basis points (bps) per month relative to the Morningstar benchmark and by 12 bps relative to the four-factor model, based on gross return. In contrast, in the context of net return, funds underperform their Morningstar benchmarks by 4 bps per month on average. This is consistent with the negative average net alpha reported by [Pástor et al. \(2015\)](#) and [Zhu \(2018\)](#). The monthly expense ratio is calculated as one-twelfth of the annual gross expense ratio reported by Morningstar. Fund size is measured by the inflation-adjusted AUM at the end of the previous month. The distribution of AUM is highly right skewed, with a median value of \$803 million, and a mean of \$3,901 million. Fund age is measured as the number of years since the fund's inception date. The turnover ratio is defined as the minimum of aggregate purchases and sales divided by the average annual AUM, expressed as a percentage. The average turnover ratio for our sample is 72.81% per year. To remove outliers, we winsorize all variables at the 1st and 99th percentiles each month.

Table 23: Summary Statistics

Panel A presents descriptive statistics on the fund/month sample from January 1993 to December 2023. Gross alpha and net alpha are benchmark-adjusted by subtracting the Morningstar designated benchmark index return from the fund's gross return and net return. AUM is the fund's real assets under management in millions of dollars. FundAge is the time in years since the fund's inception date. Turnover is defined as the minimum of aggregate purchases and sales divided by the average annual AUM in percentage. Assuming the functional form of decreasing alpha to scale to be $r = a - b \times q$, Panel B presents the parameter b (decreasing return to scale) and a (the alpha earned on the first dollar invested) for size deciles based on funds average AUM over each fund's life. The parameter b is estimated using the panel estimator RD2 proposed by [Zhu \(2018\)](#). The parameter a for the i-th fund is estimated as $\hat{a}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} (r_{it} + \hat{b}_i q_{it-1})$. The parameter a and b are estimated using gross alpha against Morningstar benchmark.

Panel A: Descriptive Statistics								
	No.Fund/ month	Mean	Std	Percentiles				
				Min	25%	50%	75%	Max
Gross Ret(%)	632,506	0.89	5.02	-31.00	-1.69	1.25	3.76	39.10
Gross alpha ^{MS} (%)	632,506	0.04	1.86	-18.28	-0.82	0.01	0.86	27.46
Gross alpha ^{4F} (%)	524,258	0.12	1.77	-14.30	-0.74	0.01	0.99	17.41
Net ret(%)	632,506	0.80	5.22	-31.08	-1.92	1.20	3.82	38.92
Net alpha ^{MS} (%)	632,506	-0.04	1.86	-18.85	-0.91	-0.07	0.77	27.30
Expense ratio (%)	632,506	0.09	0.06	0.00	0.08	0.10	0.12	0.41
AUM (\$ mil)	632,506	3901	9781	21	209	803	2935	242006
FundAge	632,506	13.81	13.41	0.00	5.00	10.00	19.00	99.00
Turnover(%)	380,951	72.81	72.64	0.00	27.00	52.00	93.00	716.00

Panel B: Decreasing returns to scale in size deciles (RD2)								
$r^{g,MS} = a - b \times q$								
AUM(\$ Mil)	b			a (%)				
	slope($\times 10^6$)	t-value	Mean	Min	25%	50%	75%	Max
1 40	50.5687	4.521	0.08	-1.33	-0.06	0.12	0.28	1.05
2 86	20.5852	6.075	0.10	-1.14	-0.01	0.12	0.25	1.47
3 155	6.6371	6.148	0.07	-0.82	-0.04	0.07	0.20	1.09
4 265	6.1779	6.734	0.14	-1.31	0.04	0.16	0.26	1.01
5 438	2.9167	5.046	0.14	-0.54	0.01	0.13	0.21	4.15
6 712	1.9002	6.672	0.16	-0.76	0.07	0.15	0.24	0.74
7 1141	0.8433	6.637	0.14	-0.92	0.06	0.14	0.24	0.93
8 1933	0.4610	6.910	0.13	-1.31	0.06	0.13	0.21	0.75
9 3715	0.3368	9.546	0.19	-0.39	0.11	0.18	0.28	0.61
10 15737	0.0321	4.647	0.11	-0.83	0.04	0.11	0.19	0.53

We estimate funds' alpha-generating ability on the first dollar invested(a) and their dis-

economies of scale (b) using the linear functional form of decreasing returns to scale specified in Eq. (8), and employ the RD2 estimator following [Zhu \(2018\)](#). Panel B reveals substantial heterogeneity in decreasing returns to scale across funds. The coefficient estimates of b range from 0.0321×10^{-6} in the top decile (the largest funds) to 50.5687×10^{-6} in the bottom decile. Panel B also provides summary statistics for estimates of a within each AUM-sorted portfolio. We find that most funds have positive a , implying that they can generate positive alpha if not constrained by diseconomies of scale. Overall, these statistics are generally consistent with [Zhu \(2018\)](#). As discussed in [Zhu \(2018\)](#), the significance of the decreasing returns to scale remains robust when controlling for industry size, and when adjusting returns using alternative risk models, such as the four-factor model.

4.3 Realized Alpha and Luck

The fundamental premise of modeling luck in the presence of naive performance-chasing investors is that realized alpha is noisy and largely driven by luck. If realized fund alpha were well explained by the expected alpha in Eq. (8), investors' performance-chasing behavior could be rationalized as learning about a fund's funding status, as argued by [Berk and Green \(2004\)](#). In this section, we document that realized fund alpha contains a substantial idiosyncratic component that cannot be explained by expected alpha, even under both linear and nonparametric specifications.

4.3.1 Linear Model

We rely on the linear model in Eq. (8), estimated using the RD2 estimator of [Zhu \(2018\)](#), for most of our empirical analysis for several reasons. First, [Barras et al. \(2022\)](#) show that 17.6% of funds appear to have no diseconomies of scale due to estimation noise. The RD2 estimator of [Zhu \(2018\)](#) addresses this issue by estimating diseconomies of scale across size deciles, which allows for cross-sectional variation in the scale parameter (b) while ensuring that all funds have valid estimates. Second, the linear specification is both interpretable and widely adopted in mutual fund research ([Barras et al., 2022; Zhu, 2018](#)). Given that variation in fund alpha is much larger than variation in fund size over a fund's life cycle, adding model

complexity provides limited gains.⁴² Finally, [Barras et al. \(2022\)](#) provide strong evidence supporting the validity of the linear model against concerns of omitted variables and model misspecification, suggesting that the linear specification is appropriate in this context.

Table 24 examines how well the documented diseconomies of scale explain the variation in realized gross alpha. Using the full sample from 1990 to 2023, we first estimate each fund's a and b following [Zhu \(2018\)](#)'s RD2 approach. The full sample is used to obtain the best possible estimates for a and b , as it captures the entire life cycle of each fund. Next, using AUM at $t - 1$, we compute the expected alpha ($\alpha_{i,t}$) for each fund at t following Eq. (8). Intuitively, this expected alpha represents the alpha that a fund, given its alpha-generating ability on the first dollar (a) and scalability (b), is expected to generate, for its current size (q) under management. Finally, we use the R^2 from a linear regression to assess how much of the variation in realized alpha over different windows can be explained for by contemporaneous expected alpha.

⁴²Indeed, we show that the luck components derived from the nonparametric and linear models are highly correlated (above 0.97).

Table 24: Realized Alpha, Expected Alpha and Luck - Linear Model

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. Panel A presents the R^2 of simple linear regressions, where the dependent variable is the average historical realized alpha ($\bar{r}_{i,t-m,t}$), and the independent variable is the average historical expected alpha ($\bar{\alpha}_{i,t-m,t}$) over window m . Panel B presents the R^2 of simple linear regressions, where the dependent variable is the average historical realized alpha ($\bar{r}_{i,t-m,t}$), and the independent variable is the average historical luck ($\bar{\epsilon}_{i,t-m,t}$) over window m . Gross alpha against Morningstar benchmark is used in Panel A.1 and B.1, whereas gross 4 factor alpha is used in Panel A.2 and B.2.

Panel A: Realized Alpha and Expected Alpha					
A.1: $\bar{r}_{i,t-m,t}^{g,MS} = \beta_0 + \beta_1 \bar{\alpha}_{i,t-m,t}^{g,MS} + e$					
m=	1m	12m	24m	36m	60m
R^2	1.11%	7.79%	11.81%	14.71%	17.57%
	Year-Month FE: No		Fund FE: No		
R^2	0.35%	1.71%	2.08%	2.06%	1.61%
	Year-Month FE: Yes		Fund FE: Yes		
A.2: $\bar{r}_{i,t-m,t}^{g,4F} = \beta_0 + \beta_1 \bar{\alpha}_{i,t-m,t}^{g,4F} + e$					
m=	1m	12m	24m	36m	60m
R^2	1.09%	8.02%	11.97%	14.45%	17.38%
	Year-Month FE: No		Fund FE: No		
R^2	0.20%	0.92%	1.19%	1.23%	1.02%
	Year-Month FE: Yes		Fund FE: Yes		
Panel B: Realized Alpha and Luck					
B.1: $\bar{r}_{i,t-m,t}^{g,MS} = \beta_0 + \beta_1 \bar{\epsilon}_{i,t-m,t}^{g,MS} + e$					
m=	1m	12m	24m	36m	60m
R^2	98.99%	87.06%	80.07%	75.52%	71.46%
	Year-Month FE: No		Fund FE: No		
R^2	99.80%	92.74%	88.41%	85.50%	82.93%
	Year-Month FE: Yes		Fund FE: Yes		
B.2: $\bar{r}_{i,t-m,t}^{g,4F} = \beta_0 + \beta_1 \bar{\epsilon}_{i,t-m,t}^{g,4F} + e$					
m=	1m	12m	24m	36m	60m
R^2	98.96%	91.81%	88.06%	86.03%	83.83%
	Year-Month FE: No		Fund FE: No		
R^2	99.96%	99.69%	99.55%	99.50%	99.49%
	Year-Month FE: Yes		Fund FE: Yes		

We find that, empirically, the expected alpha for a one-month window explains only 1.11% of the cross-sectional variation in realized alpha during the same period. When the

window length increases to five years, the explanatory power of expected alpha rises to 17.57%, indicating that a fund’s funding status has a greater long-term impact on realized alpha. Next, after controlling for fund fixed effects and year-month fixed effects, we find that the funding status explains at most 2.08% of realized alpha. This occurs because AUM exhibits significant across-funds variation but much smaller variation over a given fund’s lifespan. The same result holds when using alpha adjusted for the 4-factor model.

Motivated by [Korteweg and Sorensen \(2017\)](#) and [Rossi \(2019\)](#), we define the idiosyncratic component ($\epsilon_{i,t}$) of realized alpha that cannot be explained by diseconomies of scale as luck, calculated as the difference between realized alpha ($r_{i,t}$) and expected alpha ($\alpha_{i,t}$). We find that luck explains 98.99% of the cross-sectional variation in monthly realized alpha and 71.46% of the variation in 5-year realized alpha. After controlling for fund and time fixed effects, luck still accounts for at least 82.93% of realized alpha. Similar results are observed when using 4-factor alpha. We also provide empirical evidence to support our definition of luck. Specifically, we show that the idiosyncratic component ($\epsilon_{i,t}$) exhibits properties consistent with the notion of luck: it averages out to zero over a fund’s life and is independent over time. The corresponding results are reported in [Appendix 3B](#) and [Appendix 3C](#).

4.3.2 Nonparametric Model

Although the linear model of diseconomies of scale has been extensively tested for robustness against nonlinearity and potential omitted variables in [Barras et al. \(2022\)](#), there are still concerns to the linear model. First, the RD2 by [Zhu \(2018\)](#) was estimated across deciles of funds. Although this approach allows for some heterogeneity in the diseconomies-of-scale parameter (b) across funds, it still produces a pooled estimate, whereas ideally the estimation should be conducted at the individual fund level. Second, the linear specification may be subject to model misspecification. Thus, we further employ a nonparametric approach at the individual fund level to enhance robustness to address these limitations.

Specifically, to examine the relationship between alpha and AUM, we adopt a two-step nonparametric procedure that allows for flexible nonlinear estimation while imposing

monotonicity. In the first step, we use kernel regression to estimate the nonlinear relation between alpha and AUM. Similar nonparametric methods have also been applied in finance, for example, the estimation of state-price density in [Aït-Sahalia and Lo \(1998\)](#). While kernel regression provides a flexible fit, it does not guarantee monotonicity in the estimated relationship. Because the nature of diseconomies of scale requires the estimated relationship to be negatively monotonic, we apply isotonic regression as a post-processing step. Details of the estimation procedure are provided in [Appendix 3A](#).

We estimate the expected alpha ($\alpha_{i,t}$) for each fund at t using the nonparametric monotone regression, and we calculate the luck ($\epsilon_{i,t}$) as the difference between realized alpha ($r_{i,t}$) and expected alpha ($\alpha_{i,t}$). [Table 25](#) is constructed in the same way as [Table 24](#), but using expected alpha and luck estimated from nonparametric regression. The results in [Table 25](#) are generally consistent with those from the linear model, but the nonparametric approach provides a slightly better fit of expected alpha to realized alpha, as it allows greater flexibility. Moreover, in this specification, expected alpha is estimated at the individual fund level rather than relying on pooled estimates. Nevertheless, even though the nonparametric method achieves a better fit relative to the linear model, it can at most explain 55% of the variation in realized alpha. In addition, the correlation between expected alpha estimated from the linear and nonparametric models is about 30%. Although this indicates some degree of divergence between the two approaches, the relatively small variation in fund size compared with the variation in alpha implies that such differences have limited impact on the estimation of luck. As expected, the correlation between the luck components derived from the two models is as high as 98%. The properties of luck under the nonparametric model are reported in [Appendix 3B](#) and [Appendix 3C](#), consistent with our earlier results, luck is zero-sum and independent over time.

Table 25: Realized Alpha, Expected Alpha and Luck - Nonparametric Model

The expected alpha α_{it}^{nonl} for fund i in month t is estimated using nonparametric kernel regression on fund alpha against AUM. The luck component ϵ_{it}^{nonl} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}^{nonl}$. Panel A presents the R^2 of simple linear regressions, where the dependent variable is the average historical realized alpha ($\bar{r}_{i,t-m,t}$), and the independent variable is the average historical expected alpha ($\bar{\alpha}_{i,t-m,t}$) over window m . Panel B presents the R^2 of simple linear regressions, where the dependent variable is the average historical realized alpha ($\bar{r}_{i,t-m,t}$), and the independent variable is the average historical luck ($\bar{\epsilon}_{i,t-m,t}$) over window m . Gross alpha against Morningstar benchmark is used in Panel A.1 and B.1, whereas gross 4 factor alpha is used in Panel A.2 and B.2.

Panel A: Realized Alpha and Expected Alpha					
A.1: $\bar{r}_{i,t-m,t}^{g,MS} = \beta_0 + \beta_1 \bar{\alpha}_{i,t-m,t}^{g,MS,nonl} + e$					
m=	1m	12m	24m	36m	60m
R^2	0.68%	4.90%	8.12%	10.70%	13.96%
	Year-Month FE: No		Fund FE: No		
R^2	1.00%	5.63%	7.69%	8.60 %	8.97%
	Year-Month FE: Yes		Fund FE: Yes		
A.1: $\bar{r}_{i,t-m,t}^{g,4F} = \beta_0 + \beta_1 \bar{\alpha}_{i,t-m,t}^{g,4F,nonl} + e$					
m=	1m	12m	24m	36m	60m
R^2	0.50%	4.97%	8.50%	11.04%	14.45%
	Year-Month FE: No		Fund FE: No		
R^2	0.64%	4.14%	5.74%	6.41 %	6.43%
	Year-Month FE: Yes		Fund FE: Yes		
Panel B: Realized Alpha and Luck					
B.1: $\bar{r}_{i,t-m,t}^{g,MS} = \beta_0 + \beta_1 \bar{\epsilon}_{i,t-m,t}^{g,MS,nonl} + e$					
m=	1m	12m	24m	36m	60m
R^2	96.39%	75.71%	65.98%	60.06%	55.40%
	Year-Month FE: No		Fund FE: No		
R^2	97.11%	80.77%	73.80%	70.00%	68.30%
	Year-Month FE: Yes		Fund FE: Yes		
B.1: $\bar{r}_{i,t-m,t}^{g,4F} = \beta_0 + \beta_1 \bar{\epsilon}_{i,t-m,t}^{g,4F,nonl} + e$					
m=	1m	12m	24m	36m	60m
R^2	95.73%	77.64%	70.96%	67.95%	64.95%
	Year-Month FE: No		Fund FE: No		
R^2	96.53%	84.63%	82.00%	81.72%	83.26%
	Year-Month FE: Yes		Fund FE: Yes		

4.4 Theoretical Framework

After documenting that luck accounts for a substantial share of realized alpha, it becomes essential to model its role in the presence of performance-chasing investors, as luck may mislead investors' capital flows. In this section, we propose a simple theoretical framework to model fund luck under diseconomies of scale, incorporating performance-chasing investors and non-optimizing managers. In a world of diseconomies of scale, and under the assumption that (i) the relationship between size and alpha is linear, and (ii) managers are non-optimizing, i.e. actively managing all assets regardless of funding status, the expected alpha (α_t) of a fund at time t is given by:

$$\alpha_t = a - bq_{t-1}, \quad (9)$$

where q is the dollar amount of the funds' AUM. The parameter a is the alpha a fund manager earns on the first dollar invested. The parameter b ($b > 0$) captures the decreasing return of scale. The realized alpha, r_t , at time t consists of the expected component α_t and a idiosyncratic component, luck ϵ_t , expressed as:

$$r_t = \alpha_t + \epsilon_t. \quad (10)$$

We model the change in fund size from time $t - 1$ to t as:

$$q_t - q_{t-1} = q_{t-1}(R_t + f(r_t)), \quad (11)$$

where R_t is realized return of the fund which equals realized benchmark return plus realized alpha, i.e. $R_t = B_t + r_t$, and $f(r_t)$ represents the capital injection into the fund from $t - 1$ to t as a percentage of its size, which is a function of contemporaneous alpha. Supported by [Spiegel and Zhang \(2013\)](#), we assume that this injection is a linear function of fund alpha, such that $f(r_t) = c_0 + c_1 r_t$, where $c_1 > 0$.

