Tilting at Windmills: Biased Benchmarks and the Risk-Taking Response of Mutual Funds *

Richard B. Evans† Sanghyun (Hugh) Kim‡ Thomas A. Maurer§

September 16, 2025

Abstract

We study how changes in third-party relative performance evaluation (RPE) shape mutual-fund managers' incentives. We exploit Morningstar's 2002 shift from a single U.S. equity peer group to size–style categories and its 2016 introduction of ESG Globe ratings to explore if biased benchmarking disadvantaged certain funds and induced risk-taking. Pre-2002, growth funds received lower star ratings and held higher-beta, more volatile portfolios—especially when the value spread was large. Similarly, before the globe rating introduction, ESG funds held higher risk stocks with lower ESG ratings. These higher-risk holding effects disappear after both the 2002 and 2016 Morningstar reclassification events.

JEL-Classification: G11, G24.

Keywords: Mutual Funds, Peer Benchmarks, Risk-Taking, Morningstar Ratings, ESG Investing.

*We thank Haoran Wang for outstanding work as our research assistant. Moreover, we are thankful to Francesco Franzoni, Laurent Fresard, Shiyang Huang, Yueting Jiang, Dunhong Jin, Thomas Lee, Pedro Matos, Roni Michaely, Jayoung Nam, Tim Riley, Andrea Rossi, Thomas Schmid, Clemens Sialm, Qifei Zhu, and seminar participants at Università della Svizzera Italiana, Florida State University, University of St. Gallen, Hanyang University, University of Hong Kong, University of Virginia and conference participants at the 2025 Asian Finance Association Annual Conference, 2025 UIC Finance Conference, 2025 Northern Finance Association, 2025 Financial Management Association, 2025 Southern Finance Association, and the 2026 American Finance Association for many helpful comments and suggestions. Finally, we are thankful for the generous funding provided by the Research Grants Council of Hong Kong (General Research Fund) and the Social Sciences and Humanities Research Council of Canada (Insight Development Grant).

†Corresponding Author, Darden School of Business, The University of Virginia, 100 Darden Blvd, Charlottesville, VA 22903, USA, evansr@darden.virginia.edu.

‡Lazaridis School of Business and Economics, Wilfrid Laurier University, shkim@wlu.ca.

§Faculty of Business and Economics, The University of Hong Kong, maurer@hku.hk.

1 Introduction

"I know who I am" said Don Quixote, "and who I may be. I will not be deterred.
I shall do battle with these windmills, for they are giants in my eyes."
— Miguel de Cervantes, Don Quixote

"...sometimes, we're like Don Quixote, tilting at windmills, convinced we can win,
even when it's clear the odds are stacked against us."
— Garth Stein, The Art of Racing in the Rain

Principal-agent theory suggests that tying a manager's pay to firm performance can motivate them to act in shareholders' best interests (e.g., Holmstrom (1979)). However, this also exposes risk-averse managers to common shocks they can't control. Relative performance evaluation (RPE) - measuring a manager's performance against a suitable benchmark - can reduce these risks and make incentives more effective (e.g., Baiman and Demski (1980), Lazear and Rosen (1981), Holmstrom (1982), and Green and Stokey (1983)). Early research focused on the benefits of representative and unbiased benchmarks[1], but later studies examined potential problems with RPE, especially when managers can influence the comparison set. For instance, Bizjak et al. (2008) show that CEOs shape peer groups in ways that boost their pay[2], while Sensoy (2009) finds that fund managers choose benchmarks that make their performance look better.[3]

While this literature focuses on biased benchmarking that arises from the agent's (i.e., CEO/fund manager) influence, these choices and the corresponding effects are endogenous in nature. In contrast, in this paper, we examine third party selection of RPE peer groups and benchmarks and changes to these choices that are plausibly exogenous from the perspective of the agent. In such a setting, we are better able to examine how changes in RPE efficacy affect

[1]For example, Holmstrom (1982) abstracts from the representativeness of the benchmark, or as he describes it, "...the optimal use of peer performance", instead assuming a sufficient statistic condition about the benchmark aggregation function.

[2]Additional papers exploring the impact of CEOs on their peer group selection include Faulkender and Yang (2010), Bizjak et al. (2011), Albuquerque et al. (2013), and Cadman and Carter (2014).

[3]Additional papers exploring the impact of fund managers on benchmark selection include Elton et al. (2014), Chen et al. (2021), Cremers et al. (2022), Mullally and Rossi (2025), and Chen et al. (2025).

agents' decisions. The metaphor of battling illusory giants that we reference above frames our study: managers respond to perceived disadvantages from benchmark construction, even if those disadvantages are artifacts of methodology. Specifically, in this paper we examine such a third party RPE provider, Morningstar, and two plausibly exogenous events where they changed their RPE methodology. The first is Morningstar's 2002 change in their star rating methodology. The second is their 2016 introduction of ESG-focused Globe ratings.

This change in Morningstar star rating methodology and the introduction of Morningstar globe ratings provide an ideal setting to examine how changes in RPE efficacy affect the actions of agents. Morningstar ratings, which are based on relative fund performance within Morningstar peer groups, have a significant and causal impact on fund flows (Del Guercio and Tkac, 2008; Reuter and Zitzewitz, 2021; Evans and Sun, 2021; Ben-David et al., 2022b). Similarly, at the time of their launch, Hartzmark and Sussman (2019) show that investor flow was highly sensitive to this ESG ratings measure.[4] This evidence that investors respond to these two different RPE measures suggest that these two changes are a compelling setting in which to examine the impact of changes in RPE peer group on fund manager actions.

In the first case, Morningstar changed the peer groups against which their RPE star rating metric was assessed. Before June 2002, a single peer group consisting of all domestic equity funds was used. In June of 2002, however, Morningstar changed to assigning peer groups by market capitalization and value/growth tilts. Ranking all U.S. domestic equity funds within a single category is tantamount to using a single-factor model (e.g., the CAPM) for risk adjustment (Evans and Sun, 2021), disadvantaging growth funds. It is well-documented in the literature that growth stocks, on average, earn a lower CAPM alpha than

[4]It is important to note that Gantchev et al. (2024) find that as the performance of ESG stocks declined, the tradeoff between globe ratings and risk-adjusted performance became more salient, causing investor flow to be less related to the globe ratings. We discuss this broader issue of time-variation in ESG returns from the perspective of Pástor et al. (2022) below.

value stocks (Fama and French, 1993), so a sizable value premium places growth funds at a significant disadvantage when competing against value funds under the single-category-group (or single-factor) framework. Morningstar began grouping U.S. domestic equity funds into nine (three-by-three) categories, classified by the size by value-growth Morningstar equity style box, partially resolving the biased benchmarking issue, with growth funds no longer being compared to value funds under this new multi-category (e.g., Fama-French three factor model) framework.

In analyzing this 2002 peer group change, we first examine the Morningstar ratings of growth and value funds. We find that before the change, growth funds have statistically lower star ratings than value funds, but there is no meaningful difference after the change. We also find that prior to 2002, Morningstar ratings are highly correlated with the past returns of the value factor (HML) of the Fama-French 3-factor model (Fama and French, 1993). The better the value factor had performed in the past 36 months, the worse the ratings for growth funds compared to value funds, a pattern that also disappears after the change.

We then examine how the fund manager (agent) responds to the change in RPE peer group. Before the change, growth fund managers purchase higher risk securities as measured by both beta and volatility than value funds, a pattern that disappears after the change. To further corroborate our mechanism, we exploit the time-variation in the expected value premium, using the value spread as a proxy. Cohen et al. (2003) find that the value spread positively predicts the value premium. Thus, we expect that when the value spread is larger, the disadvantage of growth funds would be more severe. Indeed, we find that disadvantaged growth funds increase risk-taking even more when the expected value spread is higher, in the pre-refinement period. In contrast, after the refinement of Morningstar peer groups, this correlation between the expected value spread and greater risk-taking of growth funds relative to value funds is no longer statistically or economically significant.

Our mechanism implies that the increase in risk-taking should be especially pronounced for growth funds that have stronger incentives to mitigate their disadvantage. Since Morningstar ratings are discrete with only five possible outcomes, growth funds that are ranked close to one of the four rating thresholds are more concerned about their disadvantage compared to funds that are far from a threshold. When we test this incentive issue, we find that the closer a fund is to a rating threshold, the greater their risk-shifting behavior.

In the second case, Morningstar introduced their ESG globe ratings. Before March 2016, investors using Morningstar to make investment decisions had to rely solely on a star rating which did not differentiate on the basis of ESG. However, after their March 2016 introduction of the globe ratings, investors had both risk/return and ESG ratings to make their decision, effectively changing the peer group against which funds were compared. While the academic literature on the relationship between ESG characteristics and stock returns is mixed, Pástor et al. (2022) provide a compelling explanation for these mixed results. Namely, time variation in environmental concerns can lead to the outperformance of green assets relative to brown, even with the existence of their estimate of a negative greenium[5] (the expected underperformance of green assets relative to brown).[6] We build upon this insight in order to assess how fund managers might respond to biased benchmarking. Given an expectation of a negative greenium, ESG fund managers would be at a disadvantage relative to non-ESG fund managers and would therefore respond by increased risk-taking. However, this behavior should be mitigated in periods where managers expect ESG stock

[5] Using one of the most comprehensive analyses to date including data across 49 countries and 23 greenness measures, Eskildsen et al. (2024) also find a negative greenium .

[6] Additionally, there is greater consensus in the literature on the reduction in risk associated with ESG. For example, Albuquerque et al. (2019) model theoretically and test empirically the decision to engage in corporate social responsibility (CSR) activities, finding that such firms exhibit decreased systematic risk. Hoepner et al. (2024) show that shareholder engagement on ESG issues reduces downside risks for the targeted firms. Focusing on the global financial crisis, Lins et al. (2017) finds that firms with higher CSR activity generated higher returns, had higher profitability, growth and sales per employee, consistent with hedging downside risk. Using options data, Ilhan et al. (2021) show that the cost of downside protection is higher for firms that are more carbon-intensive.

outperformance. Our empirical tests, described below, are designed to capture both return aspects of ESG as elucidated by Pástor et al. (2022).

With respect to investor flows, we find they are highly sensitive to Morningstar stars throughout. However, after the introduction of the Globe ratings in March of 2016, this star-flow sensitivity decreases between and 27% and 31%, but only for the higher globe rated funds. There is no statistically or economically significant difference for the low globe rated funds. With respect to manager security selection decisions, we also find that those funds that ultimately receive a high globe rating in March of 2016, select stocks with higher risk - beta and volatility - and lower ESG ratings, before the globe ratings are introduced. For those funds whose names clearly indicate they are ESG-focused, we find no such pattern, consistent with their fund name already creating an alternative peer group comparison. Finally, with respect to time-varying expectations of ESG concerns, we proxy for expectations with future realized ESG stock performance. Constructing an ESG factor with the Sustainalytics ESG ratings that Morningstar globes are based upon, we find that the managerial risk-taking behavior is mitigated when there is expected ESG stock outperformance.

Our paper is related to a large literature on mutual fund risk-taking. Generally speaking, the literature attributes variations in risk-taking either to fund manager skills, risk-shifting due to non-linearities in the relationship between performance and manager compensation, or excessive risk-taking of past losers who try to catch up with past winners. Brown et al. (1996) and Chevalier and Ellison (1997) initiated a large literature showing that past losers take more risks than winners in the remainder of the year.[7] Among these papers, Han et al. (2024) is closest to our setting, in that they use the 2002 Morningstar ratings methodology change to explore the role of mutual fund loser risk-taking in propagating the high beta, idiosyncratic volatility, and skewness anomalies. The economic channel

[7] This literature includes Chen and Pennacchi (2009), Basak et al. (2007), Schwarz (2012), Busse (2001), Basak et al. (2008), Cullen et al. (2012), Spiegel and Zhang (2013), Kim (2019), Lee et al. (2019), Ma and Tang (2019), Ma et al. (2019), Del Guercio and Tkac (2002), Huang et al. (2011).

in our study differs from the extant literature on risk-shifting. We are not investigating past losers. Rather, we focus on the effects of peer group assignments that are relevant for mutual fund investors when allocating their wealth to funds (i.e., fund flows). Specifically, funds that expect to be at a disadvantage due to an unfair peer group assignment anticipate a systematic disadvantage when competing against their advantaged peers. This incentivizes them to take more risks in an attempt to compete with their advantaged peers.

Our paper is also related to the literature on manager incentives and compensation. For example, Evans et al. (2024) contrast the use of pure versus peer benchmarks and find that peer benchmarks incentivize managers to be more active and exert more effort, resulting in stronger performance. Importantly, they point out that benchmarks used to determine a manager's bonus may differ from prospectus benchmarks. While Evans et al. (2024) focus on benchmarks used for bonus payments of fund managers, we investigate the implications of benchmarks used by fund investors to guide their wealth allocation to funds. That is, mutual fund investors may use performance metrics and benchmarks to rank managers and decide in which fund to invest, which may differ from prospectus benchmarks or the evaluation methods used to determine manager bonus payments.

Similarly, there is recent work documenting that the compensation of fund managers crucially hinges on the assets under management (AUM) and the fee revenue generated by funds (Ibert et al., 2018; Cen et al., 2023; Bai et al., 2024). Our paper relates to this literature because the choice of peer groups for relative performance evaluation by Morningstar can affect fund flows. If that peer group is heterogeneous in its systematic risk exposures (e.g., grouping value and growth funds), this may generate undesirable incentives for fund managers leading to unintended distortions in fund behavior. In particular, if certain funds are disadvantaged (e.g. growth funds) relative to other funds (e.g. value funds) due to these differences in their systematic risk exposures, they may have an incentive to increase risk-taking to compete with their advantaged peers. This increase in risk-taking may result in an inefficient outcome for fund investors as a fund's risk profile may deviate from the desired level.

2 Data and Construction of Variables

Our primary data come from merging the CRSP Survivorship-Bias-Free Mutual Funds Database with Morningstar Direct for all open-ended U.S. domestic equity mutual funds from 1988 to 2022. From CRSP, we obtain information on returns and total net assets (TNA), expense ratios, turnover ratios, and historical fund names and family identifiers. From Morningstar Direct, we obtain information on Morningstar star ratings, Morningstar globe ratings and their underlying ESG scores, Morningstar categories (investment styles defined along the size–value/growth dimensions), and prospectus benchmark indices. In addition, we obtain fund holdings directly from Morningstar.

To merge CRSP and Morningstar data, we match funds across the two databases using CUSIP, ticker, and fund name through a matching algorithm (Berk and van Binsbergen, 2015; Pástor et al., 2015) that cross-checks time series of funds' monthly returns and TNAs. We also obtain the text descriptions of mutual funds' principal investment strategies from the SEC's Mutual Fund Prospectus Risk/Return Summary Data Sets, which we merge with our main dataset using the CRSP–CIK map. As is standard in the literature, we aggregate share-class-level variables to the fund level across all share classes belonging to the same fund.