To obtain a rigorous mathematical derivation, we model contemporaneous performance chasing ($flow_t \equiv f(r_t)$). If performance chasing were instead modeled on lagged alpha ($flow_t \equiv f(r_{t-1})$), the same intuition would hold, however, a formally correct differentiation would require additional assumptions about the joint behavior of luck at t and $t - 1$ (e.g., the dependence between ϵ_t and ϵ_{t-1}). In contrast, modeling contemporaneous performance chasing avoids these auxiliary assumptions. This approach is also empirically reasonable when alpha is computed over a long horizon. In our setting, most propositions are tested with one-year measures of luck or alpha, for which contemporaneous chasing provides a close empirical approximation.

Proposition 1: While naive performance-chasing investors pursue the realized alpha, r_t , they unintentionally allocate capitals in line with the BG model by effectively chasing the expected alpha, α_t . At the same time, they inadvertently chase luck ϵ_t .

Proof:

$$flow_t \equiv f(r_t) = c_0 + c_1 r_t \quad (12)$$

$$\frac{dflow_t}{dr_t} = \frac{dflow_t}{d\alpha_t} = \frac{dflow_t}{d\epsilon_t} = c_1 > 0 \quad (13)$$

Eq. (13) indicates that when investors naively chase performance (r_t), which comprises both expected alpha and luck, they unintentionally allocate capital to funds with positive expected alpha, consistent with the BG model as documented in [Zhu \(2018\)](#), while also inadvertently pursuing luck.

Proposition 2: There is a positive relation between fund luck and fund AUM.

Proof:

The first-order condition of the fund AUM in Eq. (11) with respect to fund luck is:

$$\frac{dq_t}{d\epsilon_t} = q_{t-1}(1 + c_1) > 0. \quad (14)$$

Eq. (14) suggests that funds experienced a period of good luck tend to have larger AUM due to an increase in the value of their stock holdings and newly attracted capital inflows both driven by luck.

Proposition 3: There is a negative relation between fund luck and expected future alpha.

Proof:

The fund's expected alpha at time $t + 1$ is given by:

$$\alpha_{t+1} = a - bq_t. \quad (15)$$

After substituting Eq. (11) into Eq. (15), the expected alpha becomes:

$$\alpha_{t+1} = a - bq_{t-1}(1 + B_t + r_t + c_0 + c_1r_t). \quad (16)$$

The first-order condition of the expected alpha with respect to fund luck is:

$$\frac{d\alpha_{t+1}}{d\epsilon_t} = -bq_{t-1}(1 + c_1) < 0. \quad (17)$$

Eq. (17) suggests that, in a world of diseconomies of scale, fund luck simultaneously increase a fund's AUM and attract new capital inflows, which, in turn, leads to a decline in their future expected alpha.

4.5 Chasing Performance and Future Expected Alpha

4.5.1 Luck, Flow and Size

Proposition 1 suggests that, while investors chase past performance, fund flows appear respond rationally to funding status, as documented by [Zhu \(2018\)](#), and also, chase luck, which is inherently non-persistent. We run simple linear regressions where the dependent variable is fund flow in the following month and the independent variables are historical

funding status (measured by average expected alpha), historical luck (measured by average luck), or historical alpha (measured by average realized alpha) over 12 to 36 months. The results of past performance chasing are reported in columns (1) to (3) of Table 26. In Panel A, the coefficients on the historical funding status are positive and significant, indicating that fund flows react positively to funding status. This provides evidence of investor rationality consistent with the BG model and the empirical findings of [Zhu \(2018\)](#). In Panel B, when the independent variable is historical luck, the coefficients are also positive and significant, suggesting substantial luck chasing. In Panel C, the positive and significant coefficients of historical performance imply performance chasing, consistent with the findings in [Chevalier and Ellison \(1997\)](#) and [Sirri and Tufano \(1998\)](#).

Table 26: Performance-Chasing, Investor Sophistication and Luck-Chasing

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. This table presents the results of simple linear regressions, where the dependent variable is the flow in next month ($flow_{i,t,t+1}$) and the average historical monthly flow ($\bar{flow}_{i,t-m,t}$) over window m . In Panel A, the independent variable is the average historical expected alpha ($\bar{\alpha}_{i,t-m,t}$) over window m . In Panel B, the independent variable is the average historical luck ($\bar{\epsilon}_{i,t-m,t}$) over window m . In Panel C, the independent variable is the average historical realized alpha ($\bar{r}_{i,t-m,t}$) over window m . Net alpha against Morningstar benchmark is used in this table. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: Flow and Expected Alpha						
$Flow = \beta_0 + \beta_1 \bar{\alpha}_{i,t-m,t}^{n,MS} + e$						
Flow=	$flow_{i,t,t+1}$			$\bar{flow}_{i,t-m,t}$		
m=	(1)	(2)	(3)	(4)	(5)	(6)
β_1	3.152***	3.527***	3.719***	2.231***	2.328***	2.394***
	(60.165)	(65.946)	(68.596)	(55.152)	(57.732)	(59.583)
R ²	0.99%	1.24%	1.37%	0.84%	0.92%	0.97%
	Year-Month FE: Yes			Year-Month FE: Yes		
	Style FE: Yes			Style FE: Yes		
Panel B: Flow and Luck						
$Flow = \beta_0 + \beta_1 \bar{\epsilon}_{i,t-m,t}^{n,MS} + e$						
Flow=	$flow_{i,t,t+1}$			$\bar{flow}_{i,t-m,t}$		
m=	(1)	(2)	(3)	(4)	(5)	(6)
β_1	1.423***	1.721***	1.817***	1.095***	1.569***	1.825***
	(71.740)	(59.268)	(51.251)	(55.860)	(52.407)	(48.946)
R ²	2.50%	2.32%	2.06%	2.47%	3.25%	3.54%
	Year-Month FE: Yes			Year-Month FE: Yes		
	Style FE: Yes			Style FE: Yes		
Panel C: Flow and Realized Alpha						
$Flow = \beta_0 + \beta_1 \bar{r}_{i,t-m,t}^{n,MS} + e$						
Flow=	$flow_{i,t,t+1}$			$\bar{flow}_{i,t-m,t}$		
m=	(1)	(2)	(3)	(4)	(5)	(6)
β_1	1.594***	1.973***	2.123***	1.267***	1.850***	2.173***
	(78.097)	(68.303)	(61.798)	(74.818)	(74.232)	(71.679)
R ²	3.20%	3.27%	3.14%	3.41%	4.72%	5.31%
	Year-Month FE: Yes			Year-Month FE: Yes		
	Style FE: Yes			Style FE: Yes		

We report the R^2 of the linear regressions to assess which independent variable best explains the variation in fund flow. The R^2 in Panel C is the highest among the three panels, indicating that past alpha is the primary driver of fund flows rather than funding status or luck. This finding aligns with the notion of naïve performance chasing (Ben-David et al. (2022a)).

Since the past performance window is relatively long (12 - 36 months), investors may also exhibit contemporaneous performance chasing. To capture this, we also construct average fund flow over the same period and use it as the dependent variable. This approach follows Chevalier and Ellison (1997), who finds that contemporaneous performance chasing has greater explanatory power than lagged performance chasing.⁴³ The results for contemporaneous performance chasing, reported in columns (4) to (6) of Table 26, are consistent with those for past performance chasing shown in columns (1) to (3). In particular, investors react positively to historical funding status, historical luck, and historical alpha, but historical alpha continues to have stronger explanatory power for contemporaneous flow than funding status and luck.

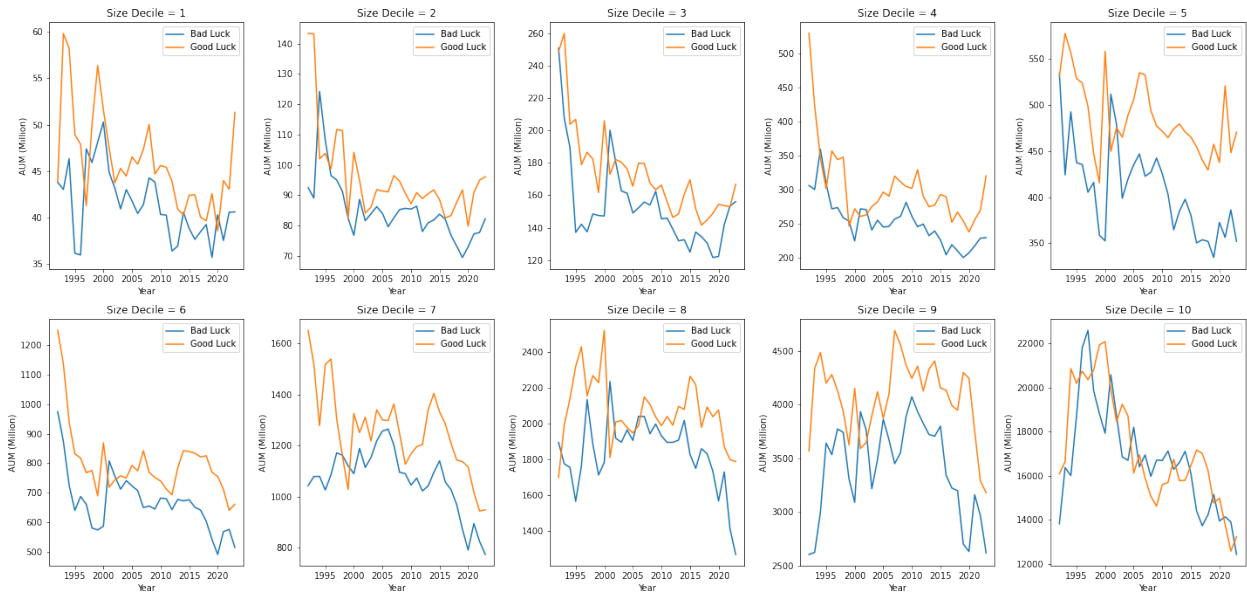
We further visualize the luck-chasing phenomenon for each year from 1990 to 2023. Each month, we calculate fund' average luck over the past 12 months ($\bar{\epsilon}_{i,t-12,t}$) using Morningstar benchmark-adjusted net alpha. Funds with positive $\bar{\epsilon}_{i,t-12,t}$ are defined as having good luck, and those with negative $\bar{\epsilon}_{i,t-12,t}$ are defined as having bad luck. We calculate the mean flow among good-luck and bad-luck funds each month, then plot the annual averages. Appendix 3D (a) plots results using the next month flow (past performance chasing), whereas Appendix 3D (b) uses the flow over the previous 12 months (contemporaneous performance chasing). As shown in Appendix 3D, in each year from 1990 to 2023, funds with good luck in the past 12 months experience higher flows than those with bad luck. Since luck is, by definition, zero-sum and independent over time, and theoretically even negatively predictive of future alpha, the observed luck-chasing phenomenon provides direct evidence of investor naivety.

⁴³See Table 2 of Chevalier and Ellison (1997)

Proposition 2 predicts that funds experiencing good luck tend to have greater AUM owing both to the holdings appreciation and to capital inflows. We plot the average AUM of funds experienced good luck and bad luck for each year from 1990 to 2023 in Figure 8. Each month, we calculate funds' average luck over the past 12 months ($\bar{\epsilon}_{i,t-12,t}$) using Morningstar benchmark-adjusted net alpha. Funds with good (bad) luck are defined as those with positive (negative) $\bar{\epsilon}_{i,t-12,t}$. For each size decile, we calculate the mean AUM among good- and bad-luck funds each month, and plot the annual averages. As shown in the Figure 8, in each year from 1990 to 2023, funds with good luck in the previous 12 months consistently have greater AUM than those funds with bad luck across all size deciles, except for decile 10. We attribute the weaker relationship between luck and size in decile 10 to the lower volatility of large funds' alphas compared with smaller funds (Spiegel and Zhang (2013)). When alpha volatility is low, the difference in past performance between lucky and unlucky large funds is small. Consequently, the AUM gap between them is limited due to smaller portfolio appreciation and reduced performance chasing.

Figure 8: Luck and AUM by Year

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. Each month, we calculate funds' average luck over the past 12 months ($\bar{\epsilon}_{i,t-12,t}$) using Morningstar benchmark adjusted net alpha. Funds with positive $\bar{\epsilon}_{i,t-12,t}$ are labeled as "Good Luck," while those with negative $\bar{\epsilon}_{i,t-12,t}$ are labeled as "Bad Luck." Funds are sorted based on their average AUM over its life into deciles. For each size decile, we plot the average AUM (in millions) of Good Luck and Bad Luck funds from 1993 to 2019.



We also report regression analysis in Appendix 3E. As shown in Panel A, a fund's past luck positively and significantly predicts its AUM. This result is robust across alternative window lengths and remains consistent after including fund and time fixed effects. The relationship becomes even stronger after controlling for fund-size deciles in Panel B.

4.5.2 Luck and Future Expected Alpha

Proposition 3 of our model predicts that funds experiencing periods of good luck tend to underperform in the future. In Table 27, Panel A, we run a simple linear regression with future net alpha as the dependent variable and average luck over the past 12 months as the independent variable. The results show that a fund's past luck significantly and negatively predicts its alpha over the next 12 and 24 months. These findings are robust

after controlling for style and time fixed effects. Additionally, we examine the impact of past luck on future value added, with similar results reported in Panel B. This suggests that the negative relationship between past luck and future performance extends beyond alpha, underscoring the importance of accounting luck in performance evaluation.

Table 27: Luck, Future Alpha and Value Added

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. Panel A presents the results of the regressions, where the dependent variable is the average future realized alpha ($\bar{r}_{i,t+m,t}$) over window m , and the independent variable is the average historical luck ($\bar{\epsilon}_{i,t-1y,t}$) over past 1 year. Panel B presents the results of the regressions, where the dependent variable is the average future realized value added ($\bar{V}_{i,t+m,t}$) over window m , and the independent variable is the average historical luck ($\bar{\epsilon}_{i,t-1y,t}$) over past 1 year. The value added ($V_{i,t}$) is the product of lagged AUM in millions ($q_{i,t-1}$) and alpha ($r_{i,t}$). Net alpha against Morningstar benchmark is used in this table. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: Luck and Future Alpha				
$\bar{r}_{i,t,t+m}^{n,MS} = \beta_0 + \beta_1 \bar{\epsilon}_{i,t-12m,t}^{n,MS} + e$				
m=	12m		24m	
β_1	-0.012*** (-4.983)	-0.031*** (-12.704)	-0.057*** (-33.309)	-0.069*** (-40.136)
Year-Month FE	No	Yes	No	Yes
Style FE	No	Yes	No	Yes
R ²	0.02%	0.08%	0.43%	0.63%
Panel B: Luck and Future Value Added				
$\bar{V}_{i,t,t+m}^{n,MS} = \beta_0 + \beta_1 \bar{\epsilon}_{i,t-12m,t}^{n,MS} + e$				
m=	12m		24m	
β_1	-49.641*** (-2.748)	-134.560*** (-7.358)	-180.302*** (-14.163)	-239.669*** (-18.071)
Year-Month FE	No	Yes	No	Yes
Style FE	No	Yes	No	Yes
R ²	0.01%	0.02%	0.07%	0.12%

To better understand the mechanism underlying the negative relation between luck and future performance, we design a simulation to numerically decompose this relationship. Simulation-based approaches have been widely applied in recent studies of mutual fund and

hedge fund performance,⁴⁴ and our procedure primarily follows Rossi (2019). The goal is to generate a panel of mutual fund data in which key variables —such as fund alpha, flow, and size —not only match the empirical distributions, but also reflect core economic mechanisms observed in real-world data, including diseconomies of scale, performance chasing, and portfolio (de)appreciation driven by fund returns. The simulated data depict a hypothetical world consistent with the framework of this paper, allowing us to examine whether the negative relation between luck and future alpha observed in real data can be explained within such a controlled setting. Moreover, the simulations enable us to isolate the contribution of each mechanism to this negative relationship.

The simulation process involves estimating two models from the actual data: one capturing diseconomies of scale and the other capturing performance chasing. First, we estimate the diseconomies of scale model following Zhu (2018), as discussed in the data section. Specifically, this model allows us to assign each fund a unique parameter a_i , representing its ability to generate abnormal returns on the first dollar invested, and a scalability parameter b_i , estimated across ten fund-size deciles. Once the parameters a_i and b_i are estimated for each fund, we use them together with lagged fund size (q_{t-1} from time $t - 1$) to extract the luck component (ϵ_{it}) of fund alpha at time t . These estimated luck components are stored for use in the bootstrapping simulation.

Second, we estimate performance chasing using a simple pooled OLS regression, where the dependent variable is the monthly fund flow and the independent variable is the average alpha over the past 12 months, expressed as:

$$flow_{i,t} = \lambda_0 + \lambda_1 \bar{r}_{i,t-13,t-1}^{n,MS} + e_{i,t}. \quad (18)$$

In this specification, each fund is associated with an intercept parameter (λ_0),⁴⁵ capturing

⁴⁴Kosowski et al. (2006), Fama and French (2010), and Harvey and Liu (2022) use simulation to study mutual fund alphas and Kosowski et al. (2007) use simulation to study hedge fund alphas.

⁴⁵Rather than relying on the constant estimate from the pooled OLS regression, we also use the average

systematic flow over the sample period, and a sensitivity parameter (λ_1), capturing the degree of performance chasing by investors. Both parameters are assumed constant across funds. After estimating λ_0 and λ_1 , we combine them with the historical alpha over the past 12 months to compute the residual flow ($e_{i,t}$) at time t following Eq. (18), which is then stored for bootstrapping.

For each fund, the simulation proceeds recursively, starting from its first observation at time t_0 . At t_0 , the fund's initial size is set to its actual size from the data, and its alpha is initialized to zero. The simulation begins at time t_1 , where fund's alpha is computed as $r_{i,t_1} = a_i - b_i \times q_{t_0} + \epsilon_{t_1}$, where q_{t_0} is the lagged size and ϵ_{t_1} is bootstrapped from the fund's actual lifetime data. Fund flow at t_1 is estimated using Eq. (18), with the residual flow ($e_{i,t}$) also bootstrapped from the same fund's data. For the first 12 observations, where a 12-month average alpha cannot be computed, we use an expanding average as a proxy. Once alpha and flow are obtained at t_1 , the fund size is updated as: $q_{i,t_1} = q_{i,t_0} \times (1 + B_{t_1} + r_{i,t_1} + \text{Flow}_{i,t_1})$, where B_{t_1} is the Morningstar benchmark return at t_1 . This recursive process continues for all subsequent time periods for each fund.

To isolate the role of different mechanisms —namely, diseconomies of scale, performance chasing, and portfolio (de)appreciation due to fund returns—that contribute to the negative relation between luck and future performance, we must mute each mechanism in turn. In Simulation 1 (Sim1), we set the parameter $b_i = 0$, effectively removing diseconomies of scale. The simulated data under this setting represent a world where fund alpha is independent of fund size. In Simulation 2 (Sim2), we set the performance chasing sensitivity parameter $\lambda_1 = 0$, eliminating investor responses to past alpha. In this scenario, flows depend only on a market-wide systematic flow component and an idiosyncratic component bootstrapped from actual data. In Simulation 3 (Sim3), we modify the size-updating equation to: $q_{i,t} =$

fund flow over each fund's lifetime in the actual data as a proxy for the regression's intercept. This approach allows each fund to have a unique $\lambda_{0,i}$ parameter. The simulation result is nearly identical to our reported result.

$q_{i,t-1} \times (1 + B_t + \text{flow}_t)$, thereby constructing a world where fund alpha does not affect portfolio size. Intuitively, this implies that managers passively invest the increment of portfolio value resulting from positive alpha, while any asset depreciation due to negative alpha is offset by the fund manager through borrowing. Simulation 4 (Sim4) is our main specification, in which all three mechanisms —diseconomies of scale, performance chasing, and alpha-induced portfolio growth —are incorporated simultaneously.

We generate simulation datasets corresponding to Simulation 1 to 4 and perform the same regression analysis as in Table 27. We then compare the regression coefficients obtained from the simulated data with those from the actual data. The results are reported in Table 28. Panel A presents results from a single simulation run. In Simulation 1, the coefficient is statistically indistinguishable from zero, suggesting that the negative relation between luck and future performance disappears in a world without diseconomies of scale. In Simulations 2 and 3, the coefficients are -0.005 and -0.008, respectively. Both are significantly negative, but are smaller in magnitude than the -0.012 estimated from actual data. This suggests that when diseconomies of scale are present, both performance chasing and alpha-induced portfolio growth independently contribute to the negative relation between luck and future performance, albeit to a lesser extent. Finally, in Simulation 4, where all three mechanisms are included, the coefficient is -0.014, closely approximating the coefficient estimated from actual data. The t -statistics are generally larger in the simulated data reflecting reduced noise in the simulated environment by design.

Table 28: Luck and Future Alpha: Simulation Evidence

This table reports the simulation results examining how diseconomies of scale, performance chasing, and portfolio (de)appreciation jointly contribute to the negative relationship between past luck and future performance. Sim1 represents a world without diseconomies of scale, Sim2 represents a world without performance chasing, Sim3 represents a world without portfolio (de)appreciation, and Sim4 is the main simulation incorporating all three mechanisms. Panel A reports the coefficient (β_1) from a single simulation under each scenario (Sim1 to Sim4), measuring the relation of historical luck, defined as the average residual return ($\bar{\epsilon}_{i,t-12m,t}$) over the past 12 months on future realized alpha ($\bar{\epsilon}_{i,t+m,t}$). These estimates are compared to coefficient from actual data. Panel B shows the distribution of β_1 across 1000 simulations for each of the four settings. Net alpha against Morningstar benchmark is used in this table. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: Result from 1 simulation					
$\bar{r}_{i,t,t+12m}^{n,MS} = \beta_0 + \beta_1 \bar{\epsilon}_{i,t-12m,t}^{n,MS} + e$					
	Year-Month FE: No	Style FE: No			
	Data	Sim1	Sim2	Sim3	Sim4
Diseconomies of scale	n.a.	No	Yes	Yes	Yes
Flow _{t+1} =f(alpha)	n.a.	Yes	No	Yes	Yes
Portfolio (de)appreciation	n.a.	Yes	Yes	No	Yes
β_1	-0.012*** (-4.983)	0.001 (0.952)	-0.005*** (-2.728)	-0.008*** (-4.554)	-0.014*** (-8.156)
Panel B: Distribution of β_1 from 1000 simulations					
Percentiles	5th	10th	50th	90th	95th
β_1 from Sim1	-0.007	-0.006	-0.000	0.006	0.008
β_1 from Sim2	-0.014	-0.012	-0.006	-0.000	0.001
β_1 from Sim3	-0.017	-0.016	-0.009	-0.004	-0.002
β_1 from Sim4	-0.023	-0.021	-0.015	-0.010	-0.009

Panel B presents the distribution of the estimated coefficients across 1,000 simulations. In Simulation 1, the 5th percentile of the coefficient distribution is -0.007, indicating that in a world without diseconomies of scale, it is highly unlikely to observe a coefficient as negative as -0.012. In Simulation 2, the -0.012 threshold lies at the 10th percentile, suggesting that even with diseconomies of scale but without performance chasing, there is only a 10% chance of observing a Sim1 coefficient of this magnitude. In Simulation 3, the -0.012 value falls roughly between the 10th and 90th percentiles, but remains larger than the median (-0.009). This indicates that alpha-induced portfolio growth contributes to the negative relation between

luck and future performance, though to a lesser extent than performance chasing. The mechanical relation between portfolio appreciation and realized alpha is 1, whereas in the actual data, the sensitivity of flow to realized alpha is greater than 1, as shown in Table 26 Panel C. Thus, the alpha-induced flow injection is indeed greater than alpha-induced portfolio appreciation in our actual data.

While the results from 1,000 simulation suggest that diseconomies of scale and performance chasing can statistically account for the observed negative relation, alpha-induced portfolio growth is an unavoidable component in real-world fund dynamics. In Simulation 4, where all three mechanisms are present, the median of the coefficient shifts to -0.015, and the -0.012 value observed in the actual data falls within the 10th to 90th percentile range. This indicates that the combined effect of all three mechanisms plausibly explain the negative relation between luck and future performance observed in real data.

Finally, we have demonstrated that a large portion of past alpha is attributed to luck, and luck negatively predicts future performance. However, distinguishing whether a fund's outperformance stems by luck or underfunding is difficult without knowing the fund's a and b parameters, which require the fund's full sample for estimation. This raises an empirical question: what are the consequences of chasing past alpha? Are performance-chasing investors more likely to pick underfunded funds, or are they merely chasing luck? Results in Appendix 3F show that, in the cross-section of funds, the impact of funding status dominates that of luck, whereas over a fund's life, the effect of luck dominates. These results are consistent with the findings in Table 24, which demonstrate that the explanatory power of funding status is significantly weaker within funds than across funds.

4.6 Capital Misallocation and Skill

4.6.1 Luck and Capital Misallocation

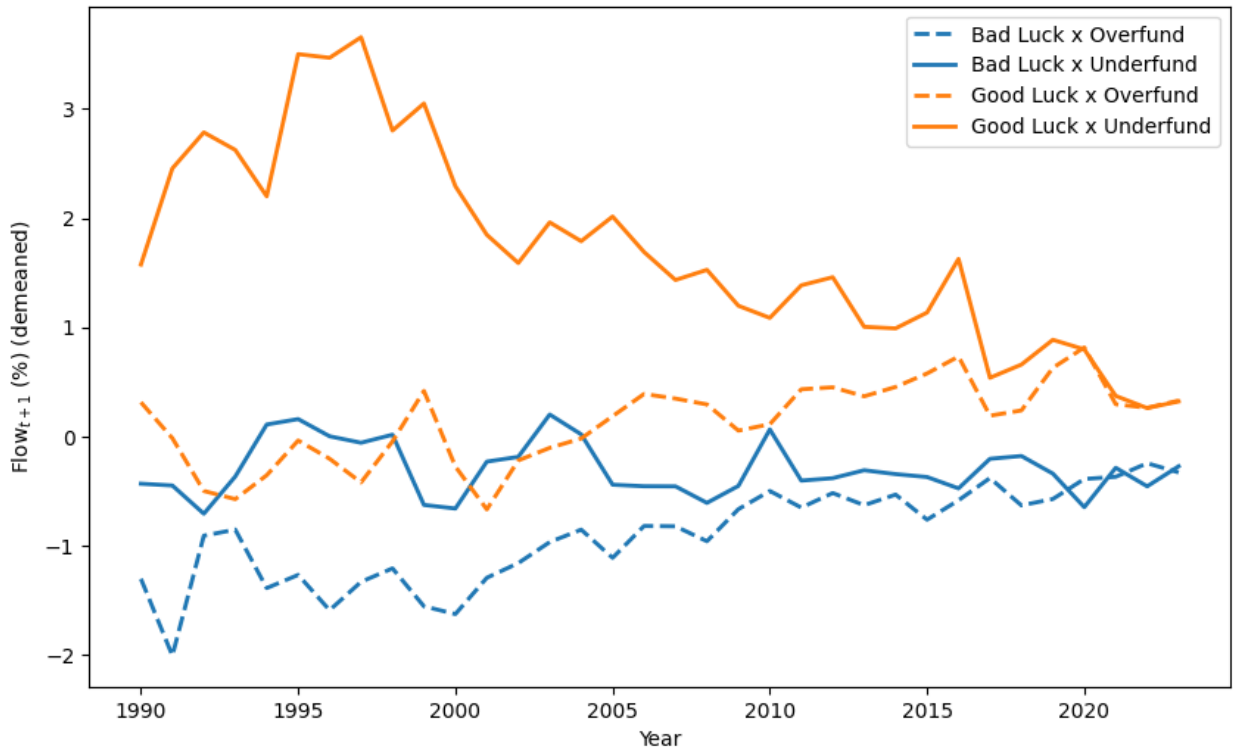
Chasing alpha leads not only to luck chasing but also to rational chasing of funding status, as documented in this paper and in Zhu (2018). Hence, funds should adjust toward their

optimal size only gradually. This slow adjustment toward optimality has been documented by [Barras et al. \(2022\)](#), who argue that such gradual convergence may be driven by investor learning, while uncertainty about managerial skill and scalability contributes to persistent capital misallocation. We argue that luck is an important source of this “slow learning,” which in turn leads to capital misallocation. As predicted by Proposition 3, luck negatively predicts future expected alpha. Recall that within the BG framework, expected alpha serves as a proxy for funding status. Therefore, Proposition 3 also implies that luck negatively affect funding status in the next period. Given luck dominates the variation of realized alpha, Proposition 3 implies potential capital misallocation. Ideally, or under BG model, overfunded(underfunded) funds at time t , i.e. funds with negative(positive) α_t , should experience capital outflows(inflows), driving to zero α_{t+1} in equilibrium. However, since luck is idiosyncratic, overfunded(underfunded) funds can be even more overfunded(underfunded) in the next period merely due to good(bad) luck.

Each month, we independently sort funds into four groups (2×2) based on luck (good luck or bad luck) and funding status (overfunded or underfunded). We then report the average flow in the subsequent month($flow_{i,t,t+1}$) in Figure 9. We find that overfunded funds still receive inflows if they experience good luck in the previous year (orange dashed line), while underfunded funds continue to experience outflows after bad luck (blue solid line). This pattern of capital misallocation driven by luck presents throughout most of our sample period. Because alpha contains a large zero-sum component of luck and only a small component attributable to funding status, naively chasing alpha can lead to short-term capital misallocation. However, such misallocation tends to dissipate overtime, resulting in rational capital allocation in the long run, as documented in [Barras et al. \(2022\)](#).

Figure 9: Luck and Capital Misallocation

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. Each month, we calculate funds' average luck over the past 12 months ($\bar{\epsilon}_{i,t-12,t}$) using Morningstar benchmark adjusted alpha. Funds with positive $\bar{\epsilon}_{i,t-12,t}$ are labeled as "Good Luck," while those with negative $\bar{\epsilon}_{i,t-12,t}$ are labeled as "Bad Luck." A fund is classified as Underfunded (Overfunded) if it has positive (negative) expected alpha α_{it} at time t . Each month, funds are independently sorted into four groups (2x2) based on luck and funding status. We then report the average flow in the subsequent month ($flow_{i,t,t+1}$). To focus on cross-sectional variation, flows are demeaned within each month. The graph plots the annual average of flows from 1990 to 2023. Net alpha against Morningstar benchmark is used in this figure.



4.6.2 Value Added as Skill

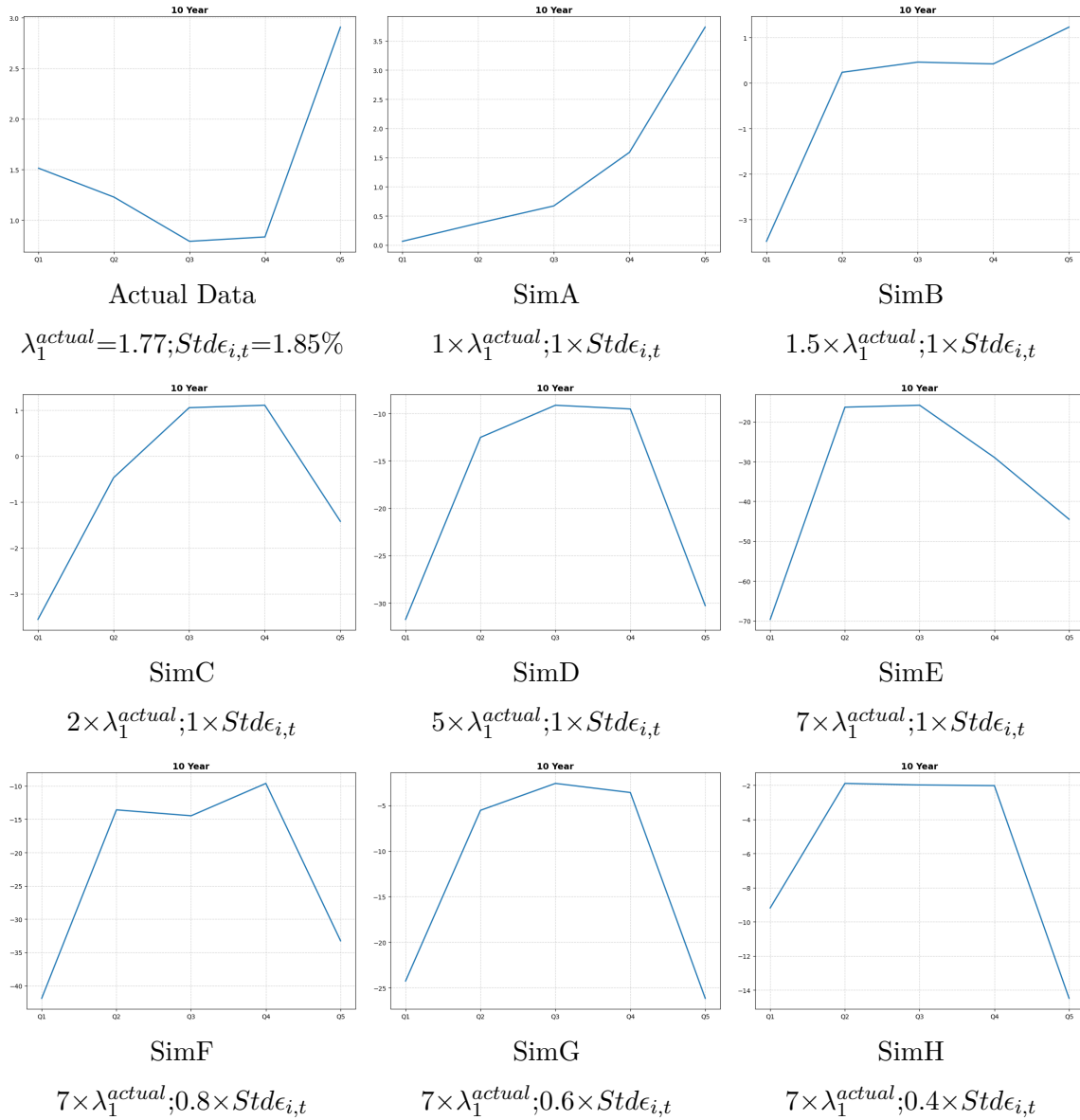
Berk and Van Binsbergen (2015) demonstrate that under the neoclassical assumptions, where managers optimize, markets are competitive, and investors are rational, value added provides a valid measure of managerial skill. They find that value added can persist for up to ten years, and argue that this persistence constitutes the strongest evidence for the

existence of skill. The persistence of value added in [Berk and Van Binsbergen \(2015\)](#) suggests that historical value added can be used to identify skilled managers, which appears to contradict our message that backward-looking measures are unreliable. To reconcile this, we test whether similar persistence can also arise in a world where investors are naive, alpha is noisy, and managers are unaware of their own limits.