We focus on actively managed mutual funds. To this end, we exclude index funds based on the index fund flags provided by CRSP and Morningstar. To mitigate incubation bias (Evans, 2010), we exclude fund-month observations where fund TNA falls below $5 million.

Risk Measures: Beta and Volatility

For risk measures, we use market beta and total volatility. To examine fund managers' security selection decisions and estimate their intended risk exposure at a given point in time, we exploit detailed information from funds' portfolio holdings and the daily returns of their stock holdings (Busse, 2001; Kempf et al., 2009). Specifically, to estimate the expected risk level of fund i in month t denoted by $\mathbb{E}_{t-1}[Risk_{i,t}]$ we proceed in three steps. First, for the portfolio construction we use the portfolio weights at the beginning of month t (equivalently, at the end of month $t-1$). Second, we compute the daily returns of this constructed portfolio over the prior three months (from $t-4$ to $t-1$), holding portfolio weights fixed. Third, we estimate the market beta and volatility using the daily portfolio returns.

Market beta is estimated from the one-factor CAPM model, using CRSP value-weighted returns as the market benchmark. Volatility is calculated as the standard deviation of the portfolio's daily returns, annualized and reported as a percentage. These holdings-based risk measures allow us to capture fund managers' active security selection decisions—such as increasing or reducing exposure to high- or low-risk (beta or volatility) stocks—while accounting for the covariance structure of the fund's stock holdings, particularly in the case of the volatility measure.

We retain fund-month observations with holdings information available at the beginning of the month. Our use of Morningstar holdings data helps minimize data loss, as Morningstar holdings are not only more frequently available than those from Thomson/Refinitiv (commonly used in the mutual fund literature) but also provide a more comprehensive representation of the actual composition of fund portfolios (e.g., Elton et al., 2011).

Style Drift Measures

To measure style drift (deviation from the investment mandate), we use *% Outside Style Box* and *Value/Growth Dispersion*. *% Outside Style Box* is calculated as the percentage of a fund's stock holdings whose Morningstar equity style assignments (based on value–growth scores) fall outside the fund's designated value or growth style box. For a growth (value) fund, *% Outside Style Box* = 100% − *Equity Style Growth (Value)%*, which is obtained from Morningstar Direct. *Value/Growth Dispersion* is calculated as the weighted variance of the value–growth scores of a fund's stock holdings, also obtained from Morningstar Direct.

Growth/Value Style Mandates

Mutual funds typically have specific investment mandates that restrict them from deviating from certain investment styles, such as value or growth. We classify a fund as having a value or growth style mandate if either its name or the name of its prospectus benchmark index contains a value or growth designation. For prospectus benchmark indices, we rely on snapshots of Morningstar data collected between December 2008 and December 2024, as in **?**, since Morningstar does not provide a historical time series of benchmark indices.

Distance to Rating Thresholds

To obtain exogenous variation in the likelihood of rating upgrades and downgrades, we use the fund's return at month $t-36$ relative to its peers with the same star rating as a proxy for expected distance to rating thresholds. The intuition is that the return at $t-36$ influences the three-year rating up to month $t-1$, but not in month t. Consequently, when an extremely poor (good) return from 36 months earlier rolls out of the sample, the fund's percentile ranking moves closer to an upper (lower) rating threshold.

We define $\mathbb{1}(Close_{i,t})$ as an indicator equal to one if the fund's return at $t-36$ falls in the bottom (top) decile relative to its peers with the same rating as of month $t-1$ (for ratings of one to four stars and two to five stars, respectively), and zero if it lies within the middle eight deciles.

To validate our approach, we estimate realized distance to rating thresholds in the post-June 2002 period, when percentile rankings and star ratings are more transparent to replicate. For each category-month, we sort funds with the same ratings at time $t-1$ into deciles based on returns from 36 months earlier R_{t-36}. The most extreme deciles (1 and 10) are assigned to the first quintile (close), deciles 2 and 9 to the second, and so forth. For each quintile, we calculate the average realized distance to rating thresholds among index funds, which are unlikely to adjust portfolios to influence ratings at month t. Table A3 in the Internet Appendix shows that the average realized distance decreases monotonically from the last quintile to the first, with the gap between extreme quintiles exceeding half a percentile.

These results indicate that our instrument successfully identifies funds near rating thresholds. While this information is available to fund managers as of month $t-1$, the percentile rankings and star ratings for month t remain uncertain, since they depend on relative performance during month t, which is the dependent variable in most of our analyses.

Finally, the return from 36 months earlier is unlikely to affect managerial decisions today other than through its impact on Morningstar star ratings. Thus, this variable is plausibly exogenous and provides a valid instrument for classifying funds as close to or far from rating thresholds.

ESG Scores

We use two closely related ESG score measures in the analysis. First, we use the March 2016 Morningstar globe ratings, denoted $ESG^{Globes\ Mar2016}$, to characterize the ESG focus of funds. The globe ratings are based on within-category percentile rankings of portfolio-level

ESG scores that aggregate Sustainalytics' company-level ESG scores. $ESG^{Globes\ Mar2016}$ was a snapshot of data obtained from Morningstar Direct prior to the methodology change in October 2019 (see, e.g., Rzeźnik et al., 2021). Second, to measure funds' ESG tilts, we compute portfolio-level ESG scores, denoted $ESG^{Sustainalytics}$, as value-weighted averages of Sustainalytics' ESG scores of stock holdings for each fund-month with holdings data available at the beginning of the month. We retain fund-month observations with at least 10 stock holdings that have valid Sustainalytics' ESG scores.

ESG Portfolio Returns

To measure the performance of ESG investing, we construct an ESG factor portfolio as follows: (1) merge Sustainalytics ESG data with CRSP stock files by CUSIP and ticker; (2) sort stocks into terciles based on their Sustainalytics ESG scores (restricting to SHRCD = 10 or 11); (3) compute the value-weighted returns of the top (high ESG) and bottom (low ESG) terciles; and (4) construct the ESG factor portfolio by going long the top tercile (high ESG) and short the bottom tercile (low ESG). The return on this ESG portfolio is denoted by R^{ESG}.

ESG Labels

An ESG fund often advertises its ESG focus in several ways. One straightforward approach is to include an ESG-related keyword in the fund's name. A more common way of signaling an ESG orientation is through the investment strategy description provided to investors via the prospectus and other marketing materials. Prior studies find that such an ESG label has a substantial impact on fund flows, above and beyond globe ratings (e.g., Baker et al., 2024). We identify whether a fund has an ESG label based on ESG-related keywords, following Fisch and Robertson (2022) and Andrikogiannopoulou et al. (2022) for fund names and strategy descriptions, respectively.

Other Variables

The rest of the variables are defined in the usual way. As is standard in the literature, we aggregate share-class-level variables to the fund level across all share classes belonging to the same fund. The summary statistics are reported in Table 1.

[Insert Table 1]

3 Value versus Growth Funds

3.1 Background on Morningstar Star Ratings

Morningstar star ratings are based on Morningstar Risk-Adjusted Returns (MRAR) over the past three, five, and ten years, depending on data availability. Each month, share classes are ranked within their peer groups. The top 10% receive five stars, the next 22.5% four stars, the middle 35% three stars, the next 22.5% two stars, and the bottom 10% one star.

An overall rating is assigned as a weighted average of available horizons. Funds less than three years old are unrated. Those with three to five years of history are rated on the three-year performance alone. Those with five to ten years use a 40/60 split between three- and five-year ratings. Those with ten or more years use a 20/30/50 split across the three-, five-, and ten-year ratings. Thus, the past three years of performance account for between 53% and 100% of the rating.

On June 30, 2002, Morningstar refined the peer groups used for ratings. Prior to the change, all U.S. equity mutual funds were ranked relative to one another. However, since style performance is a significant driver of fund performance but out of control of fund managers, Morningstar ratings were more a reflection of style performance rather than manager skills (Evans and Sun, 2021; Ben-David et al., 2022b). This issue became once again prevalent with the dot-com crash, resulting in a drastic underperformance of growth stocks and significantly lower ratings of growth funds. In June 2002 Morningstar started to rank funds relative to only their peers within nine style categories (Large, Mid, Small $\times$ Growth, Blend, Value), e.g., small growth relative to other small growth, and large value relative to other large value funds.

On top of the issue that ratings were largely driven by style performance, we further argue that the peer group choice prior to June 2002 has put growth funds at a significant disadvantage when competing against value funds due to the existence of the value premium (Fama and French, 1993). This is an issue that has not been studied in the literature.

Following prior studies (Evans and Sun, 2021; Ben-David et al., 2022a; Kim, 2022; Han et al., 2024), we exploit the June 2002 change in peer groups (benchmarks) for identification. We first document that growth funds were disadvantaged while value funds were advantaged prior to June 2002 (Sections 3.2 and 3.3). Second, we show that this disadvantage has provided incentives for growth funds to increase risk-taking (Sections 3.4 to 3.8). Third, we provide evidence that risk-taking is effective in raising a fund's star rating.

3.2 Star Rating Disadvantage of Growth Funds

We begin our empirical analysis by examining whether growth funds were disadvantaged relative to value funds when differences in investment styles were not adequately incorporated into Morningstar's relative performance evaluations prior to June 2002.

In Figure 1, we visually assess whether growth funds were disadvantaged relative to value funds by plotting the monthly time series of Morningstar star ratings, averaged across all funds in the growth categories (Large Growth, Mid-Cap Growth, and Small Growth) and the value categories (Large Value, Mid-Cap Value, and Small Value) from January 1988 to December 2022. Prior to the 2002 change in Morningstar's peer groups for star ratings, growth funds on average received lower ratings than value funds. After the change, this disadvantage was completely eliminated by design, with the average ratings for both growth and value funds converging toward three stars, the midpoint of the Morningstar scale.

[Insert Figure 1]

Next, we estimate the extent to which growth funds were disadvantaged relative to value funds. Table 2 shows that, prior to the 2002 change, growth funds earned an average of 2.98 Morningstar stars compared to 3.18 for value funds, a statistically significant difference of −0.20 stars. This gap is also economically meaningful, representing 0.19 standard deviations in star ratings. After the change, the gap was completely eliminated, with the difference declining to virtually zero (0.003) and becoming statistically insignificant.

[Insert Table 2]

The true extent of the disadvantage faced by growth funds is likely more severe than what is shown in Table 2. In response to the benchmark bias in star ratings, growth funds can increase risk-taking in an effort to catch up with value funds and counteract their disadvantage. Because of such behavior, the observed gap in star ratings between growth and value funds is smaller than it would have been had growth funds not adjusted the risk profiles of their portfolios. Quantifying this effect is challenging, since we do not observe ratings for disadvantaged growth funds that refrained from increasing risk-taking. Nevertheless, in Section 3.9, we exploit plausibly exogenous variation in incentives—measured by the distance of fund rankings to rating cutoffs—and present evidence that the effect is economically substantial.

3.3 Time-Varying Star Rating Disadvantage of Growth Funds

We argue that the disadvantage of growth funds relative to value funds is driven by factors unrelated to managerial skill and therefore beyond the control of fund managers.

Rather, it arises from differences in investment styles, specifically risk factor exposures to the value premium. To be consistent with this explanation, in the pre-change period the gap in star ratings between value and growth funds should covary with the realized value premium (HML).

To test this prediction, we estimate the following linear regression model:

$$Stars_{i,t} = \delta \mathbb{1}(\text{Growth}_i) \times \text{HML}_{t-35,t} + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}. \tag{1}$$

where i indexes mutual funds, and t indexes time (in months). $Stars_{i,t}$ denotes fund i's Morningstar star rating at time t. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value).[8] $HML_{t-35,t}$ is the 36-month cumulative return on the value factor from the Fama–French three-factor model (Fama and French, 1993). $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and θ_t denote fund and time fixed effects, respectively. Our sample covers the period 1988–2022, and standard errors are double-clustered by fund and time.

We use the past 3-year window for the realized value premium (HML), since this horizon carries the greatest weight in Morningstar's star rating methodology. As described in Section 3.1, Morningstar star ratings are based on weighted averages of past 3-, 5-, and 10-year performance, depending on fund age. The 3-year performance is the most influential, accounting for 53–100% of the rating.

For identification, we rely on fund fixed effects, exploiting within-fund variation in star ratings that is correlated with time-series variation in the realized value premium (HML).

[8] We exclude funds in the Morningstar blend categories (Large Blend, Mid-Cap Blend, and Small Blend), as well as a small number of funds that experience investment style drift, switching from value to growth or vice versa. As a result, the uninteracted term $\mathbb{1}(Growth_i)$ is absorbed by the fund fixed effects.

To examine how the star rating disadvantage of growth funds changes with the change in peer-group definitions (i.e., benchmarks), we estimate Equation (1) separately for the periods before and after the June 2002 change.

We present the results in Table 3. The key coefficient of interest is δ, which captures the interaction between the growth-fund indicator and the realized value premium (HML). As predicted, δ is negative and statistically significant in the pre-change period (columns (1) and (2)). In other words, a higher realized value premium amplified the disadvantage of growth funds relative to value funds, widening the star rating gap. Including or excluding lagged fund characteristics has no material effect on the estimates. For brevity, we focus on the full specification (column (2)).

The estimated slope coefficient, $\delta = -2.90$, implies that a one–standard deviation (0.23) increase in the 36-month HML return is associated with a reduction of 0.67 stars for growth funds, corresponding to 0.64 standard deviations in star ratings.

By contrast, δ becomes statistically insignificant and economically negligible in the post-change period (columns (3) and (4)), indicating that the June 2002 change in Morningstar's peer-group definitions effectively eliminated the disadvantage faced by growth funds when competing against value funds.

[Insert Table 3]

3.4 Risk-Taking Response: Value versus Growth Funds

We now test the hypothesis that growth funds increased risk-taking when they were disadvantaged by Morningstar's peer-group assignments, which prior to June 2002 did not account for the value premium.

We examine two measures of risk-taking: beta and volatility. Volatility captures total risk, while beta reflects market risk exposure. For each fund–month (i, t), we construct ex-ante measures of risk-taking, $\mathbb{E}_{t-1}[Risk_{i,t}]$, using fund holdings data at the end of month $t-1$ and daily stock returns from the preceding three months (end of month $t-4$ through $t-1$). Details are provided in Section 2.

We remain agnostic about the economic interpretation of these measures, and specifically whether increased risk-taking improves or worsens investor welfare. In Section 3.9, however, we show that greater risk-taking leads to higher star ratings. This supports our mechanism: when disadvantaged by biased benchmarking, fund managers increase risk-taking to boost their star ratings, even at the cost of deviating from their investment style mandates (see Section 3.6).

As a first step, we plot measures of portfolio risk for growth funds and value funds in the pre-2002 period by plotting the monthly time series of beta and volatility, averaged across all funds in the growth categories (Large Growth, Mid-Cap Growth, and Small Growth) and the value categories (Large Value, Mid-Cap Value, and Small Value), as shown in Figures 2 and 3. Growth funds exhibited higher beta and volatility than value funds prior to the 2002 change in Morningstar peer groups. Since the change, the gap in beta and volatility between the two investment styles has nearly disappeared.