First, we replicate the key findings on the persistence of value-added reported by [Berk and Van Binsbergen \(2015\)](#). At time t , we calculate the historical average of value added from each fund's first observation and sort funds into quintiles. We include only funds with at least ten years of observations to ensure that luck averages out to zero within the estimation window. We then report the average value added (in 2023 dollars, millions) over the subsequent ten years for each quintile. We find that funds with the highest historical value added (quintile 5) are indeed more likely to add value over the next ten years, consistent with the findings in [Berk and Van Binsbergen \(2015\)](#). This result is presented in the first sub-figure ("Actual Data") of [Figure 10](#).

Figure 10: Value Added in Simulations

This table examines the persistence of value added and overall value added in mutual fund industry in both actual and simulated data. We incorporate all three mechanisms in these simulations: diseconomies of scale, performance chasing, and portfolio (de)appreciation. We estimate Eq. (18) using actual data and then use multiples of λ_1 to generate simulations SimA through SimH. We also down-scale the luck($\epsilon_{i,t}$) in SimF to SimH. For both actual and simulated data, at time t we calculate the historical average of value added and sort funds into quintiles. The average value added (2023 dollar, in millions) from $t + 1$ to $t + 121$ (10 years) is reported for each quintile. Gross alpha against Morningstar benchmark is used.



Next, we use bootstrapping simulations to replicate the actual data while maintaining the same distributions and underlying mechanisms. In the simulated data, managers face the same diseconomies of scale as in the actual data, and investors exhibit the same performance-chasing behavior (λ_1 in Eq.(18)). Importantly, by design, each manager simply invests the amount of capital allocated by naive, performance-chasing investors —managers do not actively optimize beyond the capital they receive. Thus, neither managers nor investors behave fully rationally, in contrast to the assumptions of [Berk and Van Binsbergen \(2015\)](#). Using the simulated data, we conduct the same persistence test as in the actual data. As shown in sub-figure SimA, we observe strong persistence of value added. However, when the performance-chasing coefficient (λ_1) is increased to 1.5, 2, 5, and 7 times in Simulations B through E, the persistence largely disappears.

While we do not claim that skilled managers do not exist, nor that value added cannot capture skill when present, our point is that in a world with naive investors and non-optimizing managers, slow performance chasing can generate the appearance of persistent value added. The intuition is as follows: a fund with persistent positive value added may indeed reflect managerial skill—implying that the manager actively chooses to operate at the optimal size given the limits imposed by a and b , thereby maximizing value extraction. Alternatively, persistent positive value added may arise from an underfunded but non-optimizing manager when investors learn about funding status only gradually and fail to reallocate capital promptly, such a manager remains persistently underfunded which mechanically leading to persistent value added.

A natural question arises: given luck averages out to zero, would more sensitive performance-chasing behavior by investors help the mutual fund industry converge more quickly to the BG equilibrium? Our answer is yes, but at the cost of short-term value destruction. This trade-off is evident in the y-axis of SimA through SimE. As the performance-chasing coefficient (λ_1) increases from 1 to 7 times its actual value, the average value added declines from positive in SimA to increasingly negative in SimB through SimE. The mechanism, discussed

in Section 4.6.1, is intuitive: When alpha chasing intensifies, a fund experiencing a lucky period receives larger inflows, inflating its size, which consequently lead to value destruction. Conversely, a fund experiencing bad luck loses substantial capital; even if it becomes underfunded, it cannot add value because of its diminished size. This mechanism is further confirmed in SimF to SimH, where we reduce the variation of luck. As the variation in luck decreases, alpha chasing becomes more efficient because alpha more accurately reflects funding status. Consistently, we find that overall value destruction diminishes when luck is less volatile.

Another interesting question is: if skilled managers truly add value, but historical value added does not reliably indicate skill, how should value added be used in practice? We argue that value added is better used as an incentive mechanism rather than as a backward-looking performance metric. If value added is tied to compensation, underfunded managers would be incentivized to either borrow capital to invest, as predicted by the neoclassical assumption, or adopt a more aggressive investment strategy with higher a but also higher b . By contrast, overfunded managers would limit fund size or passively manage capital once the capacity constraint is reached. Overall, using value added as an incentive mechanism would motivate underfunded funds to add more value and prevent overfunded funds from destroying value, aligning managerial incentives with optimal fund size.

Furthermore, how many managers actually possess skill? i.e. those with positive a who properly manage fund size based on their awareness of a and b . Although this is not the main focus of our paper, Zhu (2018) provides some insight. Among 3,077 funds studied, 1,559 destroy value due to overfunding. Of the 1,199 funds that create value, 457 do so likely because they are underfunded, and only 210 funds (6.8%) are able to deliver positive value even when overfunded. However, these statistics are ex post and based on the full-sample information. How, then, can skill be assessed ex ante? We argue that qualitative evaluation is necessary, a point we discuss further in Section 4.7.3.

4.7 Luck and Ratings

The financial industry features numerous rating systems that investors rely on to guide investment decisions and allocate capital. Prior research has investigated the predictive power of various ratings for future fund performance. For instance, [Blake and Morey \(2000\)](#) and [Morey \(2003\)](#) finds that Morningstar’s star ratings are not reliable indicators of future outperformance, while [Armstrong et al. \(2019\)](#) argues that qualitative evaluations are essential, showing that Morningstar’s analyst ratings, particularly the “gold medal”, can predict future outperformance. However, existing studies primarily focus on the empirical predictability of different rating methods, while offering limited insights into why these ratings succeed or fail in forecasting performance. In this section, we decompose fund ratings into their exposure to funding status and luck across various rating methods to examine how this components influence predictive performance. We find that the exposure of a rating method to funding status and luck is a crucial determinant of its success or failure in predicting future performance. In this paper, we focus on three rating products produced by Morningstar: the star rating, the medalist rating assigned by human analysts (known as the analyst rating), and the medalist rating generated by machine learning algorithms (known as the quantitative rating).

4.7.1 Rating Statistics

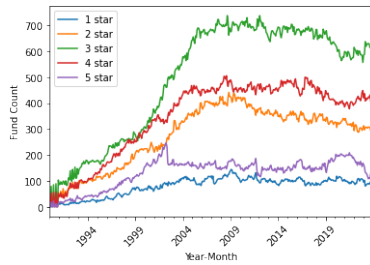
Morningstar offers three primary rating systems to evaluate mutual funds: the star rating, the analyst rating, and the quantitative rating. The star rating, introduced in 1985, is a backward-looking, purely quantitative measure based on a fund’s risk-adjusted past performance relative to its peers. Funds are rated on a scale of one to five stars, with five-star funds representing the top 10% of performers in their category. The analyst rating, launched in 2011, provides a forward-looking, qualitative assessment conducted by Morningstar’s experienced research analysts. This rating evaluates a fund’s long-term potential for delivering superior risk-adjusted returns by analyzing factors such as management quality, investment

process, and fees. Analyst ratings are assigned as Gold, Silver, Bronze, Neutral, or Negative. The coverage of analyst ratings remains limited, as conducting qualitative evaluations of funds is costly and resource-intensive. To improve coverage, in 2017, Morningstar trained a machine learning model using quantitative fund characteristics as inputs from funds with analyst rating, aiming to mimic analyst judgments. Morningstar refers to this new approach as the “quantitative rating” . Morningstar contends that this captures the methodology and reasoning of its analyst team, thereby providing a scalable, forward-looking evaluation.

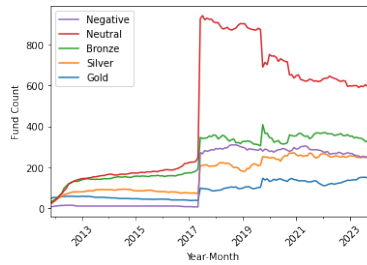
Figure 11 (a) presents statistics on the distribution of the star ratings from 1990 to 2023. The distribution of star ratings is relatively stable: the top 10% of funds (based on past performance) receive 5 stars, the next 22.5% receive 4 stars, the middle 35% receive 3 stars, the following 22.5% receive 2 stars, and the bottom 10% receive 1 star. Figure 11 (b) illustrates the statistics of the medalist ratings (a combination of analyst ratings and quantitative ratings) from 2011 to 2023. Coverage of medalist ratings expanded sharply from 2017 following the introduction of quantitative ratings. Figure 11 (c) focuses on the statistics of analyst medalist ratings from 2011 to 2023. Compared to the star ratings shown in (a), the coverage of analyst medalist ratings is substantially lower and constitutes only a small subset of the overall medalist ratings depicted in (b).

Figure 11: Rating Statistics

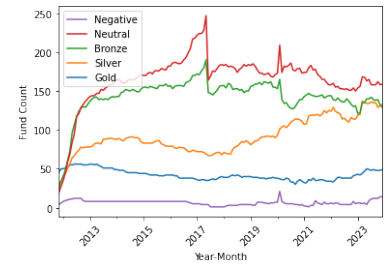
This figure illustrates the number of funds with each Star rating (Panel A) and Medalist rating (Panel B) over time, categorized by year and month. Star rating data is available starting in 1990, while Medalist rating data begins in 2011. The original Medalist rating is a qualitative, forward-looking assessment provided by fund analysts. In 2017, Morningstar introduced the Quantitative Medalist rating, which employs a machine learning algorithm trained on analyst ratings to replicate their evaluations. As a result, the Medalist ratings shown in Panel B include both analyst-assigned ratings and those generated by the machine learning algorithm. Panel C specifically displays the Medalist ratings assigned solely by human analysts.



(a) Star Rating



(b) Medalist Rating (Analyst + Quantitative)



(c) Medalist Rating (Analyst Only)

We report the correlations of ratings with fund characteristics in Table 29. The star rating exhibits a strong correlation with past performance (0.31). In contrast, the analyst rating shows a much weaker correlation (0.06), consistent with Morningstar’s claim that the analyst rating is forward-looking. Notably, the analyst rating has a stronger negative correlation with turnover (-0.27) and a stronger positive correlation with AUM (0.39) compared to the star rating. The quantitative rating demonstrates a stronger correlation with past performance (0.16) than the analyst rating. As expected, due to its higher correlation with past performance, the quantitative rating aligns more closely with the star rating. Across various fund characteristics, the quantitative rating resembles the star rating, differing mainly in its slightly lower correlation with past performance.

Table 29: Correlation between Ratings and Fund Characteristics

This table reports the correlations between three Morningstar ratings—Star Rating, Analyst Medalist Rating, and Quantitative Medalist Rating—and various fund characteristics. For the calculation, ratings are converted to numerical values ranging from 1 to 5, where the lowest rating (1 star for the Star Rating or Negative for Medalist Ratings) is assigned a value of 1, and the highest rating (5 stars for the Star Rating or Gold for Medalist Ratings) is assigned a value of 5.

	Star rating	Medalist rating (analyst)	Medalist rating (quantitative)
Star rating	1	0.34	0.52
AUM	0.19	0.39	0.19
Past 12m net alpha	0.31	0.06	0.16
expense ratio	-0.21	-0.22	-0.21
age	-0.03	-0.03	-0.08
turnover	-0.10	-0.27	-0.11

4.7.2 Star Rating, Funding Status and Luck

In Panel A of Table 30, we report the proportions of funds with positive and negative luck over the past year, along with the corresponding differences within each star-rating category. Among 1-star funds, 30.9% experienced good luck while the remaining 69.1% experienced bad luck, resulting in a luck imbalance of -38.2%. This indicates that bad-luck funds dominate good-luck funds by 38.2% within the 1-star category. Conversely, among 5-star funds, 70.2% experienced good luck and 29.8% experienced bad luck, yielding a luck imbalance of 40.4%, where good-luck funds dominate by this margin. Overall, the star rating exhibits an exposure to past one-year luck of 78.6% (40.4% - (-38.2%)).

Table 30: Star Rating, Luck and Funding Status

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. Panel A presents the proportions of funds with positive and negative luck over the past year ($\bar{\epsilon}_{i,t-12m,t}$), along with the difference within each star rating. Panel B shows the proportions of funds with positive and negative expected alpha (α_{it}) and their differences within each star rating. A fund is classified as Underfunded (Overfunded) if it has positive (negative) expected alpha at time t . Net alpha against Morningstar benchmark is used in this table.

Panel A: Star rating and luck			
	Positive Luck $\bar{\epsilon}_{i,t-12m,t}^{n,MS} > 0$	Negative Luck $\bar{\epsilon}_{i,t-12m,t}^{n,MS} < 0$	Positive-Negative
1 Star	30.9%	69.1%	-38.2%
2 Star	38.3%	61.7%	-23.4%
3 Star	47.4%	52.6%	-5.2%
4 Star	57.8%	42.2%	15.6%
5 Star	70.2%	29.8%	40.4%
Panel B: Star rating and funding status			
	Underfunded $\alpha_{i,t}^{n,MS} > 0$	Overfunded $\alpha_{i,t}^{n,MS} < 0$	Over-Under
1 Star	39.8%	60.2%	20.4%
2 Star	42.6%	57.4%	14.8%
3 Star	47.4%	52.6%	5.2%
4 Star	55.3%	44.7%	10.6%
5 Star	64.0%	36.0%	-28.0%

Panel B of Table 30 reports the proportions of underfunded and overfunded funds within each star-rating category. Among 1-star funds, 39.8% are underfunded while 60.2% are overfunded, resulting in a funding status imbalance of 20.4%. This indicates that overfunded funds dominate underfunded funds by 20.4% within the 1-star category. Among 5-star funds, 64.0% are underfunded while 36.0% are overfunded, leading to a funding-status imbalance of -28.0%, where underfunded funds dominate. Consequently, the star rating displays an overall exposure to funding status of -48.4% (-28.0% - 20.4%).

Since the exposure of star ratings to funding status is smaller in absolute magnitude than their exposure to luck, as shown in Table 30, Proposition 3 predicts that star ratings

should negatively predict future alpha. To test this, we monitor funds' future average alpha and value added over 12 months and 24 months for each star-rating category. As shown in Panel A of Table 31, funds with 5-star ratings significantly underperform 1-star funds in terms of future 12-month and 24-month Morningstar benchmark-adjusted alpha and value added. To control for additional fund characteristics, we apply a regression approach, and results reported in Panel B of Table 31 remain consistent and robust. These results are broadly in line with Blake and Morey (2000) and Morey (2003), confirming that backward-looking star ratings fail to predict future outperformance, and, in fact, exhibit a negative relation with subsequent returns.

Table 31: Star Ratings and Future Performance

Panel A presents the future monthly alpha and value added over the next 1 and 2 years for each rating category (1 to 5 star). Each month, we calculate the average monthly alpha ($\bar{r}_{i,t,t+12m}$ and $\bar{r}_{i,t,t+24m}$) and value added ($\bar{V}_{i,t,t+12m}$ and $\bar{V}_{i,t,t+24m}$) over the upcoming 1-year and 2-year periods and then average these monthly future values for fund portfolios with 1- to 5-star ratings. The value added ($V_{i,t}$) is the product of lagged AUM in millions ($q_{i,t-1}$) and alpha ($r_{i,t}$). The table reports the time series average of monthly alpha and value added for fund portfolios across the 1- to 5-star ratings. Panel B presents the regression results of the relation between star rating and future performance. Net alpha against Morningstar benchmark is used in this table. The sample covers from 1990 to 2023. The controls are expense ratio, fund age and turnover. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: Portfolio sorting				
	Future Alpha		Future Value Added	
	$\bar{r}_{i,t,t+12m}^{n,MS}$ (%)	$\bar{r}_{i,t,t+24m}^{n,MS}$ (%)	$\bar{V}_{i,t,t+12m}^{n,MS}$ (mil)	$\bar{V}_{i,t,t+24m}^{n,MS}$ (mil)
1 Star	0.025	0.044	-0.267	0.058
2 Star	-0.031	-0.015	-0.993	-0.670
3 Star	-0.050	-0.036	-2.717	-2.538
4 Star	-0.055	-0.055	-6.290	-7.213
5 Star	-0.050	-0.056	-13.768	-11.222
5-1 Star	-0.075***	-0.100***	-8.583***	-9.266***
	(-4.336)	(-7.109)	(-4.755)	(-5.350)
Panel B: Regression				
$Perf = \beta_0 + \beta_1 STAR_{i,t} + e$				
	Future Alpha		Future Value Added	
Perf =	$\bar{r}_{i,t,t+12m}^{n,MS}$ (%)	$\bar{r}_{i,t,t+24m}^{n,MS}$ (%)	$\bar{V}_{i,t,t+12m}^{n,MS}$ (mil)	$\bar{V}_{i,t,t+24m}^{n,MS}$ (mil)
β_1	-0.031	-0.032	-1.806	-1.810
	(-25.635)	(-33.575)	(-13.253)	(-18.253)
Control	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes
Style FE	Yes	Yes	Yes	Yes
R ²	0.35%	0.49%	0.09%	0.16%

4.7.3 Quantitative versus Qualitative

Similar to Table 30, we analyze the exposure of analyst ratings and quantitative ratings to luck and funding status in Table 32. Given the extremely small number of funds rated as “negative”, as reported in Table 11, we combine “negative” and “neutral” categories into a single group labeled “not recommended”, following Armstrong et al. (2019). Panel A

of Table 32 reports the proportions of funds with positive and negative luck within each analyst rating. Among “not recommended” funds, 48.5% experienced bad luck while 51.5% experienced good luck, resulting in a luck imbalance of -3.0%, indicating that bad-luck funds slightly dominate. In contrast, among “gold” funds, 52.9% experienced good luck and 47.1% experienced bad luck, yielding a luck imbalance of 5.8%, where good-luck funds dominate. Overall, the analyst rating exhibits an exposure to past one-year luck of 8.8% (5.8% - (-3.0%)).

Table 32: Medalist Rating, Luck and Funding Status

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i + b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. Funds are classified into three groups: funds with a medalist rating assigned by analysts, funds with a medalist rating assigned by machine learning, and funds without a medalist rating. Fund luck and expected alpha are demeaned within each group in each month. Panel A presents the proportions of funds with positive and negative luck over the past year ($\bar{\epsilon}_{i,t-12m,t}$), along with the difference within each medalist rating by analysts. Panel B shows the proportions of funds with positive and negative expected alpha (α_{it}) and their differences within each medalist rating by analysts. A fund is classified as Underfunded (Overfunded) if it has positive (negative) expected alpha at time t . Panels C and D mirror Panels A and B, respectively, but are based on the subsample of medalist ratings generated by the machine learning algorithm. Net alpha against Morningstar benchmark is used in this table.

Panel A: Medalist rating (analyst) and luck			
	Positive Luck $\bar{\epsilon}_{i,t-12m,t}^{n,MS} > 0$	Negative Luck $\bar{\epsilon}_{i,t-12m,t}^{n,MS} < 0$	Positive-Negative
Not Recommend	48.5%	51.5%	-3.0%
Bronze	50.7%	49.3%	1.4%
Silver	50.3%	49.7%	0.6%
Gold	52.9%	47.1%	5.8%
Panel B: Medalist rating (analyst) and funding status			
	Underfunded $\alpha_{i,t}^{n,MS} > 0$	Overfunded $\alpha_{i,t}^{n,MS} < 0$	Over-Under
Not Recommend	46.3%	53.7%	7.4%
Bronze	49.4%	50.6%	-1.2%
Silver	54.8%	45.2%	-9.6%
Gold	56.9%	43.1%	-13.8%
Panel C: Medalist rating (quantitative) and luck			
	Positive Luck $\bar{\epsilon}_{i,t-12m,t}^{n,MS} > 0$	Negative Luck $\bar{\epsilon}_{i,t-12m,t}^{n,MS} < 0$	Positive-Negative
Not Recommend	46.8%	53.2%	-6.4%
Bronze	54.7%	45.3%	9.4%
Silver	57.5%	42.5%	15.0%
Gold	59.4%	40.6%	18.8%
Panel D: Medalist rating (quantitative) and funding status			
	Underfunded $\alpha_{i,t}^{n,MS} > 0$	Overfunded $\alpha_{i,t}^{n,MS} < 0$	Over-Under
Not Recommend	48.1%	51.9%	3.8%
Bronze	53.0%	47.0%	-6.0%
Silver	54.0%	46.0%	-8.0%
Gold	56.0%	44.0%	-12.0%

Panel B of Table 32 examines the proportions of underfunded and overfunded funds across analyst ratings. Within the “not recommended” category, 46.3% of funds are underfunded, while 53.7% are overfunded, leading to a funding status imbalance of 7.4%, indicating overfunded funds dominate. For “gold” funds, 56.9% are underfunded and 43.1% are overfunded, resulting in a funding status imbalance of -13.8%, where underfunded funds dominate. Consequently, the analyst rating demonstrates an overall exposure to funding status of -21.2% (-13.8% - 7.4%).

Since the exposure of analyst ratings to funding status is greater than their exposure to luck in absolute magnitude, as shown in Table 32, Proposition 3 predicts that analyst ratings might positively predict future alpha. To test this, we analyze funds’ future 12-month and 24-month alpha and value added for ratings ranging from “not recommended” to “gold”. As shown in Panel A of Table 33, funds with a “gold medal” significantly outperform “not recommended” funds in terms of future 12-month and 24-month, as measured by Morningstar benchmark-adjusted alpha. However, no such pattern is observed when future performance is measured by value added, suggesting that analyst ratings are less effective among large funds.

Table 33: Medalist Ratings (analyst) and Future Performance

Panel A presents the future monthly alpha and value added over the next 1 and 2 years for each analyst Medalist rating category (Not Recommend (Negative + Neutral), Bronze, Silver, Gold). Each month, we calculate the average monthly alpha ($\bar{r}_{i,t,t+12m}$ and $\bar{r}_{i,t,t+24m}$) and value added ($\bar{V}_{i,t,t+12m}$ and $\bar{V}_{i,t,t+24m}$) over the upcoming 1-year and 2-year periods and then average these monthly future values for fund portfolios. The value added ($V_{i,t}$) is the product of lagged AUM in millions ($q_{i,t-1}$) and alpha ($r_{i,t}$). The table reports the time series average of monthly alpha and value added for fund portfolios based on Medalist ratings. Panel B presents the portfolio sorting results in subsamples (Large Funds vs Others). We define Large funds as those with AUM in top quintile each month. Net alpha against Morningstar benchmark is used in this table. The sample covers from 2011 to 2023, excluding 2020 due to COVID. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: Medalist rating (analyst) vs Performance				
	Future Alpha		Future Value Added	
	$\bar{r}_{i,t,t+12m}^{n,MS}$ (%)	$\bar{r}_{i,t,t+24m}^{n,MS}$ (%)	$\bar{V}_{i,t,t+12m}^{n,MS}$ (mil)	$\bar{V}_{i,t,t+24m}^{n,MS}$ (mil)
Not Recommend	-0.047	-0.027	-3.600	-3.108
Bronze	-0.038	-0.014	-5.105	-4.548
Silver	-0.042	-0.016	-9.974	-8.395
Gold	-0.004	0.010	-5.972	-2.934
G - NR	0.044***	0.037***	-2.386	-0.174
	(5.422)	(5.242)	(-1.166)	(0.138)
Panel B: Medalist rating (analyst) vs Performance in AUM subsamples				
	Large Funds		Others	
	$\bar{V}_{i,t,t+12m}^{n,MS}$ (mil)	$\bar{V}_{i,t,t+24m}^{n,MS}$ (mil)	$\bar{V}_{i,t,t+12m}^{n,MS}$ (mil)	$\bar{V}_{i,t,t+24m}^{n,MS}$ (mil)
Not Recommend	-7.939	-8.177	-2.067	-1.855
Bronze	-9.029	-8.191	-2.949	-2.283
Silver	-15.535	-12.792	-4.316	-4.012
Gold	-8.614	-5.096	-0.149	0.508
G - NR	-0.675	3.080	1.893***	2.364***
	(-0.207)	(1.094)	(2.969)	(4.034)

Further, we test the predictability of analyst ratings on future value added separately for large funds and other funds. We define “large funds” as those with AUM in the top quintile each month. The results in Panel B of Table 33 show that the positive predictability of analyst ratings is robust among non-large funds but disappears for large funds. We suspect this is because large funds are typically more passive, making it harder for them to significantly outperform or underperform their benchmarks.

Panel C of Table 32 reports the proportions of funds with positive and negative luck within each quantitative rating. Among “not recommended” funds, 46.8% experienced good luck while 53.2% experienced bad luck, resulting in a luck imbalance of -6.4%, indicating that bad-luck funds slightly dominate. In contrast, among gold medal funds, 59.4% experienced good luck and 40.6% experienced bad luck, yielding a luck imbalance of 18.8%, where good-luck funds dominate. Overall, the analyst rating exhibits an exposure to past one-year luck of 25.2% (18.8% - (-6.4%)). Panel D of Table 32 examines the proportions of underfunded and overfunded funds across quantitative ratings. Within the “not recommended” category, 48.1% of funds are underfunded, while 51.9% are overfunded, leading to a funding status imbalance of 3.8%, indicating overfunded funds dominate. For “gold medal” funds, 56.0% are underfunded and 44.0% are overfunded, resulting in a funding status imbalance of -12.0%, where underfunded funds dominate. Consequently, the quantitative rating demonstrates an overall exposure to funding status of -15.8% (-12.0% - 3.8%).

Since the exposure of quantitative ratings to funding status is smaller than their exposure to luck in absolute magnitude, we suspect that, similar to star ratings, quantitative ratings may fail or even negatively predict future alpha. As shown in Table 34, funds with a “gold medal” in quantitative ratings significantly underperform “not recommended” funds in terms of future 12-month and 24-month Morningstar benchmark-adjusted alpha. Additionally, we also find weaker, though consistent evidence of the failure of quantitative ratings when future value added is used as the performance measure.

Table 34: Medalist Ratings (quantitative) and Future Performance

This table presents the future monthly alpha and value added over the next 1 and 2 years for each machine learning generated Medalist rating category (Not Recommend (Negative + Neutral), Bronze, Silver, Gold). Each month, we calculate the average monthly alpha ($\bar{r}_{i,t,t+12m}$ and $\bar{r}_{i,t,t+24m}$) and value added ($\bar{V}_{i,t,t+12m}$ and $\bar{V}_{i,t,t+24m}$) over the upcoming 1-year and 2-year periods and then average these monthly future values for fund portfolios. The value added ($V_{i,t}$) is the product of lagged AUM in millions ($q_{i,t-1}$) and alpha ($r_{i,t}$). The table reports the time series average of monthly alpha and value added for fund portfolios based on Medalist ratings. Net alpha against Morningstar benchmark is used in this table. The sample covers from 2011 to 2023, excluding 2020 due to COVID. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

	Future Alpha		Future Value Added	
	$\bar{r}_{i,t,t+12m}^{n,MS}$ (%)	$\bar{r}_{i,t,t+24m}^{n,MS}$ (%)	$\bar{V}_{i,t,t+12m}^{n,MS}$ (mil)	$\bar{V}_{i,t,t+24m}^{n,MS}$ (mil)
Not Recommend	0.004	0.027	-0.528	-0.675
Bronze	-0.019	-0.014	-0.953	-0.730
Silver	-0.011	0.017	-0.510	-0.273
Gold	-0.060	-0.014	-1.302	-1.679
G - NR	-0.065***	-0.042**	-0.774	-1.004**
	(-3.277)	(-2.403)	(-1.264)	(-2.383)

4.8 Conclusion

It is conceptually challenging to reconcile Berk and Green (2004) with a growing body of research documenting market irrationalities, including evidence of fund managers' lack of skill (Jensen (1968), Fama and French (2010), Barras et al. (2010)), non-optimizing managers (Zhu (2018), Barras et al. (2022)), and naive investors (Ben-David et al. (2022a)). As a result, Fama and French (2010) rejects the equilibrium implications of the BG model for fund returns, while Zhu (2018) and Barras et al. (2022) find that the mutual fund industry has not yet reached the BG equilibrium, although it appears to be gradually converging toward it.

Building on Berk and Green (2004) and Berk and Van Binsbergen (2015), we introduce "luck" into the system of fund dynamics and examines how luck and skill interact in a world characterized by diseconomies of scale, non-optimizing managers, naive investors, and noisy realized alpha beyond managers' control. Our framework bridges the gap between the

idealized BG world and the empirical reality. We show that, in practice, fund managers do possess alpha-generating ability, but often fail to manage fund size optimally because they are unaware of their capacity limits and they cannot control most of their realized alpha due to the influence of luck. Similarly, investors allocate capital in ways that appear sophisticated within the BG framework, but are, in fact, driven largely by naive performance chasing - unintentionally channeling capital towards funds that have been lucky recently.

Due to diseconomies of scale, funds experiencing lucky periods tend to become oversized through portfolio appreciation and inflows, which in turn makes it increasingly difficult for them to outperform. We apply this insight to fund ratings and find that most backward-looking, performance-based ratings primarily capture luck, explaining their limited predictive power. In addition, we show that investors' naive performance-chasing behavior leads to short-term capital misallocation and consequent value destruction. Skill, as argued by [Berk and Van Binsbergen \(2015\)](#), should be proxied by value added as it shows strong persistence. However, we find such persistence can arise in a semi-rational world with non-optimizing managers and naive investors, highlighting the need for caution when using value added as a proxy for skill. We argue that a qualitative, forward-looking approach is essential when evaluating both funding status and managerial skill, as backward-looking proxies based on historical data are inherently unreliable.

As investors have increasingly shift toward passive investing in recent decades, our findings offer important implications for the active mutual fund industry. For investors, the objective is to identify funds with genuine alpha, but it is crucial to recognize that alpha fundamentally depends on funding status and that its realization is heavily influenced by luck. Since luck negatively predicts future performance, investors should avoid relying solely on past performance when selecting underfunded funds. Instead, forward-looking, qualitative assessments can enhance the identification of funds with strong future alpha while promoting more efficient capital allocation toward underfunded funds. From a practical perspective, our results shed light on mutual fund ratings. We show that backward-looking star ratings

primarily capture luck, leading to poor predictability of future performance. By contrast, analyst ratings are genuinely forward-looking, and largely neutral to luck. Although Morningstar's quantitative ratings aim to replicate analyst judgments using machine learning, we find that they remain backward-looking due to their reliance on historical quantitative inputs.

At the industry level, most fund managers possess alpha-generating ability, making active management a worthwhile profession. However, managers are often failed to operate at their optimal size and consequently do not maximize value added, either because they are unaware of their capacity limits or because industry incentives —such as the pursuit of AUM growth —push fund sizes away from their optimal levels. We therefore advocate using value added as an incentive mechanism, rather solely as a performance metric, to encourage managers to operate funds at their optimal scale. By reducing investors' reliance on historical performance and aligning managerial incentives with value creation, the mutual fund industry can progress toward a healthier and more efficient equilibrium.

Appendix 1A: Variable Definition

This table reports the construction of variables used in our study.

Variables Name	Description
VS_m	Visual Scaling measured by average price/ (high price - low price), where average price, high price and low price are estimated from t-m to t.
TO	The average monthly share trading volume divided by the average number of shares outstanding during a 1-year period.
Size (mil)	Market capitalization in millions calculated by multiplying share price with number of share outstanding.
Price	Stock price in dollar.
M/B	Market-to-book ratio measured by Market capitalization/book value.
IO	Institutional ownership is calculated by dividing the number of shares held by institutional investors by the total shares outstanding, ranging from 0% to 100%, from 13-F filings.
Mom	Buy-and-hold returns from past 12 month.
$High_{52w}(Low_{52w})$	$(Price - 52\text{-week High(Low)}) / Price$.
RVol	Return volatility using monthly return from past 12 months.
IVol	Idiosyncratic volatility using monthly return from past 12 months under market model.
Beta	Market beta using monthly return from past 12 months under market model.
MAX	Maximum daily return over the past month following Bali et al. (2011) .
ISkew	Idiosyncratic skewness using daily return from past 1 month under market model.
ST	Salience Theory measure constructed with close-to-close daily returns in a 1-month window following Cosemans and Frehen (2021) .

Appendix 1B: Summary Statistics of Retail Trades

This table reports the summary statistics of the key characteristics of buy trades from the brokerage firm. Trade size is defined as price times quantity. PC is a dummy variable that equals 1 if the trade is computer-based and 0 otherwise. Male is a dummy variable indicating male investors. Young is a dummy variable indicating whether the household head is under 40. Single is a dummy variable for single investors. VS_{6m} (VS_{12m}) is the visual scaling multiplier measured by average price/ (high price - low price), where average price, high price and low price are estimated from t-6m (t-12m) to t. VSE is the mean of VS_{6m} and VS_{12m} .

	No. Obs	Mean	Std Dev	Percentiles		
				25th	Median	75th
Trade Size	179,137	9625.12	21509.93	2425.00	4650.00	9487.50
Price/share	179,137	30.17	112.99	11.50	22.75	40.00
Quantity	179,137	564.16	1444.34	100.00	200.00	500.00
PC	179,137	0.07	0.26	0.00	0.00	0.00
Male	179,137	0.74	0.44	0.00	1.00	1.00
Young	179,137	0.80	0.40	1.00	1.00	1.00
Single	179,137	0.22	0.42	0.00	0.00	0.00
VS_{6m}	179,137	2.63	1.74	1.43	2.15	3.35
VS_{12m}	179,137	1.74	1.15	0.99	1.39	2.18
VSE	179,137	2.18	1.30	1.26	1.80	2.74

Appendix 1C: Model

I introduce a equilibrium model with poorly diversified investors with a portion of them whose risk evaluation depends on depicted volatility. To simplify the model, I consider a market populated by n single-holding investors. This assumption is reasonable, as [Goetzmann and Kumar \(2008\)](#) find that more than 25% of retail investors hold only a single stock, and even when investors attempt to diversify their portfolios by adding more stocks, they fail to select uncorrelated stocks. This suggests that investors struggle to assess stock correlations and instead tend to treat each stock in isolation. In addition, [Barber et al. \(2008\)](#) provide evidence that retail investors' trade impact market prices, supporting modeling naive, single-holding investors as a force in determining asset prices in this study.