[Insert Figures 2 and 3]

More formally, we estimate the extent to which disadvantaged growth funds increased risk-taking using the following difference-in-differences (DiD) regression:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta \mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}. \tag{2}$$

where $\mathbb{E}_{t-1}[Risk_{i,t}]$ is as defined above. $\mathbb{1}(Pre_t^{Jun2002})$ is an indicator variable equal to one if time t falls before the change in the Morningstar peer groups used for the star ratings in June 2002, and zero otherwise. All other variables are as specified in Equation (1). Our sample covers the period 1988–2022, and standard errors are double-clustered by fund and time.

We present the results in Table 4. In columns (1) and (3), we estimate the DiD specification in Equation (2), controlling only for fund and time fixed effects. The coefficient of interest is δ, representing the interaction between the growth fund indicator and the pre-June 2002 period indicator. δ is positive and significant for both risk measures, implying that growth funds took on more risk compared to value funds before June 2002, when disadvantaged by biased benchmarking due to the value premium.

The estimates are not only statistically significant but also economically meaningful. We find that, prior to June 2002, the beta and annualized volatility of disadvantaged growth funds were 0.29 and 6.91 percentage points higher, respectively, than those of value funds. This corresponds to more than one-quarter of the median beta (1.06) and about two-fifths of the median volatility (16.66%) across all fund-month observations in our sample. Adding lagged fund characteristics as controls does not materially affect the results (columns (2) and (4)).

[Insert Table 4]

Identification comes from a DiD framework that compares fund risk-taking across investment style categories before and after June 2002. A potential concern is that the long sample window (1988–2022) may be too broad to attribute the observed risk-taking patterns solely to the Morningstar peer-group change in June 2002. In other words, other developments over this period—such as the gradual weakening of the value factor (HML)—could confound our findings. To address this concern, we conduct robustness checks using shorter and alternative time windows.

Table A1 reports estimates of the specification in Equation (2), including fund controls and fund and time fixed effects, for alternative time windows of ±1 and ±2 years around the

June 2002 Morningstar peer-group change. These narrow windows minimize the likelihood that other events confound our results. Reassuringly, the results are virtually identical across subperiods, with δ consistently significant and point estimates closely matching the baseline estimates in Table 4. Overall, these results suggest that the choice of peer benchmarks can distort mutual fund behavior - prior to the June 2002 Morningstar peer-group change, disadvantaged growth funds exhibited significantly higher beta and volatility.

3.5 Time-Varying Value Premium and Risk-Taking

In examining risk-taking behavior, we have thus far implicitly assumed that the value premium—which disadvantages growth funds relative to value funds—is constant over time. However, the asset pricing literature shows that the value premium is time-varying (Cohen et al., 2003). As documented in Section 3.3, star ratings co-varied strongly with the realized value premium in the pre-June 2002 period. Accordingly, when the value premium is expected to be larger and growth funds anticipate a greater disadvantage, they should increase risk-taking to offset the larger disadvantage when competing with value funds.

Building on Cohen et al. (2003), we use the value spread, $\log(B/M)_{t-1}^{\text{Hi–Lo}}$, as a proxy for the expected value premium $\mathbb{E}_{t-1}[HML_t]$ (conditional on information at time $t-1$). The value spread is defined as the log difference between the book-to-market ratio of the value portfolio (top decile of book-to-market) and that of the growth portfolio (bottom decile of book-to-market) at the end of the previous June.[9]

To test our prediction, we estimate the following linear regression model:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta \mathbb{1}(Growth_i) \times \mathbb{E}_{t-1}[HML_t] + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}. \tag{3}$$

where $\mathbb{E}_{t-1}[HML_t]$ is defined as above. All other variables are as specified in Equations (1)

[9] We thank Ken French for making these data available on his website.

and (2). Our sample covers the period 1988–2022, and standard errors are double-clustered by fund and time.

For identification, we rely on fund fixed effects, exploiting within-fund variation in beta and volatility driven by time-series variation in the expected value premium. To examine how risk-taking behavior changes with the peer-group definitions (i.e., benchmarks), we estimate Equation (3) separately for the periods before and after the June 2002 change in Morningstar peer groups.

We present the results in Table 5. The key coefficient is δ, representing the interaction between the growth fund indicator and the expected value premium. As predicted, δ is positive and statistically significant for both beta and volatility in the pre–June 2002 period (columns (1) and (3)). When the value spread was larger (and fund managers would expect the value premium to be higher, with value stocks outperforming growth stocks by a wider margin), growth funds tended to take on even more risk to offset their greater disadvantage.

For beta, we estimate $\delta = 0.24$, implying that a one–standard deviation increase in the value spread (0.44) widened the beta gap between growth and value funds by 0.11, which is economically significant given the median fund beta of 1.06. For volatility, we estimate $\delta = 8.72$, implying that a one–standard deviation increase in the value spread (0.44) raised volatility by 3.84 percentage points, which is also economically meaningful relative to the median volatility of 16.66%.

In contrast to the pre-change period, δ is indistinguishable from zero for both risk measures in the post-change period (columns (2) and (4)). This suggests that the June 2002 change in Morningstar peer groups effectively eliminated the benchmark bias disadvantaging growth funds in relative performance evaluations.

[Insert Table 5]

3.6 Value/Growth Style Drift

Our economic channel posits that prior to June 2002, growth funds faced a relative disadvantage compared with value funds. This disadvantage stems from the fact that growth funds tend to invest in growth stocks, which on average have lower expected returns than value stocks (Fama and French, 1993). Faced with this disadvantage, growth funds may be tempted to tilt their portfolios to value stocks. Accordingly, if growth funds were concerned about their relative disadvantage before 2002 but not afterward, we expect to observe a reduction in style drift outside their designated style boxes following the Morningstar rating change.

We consider two complementary measures of style drift: % Outside Style Box and Value/Growth Dispersion. The first measure is the percentage of a fund's holdings whose Morningstar equity style assignments (based on value–growth scores) fall outside its designated category (value or growth). The second measure is the weighted variance of value–growth scores within a fund's portfolio, with higher values indicating greater deviation from the mandate.

To formally test for changes in style drift, we estimate the following DiD specification:

$$Drift_{i,t} = \delta \mathbb{1}(Growth_i) \times \mathbb{1}(Post_t^{Jun2002}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}, \tag{4}$$

where $Drift_{i,t}$ is defined as above, and $\mathbb{1}(Post_t^{Jun2002})$ equals one if time t falls on or after the change in Morningstar peer groups in June 2002. All other variables are specified in Equations (1) and (2). Our sample spans the period from 1988 to 2022, and standard errors are double-clustered by fund and time.

The results are reported in Table 6. Columns (1) and (2) report results for % Outside Style Box, while columns (3) and (4) report results for Value/Growth Dispersion. Across all specifications, the interaction term δ is negative and statistically significant. This finding supports our hypothesis that, prior to June 2002, growth funds invested relatively more outside their style mandate than value funds, reflecting concerns that investing in growth stocks placed them at a disadvantage. After the change in peer benchmarks, however, this incentive to deviate from the mandate disappeared, and growth funds' portfolios became more closely aligned with their designated styles.

The estimates are economically meaningful. Growth funds reduced their holdings outside the style box by about 4.5-4.7 percentage points, corresponding to roughly one-third of a standard deviation in the % Outside Style Box measure. Similarly, the decline of 2.7-3.0 units in Value/Growth Dispersion represents about 0.15-0.17 of a standard deviation. Taken together, these results suggest that the 2002 revision to Morningstar's RPE methodology substantially curtailed style drift by growth funds.

[Insert Table 6]

3.7 Role of Style Mandates

Growth managers facing a disadvantage due to the value premium could, in principle, attempt to mitigate their disadvantage by tilting their portfolios toward value stocks. In practice, however, this is rarely feasible because most funds operate under explicit investment style mandates along the value-growth dimension, which constrain them from deviating substantially from their designated style. Consequently, increasing risk while remaining close to the mandated style represents a more practical strategy for growth funds to offset their disadvantage.

We test this hypothesis by splitting funds into two subgroups: those with explicit style mandates and those without. A fund is classified as having a style mandate if its name or its prospectus benchmark name includes a value or growth designation. The vast majority of funds in our sample—nearly 80% of fund–month observations—do have value/growth style designations, underscoring the central role of style mandates in constraining portfolio choices. The premise is that funds with style mandates face stricter restrictions on portfolio choice and are therefore less able to tilt away from their designated style. Table A2 in the Online Appendix supports this assumption: both of our style drift measures are significantly smaller for funds with style mandates (see also Section 3.6).

Consider a growth fund without a specific growth mandate. Such a fund has some flexibility to allocate part of its portfolio to value stocks, thereby partially mitigating the benchmark-induced disadvantage prior to June 2002. In contrast, a growth fund with an explicit growth mandate is more constrained to invest in growth stocks and has limited ability to tilt toward value. As a result, growth funds with mandates are more likely to respond to the pre-2002 disadvantage by investing in relatively riskier stocks.

We test the role of style mandates in shaping the risk-taking behavior of disadvantaged growth funds in two steps. First, we re-estimate the DiD specification in (2) separately for the subsamples of funds with and without style mandates. Consistent with our argument, we predict that δ will be larger for the subsample of funds with mandates and smaller for those without mandates. Second, we further test this prediction by estimating a full-sample regression that includes a triple interaction among $\mathbb{1}(\text{Growth}_i)$, $\mathbb{1}(\text{Pre}_t^{\text{Jun2002}})$, and $\mathbb{1}(\text{Mandate}_{i,t-1})$:

$$\begin{aligned}\mathbb{E}_{t-1}[Risk_{i,t}] = {} & \rho\mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) \times \mathbb{1}(Mandate_{i,t-1}) + \delta_1\mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) \\ & + \delta_2\mathbb{1}(Growth_i) \times \mathbb{1}(Mandate_{i,t-1}) + \delta_3\mathbb{1}(Pre_t^{Jun2002}) \times \mathbb{1}(Mandate_{i,t-1}) \\ & + \beta\mathbb{1}(Mandate_{i,t-1}) + \gamma\Gamma_{i,t-1} + \theta_i + \theta_t\ (\theta_{c,t-1}) + \varepsilon_{i,t}.\end{aligned} \tag{5}$$

where $\mathbb{1}(Mandate_{i,t-1})$ is an indicator variable equal to one if fund i has a value or growth mandate at time $t-1$. All other variables are specified in Equations (1) and (2). Our sample covers the period from 1988 to 2022, and standard errors are double-clustered by fund and time. In some specifications we replace time fixed effects θ_t with lagged category-by-time fixed effects $\theta_{c,t-1}$ to explicitly control for time-varying differences in the risk profiles of fund holdings across size-by-value/growth categories. The coefficient of interest is ρ, which provides a formal test of whether disadvantaged growth funds with mandates take more risk than those without, relative to their value counterparts.

Table 7 reports the results. In Panel A, columns (1) and (3) present the results for funds with style mandates, while columns (2) and (4) show the corresponding results for funds without mandates. When risk-taking is measured by beta, δ is 0.31 for funds with mandates, whereas δ is 0.17 for funds without mandates. Thus, although disadvantaged growth funds increase risk-taking in both groups, the effect is nearly twice as large for funds with mandates. Consistent with this, the coefficient on the triple interaction term, $\rho = 0.12$ (Panel B, column (1)), indicates that the heightened risk-taking of disadvantaged growth funds in the pre–June 2002 period is particularly pronounced among funds with style mandates, which are less able to mitigate the benchmark-induced disadvantage by tilting toward value. Importantly, the triple interaction regression shows that this difference between funds with and without mandates is statistically significant.

Similarly, when risk-taking is measured by volatility, δ is 7.31 for funds with mandates and 4.65 for funds without mandates (columns (3) and (4) in Panel A). The difference is economically large, indicating that style mandates impose meaningful constraints on funds' ability to invest outside their mandated universe of stocks. Furthermore, the triple interaction coefficient, $\rho = 2.65$ (Panel B, column (3)), is again statistically significant.

[Insert Table 7]

A potential concern in our DiD specification in Equation (2) is that our results might be driven by changes in risk profiles of the underlying stocks rather than changes in funds' risk-taking behaviors. Indeed, the average growth stock was more volatile relative to the average value stock prior to June 2002. A possible explanation of the differential change in volatilities of growth versus value stocks around 2002 might be that many risky growth stocks have vanished after the dot-com crash.

We address this concern by exploiting time-varying within-style differences in our triple interaction framework. This approach provides an ideal setting to control for time variation in style-specific characteristics–such as, but not limited to, differences in the risk profiles of value and growth stocks–and allows us to isolate the risk-taking behavior of funds. More specifically, we replace time fixed effects with lagged category-by-time fixed effects in the triple interaction model in Equation (6). This approach explicitly controls for time-varying unobservables common to funds (with or without style mandates) within a given category, while allowing such unobservables to differ across the six possible size–value/growth categories. As a result, our results are not driven by differential changes in the risk profiles of value and growth stocks.

Columns (2) and (4) of Panel B in Table 7 present the results of the triple interaction test with lagged category-by-time fixed effects. The coefficient ρ remains significant across both risk-taking measures. For beta, ρ decreases only slightly from 0.12 to 0.09, indicating that time-varying style-specific unobservables explain little of our baseline effect. Similarly, for volatility, ρ changes only modestly from 2.65 to 2.25. Thus, accounting for style-specific time-varying unobservables has no material impact on our results.

3.8 Risk-Taking Incentives Near Ratings Thresholds

We posit that disadvantaged growth funds are particularly concerned about their relative performance, as it directly determines their Morningstar star ratings. Since star ratings are discrete with only five possible values, the incentives to attempt to influence star ratings via increased risk-taking are strongest for funds positioned near one of the four cutoffs. To illustrate, consider two disadvantaged growth funds that both currently hold a four-star rating. The boundaries for four stars correspond to the 67.5th and 90th percentiles in the ranking of Morningstar risk-adjusted returns. Suppose the first fund is ranked just below the cutoff for five stars (e.g., at the 89.9th percentile), while the second fund lies far from either cutoff (e.g., at the 78.75th percentile). Risk-taking by the first fund is more likely to make a difference in star ratings than for the second fund. Accordingly, the first fund has much stronger incentives to increase risk than the second.

We build on this insight and test whether disadvantaged growth funds that are likely to be ranked closer to a Morningstar rating cutoff take more risk relative to their value counterparts. Since Morningstar does not provide historical percentile rankings based on the rating methodology in effect at the time, we cannot directly compute the distance between a fund's percentile ranking and the nearest rating cutoff without first estimating these rankings. Estimating percentile rankings of Morningstar's risk-adjusted ratings, however, is particularly challenging, especially prior to June 2002 (Evans and Sun, 2021). A further complication is that fund managers may not know their percentile rank in real time, as it is difficult to track all peers, and their rankings and ratings for month t depend on performance relative to all peer funds in that month. Moreover, a fund's recent relative performance may itself influence its current risk-taking behavior, so using rankings based on recent performance may not precisely identify the effect of a fund's expected distance to a rating cutoff.