Investors can choose to invest stocks indexed by i . n_1 investors correctly observe the true variance σ_i^2 of stock i 's return, while n_2 investors perceive a visually-scaled variance $(VS_{ij}\sigma_i)^2$, where VS_{ij} represents investor j 's degree of visual scaling in assessing the risk of stock i .

Following [Merton et al. \(1987\)](#) and [Malkiel and Xu \(2002\)](#), the utility function for investor j is given by:

$$u(W_j) = \mathbb{E}(W_j) - \frac{1}{2\tau} \mathcal{F}(\text{Var}(W_j)), \quad (19)$$

where $\tau > 0$ is the coefficient of risk aversion and $\mathcal{F}(\cdot)$ represents the subjective distortion function.

The corresponding budget constraint of investor j can be written as:

$$W_j = W_{0,j}(1 + r) + X_{ij}(R_i - r), \quad (20)$$

where $W_{0,j}$ is the initial endowment, r is the risk-free rate, X_{ij} is the position in stock i , and R_i is the return of stock i .

Group 1: Investors with Correct Risk Perception

For the n_1 investors who correctly observe σ_i^2 , their utility maximization problem is:

$$\max_{X_{ij}} \left[W_{0,j}(1 + r) + X_{ij}(\mathbb{E}(R_i) - r) - \frac{1}{2\tau} X_{ij}^2 \sigma_i^2 \right]. \quad (21)$$

Taking the first-order condition (FOC) with respect to X_{ij} and solving for X_{ij} , I obtain the optimal demand:

$$X_{i,1} = \frac{\tau(\mathbb{E}(R_i) - r)}{\sigma_i^2}. \quad (22)$$

Appendix 1C: continue

Group 2: Investors with Visual Risk Perception

For the n_2 investors, who perceive the variance as $(VS_{ij}\sigma_i)^2$, their utility maximization problem is:

$$\max_{X_{ij}} \left[W_{0,j}(1+r) + X_{ij}(\mathbb{E}(R_i) - r) - \frac{1}{2\tau} X_{ij}^2 (VS_{ij}\sigma_i)^2 \right]. \quad (23)$$

Taking the FOC and solving for X_{ij} , I obtain:

$$X_{ij} = \frac{\tau(\mathbb{E}(R_i) - r)}{VS_{ij}^2\sigma_i^2}. \quad (24)$$

Market Demand and Equilibrium

At market equilibrium, the total supply of stock i , denoted as S_i , equals the total demand from both groups:

$$S_i = \sum_{j=1}^{n_{1,i}} X_{i,1} + \sum_{j=1}^{n_{2,i}} X_{ij}. \quad (25)$$

Substituting the expressions for $X_{i,1}$ and X_{ij} and rearranging:

$$\mathbb{E}(R_i) = r + \frac{\sigma_i^2 S_i}{\tau \left[n_{1,i} + \sum_{j=1}^{n_{2,i}} \frac{1}{VS_{ij}^2} \right]}. \quad (26)$$

When $n_{2,i} \rightarrow n_i$

If the visual investors dominate the market such that $n_{2,i} \rightarrow n_i$ and the distortion (visual scaling in my case) VS_{ij} are relatively homogeneous and can be estimated as VSE_i , then the expected return is:

$$\mathbb{E}(R_i) \approx r + VSE_i^2 \frac{S_i}{\tau n_i} \sigma_i^2. \quad (27)$$

In this case, the visual distorted investors determine the equilibrium expected return, thus the equilibrium expected return compensates the depicted risk.

Proposition 1: When visual investors are present, they demand premium for the depicted volatility ($\sigma_i \times VSE$).⁴⁶ Thus, there is a positive relationship between depicted volatility and expected returns.

When $n_{2,i} \rightarrow 0$

If there are no visual investors for stock i , i.e., $n_{2,i} \rightarrow 0$, the expected return simplifies to:

$$\mathbb{E}(R_i) = r + \frac{S_i}{\tau n_i} \sigma_i^2. \quad (28)$$

Appendix 1C: continue

In this case, only the n_1 investors with correct risk perception determine the equilibrium expected return, and it depends solely on their risk-bearing capacity.

Proposition 2: When visual investors are absent, they demand premium for the actual volatility (σ_i), resulting in a positive relationship between actual volatility and expected returns, but there is no premium for depicted volatility.

The covariance between $E(R_i)$ and σ^2 is:

$$\text{Cov}(E(R_i), \sigma^2) = \frac{S}{\tau n} \cdot \text{Cov}(VSE^2 \cdot \sigma^2, \sigma^2), \quad (29)$$

where

$$\text{Cov}(VSE^2 \cdot \sigma^2, \sigma^2) = \mathbb{E}[VSE^2 \cdot \sigma^4] - \mathbb{E}[VSE^2 \cdot \sigma^2] \cdot \mathbb{E}[\sigma^2]. \quad (30)$$

When VSE^2 and σ^2 are independent

If VSE^2 and σ^2 are independent:

$$\mathbb{E}[VSE^2 \cdot \sigma^4] = \mathbb{E}[VSE^2] \cdot \mathbb{E}[\sigma^4]. \quad (31)$$

Thus:

$$\text{Cov}(E(R_i), \sigma^2) = \frac{S}{\tau n} (\mathbb{E}[VSE^2] \cdot \mathbb{E}[\sigma^4] - \mathbb{E}[VSE^2] \cdot \mathbb{E}[\sigma^2]^2). \quad (32)$$

By Jensen's inequality, $\mathbb{E}[\sigma^4] \geq \mathbb{E}[\sigma^2]^2$, ensuring $\text{Cov}(E(R_i), \sigma^2) \geq 0$.

Therefore, when the VSE is independently assigned to stocks, there is positive premium to volatility even with the existence of visual investors.

When VSE^2 and σ^2 are negatively correlated

If VSE^2 and σ^2 are negatively correlated:

$$\mathbb{E}[VSE^2 \cdot \sigma^4] < \mathbb{E}[VSE^2] \cdot \mathbb{E}[\sigma^4]. \quad (33)$$

Substituting this:

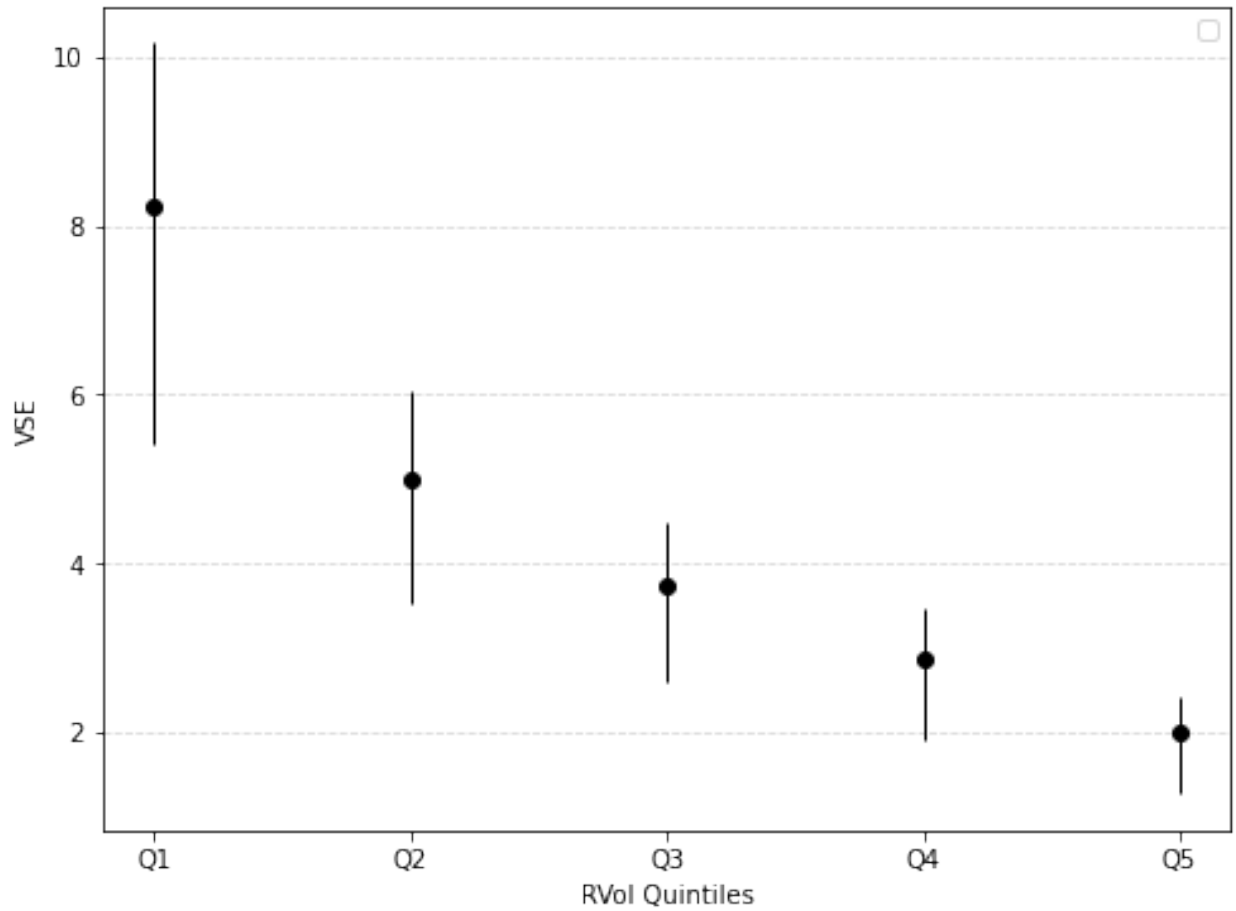
$$\text{Cov}(E(R_i), \sigma^2) < \frac{S}{\tau n} (\mathbb{E}[VSE^2] \cdot \mathbb{E}[\sigma^4] - \mathbb{E}[VSE^2] \cdot \mathbb{E}[\sigma^2]^2). \quad (34)$$

This implies $\text{Cov}(E(R_i), \sigma^2)$ may no longer be positive or even become negative.

Proposition 3: When visual investors are present and the volatility (σ_i) is negatively related to visual scaling (VSE), the risk premium associated with the actual volatility is smaller.

Figure 12: Appendix 1D: Volatility and Visual Scaling Exposure

This figure plots the 25th percentile, mean and 75 percentile of VSE against return volatility quintiles. VSE is the mean of VS_{3m} , VS_{6m} and VS_{12m} .



Appendix 2A: Risk Taking Variables (Part 1)

This table reports the construction of risk taking variables used in our study.

Variables Name	Description
ΔStd	<p>The logarithmic difference of the standard deviation between the future 12-month fund gross return (monthly) and the same Std estimated using past 12 months data.</p> $\Delta Std_{i,t} = \log\left(\frac{\sigma_{i,t+1,t+12}}{\sigma_{i,t-11,t}}\right)$
$\Delta Beta_{MS}$	<p>The logarithmic difference of the beta between the future 12 months fund gross returns with the Morningstar category index and the same beta estimated using past 12 months data.</p> $\Delta Beta_{i,t} = \log\left(\frac{\beta_{i,t+1,t+12}}{\beta_{i,t-11,t}}\right)$
$\Delta IVol_{MS}$	<p>The logarithmic difference of the standard deviation of the future 12 months fund excess returns over the Morningstar category index and the standard deviation estimated using past 12 months excess return.</p> $\Delta IVol_{i,t} = \frac{\sigma_{ben. \text{ excess return},i,t+1,t+12}}{\sigma_{ben. \text{ excess return},i,t-11,t}}$
$\Delta Beta_{HB}$	<p>The logarithmic difference of the beta of the future 12 months fund holding-based returns against the Morningstar category index and the same beta estimated using past 12 months holding-based return. We calculate the holding-based return each month as the value-weighted monthly return of the latest available holdings.</p>
$\Delta Beta_{Average}$	<p>Each month, we acquire funds' latest holding data and next available holding data. We calculate the value-weighted average of stock betas against Morningstar category index, referred to as $Beta_{Average}$ for the two holding data. We compute the logarithmic difference between the two $Beta_{Average,i}$.</p>

Appendix 2A: Risk Taking Variables (Part 2)

Variables Name	Description
$\Delta HHI_{industry}$	Each month, we acquire funds' latest holding data and next available holding data. We calculate the sum of squared dollar weight of each industry (SIC 2-digit), referred to as $HHI_{industry,i}$ for the two holdings. We compute the logarithmic difference between the two $HHI_{Industry,i}$.
$\Delta KSZ_{industry}$	Each month, we acquire funds' latest holding data and next available holding data. We calculate the sum of squared difference between funds industry (SIC 2-digit) weight and market industry weight, referred to as $KSZ_{industry,i}$, for the two holdings. We compute the logarithmic difference between the two $KSZ_{Industry,i}$.
ΔGap_{abs}	Each month, we acquire funds' latest holding data and calculate the absolute difference between fund value-weighted holding-based returns and fund's gross return, referred as $Gap_{i,t}$. Subsequently, we compute the logarithmic difference between the sum of $Gap_{i,t}$ over the next 12 months and the sum of $Gap_{i,t}$ over the past 12 months. $\Delta Gap_{abs,i,t} = \log\left(\frac{\sum_{t=1}^{12} Gap_{i,t}}{\sum_{t=-11}^0 Gap_{i,t}}\right)$
$\Delta Turnover$	The logarithmic difference of the fund's total turnover from the future 12 months and the total turnover from past 12 months. $\Delta Turnover_{i,t} = \log\left(\frac{\sum_{t=1}^{12} Turnover_{i,t}}{\sum_{t=-11}^0 Turnover_{i,t}}\right)$
$\Delta R2$	The logarithmic difference of the R2 between the future 12 months fund gross returns with the Morningstar category index and the same R2 estimated using past 12 months gross return. $\Delta R2_{i,t} = \log\left(\frac{R2_{i,t+1,t+12}}{R2_{i,t-11,t}}\right)$

Appendix 2B: Ranking Variables

This table reports the construction of ranking variables used in our study.

Variables Name	Description
Rank_{new}	Each month, for each Morningstar Category (3×3), we sort each fund based on their past 3-year Morningstar risk-adjusted return (MRAR) in ascending order into quintiles from 1 to 5 within each Overall rating (1 Star to 5 Star).
$\text{Rank}_{new,3}$	Each month, for each Morningstar Category (3×3), we sort each fund based on their past 3-year Morningstar risk-adjusted return (MRAR) in ascending order into quintiles from 1 to 5 within each 3-year rating (1 Star to 5 Star).
$\text{Rank}_{new,5}$	Each month, for each Morningstar Category (3×3), we sort each fund based on their past 5-year Morningstar risk-adjusted return (MRAR) in ascending order into quintiles from 1 to 5 within each 5-year rating (1 Star to 5 Star).
$\text{Rank}_{new,10}$	Each month, for each Morningstar Category (3×3), we sort each fund based on their past 10-year Morningstar risk-adjusted return (MRAR) in ascending order into quintiles from 1 to 5 within each 10-year rating (1 Star to 5 Star).
Rank_{old}	Each month, we sort each fund based on their past 3-year Morningstar risk-adjusted return (MRAR) in ascending order into quintiles from 1 to 5 within each Overall rating (1 Star to 5 Star).

Appendix 2C: Market Volatility and Unemployment Risk

This table examines time series variation of rating-driven risk-shifting from reform-2022. Top_{new} ($Bottom_{new}$) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $Rank_{new} = 5$ ($=1$). The $Rank_{new}$ represents quintile sorting of equity funds based on their 3-year Morningstar risk-adjusted return (MRAR) from 1 to 5 within each overall rating for each 3×3 Morningstar category. ‘ $Low_{past\ 12m}$ ’ (and ‘ $Low_{around\ 24m}$ ’) takes the value 1 for months with the bottom quintile of past 12-month (around 24-month) standard deviation of the monthly returns of the S&P 500 market index from reform to 2022, 0 otherwise. ‘Negative Full-Year Return’ takes the value 1 for year with negative market return and 0 otherwise, while ‘Crisis Period’ takes the value 1 for year 2008, 2018, 2019 and 2020, 0 otherwise. Control variables are past 1-month, 3-month, 6-month, and 12-month gross returns, past 1-month, 3-month, 6-month, and 12-month flows, the logarithm of fund size, turnover ratio, fund gross expense, and overall rating with style fixed effects and month fixed effects. Standard errors are adjusted for heteroskedasticity and clustered by style and month. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Dep.Var(%) = ΔStd	(1)	(2)	(3)	(4)
Intercept	-3.403*** (-2.631)	-3.450*** (-2.701)	-3.342*** (-2.577)	-3.339*** (-2.602)
Top_{new}	0.672*** (3.869)	0.684*** (3.594)	0.668** (3.545)	0.812*** (4.179)
$Bottom_{new}$	-0.873*** (-4.278)	-1.091*** (-4.982)	-1.084*** (-4.539)	-1.011*** (-4.995)
Top_{new} × $Low_{past\ 12m}$	-1.457** (-2.414)			
$Bottom_{new}$ × $Low_{past\ 12m}$	-0.755 (-1.527)			
Top_{new} × $Low_{around\ 24m}$		-1.555** (-2.116)		
$Bottom_{new}$ × $Low_{around\ 24m}$		0.308 (0.502)		
Top_{new} × Negative Full-Year Return			-1.376** (-3.099)	
$Bottom_{new}$ × Negative Full-Year Return			0.264 (0.398)	
Top_{new} × Crisis Period				-3.198*** (-4.846)
$Bottom_{new}$ × Crisis Period				0.690 (1.632)
Controls	YES	YES	YES	YES
Style Fixed Effect	YES	YES	YES	YES
Month Fixed Effect	YES	YES	YES	YES
adj.R squared	1.5%	1.5%	1.5%	1.5%

Appendix 2D: Rating Change for Top-tier and Bottom-tier Funds

This table presents the time-series average of characteristics for top-tier funds and bottom-tier funds in the post-reform period. Top_{new} ($Bottom_{new}$) is an indicator variable for top-tier funds (bottom-tier funds) under the new methodology, which takes a value of 1 if $Rank_{new} = 5$ ($=1$). $\Delta rating_{t-n,t}$ is funds' rating change from month $t-n$ to month t . Pos (Neg) $\Delta rating_{t-n,t}$ equals 1 if $\Delta rating_{t-n,t}$ is positive (negative) and 0 otherwise. The t-statistics in parentheses are Newey-West adjusted, with significance levels indicated by *, **, and ***, representing 10%, 5%, and 1%, respectively.

	$Bottom_{new}$	Top_{new}	Top_{new} - $Bottom_{new}$
$\Delta rating_{t-1,t}$	0.056	-0.060	-0.116*** (16.047)
$Pos\Delta rating_{t-1,t}$	0.096	0.037	-0.058*** (-21.714)
$Neg\Delta rating_{t-1,t}$	0.040	0.097	0.057*** (10.371)
$\Delta rating_{t-6,t}$	0.017	-0.033	-0.050* (-1.797)
$Pos\Delta rating_{t-6,t}$	0.165	0.135	-0.030*** (-2.639)
$Neg\Delta rating_{t-6,t}$	0.147	0.171	0.024 (1.547)

Appendix 3A: Nonparametric Monotone Regression

To examine the relationship between q (size) and r (alpha), we employ a two-step nonparametric procedure that allows for flexible nonlinear estimation while enforcing monotonicity.

Step 1: Kernel Regression (Unconstrained Nonparametric Estimation). We first estimate the conditional mean function of r given q using a Nadaraya–Watson kernel estimator. Specifically, for each observation q , the estimator is given by

$$\hat{m}(q) = \frac{\sum_{i=1}^n K\left(\frac{q-q_i}{h}\right) r_i}{\sum_{i=1}^n K\left(\frac{q-q_i}{h}\right)},$$

where $K(\cdot)$ is a kernel function (we use the Gaussian kernel) and $h > 0$ is the bandwidth parameter that controls the degree of smoothing. A larger bandwidth produces a smoother but potentially biased fit, while a smaller bandwidth yields a more flexible but noisier estimate. The choice of bandwidth is crucial in kernel regression. Common practice is to determine h using cross-validation methods such as least-squares cross-validation, which adaptively selects the bandwidth that minimizes prediction error. However, these data-driven methods are computationally intensive, especially in large samples. To reduce computational complexity, we choose a fixed bandwidth of $h = 0.5$, which balances smoothness and flexibility without requiring repeated resampling.

Step 2: Monotonicity Adjustment via Isotonic Regression. While kernel regression provides a flexible nonparametric fit, it does not guarantee monotonicity in the estimated relationship. To enforce monotonicity, we apply isotonic regression as a post-processing step. Given the unconstrained kernel estimates $\{\hat{m}(q_i)\}_{i=1}^n$, isotonic regression solves

$$\min_{\theta_1 \leq \theta_2 \leq \dots \leq \theta_n} \sum_{i=1}^n (\hat{m}(q_i) - \theta_i)^2,$$

which projects the kernel regression estimates onto the space of monotone functions. The resulting fitted values $\{\hat{m}_{\text{mono}}(q_i)\}$ preserve the smoothness of kernel regression while ensuring that the estimated function is non-decreasing in q .

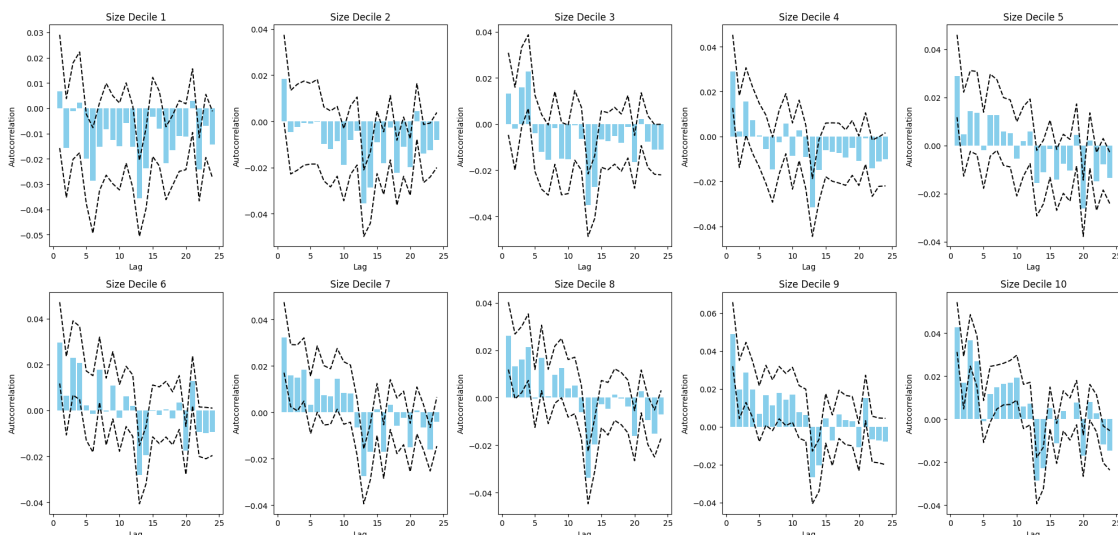
Appendix 3B: Distribution of Luck

For linear model in panel A and B, assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. For nonparametric model in panel C and D, the expected alpha α_{it}^{nonl} for fund i in month t is estimated using nonparametric kernel regression on fund alpha against AUM. The luck component ϵ_{it}^{nonl} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}^{nonl}$. This table presents the distribution of funds' average luck and Luck Ratio over their lives for the 3662 sample funds. Gross alpha against Morningstar benchmark is used in Panel A and C, whereas gross 4 factor alpha is used in Panel B and D.

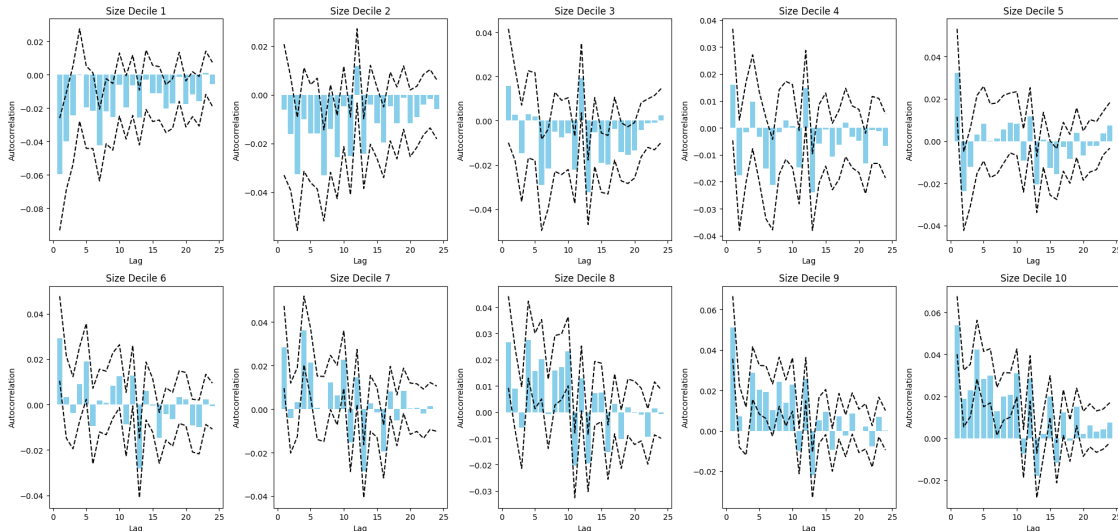
Panel A: Linear Model & Morningstar benchmark								
A.1: Sum of luck								
$\sum^T \epsilon_i^{g,MS}$	Count	Mean	Std	Min	25%	50%	75%	Max
	3662	0.000	0.000	-0.000	-0.000	0.000	0.000	0.000
A.2: Luck ratio								
$\frac{\bar{\epsilon}_i^{g,MS}}{Std\epsilon_i^{g,MS}/\sqrt{T}}$	Count	Mean	Std	Min	25%	50%	75%	Max
	3662	0.000	0.000	-0.000	-0.000	0.000	0.000	0.000
Panel B: Linear Model & 4 factor model								
B.1: Sum of luck								
$\sum^T \epsilon_i^{g,4F}$	Count	Mean	Std	Min	25%	50%	75%	Max
	3662	0.000	0.000	-0.000	-0.000	0.000	0.000	0.000
B.2: Luck ratio								
$\frac{\bar{\epsilon}_i^{g,4F}}{Std\epsilon_i^{g,4F}/\sqrt{T}}$	Count	Mean	Std	Min	25%	50%	75%	Max
	3662	0.000	0.000	-0.000	-0.000	0.000	0.000	0.000
Panel C: Nonparametric Model & Morningstar benchmark								
B.1: Sum of luck								
$\sum^T \epsilon_i^{g,MS,nonl}$	Count	Mean	Std	Min	25%	50%	75%	Max
	3662	0.001	0.053	-0.803	-0.000	0.000	0.000	1.872
B.2: Luck ratio								
$\frac{\bar{\epsilon}_i^{g,MS,nonl}}{Std\epsilon_i^{g,MS,nonl}/\sqrt{T}}$	Count	Mean	Std	Min	25%	50%	75%	Max
	3662	0.000	0.002	-0.041	-0.000	0.000	0.000	0.058
Panel D: Nonparametric Model & 4 factor model								
B.1: Sum of luck								
$\sum^T \epsilon_i^{g,4F,nonl}$	Count	Mean	Std	Min	25%	50%	75%	Max
	3662	0.000	0.042	-0.978	-0.000	0.000	0.000	0.758
B.2: Luck ratio								
$\frac{\bar{\epsilon}_i^{g,4F,nonl}}{Std\epsilon_i^{g,4F,nonl}/\sqrt{T}}$	Count	Mean	Std	Min	25%	50%	75%	Max
	3662	0.000	0.002	-0.056	-0.000	0.000	0.000	0.032

Appendix 3C: Persistence of Luck

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. Funds are sorted based on their average AUM over its life into deciles. This figure displays the autocorrelation function with 24-month lag for funds across each size decile. Gross alpha against Morningstar benchmark is used in Panel A, whereas gross 4 factor alpha is used in Panel B.



(a) Linear Model & Morningstar benchmark



(b) Linear Model & 4 factor model

Appendix 3D: Luck-Chasing by Year

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. Each month, we calculate funds' average luck over the past 12 months ($\bar{\epsilon}_{i,t-12,t}$) using Morningstar benchmark adjusted alpha. Funds with positive $\bar{\epsilon}_{i,t-12,t}$ are labeled as "Good Luck," while those with negative $\bar{\epsilon}_{i,t-12,t}$ are labeled as "Bad Luck." Net alpha against Morningstar benchmark is used in this figure. From 1990 to 2023, this figure plots the annual average Good Luck and Bad Luck funds' flow in next month ($flow_{i,t,t+1}$) in (a) and past 12-month contemporaneous flow ($flow_{i,t-12,t}$) in (b).



(a) Luck and Flow in next month



(b) Luck and Contemporaneous Flow

Appendix 3E: Luck and AUM

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. Luck is calculated using Morningstar benchmark adjusted net alpha. Panel A presents the results of the regressions, where the dependent variable is the AUM ($q_{i,t}$) of fund i in month t , and the independent variable is the average historical luck ($\bar{\epsilon}_{i,t-m,t}$) over window m . Panel B further controls for the AUM ($SizeDeciles_i$), where $SizeDeciles_i$ is size deciles based on fund i 's average AUM over its life. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: Size and Luck						
$q_{i,t} = \beta_0 + \beta_1 \bar{\epsilon}_{i,t-m,t}^{n,MS} + e$						
m=	12m		24m		36m	
$\beta_1 (\times 10^6)$	0.012*** (6.977)	0.007*** (4.021)	0.030*** (14.312)	0.027*** (13.832)	0.048*** (21.154)	0.045*** (21.613)
Year-Month FE	No	Yes	No	Yes	No	Yes
Style FE	No	Yes	No	Yes	No	Yes
R ²	0.01%	0.01%	0.03%	0.02%	0.05%	0.04%
Panel B: Size and Luck with Size Control						
$q_{i,t} = \beta_0 + \beta_1 \bar{\epsilon}_{i,t-m,t}^{n,MS} + \beta_2 SizeDeciles_i + e$						
m=	12m		24m		36m	
$\beta_1 (\times 10^6)$	0.014*** (8.226)	0.014*** (8.379)	0.035*** (18.139)	0.037*** (20.082)	0.054*** (25.864)	0.059*** (28.817)
Year-Month FE	No	Yes	No	Yes	No	Yes
Style FE	No	Yes	No	Yes	No	Yes
R ²	17.97%	16.46%	17.99%	16.77%	18.51%	16.80%

Appendix 3F: Past Alpha, Future Alpha and Value Added

Assuming a functional form of decreasing returns to scale, $r = a - b \cdot q$, the expected alpha α_{it} for fund i in month t is estimated as $a_i - b_i \cdot q_{t-1}$. The luck component ϵ_{it} for fund i in month t is therefore defined as $r_{it} - \alpha_{it}$. Panel A presents the results of the regressions, where the dependent variable is the average future realized alpha ($\bar{\epsilon}_{i,t+m,t}$) over window m , and the independent variable is the average historical alpha ($\bar{r}_{i,t-1y,t}$) over past 1 year. Panel B presents the results of the regressions, where the dependent variable is the average future realized value added ($\bar{V}_{i,t+m,t}$) over window m , and the independent variable is the average historical alpha ($\bar{r}_{i,t-1y,t}$) over past 1 year. The value added ($V_{i,t}$) is the product of lagged AUM in millions ($q_{i,t-1}$) and alpha ($r_{i,t}$). Net alpha against Morningstar benchmark is used in this table. Control variables include expense ratio, turnover, fund age and number of managers. Standard errors are adjusted for heteroskedasticity. Significance levels at 10%, 5%, and 1% are indicated by *, **, and ***, respectively.

Panel A: Past Alpha and Future Alpha						
$\bar{r}_{i,t,t+m}^{n,MS} = \beta_0 + \beta_1 \bar{r}_{i,t-12m,t}^{n,MS} + e$						
m=	12m			24m		
	(1)	(2)	(3)	(4)	(5)	(6)
β_1	0.069*** (27.539)	-0.006*** (-2.916)	-0.03*** (-10.183)	0.011*** (6.296)	-0.048*** (-29.139)	-0.053*** (-34.782)
Control	No	No	Yes	No	No	Yes
Year-month FE	No	Yes	Yes	No	Yes	Yes
Fund FE	No	Yes	Yes	No	Yes	Yes
R ²	0.41%	0.02%	0.34%	0.02%	0.35%	1.01%
Panel B: Past Alpha and Future Value Added						
$\bar{V}_{i,t,t+m}^{n,MS} = \beta_0 + \beta_1 \bar{r}_{i,t-12m,t}^{n,MS} + e$						
m=	12m			24m		
	(1)	(2)	(3)	(4)	(5)	(6)
β_1	102.885*** (6.339)	-84.717*** (-4.742)	-173.450*** (-5.242)	-31.291*** (-2.750)	-197.405*** (-16.033)	-342.940*** (-16.627)
Control	No	No	Yes	No	No	Yes
Year-month FE	No	Yes	Yes	No	Yes	Yes
Fund FE	No	Yes	Yes	No	Yes	Yes
R ²	0.01%	0.01%	0.03%	0.01%	0.08%	0.19%

References

- Yacine Aït-Sahalia and Andrew W Lo. Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, 53(2):499–547, 1998.
- Manuel Ammann and Michael Verhofen. Prior performance and risk-taking of mutual fund managers: a dynamic bayesian network approach. *Journal of Behavioral Finance*, 8(1):20–34, 2007.
- Li An, Joseph Engelberg, Matthew Henriksson, Baolian Wang, and Jared Williams. The portfolio-driven disposition effect. *The Journal of Finance*, 79(5):3459–3495, 2024.
- Kenechukwu Anadu, Mathias Kruttli, Patrick McCabe, and Emilio Osambela. The shift from active to passive investing: Risks to financial stability? *Financial Analysts Journal*, 76(4):23–39, 2020.
- Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. The cross-section of volatility and expected returns. *The journal of finance*, 61(1):259–299, 2006.
- Andrew Ang, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang. High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics*, 91(1):1–23, 2009.
- Will J Armstrong, Egemen Genc, and Marno Verbeek. Going for gold: An analysis of morningstar analyst ratings. *Management Science*, 65(5):2310–2327, 2019.
- Malcolm Baker, Brendan Bradley, and Jeffrey Wurgler. Benchmarks as limits to arbitrage: Understanding the low-volatility anomaly. *Financial Analysts Journal*, 67(1):40–54, 2011.
- Turan G Bali, Nusret Cakici, and Robert F Whitelaw. Maxing out: Stocks as lotteries and the cross-section of expected returns. *Journal of Financial Economics*, 99(2):427–446, 2011.
- Turan G Bali, Stephen J Brown, Scott Murray, and Yi Tang. A lottery-demand-based explanation of the beta anomaly. *Journal of Financial and Quantitative Analysis*, 52(6):2369–2397, 2017.