To obtain exogenous variation in distance to rating thresholds, we use the one-month return from three years earlier (i.e., $r_{i,t-36}$) and define an indicator variable, $\mathbb{1}(\text{Close}_{i,t})$, that equals one if $r_{i,t-36}$ falls in either the bottom or top decile relative to peers with the same rating at month $t-1$. The intuition is as follows. Morningstar star ratings place between 53% and 100% weight on the most recent 36 months of returns (Section 3.1). Accordingly, the one-month return three years ago is relevant for the rating at $t-1$ but becomes much less influential for the rating at t, and in fact becomes irrelevant for funds younger than five years when the return drops out of the 36-month window. A large negative (positive) return three years ago is therefore likely to shift a fund's ranking closer to a rating threshold, increasing the likelihood of crossing into a higher (lower) star rating at t. In the special case of funds currently holding a one- or five-star rating, where the only possible change is an upgrade or downgrade, we adjust the classification: a fund is considered close to a cutoff if $r_{i,t-36}$ is in the bottom decile (for a one-star fund) or the top decile (for a five-star fund).

To confirm the validity of our identification strategy, Table A3 shows that funds with large absolute past returns $r_{i,t-36}$ are indeed more likely to rank closer to a rating threshold. We conduct this validity check using the subsample of index funds, which are unlikely to make active investment decisions aimed at influencing star ratings at month t. In addition, we restrict the sample to the post-June 2002 period, when Morningstar's methodology became more transparent and risk-adjusted returns were more straightforward to compute, allowing us to obtain accurate fund rankings.

Finally, note that the one-month return from three years earlier is unlikely to affect current risk-taking except through the Morningstar rating channel. Accordingly, $r_{i,t-36}$ serves as a valid instrument for identifying a fund's expected distance to a rating cutoff, as it is unlikely to influence current risk-taking through any channel other than the one we aim to capture.

We follow the approach in Section 3.7 and examine the effect of a fund's expected distance to a rating cutoff on the risk-taking of disadvantaged growth funds in two steps. First, we re-estimate the DiD specification in (2) separately for the subsamples of funds that are close to and distant from a rating cutoff. We expect δ to be larger (smaller) for funds that are close to (distant from) a cutoff. In the second step, we estimate a full-sample regression that includes a triple interaction among $\mathbb{1}(\text{Growth}_i)$, $\mathbb{1}(\text{Pre}_t^{\text{Jun2002}})$, and $\mathbb{1}(\text{Close}_{i,t})$:

$$\begin{aligned}\mathbb{E}_{t-1}[Risk_{i,t}] = {} & \rho \mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) \times \mathbb{1}(Close_{i,t}) + \delta_1 \mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) \\ & + \delta_2 \mathbb{1}(Growth_i) \times \mathbb{1}(Close_{i,t}) + \delta_3 \mathbb{1}(Pre_t^{Jun2002}) \times \mathbb{1}(Close_{i,t}) \\ & + \beta \mathbb{1}(Close_{i,t}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t \ (\theta_{c,t-1}) + \varepsilon_{i,t}.\end{aligned} \tag{6}$$

where $\mathbb{1}(Close_{i,t})$ is as defined above. All other variables are specified in Equations (1) and (2). Our sample spans the period from 1988 to 2022, and standard errors are double-clustered by fund and time.

As in Section 3.7, the triple-interaction specification yields an identification strategy that controls for all time-varying characteristics common to a fund style (such as the risk profiles of value and growth stocks) through (lagged) category-by-time fixed effects. This framework isolates the causal effect of expected distance to a rating cutoff on the risk-taking of disadvantaged growth funds, whose only source of variation arises from their differential incentives to attempt to mitigate the benchmark-induced disadvantage prior to June 2002.

Table 8 reports the results and confirms our expectations. In Panel A, columns (1) and (3) present the estimates for beta and volatility for funds close to a rating cutoff, while columns (2) and (4) show the corresponding estimates for funds farther away. For funds near a cutoff, the beta of growth relative to value funds was on average 0.29 higher in the pre–June 2002 period (Panel A, column (1)). Consistent with our expectation, this estimate decreases (albeit only slightly) to 0.27 for funds more distant from a cutoff (Panel A, column (2)). A similar pattern emerges for volatility, with δ equal to 7.20 for funds close to a cutoff compared to 6.57 for those farther away.

Panel B of Table 8 shows that ρ is positive and significant for both risk-taking measures, regardless of whether we include time fixed effects (columns (1) and (3)) or lagged category-by-time fixed effects (columns (2) and (4)). First, this finding supports our hypothesis that disadvantaged growth funds with stronger incentives to improve their Morningstar ratings engaged in relatively more risk-taking prior to June 2002. Second, the specification with lagged category-by-time fixed effects (columns (2) and (4)) demonstrates that our risk-taking mechanism remains robust even after controlling for time-varying, style-specific unobservables. Notably, with lagged category-by-time fixed effects, ρ increases slightly from 0.017 to 0.018 in the case of beta and from 0.511 to 0.518 in the case of volatility.

[Insert Table 8]

3.9 Closing the Gap in Star Ratings

An important question is whether the increase in risk-taking by disadvantaged growth funds was effective in mitigating their inherent disadvantage and making them more competitive relative to advantaged value funds in the pre–June 2002 period. If so, and fund ratings did improve, this would validate the central assumption of our economic mechanism that disadvantaged funds had an incentive to increase risk-taking. The challenge, however, is that we cannot directly observe the counterfactual performance of disadvantaged growth funds had they not increased their risk prior to June 2002.

However, we can compare the performance of growth funds that differ in their risk-taking due to plausibly exogenous differences in incentives—specifically, being close to versus far from Morningstar rating cutoffs. Recall from Section 3.8 that disadvantaged growth funds closer to a cutoff take relatively more risk than those farther away, compared with their value counterparts. Accordingly, if disadvantaged growth funds close to a cutoff are more successful in closing the gap in star ratings relative to value funds, this would indicate that the increase in risk-taking is effective; if not, it would suggest that such risk-taking is counterproductive.

Table 9 compares the Morningstar ratings of growth versus value funds that are close to versus far from a rating cutoff in both the pre- and post–June 2002 periods. In the pre–June 2002 period, growth funds close to a cutoff had ratings 0.12 stars lower than comparable value funds. However, this gap nearly doubles to 0.23 stars for growth funds that were far from a cutoff. Moreover, the difference of 0.11 stars ($= 0.23 - 0.12$) is statistically significant. As a sanity check, we confirm that in the post–June 2002 period there is no meaningful difference in star ratings between value and funds, regardless of their proximity to a rating cutoff.

[Insert Table 9]

Recall that the observed difference in average star ratings between value and growth funds was 0.2 stars before the 2002 methodology change (Table 2). If anything, this likely underestimates the true disadvantage of growth funds, as they increased risk-taking to mitigate their benchmark disadvantage—by 0.28 in beta or 6.75% in volatility (Table 8). To obtain a corrected estimate of the disadvantage growth funds faced in the pre–June 2002 period, we conduct a back-of-the-envelope calculation. Specifically, we construct a counterfactual difference in ratings between value and growth funds under the hypothetical scenario in which growth funds did not increase risk-taking in response to their disadvantage.

To do so, we first estimate the rating increase per unit of risk-taking. We use the 0.08-star rating difference between growth funds close to versus distant from rating cutoffs (Table 9)[10]. In addition, we use the 0.018 difference in beta and the 0.518% difference in volatility across these two types of growth funds (Table 8). These estimates imply a 4.4-star rating increase per unit of beta and a 0.15-point increase per 1% volatility. Multiplying by the pre-2002 increase in growth funds' risk-taking reported in Table 4 yields an implied rating increase of $0.28 \times 4.4 = 1.2$ stars based on beta, or $6.75 \times 0.15 = 1.0$ stars based on volatility.

[10]The 0.08-star estimate is conservative. Controlling for the rating difference between equivalent value funds yields an estimate of 0.11 stars.

Adding this to the observed average value–growth gap of 0.2 stars (Table 2), we estimate that the counterfactual disadvantage of growth funds prior to June 2002 was between 1.2 and 1.4 stars. Thus, biased benchmarking before the 2002 methodology change gave rise to an economically significant disadvantage for growth funds.

We conclude with two final notes. First, we cannot conduct a comparable efficacy analysis for the sample split by style mandates. The reason is that while both growth funds with and without style mandates faced the same incentives to mitigate the pre-2002 disadvantage, the strategies available to them differed. Growth funds without a style mandate had greater flexibility to tilt their portfolios toward value, which was likely a more effective approach than increasing risk-taking in the presence of the value premium. By contrast, growth funds with a style mandate were constrained in their admissible investment universe and thus had to rely more heavily on risk-taking to address the disadvantage. Because we cannot quantify the efficacy of tilting toward value, we cannot draw conclusions from comparing the star ratings of funds with versus without style mandates.

Second, it is instructive to compare our results to the large literature initiated by Brown et al. (1996) and Chevalier and Ellison (1997), which shows that past-loser funds increase risk-taking in an attempt to catch up with winners. The mechanism in their setting is fundamentally different from ours. In their framework, poor past performance—whether due to bad luck or low skill—creates ex-post incentives to take on additional risk. By contrast, our mechanism is novel in that it captures ex-ante risk-taking incentives arising as an unintended consequence of the biased benchmark imposed by Morningstar's original peer grouping. Biased benchmarking disadvantaged certain funds from the outset, thereby inducing them to increase risk-taking. To further distinguish our mechanism from the past-loser/winner channel, we provide evidence in this section that the additional risk-taking of disadvantaged growth funds was at least partially effective in mitigating their underperformance relative to value funds. In contrast, Huang et al. (2011) show that the risk-shifting of past-loser funds is inefficient, leading to subsequent underperformance.

4 ESG versus Non-ESG Funds

In Section 4, we revisit our economic question of interest, the impact of biased benchmarks, in an ESG context. While the relationship between a manager's value/growth tilt and their risk-taking incentives is more established than the analog for ESG tilts, we use the insights of Pástor et al. (2022) as a guide for structuring our empirical tests. They first find evidence of a negative greenium.[11] In the presence of a negative greenium, ESG funds are at a disadvantage when fund flows are driven primarily by Morningstar star ratings, which evaluate relative performance solely on risk–return dimensions. Consequently, ESG funds have an incentive to increase risk-taking in order to enhance their risk–return performance, potentially at the cost of ESG performance, in an effort to offset the disadvantage arising from biased benchmarking. However, the introduction of the Morningstar ESG globe ratings in March 2016 raised the salience of ESG performance for these funds, enabling them to differentiate themselves from non-ESG funds when competing for flows. This effectively alters the peer grouping against which ESG and non-ESG funds are compared, thereby mitigating the earlier disadvantage of ESG funds. Our first set of results tests this economic hypothesis.

Pástor et al. (2022) also find evidence consistent with time-varying expectations in ESG concerns. Even with a persistent negative greenium, time-varying ESG concerns can result in time variation in the relative performance of high and low ESG-rated stocks. In periods with increasing ESG concerns, if high ESG-rated stocks are expected to outperform low, than the risk-taking motivation described above would be mitigated or possibly even reversed. Our second set of results revisits manager risk-taking in the context of this time-variation.

[11] Additional evidence of a negative greenium in bonds and stocks is documented by Baker et al. (2022), Moalla and Dammak (2024), Zerbib (2019), Immel et al. (2021), Arat et al. (2023), Stotz (2021), and Eskildsen et al. (2024).

4.1 Background on Morningstar Globe Ratings

Morningstar introduced the ESG (initially called Sustainability) 'Globe' Rating for funds in March 2016. This holdings-based measure aggregates Sustainalytics' company-level ESG scores to the portfolio level and assigns one to five globes based on percentile rankings within Morningstar categories, analogous to the star ratings.

The seminal natural-experiment study is Hartzmark and Sussman (2019), who exploit the 2016 launch of globe ratings as a salience shock and document large causal flow responses to globe ratings. They further show that investors respond primarily to the discrete globe labels rather than the underlying continuous ESG scores, and that high globe-rated funds do not outperform low globe-rated funds—consistent with salience and non-pecuniary motives driving demand.

In related work, Gantchev et al. (2024) show that managers tilt portfolios toward higher-ESG holdings to improve globe ratings and attract flows, but that such trades underperform, creating an ESG–performance trade-off and helping to explain why long-run flow sensitivity to globe ratings weakens. Building on these studies, we focus on a two year window around the March 2016 introduction of globe ratings for identification.[12]

4.2 Flow Sensitivity to Star Ratings

We first examine how the introduction of ESG globe ratings affected investors' capital allocation across mutual funds. Specifically, we test whether the launch of Morningstar globe ratings in March 2016 altered the sensitivity of investor flows to Morningstar star ratings. Prior to their introduction, ESG and non-ESG funds were evaluated solely on star ratings, placing ESG funds at a disadvantage because their sustainability focus was less

[12] Because Morningstar's globe rating methodology was changed in 2019, the previous ratings were overwritten. Fortunately, we have a prior snapshot of the original ratings from March 2016 to February 2018. In the Appendix, we examine robustness of our result over longer event windows.

salient to investors. We therefore expect that the introduction of globe ratings attenuated the sensitivity of flows to star ratings, thereby reducing the competitive pressure ESG funds faced along the risk–return dimension.

To this end, we estimate the following regression:

$$
\begin{aligned}
Flow_{i,t} = {} & \delta Stars_{i,t-1} \times \mathbb{1}(Post_t^{Mar2016}) \\
& + \beta Stasr_{i,t-1} + \sum_{s=1}^{3} \rho_s Flow_{i,t-s} + \gamma \Gamma_{i,t-1} + \theta_{c,t-1} + \varepsilon_{i,t}
\end{aligned}
\tag{7}
$$

where $Flow_{i,t}$ denotes net capital flow at time t (as a percentage of lagged TNA), $Stars_{i,t-1}$ is the lagged Morningstar star rating, $\mathbb{1}(Post_t^{Mar2016})$ is an indicator for the post-globe period, $\Gamma_{i,t-1}$ includes lagged fund characteristics as specified in Equation (1), and $\theta_{c,t-1}$ denotes lagged category-by-time fixed effects. The sample covers the period from March 2015 to February 2017, and standard errors are double-clustered by fund and time. We estimate the regression separately for funds above and below the median value of the fund's Morningstar ESG score used to determine its initial globe rating in March 2016.

Table 10 reports the results. Our coefficient of interest is δ. For high globe-rated (i.e., ESG) funds, δ is negative and statistically significant, consistent with our hypothesis that the introduction of globe ratings substantially reduced the sensitivity of flows to star ratings for ESG funds. The point estimate of –0.15 is economically meaningful, offsetting more than one-quarter of the baseline flow sensitivity ($\beta = 0.56$). By contrast, for low globe-rated funds, δ is close to zero and statistically insignificant, indicating no change in star–flow sensitivity.