- Brad M Barber and Terrance Odean. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance*, 55(2):773–806, 2000.
- Brad M Barber and Terrance Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics*, 116(1):261–292, 2001.
- Brad M Barber and Terrance Odean. Online investors: do the slow die first? *The Review of financial studies*, 15(2):455–488, 2002.
- Brad M Barber, Terrance Odean, and Ning Zhu. Do retail trades move markets? *The Review of Financial Studies*, 22(1):151–186, 2008.
- Brad M Barber, Xing Huang, and Terrance Odean. Which factors matter to investors? evidence from mutual fund flows. *The Review of Financial Studies*, 29(10):2600–2642, 2016.
- Laurent Barras, Olivier Scaillet, and Russ Wermers. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The journal of finance*, 65(1):179–216, 2010.
- Laurent Barras, Patrick Gagliardini, and Olivier Scaillet. Skill, scale, and value creation in the mutual fund industry. *The Journal of Finance*, 77(1):601–638, 2022.
- William J Bazley, Henrik Cronqvist, and Milica Mormann. Visual finance: The pervasive effects of red on investor behavior. *Management science*, 67(9):5616–5641, 2021.
- Lucian A Bebchuk, Yaniv Grinstein, and Urs Peyer. Lucky ceos and lucky directors. *The Journal of Finance*, 65(6):2363–2401, 2010.
- Itzhak Ben-David, Jiacui Li, Andrea Rossi, and Yang Song. What do mutual fund investors really care about? *The Review of Financial Studies*, 35(4):1723–1774, 2022a.
- Itzhak Ben-David, Jiacui Li, Andrea Rossi, and Yang Song. Ratings-driven demand and systematic price fluctuations. *The Review of Financial Studies*, 35(6):2790–2838, 2022b.
- Jonathan B Berk and Richard C Green. Mutual fund flows and performance in rational markets. *Journal of political economy*, 112(6):1269–1295, 2004.

- Jonathan B Berk and Jules H Van Binsbergen. Measuring skill in the mutual fund industry. *Journal of financial economics*, 118(1):1–20, 2015.
- Marianne Bertrand and Sendhil Mullainathan. Are ceos rewarded for luck? the ones without principals are. *The Quarterly Journal of Economics*, 116(3):901–932, 2001.
- Christopher R Blake and Matthew R Morey. Morningstar ratings and mutual fund performance. *Journal of financial and Quantitative Analysis*, 35(3):451–483, 2000.
- David Blitz. The quant crisis of 2018–2020: Cornered by big growth. *The Journal of Portfolio Management*, 2021.
- Marshall E Blume. An anatomy of morningstar ratings. *Financial Analysts Journal*, 54(2):19–27, 1998.
- Devdeepa Bose, Henning Cordes, Sven Nolte, Judith Christiane Schneider, and Colin Farrell Camerer. Decision weights for experimental asset prices based on visual salience. *The Review of Financial Studies*, 35(11):5094–5126, 2022.
- David P Brown and Youchang Wu. Mutual fund flows and cross-fund learning within families. *The Journal of Finance*, 71(1):383–424, 2016.
- Keith C Brown, W Van Harlow, and Laura T Starks. Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry. *The Journal of Finance*, 51(1):85–110, 1996.
- Jeffrey A Busse. Another look at mutual fund tournaments. *Journal of financial and Quantitative Analysis*, 36(1):53–73, 2001.
- James P Byrnes, David C Miller, and William D Schafer. Gender differences in risk taking: A meta-analysis. *Psychological bulletin*, 125(3):367, 1999.
- Jorge D Camba, Pedro Company, and Vetria L Byrd. Identifying deception as a critical component of visualization literacy. *IEEE Computer Graphics and Applications*, 42(1):116–122, 2022.
- Mark M Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82, 1997.

- Joseph Chen, Harrison Hong, Ming Huang, and Jeffrey D Kubik. Does fund size erode mutual fund performance? the role of liquidity and organization. *American Economic Review*, 94(5):1276–1302, 2004.
- Judith Chevalier and Glenn Ellison. Risk taking by mutual funds as a response to incentives. *Journal of political economy*, 105(6):1167–1200, 1997.
- Judith Chevalier and Glenn Ellison. Career concerns of mutual fund managers. *The Quarterly Journal of Economics*, 114(2):389–432, 1999.
- Darwin Choi, Bige Kahraman, and Abhiroop Mukherjee. Learning about mutual fund managers. *The Journal of Finance*, 71(6):2809–2860, 2016.
- Lauren Cohen, David S Kim, and Eric C So. Box jumping: Portfolio recompositions to achieve higher morningstar ratings. *Available at SSRN*, 2024.
- Mathijs Cosemans and Rik Frehen. Saliency theory and stock prices: Empirical evidence. *Journal of Financial Economics*, 140(2):460–483, 2021. ISSN 0304-405X.
- Julia Deakin, Michael Aitken, Trevor Robbins, and Barbara J Sahakian. Risk taking during decision-making in normal volunteers changes with age. *Journal of the International Neuropsychological Society*, 10(4):590–598, 2004.
- Diane Del Guercio and Paula A Tkac. Star power: The effect of monrnningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis*, 43(4):907–936, 2008.
- Jefferson Duarte, Stephan Siegel, and Lance Young. Trust and credit: The role of appearance in peer-to-peer lending. *The Review of Financial Studies*, 25(8):2455–2484, 2012.
- Edwin J Elton, Martin J Gruber, and Christopher R Blake. A first look at the accuracy of the crsp mutual fund database and a comparison of the crsp and morningstar mutual fund databases. *The Journal of Finance*, 56(6):2415–2430, 2001.
- Edwin J Elton, Martin J Gruber, and Christopher R Blake. Incentive fees and mutual funds. *The Journal of Finance*, 58(2):779–804, 2003.
- Richard B Evans and Yang Sun. Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. *The Review of Financial Studies*, 34(1):67–107,

2021.

- Eugene F Fama and Kenneth R French. Luck versus skill in the cross-section of mutual fund returns. *The journal of finance*, 65(5):1915–1947, 2010.
- Arlen Fan, Yuxin Ma, Michelle Mancenido, and Ross Maciejewski. Annotating line charts for addressing deception. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2022.
- Javier Gil-Bazo and PABLO Ruiz-Verdú. The relation between price and performance in the mutual fund industry. *The Journal of Finance*, 64(5):2153–2183, 2009.
- Markus Glaser, Zwetelina Iliewa, and Martin Weber. Thinking about prices versus thinking about returns in financial markets. *The Journal of Finance*, 74(6):2997–3039, 2019.
- William N Goetzmann and Alok Kumar. Equity portfolio diversification. *Review of Finance*, 12(3):433–463, 2008.
- Paul A Gompers and Andrew Metrick. Institutional investors and equity prices. *The quarterly journal of Economics*, 116(1):229–259, 2001.
- Mark Grinblatt and Matti Keloharju. Sensation seeking, overconfidence, and trading activity. *The Journal of Finance*, 64(2):549–578, 2009.
- Mark Grinblatt and Sheridan Titman. Adverse risk incentives and the design of performance-based contracts. *Management science*, 35(7):807–822, 1989.
- Martin Halek and Joseph G Eisenhauer. Demography of risk aversion. *Journal of Risk and Insurance*, pages 1–24, 2001.
- Yuval Noah Harari. *Sapiens: A brief history of humankind*. Random House, 2014.
- Campbell R Harvey and Yan Liu. Luck versus skill in the cross section of mutual fund returns: Reexamining the evidence. *The Journal of Finance*, 77(3):1921–1966, 2022.
- Kewei Hou and Roger K Loh. Have we solved the idiosyncratic volatility puzzle? *Journal of Financial Economics*, 121(1):167–194, 2016.
- Jennifer Huang, Kelsey D Wei, and Hong Yan. Participation costs and the sensitivity of fund flows to past performance. *The journal of finance*, 62(3):1273–1311, 2007.

- Jennifer Huang, Clemens Sialm, and Hanjiang Zhang. Risk shifting and mutual fund performance. *The Review of Financial Studies*, 24(8):2575–2616, 2011.
- Michael C Jensen. The performance of mutual funds in the period 1945-1964. *The Journal of finance*, 23(2):389–416, 1968.
- Marcin Kacperczyk, Clemens Sialm, and Lu Zheng. On the industry concentration of actively managed equity mutual funds. *The Journal of Finance*, 60(4):1983–2011, 2005.
- Marcin Kacperczyk, Clemens Sialm, and Lu Zheng. Unobserved actions of mutual funds. *The Review of Financial Studies*, 21(6):2379–2416, 2008.
- Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.
- Alexander Kempf and Stefan Ruenzi. Tournaments in mutual-fund families. *The Review of Financial Studies*, 21(2):1013–1036, 2008.
- Alexander Kempf, Stefan Ruenzi, and Tanja Thiele. Employment risk, compensation incentives, and managerial risk taking: Evidence from the mutual fund industry. *Journal of Financial Economics*, 92(1):92–108, 2009.
- Sanghyun Hugh Kim. Do mutual funds manipulate star ratings? evidence from portfolio pumping. *Evidence from Portfolio Pumping (January 5, 2022)*, 2022.
- Arthur Korteweg and Morten Sorensen. Skill and luck in private equity performance. *Journal of Financial Economics*, 124(3):535–562, 2017.
- Jennifer Lynch Koski and Jeffrey Pontiff. How are derivatives used? evidence from the mutual fund industry. *The journal of finance*, 54(2):791–816, 1999.
- Robert Kosowski, Allan Timmermann, Russ Wermers, and Hal White. Can mutual fund “stars” really pick stocks? new evidence from a bootstrap analysis. *The Journal of finance*, 61(6):2551–2595, 2006.
- Robert Kosowski, Narayan Y Naik, and Melvyn Teo. Do hedge funds deliver alpha? a bayesian and bootstrap analysis. *Journal of financial economics*, 84(1):229–264, 2007.

- Alok Kumar. Who gambles in the stock market? *The journal of finance*, 64(4):1889–1933, 2009.
- Ilyana Kuziemko, Michael I Norton, Emmanuel Saez, and Stefanie Stantcheva. How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review*, 105(4):1478–1508, 2015.
- Toomas Laarits and Jeffrey Wurgler. The research behavior of individual investors. Technical report, National Bureau of Economic Research, 2025.
- Claire Lauer and Shaun O’Brien. How people are influenced by deceptive tactics in everyday charts and graphs. *IEEE Transactions on Professional Communication*, 63(4):327–340, 2020.
- Claire Lauer and Christopher A Sanchez. What eye tracking can show us about how people are influenced by deceptive tactics in line graphs. *IEEE Transactions on Professional Communication*, 2023.
- Bruce N Lehmann. Residual risk revisited. *Journal of Econometrics*, 45(1-2):71–97, 1990.
- Haim Levy. Equilibrium in an imperfect market: A constraint on the number of securities in the portfolio. *The American Economic Review*, 68(4):643–658, 1978.
- John Lintner. Security prices, risk, and maximal gains from diversification. *The journal of finance*, 20(4):587–615, 1965.
- Chengwei Liu and Mark De Rond. Good night, and good luck: Perspectives on luck in management scholarship. *Academy of Management Annals*, 10(1):409–451, 2016.
- Yunting Liu. The short-run and long-run components of idiosyncratic volatility and stock returns. *Management Science*, 68(2):1573–1589, 2022.
- Linlin Ma and Yuehua Tang. Portfolio manager ownership and mutual fund risk taking. *Management Science*, 65(12):5518–5534, 2019.
- Burton G Malkiel and Yexiao Xu. Idiosyncratic risk and security returns. *University of Texas at Dallas (November 2002)*, 15, 2002.
- Harry Markowitz. The utility of wealth. *Journal of political Economy*, 60(2):151–158, 1952.

- Karl Marx. *Das kapital*. e-artnow, 2018.
- Winter Mason and Siddharth Suri. Conducting behavioral research on amazon’s mechanical turk. *Behavior research methods*, 44(1):1–23, 2012.
- Massimo Massa and Rajdeep Patgiri. Incentives and mutual fund performance: higher performance or just higher risk taking? *The Review of Financial Studies*, 22(5):1777–1815, 2009.
- Robert C Merton et al. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 1987.
- José Luis Montiel Olea and Tomasz Strzalecki. Axiomatization and measurement of quasi-hyperbolic discounting. *The Quarterly Journal of Economics*, 129(3):1449–1499, 2014.
- Matthew R Morey. Kiss of death: A 5-star morningstar mutual fund rating? *Available at SSRN 455240*, 2003.
- Shaun O’Brien and Claire Lauer. Testing the susceptibility of users to deceptive data visualizations when paired with explanatory text. In *Proceedings of the 36th ACM International Conference on the Design of Communication*, pages 1–8, 2018.
- Allan Paivio, Timothy B Rogers, and Padric C Smythe. Why are pictures easier to recall than words? *Psychonomic science*, 11(4):137–138, 1968.
- Anshul Vikram Pandey, Katharina Rall, Margaret L Satterthwaite, Oded Nov, and Enrico Bertini. How deceptive are deceptive visualizations? an empirical analysis of common distortion techniques. In *Proceedings of the 33rd annual acm conference on human factors in computing systems*, pages 1469–1478, 2015.
- John A Parnell, Eric B Dent, Nicholas O’Regan, and Tim Hughes. Managing performance in a volatile environment: contrasting perspectives on luck and causality. *British Journal of Management*, 23:S104–S118, 2012.
- L’uboš Pástor, Robert F Stambaugh, and Lucian A Taylor. Scale and skill in active management. *Journal of Financial Economics*, 116(1):23–45, 2015.
- Jiaping Qiu. Termination risk, multiple managers and mutual fund tournaments. *Review of*

- Finance*, 7(2):161–190, 2003.
- Jonathan Reuter and Eric Zitzewitz. How much does size erode mutual fund performance? a regression discontinuity approach. *Review of Finance*, 25(5):1395–1432, 2021.
- Andrea Rossi. Decreasing returns or reversion to the mean? the case of private equity fund growth. *SSRN*, 2019.
- Paul Schneider, Christian Wagner, and Josef Zechner. Low-risk anomalies? *The Journal of Finance*, 75(5):2673–2718, 2020.
- Christopher Schwarz and Zheng Sun. How fast do investors learn? asset management investors and bayesian learning. *The Review of Financial Studies*, 36(6):2397–2430, 2023.
- Christopher G Schwarz. Mutual fund tournaments: The sorting bias and new evidence. *The Review of Financial Studies*, 25(3):913–936, 2012.
- Yezhou Sha. The devil in the style: Mutual fund style drift, performance and common risk factors. *Economic Modelling*, 86:264–273, 2020.
- William F Sharpe. Mutual fund performance. *The Journal of business*, 39(1):119–138, 1966.
- William F Sharpe. Morningstar’s risk-adjusted ratings. *Financial Analysts Journal*, 54(4): 21–33, 1998.
- Tao Shu, Johan Sulaeman, and P Eric Yeung. Local religious beliefs and mutual fund risk-taking behaviors. *Management Science*, 58(10):1779–1796, 2012.
- Kelly Shue and Richard R Townsend. Can the market multiply and divide? non-proportional thinking in financial markets. *The Journal of Finance*, 76(5):2307–2357, 2021.
- Erik R Sirri and Peter Tufano. Costly search and mutual fund flows. *The journal of finance*, 53(5):1589–1622, 1998.
- Matthew Spiegel and Hong Zhang. Mutual fund risk and market share-adjusted fund flows. *Journal of Financial Economics*, 108(2):506–528, 2013.
- Georg Stenberg, Karl Radeborg, and Leif R Hedman. The picture superiority effect in a cross-modality recognition task. *Memory & Cognition*, 23:425–441, 1995.
- Seha M Tinic and Richard R West. Risk, return, and equilibrium: A revisit. *Journal of*

- Political Economy*, 94(1):126–147, 1986.
- Huijun Wang, Jinghua Yan, and Jianfeng Yu. Reference-dependent preferences and the risk–return trade-off. *Journal of Financial Economics*, 123(2):395–414, 2017.
- Elke U Weber, Niklas Siebenmorgen, and Martin Weber. Communicating asset risk: How name recognition and the format of historic volatility information affect risk perception and investment decisions. *Risk Analysis: An International Journal*, 25(3):597–609, 2005.
- Russ Wermers. Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses. *The Journal of Finance*, 55(4):1655–1695, 2000.
- Brenda W Yang, Camila Vargas Restrepo, Matthew L Stanley, and Elizabeth J Marsh. Truncating bar graphs persistently misleads viewers. *Journal of Applied Research in Memory and Cognition*, 10(2):298–311, 2021.
- Xinge Zhao. Exit decisions in the us mutual fund industry. *The Journal of Business*, 78(4): 1365–1402, 2005.
- Min Zhu. Informative fund size, managerial skill, and investor rationality. *Journal of Financial Economics*, 130(1):114–134, 2018.