[Insert Table 10]

4.3 Risk-Taking Response: ESG versus Non-ESG Funds

We next examine whether the introduction of Morningstar globe ratings altered the risk-taking behavior of ESG versus non-ESG funds. Prior to March 2016, both types of funds were evaluated solely on their star ratings, placing ESG funds at a disadvantage when their portfolios tilted toward high-ESG stocks with lower expected returns. We hypothesize that ESG funds respond by increasing risk-taking in order to offset this disadvantage. We further examine if the launch of globe ratings resulted in investors evaluating funds on the ESG dimension, mitigating the incentive of ESG fund managers to take on additional risk.

To test these hypotheses, we estimate the following regression:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) + \gamma\Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}, \tag{8}$$

where $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes the expected risk-taking, measured by beta or volatility of fund i in month t, as defined in Section 2. $ESG_i^{Globe\ Mar2016}$ is fund i's Morningstar ESG score used to determine the initial globe ratings at their introduction in March 2016, standardized for ease of interpretation. Identifying ESG funds based on the March 2016 globe rating and holding this classification constant throughout the sample period helps mitigate endogeneity concerns. $\mathbb{1}(Pre_t^{Mar2016})$ is an indicator equal to one if time t is prior to March 2016. All other variables are as defined in Equation (7). Our sample covers March 2015 to February 2017, and standard errors are double-clustered by fund and time.

Table 11 reports the results. Our coefficient of interest is δ. Across both risk-taking measures, δ is positive and statistically significant. This supports our hypothesis that, relative to non-ESG funds, ESG funds engaged in greater risk-taking prior to the introduction of globe ratings—consistent with efforts to offset their disadvantage when competing solely on star ratings. Specifically, a one-standard deviation increase in the ESG score is associated with an increase in beta of about 0.017 and an increase in volatility of about 0.27% before the introduction of globe ratings. Once globe ratings made sustainability more salient to investors, this incentive to take additional risk partially dissipated.

[Insert Table 11]

We further assess the robustness of these results by extending the time windows around the March 2016 introduction of globe ratings. Table A4 in the Online Appendix reports estimates for the periods September 2014 to August 2017 and March 2014 to February 2018. The magnitude of δ remains comparable to the baseline estimate and statistically significant across all specifications. These robustness checks confirm that our findings are not driven by the specific choice of sample window: the elevated risk-taking of ESG funds in the pre-globe period (relative to non-ESG funds) is consistently observed in the data.

In additional robustness tests, reported in Table A5 in the Online Appendix, we replace the continuous ESG scores with the discrete globe ratings initially assigned at their introduction. The results remain robust under this alternative specification.

4.4 ESG Labels (Placebo Tests)

Our mechanism assumes that the introduction of globe ratings in March 2016 acted as a salience shock, enabling ESG funds to differentiate themselves from other funds. This mechanism is less applicable to funds that were already explicitly labeled as ESG prior to March 2016. For these funds, investors could readily identify their sustainability focus, meaning they should have been able to differentiate from other funds even before the introduction of globe ratings. If salience did not materially change for these funds around March 2016, then their relative disadvantage (if any) should also have remained unchanged, and we would not expect to observe a shift in their risk-taking behavior.

To test this prediction, we conduct placebo tests using funds with explicit ESG labels. We classify a fund as ESG-labeled if its name or the description of its principal investment strategy in the prospectus contains ESG-related keywords. We then re-estimate the regression in Equation (8) for the subsample of ESG-labeled funds. Panel A of Table 12 shows that for these funds, δ is not statistically significant. This finding is consistent with our conjecture that the introduction of globe ratings in March 2016 did not materially alter the risk-taking incentives of funds explicitly labeled ESG prior to March 2016.

In addition, we estimate the following triple-interaction specification to compare funds with and without ESG labels:

$$\begin{aligned}\mathbb{E}_{t-1}[Risk_{i,t}] = {} & \rho ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) \times \mathbb{1}(No\ ESG\ Label_i) \\ & + \delta_1 ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) + \delta_2 \mathbb{1}(Pre_t^{Mar2016}) \times \mathbb{1}(No\ ESG\ Label_i) \\ & + \gamma \Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}\end{aligned}$$

where $\mathbb{1}(No\ ESG\ Label_i)$ is an indicator for funds without an ESG label. All other variables are defined as in Equation (8). Our sample covers the period from March 2015 to February 2017, and standard errors are double-clustered by fund and time.

Table 12, Panel B reports the results. Our coefficient of interest is ρ, which is positive and statistically significant. A one-standard deviation increase in the ESG score is associated with an increase in beta of about 0.036 or in volatility of about 0.52 before March 2016. These magnitudes are comparable to those in Table 11, confirming that the elevated risk-taking prior to globe ratings was driven by ESG funds without explicit labels.

To summarize, the placebo tests reinforce our mechanism. Only ESG funds whose sustainability orientation was not already apparent to investors prior to the introduction of globe ratings experienced a salience shock and a meaningful change in their disadvantage. Accordingly, it was these funds that engaged in comparatively greater risk-taking before

March 2016.

[Insert Table 12]

4.5 Time-Varying ESG Expectations and Risk-Taking

Given an expected negative greenium, ESG funds are disadvantaged because ESG stocks earn relatively lower expected returns. However, the evidence from Pástor et al. (2022) on the time-variation in the expected returns of ESG stocks implies time-variation in both the disadvantage of ESG funds and their incentives to take additional risk. Moreover, the salience shock from the introduction of globe ratings should dampen ESG funds' incentives to adjust risk-taking in response to the time-variation in expected ESG stock returns after March 2016.

To test this prediction, we construct a representative ESG stock index. Each month, we sort stocks by their Sustainalytics ESG scores and form a portfolio that is long the top tercile and short the bottom tercile. The return of this index in month t is denoted by R_t^{ESG}. We use Sustainalytics ESG scores as the sorting variable because they are also the inputs underlying Morningstar's globe ratings.

First, to establish that ESG performance—as measured by our ESG index—drives the disadvantage of high globe-rated funds relative to low globe-rated funds (analogous to our analysis in Section 3.3), we estimate the following model:

$$Stars_{i,t} = \delta ESG_i^{Globes\ Mar2016} \times R_{t-35,t}^{ESG} + \gamma \Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}. \tag{9}$$

where $Stars_{i,t}$ denotes fund i's Morningstar star rating at time t. $R_{t-35,t}^{ESG}$ denotes the cumulative return of our ESG index over the preceding 36 months. All other variables are defined in Equation (8). Our sample covers the period from March 2015 to February 2017, and standard errors are double-clustered by fund and time.

Table 13 reports the results. Our coefficient of interest is δ, which is positive and highly statistically significant. This suggests that funds with high globe ratings receive higher (lower) star ratings following strong (weak) past performance of our ESG index. In other words, the performance of ESG funds is well explained by the returns of our ESG index. This result is not surprising, as funds assigned high globe ratings must have held stocks with high Sustainalytics scores—the same scores used as the sorting variable in constructing our ESG index. By design, therefore, the portfolio holdings of high globe-rated funds and our ESG index should largely overlap.

Next, we test the hypothesis that ESG funds take more (less) risk when their disadvantage is expected to be larger (smaller), and that this effect should be stronger prior to the salience shock from the introduction of globe ratings in March 2016. We implement the following triple-interaction specification:

$$\begin{aligned}\mathbb{E}_{t-1}[Risk_{i,t}] = {} & \rho ESG_i^{Globes\ Mar2016} \times \mathbb{E}_{t-1}[R_t^{ESG}] \times \mathbb{1}(Pre_t^{Mar2016}) \\ & + \delta_1 ESG_i^{Globes\ Mar2016} \times \mathbb{E}_{t-1}[R_t^{ESG}] + \delta_2 ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) \\ & + \gamma \Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}\end{aligned} \tag{10}$$

Because our focus in on expected returns, we use the future realized return R_t^{ESG} in month t as a proxy for its conditional expectation at $t-1$. Although the realized one-month-ahead return is a noisy proxy for the expected return, it yields a consistent and unbiased estimate in our specification. The key insight is that all other variables are measured at $t-1$ and are therefore orthogonal to the unexpected component of the realized return. All other variables are defined in Equation (8). Our sample covers the period from March 2015 to February 2017, and standard errors are double-clustered by fund and time.

Table 14 reports the results. Our coefficient of interest is ρ, which is negative and statistically significant across all specifications. This supports our conjecture that high globe-rated funds take more (less) risk when the expected return of ESG stocks is low (high), with the effect being stronger in the pre-March 2016 period.

4.6 Change in ESG Tilts

Finally, we examine whether the introduction of globe ratings affected the ESG orientation of fund portfolios. Prior to March 2016, in balancing ESG and return concerns, we expect that ESG funds hold lower-ESG stocks. This would reflect a strategic response to their disadvantage under star ratings as the sole RPE system. Because ESG stocks were typically associated with lower risk and correspondingly lower expected returns, funds had an incentive to tilt toward riskier, lower-ESG stocks to boost performance and compete for stars. Once globe ratings made ESG performance salient to investors, however, ESG funds gained an explicit advantage from holding higher-ESG stocks consistent with their sustainability focus, and their portfolio incentives shifted accordingly.

To test this prediction, we estimate the following regression:

$$ESG_{i,t}^{Sustainalytics} = \delta\, ESG_i^{Globe\ Mar2016} \times \mathbb{1}(Post_t^{Mar2016}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}, \quad (11)$$

where $ESG_{i,t}^{Sustainalytics}$ is the value-weighted average of Sustainalytics ESG scores of fund i's stock holdings at month t. All other variables are defined in Equation (8). Our sample covers the period from March 2015 to February 2017, and standard errors are double-clustered by fund and time.

Table 15 reports the results. Our coefficient of interest is δ, which is positive and statistically significant across specifications. A one-standard deviation increase in the Morningstar ESG score at the time of the globe ratings' introduction corresponds to an increase of about 0.20 points in portfolio-level ESG scores in the post–March 2016 period. This magnitude is economically meaningful relative to the within-fund variation in ESG tilts over time.

These findings indicate that after the introduction of globe ratings, funds with higher baseline ESG scores prior to the March 2016 salience shock shifted their portfolios further toward firms with stronger ESG characteristics. This pattern is consistent with managers responding to the new performance dimension introduced by Morningstar by tilting their portfolios toward higher-ESG stocks.

[Insert Table 15]

5 Conclusion

While relative performance evaluation can theoretically provide more efficient contracting between a principal and the agent they select to act on their behalf, this depends on the efficacy of the peer group or benchmark used in the performance evaluation. The prior literature examines incomplete benchmarking due to the influence agents have over the relative comparison - such as the strategic selection of peer groups and benchmarks by CEOs and fund managers respectively. In contrast, we explore changes in third party RPE practices and the response of both principals and agents.

We examine these issues in two settings: the Morningstar's 2002 ratings methodology change and 2016 ESG globe rating introduction. In both cases, certain fund managers (i.e., growth and ESG funds) are disadvantaged relative to other fund managers (i.e., value and non-ESG funds) due to the peer group assignment. As a result, both growth and ESG funds select securities with higher risk before the change. However, after the 2002 ratings methodology change, growth funds are no longer compared to value funds, and with the 2016 ESG globe ratings introduction, investors can distinguish between risk/return and ESG characteristics. As a result, we find that growth and ESG funds (i.e., those with high globe ratings) no longer select higher risk securities. Additionally, we find that ESG funds start selecting higher ESG holdings after the change.

Our results using a third party assignment of RPE peer groups/benchmarks and plausibly exogenous changes in those assignments, provide important evidence on the limits of RPE efficacy. In the case of the 2002 Morningstar methodology change, we see that when peer groups/benchmarks compare agents with systematically different risk characteristics, the lower risk agent chooses to increase the risk of their strategy to compete. However, when the relative comparison correctly separates the two groups (i.e., growth and value funds), this risk-taking behavior declines.

Similarly, in the case of the 2016 globe ratings introduction, when an alternative dimension of security selection (i.e., ESG) is not incorporated into the RPE approach, and that dimension is negatively correlated with risk, agents will similarly increase risk to compete. However when the RPE approach is modified to incorporate that alternative dimension (i.e., both ESG and risk/return dimensions of funds are rated), this risk-taking behavior disappears. These results provide important insights into optimal construction of RPE comparison groups.

References

Albuquerque, A. M., Franco, G. D., Dikolli, S. S., Krishnan, R., 2013. Ceo power and relative performance evaluation. Contemporary Accounting Research 30, 1429–1462.

Albuquerque, R., Koskinen, Y., Zhang, C., 2019. Corporate social responsibility and firm risk: Theory and empirical evidence. Management Science 65, 4451–4949.

Andrikogiannopoulou, A., Krueger, P., Mitali, S. F., Papakonstantinou, F., 2022. Empty ESG promises: A text analysis of mutual fund prospectuses. Working paper.

Arat, E., Hachenberg, B., Kiesel, F., Schiereck, D., 2023. Greenium, credit rating, and the COVID-19 pandemic. Journal of Asset Management 24, 547–557.

Bai, J. J., Ma, L., Mullally, K., Tang, Y., 2024. What determines mutual fund managers' compensation? Evidence from US administrative earnings data. Working paper.

Baiman, S., Demski, J., 1980. Economically optimal performance evaluation and control-systems. Journal of Accounting Research 18, 184–220.

Baker, M., Bergstresser, D., Serafeim, G., Wurgler, J., 2022. The pricing and ownership of U.S. green bonds. Annual Review of Financial Economics 14, 415–437.

Baker, M. P., Egan, M., Sarkar, S. K., 2024. Demand for ESG. Working paper.

Basak, S., Pavlova, A., Shapiro, A., 2007. Optimal asset allocation and risk shifting in money management. Review of Financial Studies 20, 1583–1621.

Basak, S., Pavlova, A., Shapiro, A., 2008. Offsetting the implicit incentives: Benefits of benchmarking in money management. Journal of Banking and Finance 32, 1882–1993.

Ben-David, I., Li, J., Rossi, A., Song, Y., 2022a. Ratings-driven demand and systematic price fluctuations. Review of Financial Studies 35, 2790–2838.

Ben-David, I., Li, J., Rossi, A., Song, Y., 2022b. What do mutual fund investors really care about? Review of Financial Studies 35, 1723–1774.

Berk, J. B., van Binsbergen, J. H., 2015. Measuring skill in the mutual fund industry. Journal of Financial Economics 118, 1–20.

Bizjak, J., Lemmon, M., Nguyen, T., 2011. Are all ceos above average? an empirical analysis of compensation peer groups and pay design. Journal of Financial Economics 100, 538–555.

Bizjak, J. M., Lemmon, M. L., Naveen, L., 2008. Does the use of peer groups contribute to higher pay and less efficient compensation? Journal of Financial Economics 90, 152–168.

Brown, K. C., Harlow, W. V., Starks, L. T., 1996. Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry. Journal of Finance 51, 85–110.

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

Cadman, B., Carter, M. E., 2014. Compensation peer groups and their relation with ceo pay. Journal of Management Accounting Research 26, 57–82.

Cen, X., Dou, W. W., Kogan, L., Wu, W., 2023. Fund flows and income risk of fund managers. Working paper.

Chen, H., Cohen, L., Gurun, U. G., 2021. Don't take their word for it: The misclassification of bond mutual funds. Journal of Finance 76, 1699–1730.

Chen, H., Evans, R., Sun, Y., 2025. Self-declared benchmarks and fund manager intent: "Cheating" or competing? Journal of Financial Economics 165, 103975.

Chen, H.-l., Pennacchi, G. G., 2009. Does prior performance affect a mutual fund's choice of risk? Theory and further empirical evidence. Journal of Financial and Quantitative Analysis 44, 745–775.

Chevalier, J., Ellison, G., 1997. Risk taking by mutual funds as a response to incentives. Journal of Political Economy 105, 1167–1200.

Cohen, R. B., Polk, C., Vuolteenaho, T., 2003. The value spread. Journal of Finance 58, 609–642.

Cremers, K. J. M., Fulkerson, J. A., Riley, T. B., 2022. Benchmark discrepancies and mutual fund performance evaluation. Journal of Financial and Quantitative Analysis 57, 543–571.

Cullen, G., Gasbarro, D., Monroe, G. S., Zumwalt, J. K., 2012. Changes to mutual fund risk: Intentional or mean reverting? Journal of Banking and Finance 36, 112–120.

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

Del Guercio, D., Tkac, P. A., 2008. Star power: The effect of morningstar ratings on mutual fund flow. Journal of Financial and Quantitative Analysis 43, 907–936.

Elton, E. J., Gruber, M. J., Blake, C. R., 2011. Holdings data, security returns, and the selection of superior mutual funds. Journal of Financial and Quantitative Analysis 46, 341–367.

Elton, E. J., Gruber, M. J., Blake, C. R., 2014. The performance of separate accounts and collective investment trusts. Review of Finance 18, 1717–1742.

Eskildsen, M., Ibert, M., Jensen, T., Pedersen, L., 2024. In search of the true greenium. SSRN Electronic Journal .

Evans, R., Gómez, J.-P., Ma, L., Tang, Y., 2024. Peer versus pure benchmarks in the compensation of mutual fund managers. Journal of Financial and Quantitative Analysis 59, 3101–3138.

Evans, R., Sun, Y., 2021. Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. Review of Financial Studies 34, 67–107.

Evans, R. B., 2010. Mutual fund incubation. Journal of Finance 65, 1581–1611.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Faulkender, M., Yang, J., 2010. Inside the black box: The role and composition of compensation peer groups. Journal of Financial Economics 96, 257–270.

Fisch, J. E., Robertson, A. Z., 2022. What's in a name? ESG mutual funds and the SEC's names rule. S. Cal. L. Rev. 96, 1417.

Gantchev, N., Giannetti, M., Li, R., 2024. Sustainability or performance? Ratings and fund managers' incentives. Journal of Financial Economics 155, 103831.

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

Han, X., Roussanov, N., Ruan, H., 2024. Mutual fund risk shifting and risk anomalies. Working paper.

Hartzmark, S. M., Sussman, A. B., 2019. Do investors value sustainability? A natural experiment examining ranking and fund flows. Journal of Finance 74, 2789–2837.

Hoepner, A. G. F., Oikonomou, I., Sautner, Z., Starks, L. T., Zhou, X. Y., 2024. ESG shareholder engagement and downside risk. Review of Finance 28, 483–510.

Holmstrom, B., 1979. Moral hazard and observability. Bell Journal of Economics 10, 74–91.

Holmstrom, B., 1982. Moral hazard in teams. Bell Journal of Economics 13, 324–340.

Huang, J., Sialm, C., Zhang, H., 2011. Risk shifting and mutual fund performance. Review of Financial Studies 24, 2575–2616.

Ibert, M., Kaniel, R., Van Nieuwerburgh, S., Vestman, R., 2018. Are mutual fund managers paid for investment skill? Review of Financial Studies 31, 715–772.

Ilhan, E., Sautner, Z., Vilkov, G., 2021. Carbon tail risk. Review of Financial Studies 34, 1540–1571.

Immel, M., Hachenberg, B., Kiesel, F., Schiereck, D., 2021. Green bonds: Shades of green and brown. Journal of Asset Management 22, 96–109.

Kempf, A., Ruenzi, S., Thiele, T., 2009. Employment risk, compensation incentives, and managerial risk taking: Evidence from the mutual fund industry. Journal of Financial Economics 92, 92–108.

Kim, M. S., 2019. Changes in mutual fund flows and managerial incentives. Working paper.

Kim, S. H., 2022. Do mutual funds manipulate star ratings? Evidence from portfolio pumping. Working paper.

Lazear, E. P., Rosen, S., 1981. Rank-order tournaments as optimum labor contracts. Journal of Political Economy 89, 841–864.

Lee, J. H., Trzcinka, C., Venkatesan, S., 2019. Do portfolio manager contracts contract portfolio management? Journal of Finance 74, 2543–2577.

Lins, K. V., Servaes, H., Tamayo, A., 2017. Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. Journal of Finance 72, 1785–1824.

Ma, L., Tang, Y., 2019. Portfolio manager ownership and mutual fund risk taking. Management Science 65, 5449–5956.

Ma, L., Tang, Y., Gomez, J.-P., 2019. Portfolio manager compensation in the U.S. mutual fund industry. Journal of Finance 74, 587–638.

Moalla, W., Dammak, T., 2024. ESG combined score effects on stock performance of S&P 500-listed firms. Finance Research Letters 66, 105686.

Mullally, K., Rossi, A., 2025. Moving the goalposts? Mutual fund benchmark changes and relative performance manipulation. Review of Financial Studies 38, 1067–1119.

Pástor, L., Stambaugh, R. F., Taylor, L. A., 2015. Scale and skill in active management. Journal of Financial Economics 116, 23–45.

Pástor, L., Stambaugh, R. F., Taylor, L. A., 2022. Dissecting green returns. Journal of Financial Economics 146, 403–424.

Reuter, J., Zitzewitz, E., 2021. How much does size erode mutual fund performance? a regression discontinuity approach. Review of Finance 25, 1395–1432.

Rzeźnik, A., Hanley, K. W., Pelizzon, L., 2021. The salience of ESG ratings for stock pricing: Evidence from (potentially) confused investors. Working paper.

Schwarz, C. G., 2012. Mutual fund tournaments: The sorting bias and new evidence. Review of Financial Studies 25, 913–936.

Sensoy, B. A., 2009. Performance evaluation and self-designated benchmark indexes in the mutual fund industry. Journal of Financial Economics 92, 25–39.

Spiegel, M., Zhang, H., 2013. Mutual fund risk and market share-adjusted fund flows. Journal of Financial Economics 108, 506–528.

Stotz, O., 2021. Expected and realized returns on stocks with high- and low-ESG exposure. Journal of Asset Management 22, 133–150.

Zerbib, O. D., 2019. The effect of pro-environmental preferences on bond prices: Evidence from green bonds. Journal of Banking & Finance 98, 39–60.

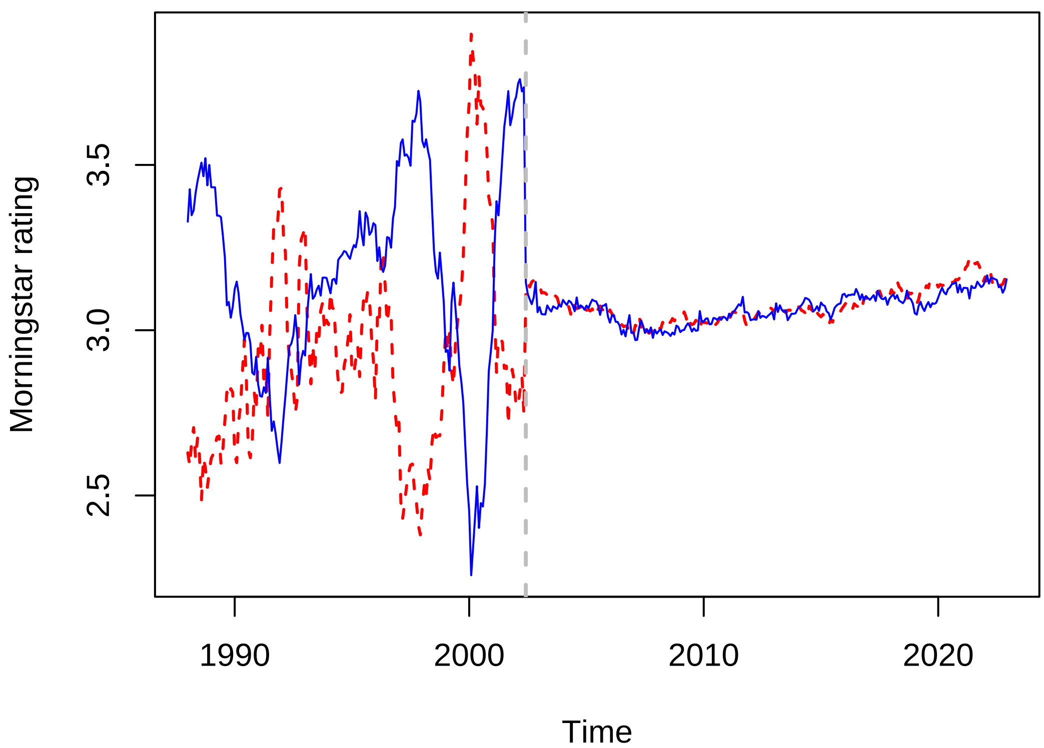

Figure 1: The figure plots the average Morningstar star rating of Value funds (blue solid line) and Growth funds (red dashed line), where a fund is classified as Value if it belongs to one of the Morningstar value categories (Large Value, Mid-Cap Value, or Small Value) and as Growth if it belongs to one of the Morningstar growth categories (Large Growth, Mid-Cap Growth, or Small Growth).

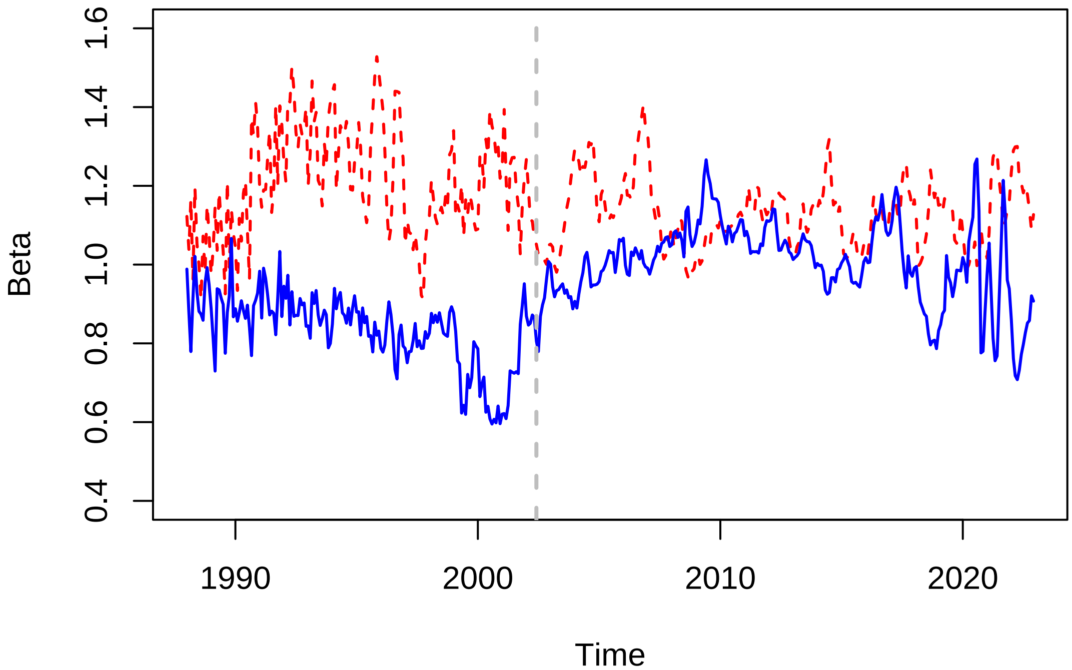

Figure 2: The figure plots the average beta of value (blue solid line) and growth (red dashed line) funds. A fund is classified as value if it belongs to one of the Morningstar value categories (Large Value, Mid-Cap Value, or Small Value), and as growth if it belongs to one of the Morningstar growth categories (Large Growth, Mid-Cap Growth, or Small Growth).

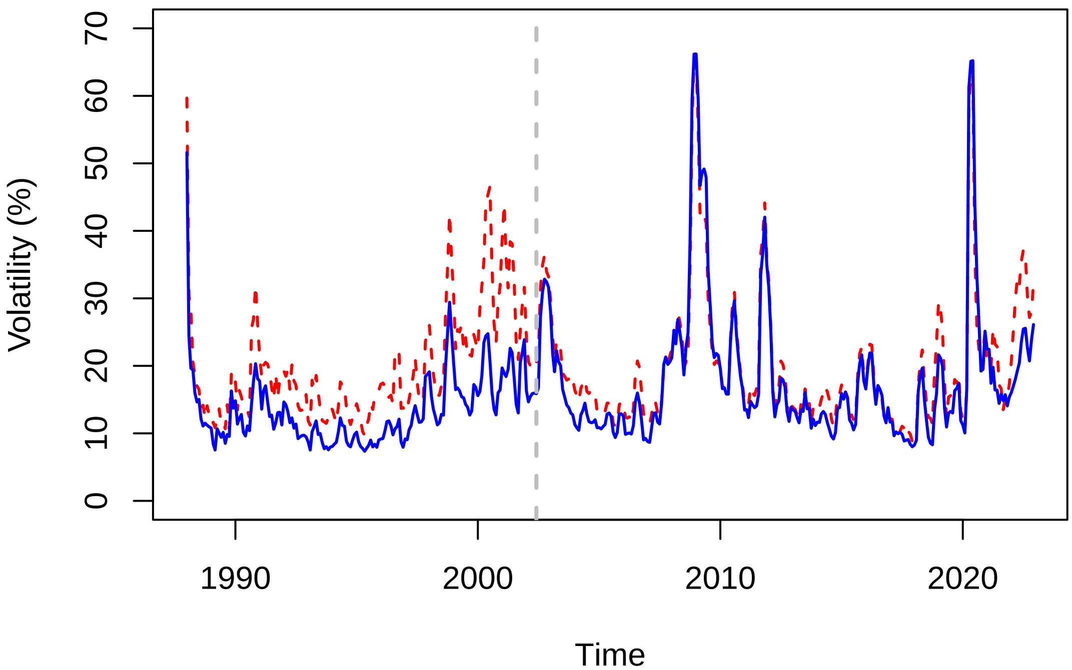

Figure 3: The figure plots the average volatility of value (blue solid line) and growth (red dashed line) funds. A fund is classified as value if it belongs to one of the Morningstar value categories (Large Value, Mid-Cap Value, or Small Value), and as growth if it belongs to one of the Morningstar growth categories (Large Growth, Mid-Cap Growth, or Small Growth).

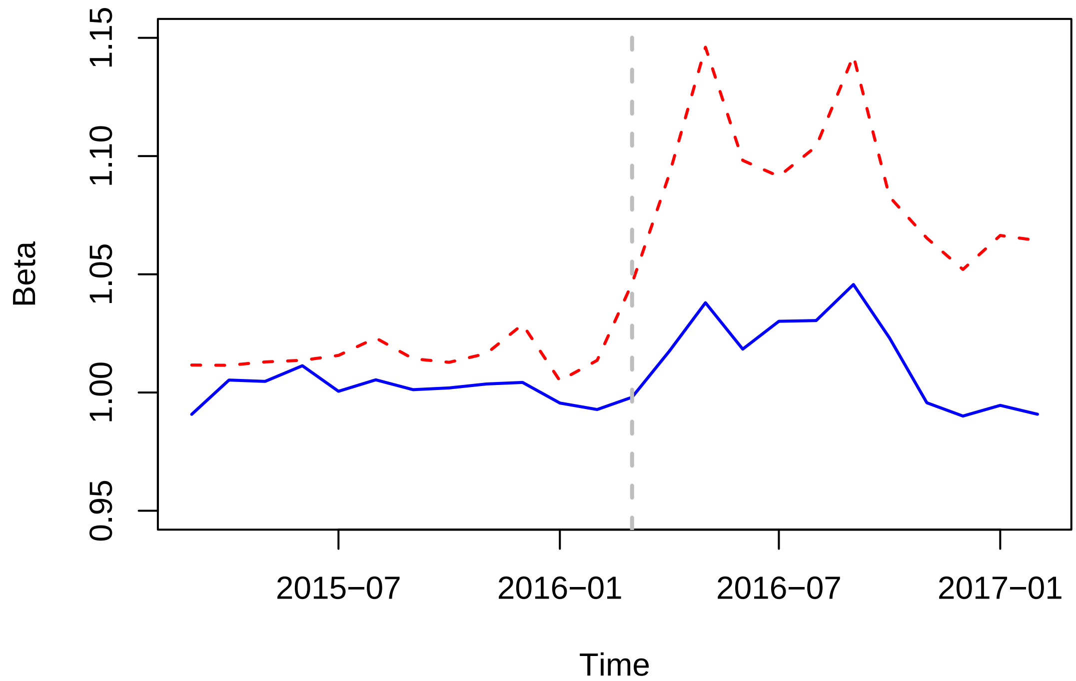

Figure 4: The figure plots the average beta of high (blue solid line) and low (red dashed line) globe-rated funds. A fund is classified as high if its Morningstar ESG score—used to determine its Morningstar globe rating at the introduction of the ratings in March 2016—was above the median, and as low otherwise.

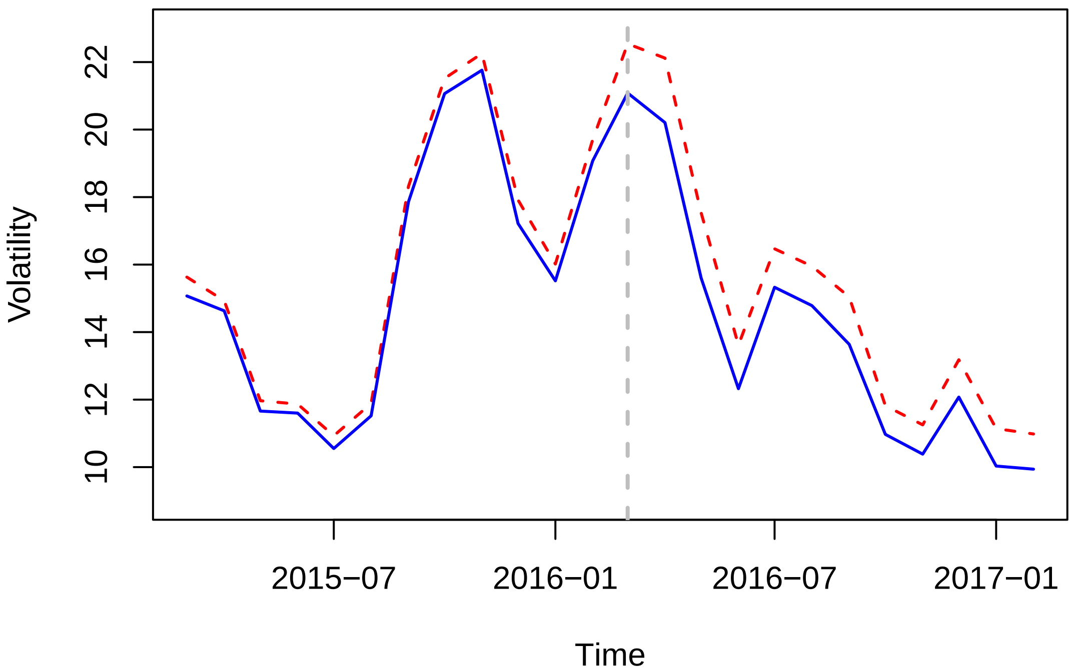

Figure 5: The figure plots the average volatility of high (blue solid line) and low (red dashed line) globe-rated funds. A fund is classified as high if its Morningstar ESG score—used to determine its Morningstar globe rating at the introduction of the ratings in March 2016—was above the median, and as low otherwise.

Table 1: Summary Statistics

Panel A of this table reports the summary statistics for the key variables used in our analyses of value versus growth Funds. All variables are indexed by fund i and time t (in months). $Stars_{i,t}$ denotes fund i's Morningstar star rating at time t. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value). $HML_{t-35,t}$ is the 36-month cumulative return on the value factor from the Fama–French three-factor model (Fama and French, 1993). $Beta_{i,t}$ denotes fund i's expected market beta at time t, estimated from stock holdings at the end of $t-1$. $Volatility_{i,t}$ denotes fund i's expected total volatility at time t, also estimated from stock holdings at the end of $t-1$. The measure is annualized and expressed as a percentage. $\log(B/M)^{\text{Hi–Lo}}_{t-1}$ is the value spread, defined as the log difference between the book-to-market ratio of the value portfolio (top decile) and that of the growth portfolio (bottom decile) at the end of the previous June. $\mathbb{1}(Mandate_{i,t-1})$ is an indicator variable equal to one if fund i has a value or growth mandate at time $t-1$. A fund is classified as having such a mandate if its name or its prospectus benchmark name contains a value or growth designation. $\mathbb{1}(Close_{i,t})$ is an indicator variable equal to one if fund i is expected to be close to a rating threshold at time t. An extreme return from 36 months earlier is used as an instrument to identify such funds. *% Outside Style Box*$_{i,t}$ is the percentage of fund i's stock holdings whose Morningstar equity style assignments (based on value–growth scores) fall outside the fund's designated value/growth style box at time t. *Value/Growth Dispersion*$_{i,t}$ is the weighted variance of the value–growth scores of fund i's stock holdings at time t. Our sample covers the period from 1988 to 2022.

Panel A: Value vs. Growth Funds

Statistic	N	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
$\mathbb{1}$(Growth)	329,500	0.60	0.49	0	1	1
Stars	329,500	3.08	1.04	2.01	3.00	4.00
$\text{HML}_{t-35,t}$	329,500	0.03	0.23	−0.10	−0.003	0.17
Beta	188,162	1.07	0.21	0.95	1.06	1.19
Volatility (%)	188,162	19.67	10.79	12.81	16.66	22.57
$\log(\text{B/M})^{\text{Hi–Lo}}$	188,162	2.64	0.44	2.35	2.65	2.92
$\mathbb{1}$(Close)	184,660	0.20	0.40	0	0	0
$\mathbb{1}$(Mandate)	179,896	0.80	0.40	1	1	1
% Outside Style Box	171,037	45.12	14.93	34.09	45.47	56.25
Value/Growth Dispersion	171,037	111.73	17.74	99.85	112.36	123.86

Table 1–*Continued*

Panel B of this table reports summary statistics for the variables used in our analysis of ESG versus non-ESG funds. All variables are indexed by fund i and time t (in months). $Flow_{i,t}$ denotes fund i's net capital flow at time t, expressed as a percentage of its lagged total net assets (TNA). $ESG_i^{Globes\ Mar2016}$ denotes fund i's Morningstar ESG score used to determine its initial globe rating in March 2016. The measure is standardized to facilitate the interpretation of results. $R_{t-35,t}^{ESG}$ denotes the 36-month cumulative return on an ESG portfolio long the top and short the bottom tercile of stocks sorted by Sustainalytics ESG scores. R_t^{ESG} denotes the monthly return an ESG portfolio long the top and short the bottom tercile of stocks sorted by Sustainalytics ESG scores. $\mathbb{1}$(No ESG Label$_i$) is an indicator variable equal to one if fund i did not have an ESG label at the time when the globe ratings were introduced, and zero otherwise. A fund is classified as having an ESG label if its name or the description of its principal investment strategy in the prospectus contains ESG-related keywords. $ESG_{i,t}^{Sustainalytics}$ is the value-weighted average of Sustainalytics ESG scores of fund i's stock holdings at time t. The remaining variables are defined as in Panel A of this table. Our sample spans the period from March 2015 to February 2017.

Panel B: ESG vs. Non-ESG Funds

Statistic	N	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
$\mathrm{ESG}^{\mathrm{Globe\ Mar2016}}$	27,429	0.00	1.00	−0.62	0.03	0.57
Stars	27,429	3.04	1.02	2	3	4
$\mathrm{R}_{t-35,t}^{\mathrm{ESG}}$	27,429	−0.08	0.04	−0.11	−0.07	−0.04
Flow (%)	25,536	−0.75	3.63	−1.62	−0.68	0.14
Beta	20,236	1.03	0.11	0.97	1.02	1.08
Volatility (%)	20,236	15.29	4.28	11.63	14.73	18.42
$\mathrm{R}_t^{\mathrm{ESG}}$	20,236	−0.002	0.014	−0.012	−0.007	0.008
$\mathbb{1}$(No ESG Label)	19,596	0.95	0.21	1	1	1
$\mathrm{ESG}^{\mathrm{Sustainalytics}}$	17,014	59.93	3.49	57.43	60.55	62.71

Table 2: Gap in Star Ratings: Value vs. Growth Funds

This table reports the average Morningstar star ratings of growth and value funds, separately for the period before the change in the Morningstar peer groups used for the star ratings (January 1988 to May 2002) and the period after the change (June 2002 to December 2022). Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

Stars					
January 1988 – May 2002			June 2002 – December 2022		
Growth	Value	Growth–Value	Growth	Value	Growth–Value
(1)	(2)	(3)	(4)	(5)	(6)
2.98	3.18	-0.20^{***}	3.07	3.07	0.003
		(-3.74)			(0.11)

Table 3: Time-Varying Gap in Star Ratings: Value vs. Growth Funds

This table reports the results of the following linear regression model:

$$Stars_{i,t} = \delta \mathbb{1}\left(\text{Growth}_i\right) \times HML_{t-35,t} + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds, and t indexes time (in months). $Stars_{i,t}$ denotes fund i's Morningstar star rating at time t. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value). $HML_{t-35,t}$ is the 36-month cumulative return on the value factor from the Fama–French three-factor model (Fama and French, 1993). $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and θ_t denote fund and time fixed effects, respectively. The regressions are estimated separately for the period before the change in the Morningstar peer groups used for the star ratings (January 1988 to May 2002) and the period after the change (June 2002 to December 2022). Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Stars			
	January 1988 – May 2002		June 2002 – December 2022	
	(1)	(2)	(3)	(4)
$\mathbb{1}$(Growth) × $\text{HML}_{t-35,t}$	-3.13^{***}	-2.90^{***}	-0.001	0.09
	(−25.49)	(−25.52)	(−0.01)	(1.02)
log(Family TNA)		−0.02		0.01
		(−0.98)		(0.37)
log(Fund TNA)		0.32^{***}		0.24^{***}
		(10.96)		(14.47)
Expense ratio		−0.09		-0.45^{***}
		(−1.39)		(−5.22)
Turnover ratio		−0.05		-0.08^{***}
		(−1.40)		(−3.50)
Fund fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	69,709	69,709	259,791	259,791
Adjusted R^2	0.50	0.53	0.45	0.49

Table 4: Risk-Taking: Value vs. Growth Funds

This table reports the results of the following linear regression model:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta \mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value). $\mathbb{1}(Pre_t^{Jun2002})$ is an indicator variable equal to one if time t falls before the change in the Morningstar peer groups used for the star ratings in June 2002, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and θ_t denote fund and time fixed effects, respectively. The sample covers the period from 1988 to 2022. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Beta		Volatility	
	(1)	(2)	(3)	(4)
$\mathbb{1}$(Growth) × $\mathbb{1}$(Pre$^{\text{Jun2002}}$)	0.29***	0.28***	6.91***	6.75***
	(10.76)	(10.62)	(9.34)	(9.29)
log(Family TNA)		0.01***		0.08*
		(3.27)		(1.94)
log(Fund TNA)		0.01***		0.24***
		(4.96)		(4.83)
Expense ratio		0.02*		0.21
		(1.68)		(1.06)
Turnover ratio		0.01***		0.32***
		(3.22)		(3.61)
Fund fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	188,162	188,162	188,162	188,162
Adjusted R^2	0.52	0.52	0.92	0.92

Table 5: Time-Varying Risk-Taking: Value vs. Growth Funds

This table reports the results of the following linear regression model:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta \mathbb{1}(Growth_i) \times \mathbb{E}_{t-1}[HML_t] + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value). $\mathbb{E}t-1[HMLt]$ denotes the expected value premium at time t. The value spread, $\log(B/M)_{t-1}^{\text{Hi–Lo}}$, is used as a proxy for this expectation. Specifically, $\log(B/M)_{t-1}^{\text{Hi–Lo}}$ is defined as the log difference between the book-to-market ratio of the value portfolio (top decile) and that of the growth portfolio (bottom decile) at the end of the previous June. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and θ_t denote fund and time fixed effects, respectively. The regressions are estimated separately for the period before the change in the Morningstar peer groups used for the star ratings (January 1988 to May 2002) and the period after the change (June 2002 to December 2022). Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Beta		Volatility	
	Pre	Post	Pre	Post
	(1)	(2)	(3)	(4)
$\mathbb{1}(\text{Growth}) \times \mathbb{E}_{t-1}[\text{HML}_t]$	0.24***	0.03	8.72***	0.06
	(6.65)	(1.08)	(6.65)	(0.10)
log(Family TNA)	−0.002	0.004**	−0.11	0.04
	(−0.51)	(2.48)	(−1.58)	(1.16)
log(Fund TNA)	0.02**	0.01***	0.40**	0.12***
	(2.30)	(3.07)	(2.23)	(2.80)
Expense ratio	−0.01	0.02**	−0.13	0.20
	(−0.59)	(2.40)	(−0.41)	(0.97)
Turnover ratio	0.0002	0.01***	−0.03	0.29***
	(0.02)	(3.06)	(−0.15)	(3.43)
Fund fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	23,244	164,918	23,244	164,918
Adjusted R^2	0.72	0.51	0.84	0.94

Table 6: Value/Growth Style Drift

This table reports the results of the following linear regression model:

$$Drift_{i,t} = \delta \mathbb{1}(Growth_i) \times \mathbb{1}(Post_t^{Jun2002}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $Drift_{i,t}$ denotes fund i's *% Outside Style Box* or *Value/Growth Dispersion* in time t. *% Outside Style Box*$_{i,t}$ is the percentage of fund i's stock holdings whose Morningstar equity style assignments (based on value–growth scores) fall outside the fund's designated value/growth style box at time t. *Value/Growth Dispersion*$_{i,t}$ is the weighted variance of the value–growth scores of fund i's stock holdings at time t. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value). $\mathbb{1}(Post_t^{Jun2002})$ is an indicator variable equal to one if time t falls on or after the change in the Morningstar peer groups used for the star ratings in June 2002, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and θ_t denote fund and time fixed effects, respectively. The sample covers the period from 1988 to 2022. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	% Outside Style Box		Value/Growth Dispersion	
	(1)	(2)	(3)	(4)
$\mathbb{1}$(Growth) × $\mathbb{1}$(Post$^{\text{Jun2002}}$)	-4.68^{***}	-4.47^{***}	-3.05^{**}	-2.69^{**}
	(−4.63)	(−4.41)	(−2.46)	(−2.19)
log(Family TNA)		−0.12		−0.02
		(−0.74)		(−0.09)
log(Fund TNA)		0.30^{*}		0.46^{*}
		(1.85)		(1.92)
Expense ratio		0.73		2.98^{***}
		(0.82)		(2.74)
Turnover ratio		0.59^{**}		1.25^{***}
		(2.04)		(3.42)
Fund fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	171,037	171,037	171,037	171,037
Adjusted R^2	0.71	0.71	0.77	0.77

Table 7: Value/Growth Style Mandate and Risk-Taking

Panel A of this table reports the results of the following linear regression model:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta \mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value). $\mathbb{1}(Pre_t^{Jun2002})$ is an indicator variable equal to one if time t falls before the change in the Morningstar peer groups used for the star ratings in June 2002, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and θ_t denote fund and time fixed effects, respectively. The regressions are estimated separately for sub-samples of funds split by an indicator variable, $\mathbb{1}(Mandate_{i,t-1})$, which equals one if fund i has a value or growth mandate at time $t-1$. A fund is classified as having such a mandate if its name or its prospectus benchmark name contains a value or growth designation. Our sample covers the period from 1988 to 2022. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

Panel A

	Beta		Volatility	
Style Mandate =	Yes	No	Yes	No
	(1)	(2)	(3)	(4)
$\mathbb{1}(\text{Growth}) \times \mathbb{1}(\text{Pre}^{\text{Jun2002}})$	0.31***	0.17***	7.31***	4.65***
	(10.96)	(4.47)	(9.38)	(5.61)
log(Family TNA)	0.004*	0.01**	0.04	0.16*
	(1.85)	(2.07)	(0.90)	(1.71)
log(Fund TNA)	0.01***	0.01**	0.25***	0.09
	(4.46)	(2.02)	(4.60)	(0.98)
Expense ratio	1.36	−1.12	21.21	−31.84
	(1.24)	(−0.55)	(0.92)	(−1.03)
Turnover ratio	0.01**	0.01*	0.29***	0.51**
	(2.52)	(1.95)	(2.98)	(2.53)
Fund fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	143,240	36,656	143,240	36,656
Adjusted R^2	0.51	0.58	0.92	0.93

Table 7–*Continued*

Panel B of this table reports the results of the following linear regression model:

$$\begin{aligned}\mathbb{E}_{t-1}[Risk_{i,t}] = &\rho \mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) \times \mathbb{1}(Mandate_{i,t-1}) + \delta_1 \mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) \\ &+ \delta_2 \mathbb{1}(Growth_i) \times \mathbb{1}(Mandate_{i,t-1}) + \delta_3 \mathbb{1}(Pre_t^{Jun2002}) \times \mathbb{1}(Mandate_{i,t-1}) \\ &+ \beta \mathbb{1}(Mandate_{i,t-1}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t \ (\theta_{c,t-1}) + \varepsilon_{i,t}\end{aligned}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value). $\mathbb{1}(Pre_t^{Jun2002})$ is an indicator variable equal to one if time t falls before the change in the Morningstar peer groups used for the star ratings in June 2002, and zero otherwise. $\mathbb{1}(Mandate_{i,t-1})$ is an indicator variable equal to one if fund i has a value or growth mandate at time $t-1$. A fund is classified as having such a mandate if its name or its prospectus benchmark name contains a value or growth designation. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and θ_t denote fund and time fixed effects, respectively. In some specifications, time fixed effects are replaced with lagged category-by-time fixed effects ($\theta_{c,t-1}$). Our sample covers the period from 1988 to 2022. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

Panel B

	Beta		Volatility	
	(1)	(2)	(3)	(4)
$\mathbb{1}$(Growth) × $\mathbb{1}$(Pre$^{\text{Jun2002}}$) × $\mathbb{1}$(Mandate)	0.12***	0.09***	2.65***	2.25***
	(3.31)	(3.86)	(3.82)	(3.69)
$\mathbb{1}$(Growth) × $\mathbb{1}$(Pre$^{\text{Jun2002}}$)	0.19***		4.69***	
	(5.23)		(6.10)	
$\mathbb{1}$(Growth) × $\mathbb{1}$(Mandate)	0.05**	0.02	0.86**	0.21
	(2.45)	(1.13)	(2.28)	(0.69)
$\mathbb{1}$(Pre$^{\text{Jun2002}}$) × $\mathbb{1}$(Mandate)	−0.03	−0.01	−0.65**	−0.16
	(−1.43)	(−1.16)	(−1.97)	(−0.80)
$\mathbb{1}$(Mandate)	−0.01	0.01	−0.26	0.15
	(−0.85)	(1.02)	(−1.09)	(0.95)
Fund characteristics	Yes	Yes	Yes	Yes
Fund fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes		Yes	
Category-by-time FEs		Yes		Yes
Observations	179,896	179,896	179,896	179,896
Adjusted R^2	0.52	0.76	0.92	0.96

Table 8: Distance to Rating Thresholds and Risk-Taking

Panel A of this table reports the results of the following linear regression model:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta \mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value). $\mathbb{1}(Pre_t^{Jun2002})$ is an indicator variable equal to one if time t falls before the change in the Morningstar peer groups used for the star ratings in June 2002, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and θ_t denote fund and time fixed effects, respectively. The regressions are estimated separately for sub-samples of funds split by an indicator variable, $\mathbb{1}(Close_{i,t})$, which equals one if fund i is expected to be close to a rating threshold at time t. An extreme return from 36 months earlier is used as an instrument to identify such funds. Our sample covers the period from 1988 to 2022. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

Panel A

	Beta		Volatility	
	Close	Distant	Close	Distant
	(1)	(2)	(3)	(4)
$\mathbb{1}$(Growth) × $\mathbb{1}$(Pre$^{\text{Jun2002}}$)	0.29***	0.27***	7.20***	6.57***
	(10.29)	(10.47)	(9.32)	(9.14)
log(Family TNA)	0.01***	0.01***	0.10*	0.06*
	(3.22)	(2.87)	(1.76)	(1.68)
log(Fund TNA)	0.01***	0.01***	0.23***	0.23***
	(3.85)	(4.85)	(3.91)	(4.67)
Expense ratio	0.02*	0.02*	0.18	0.24
	(1.76)	(1.68)	(0.77)	(1.21)
Turnover ratio	0.02***	0.01***	0.33***	0.32***
	(3.39)	(2.89)	(3.13)	(3.56)
Fund fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	44,938	139,722	44,938	139,722
Adjusted R^2	0.53	0.52	0.92	0.92

Table 8–*Continued*

Panel B of this table reports the results of the following linear regression model:

$$\begin{aligned}\mathbb{E}_{t-1}[Risk_{i,t}] = {} & \rho\mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) \times \mathbb{1}(Close_{i,t}) + \delta_1\mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) \\ & + \delta_2\mathbb{1}(Growth_i) \times \mathbb{1}(Close_{i,t}) + \delta_3\mathbb{1}(Pre_t^{Jun2002}) \times \mathbb{1}(Close_{i,t}) \\ & + \beta\mathbb{1}(Close_{i,t}) + \gamma\Gamma_{i,t-1} + \theta_i + \theta_t\ (\theta_{c,t-1}) + \varepsilon_{i,t}\end{aligned}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value). $\mathbb{1}(Pre_t^{Jun2002})$ is an indicator variable equal to one if time t falls before the change in the Morningstar peer groups used for the star ratings in June 2002, and zero otherwise. $\mathbb{1}(Close_{i,t})$ is an indicator variable equal to one if fund i is expected to be close to a rating threshold at time t. An extreme return from 36 months earlier is used as an instrument to identify such funds. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and θ_t denote fund and time fixed effects, respectively. In some specifications, time fixed effects are replaced with lagged category-by-time fixed effects ($\theta_{c,t-1}$). For details, see Table A3 in the Internet Appendix. Our sample covers the period from 1988 to 2022. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

Panel B

	Beta		Volatility	
	(1)	(2)	(3)	(4)
$\mathbb{1}$(Growth) × $\mathbb{1}$(Pre$^{\text{Jun2002}}$) × $\mathbb{1}$(Close)	0.017*	0.018**	0.511**	0.518**
	(1.71)	(2.13)	(2.13)	(2.49)
$\mathbb{1}$(Growth) × $\mathbb{1}$(Pre$^{\text{Jun2002}}$)	0.27***		6.56***	
	(10.46)		(9.09)	
$\mathbb{1}$(Growth) × $\mathbb{1}$(Close)	−0.002	0.001	−0.08*	−0.02
	(−0.83)	(0.72)	(−1.80)	(−0.68)
$\mathbb{1}$(Pre$^{\text{Jun2002}}$) × $\mathbb{1}$(Close)	−0.01	0.001	−0.05	0.05
	(−1.39)	(0.33)	(−0.27)	(0.63)
$\mathbb{1}$(Close)	0.002	−0.001	0.06*	0.0003
	(0.94)	(−1.22)	(1.72)	(0.01)
Fund characteristics	Yes	Yes	Yes	Yes
Fund fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes		Yes	
Category-by-time FEs		Yes		Yes
Observations	184,660	184,660	184,660	184,660
Adjusted R^2	0.52	0.76	0.92	0.95

Table 9: Closing the Gap in Star Ratings

This table reports the average Morningstar star ratings of growth funds and value funds, separately for the period before the change in the Morningstar peer groups used for the star ratings (January 1988 to May 2002) and the period after the change (June 2002 to December 2022), as well as for sub-samples of funds split by an indicator variable, $\mathbb{1}(Close_{i,t})$, which equals one if fund i is expected to be close to a rating threshold at time t. An extreme return from 36 months earlier is used as an instrument to identify such funds. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Star					
	January 1988 – May 2002			June 2002 – December 2022		
	Growth	Value	Growth–Value	Growth	Value	Growth–Value
	(1)	(2)	(3)	(4)	(5)	(6)
Close	3.03	3.15	-0.12^{*}	3.05	3.05	-0.001
			(-1.77)			(-0.04)
Distant	2.95	3.18	-0.23^{***}	3.08	3.08	0.002
			(-4.19)			(0.06)
Close–Distant			0.11^{***}			-0.003
			(3.10)			(-0.17)

Table 10: Flow Sensitivity to Star Ratings

This table reports the results of the following linear regression model:

$$Flow_{i,t} = \delta Stars_{i,t-1} \times \mathbb{1}(Post_t^{Mar2016}) + \beta Stars_{i,t-1} + \sum_{s=1}^{3} \rho_s Flow_{i,t-s} + \gamma \Gamma_{i,t-1} + \theta_{c,t-1} + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $Flow_{i,t}$ denotes fund i's net capital flow at time t, expressed as a percentage of its lagged total net assets (TNA). $Stars_{i,t-1}$ denotes fund i's Morningstar star rating at time $t-1$. $\mathbb{1}(Post_t^{Mar2016})$ is an indicator variable equal to one if time t falls on or after the introduction of Morningstar globe ratings in March 2016, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. $\theta_{c,t-1}$ denotes lagged category-by-time fixed effects. The regressions are estimated separately for subsamples of funds split by the median value of $ESG^{Globes\ Mar2016}$, the fund's Morningstar ESG score used to determine its initial globe rating in March 2016. Our sample covers the period from March 2015 to February 2017. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Flow			
$ESG^{\text{Globes Mar2016}}$ =	High		Low	
	(1)	(2)	(3)	(4)
Stars × $\mathbb{1}(\text{Post}_t^{\text{Mar2016}})$	-0.14^{*}	-0.15^{**}	0.06	0.06
	(-1.99)	(-2.09)	(0.76)	(0.75)
Stars	0.45^{***}	0.56^{***}	0.44^{***}	0.55^{***}
	(8.72)	(9.33)	(7.50)	(8.51)
Flow_{t-1}	0.26^{***}	0.26^{***}	0.23^{***}	0.22^{***}
	(10.30)	(10.17)	(10.44)	(10.31)
Flow_{t-2}	0.15^{***}	0.15^{***}	0.17^{***}	0.16^{***}
	(10.23)	(9.99)	(11.27)	(11.21)
Flow_{t-3}	0.15^{***}	0.15^{***}	0.11^{***}	0.10^{***}
	(9.79)	(9.36)	(7.84)	(7.78)
log(Family TNA)		0.03		0.03
		(1.06)		(1.33)
log(Fund TNA)		-0.10^{**}		-0.18^{***}
		(-2.47)		(-6.23)
Expense ratio		0.37^{**}		0.14
		(2.78)		(0.71)
Turnover ratio		-0.22^{*}		-0.24^{***}
		(-1.77)		(-3.77)
Category-by-time FEs	Yes	Yes	Yes	Yes
Observations	12,657	12,657	12,879	12,879
Adjusted R^2	0.25	0.25	0.21	0.21

Table 11: Risk-Taking: ESG vs. Non-ESG Funds

This table reports the results of the following linear regression model:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) + \gamma\Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $ESG_i^{Globes\ Mar2016}$ denotes fund i's Morningstar ESG score used to determine its initial globe rating in March 2016. The measure is standardized to facilitate the interpretation of results. $\mathbb{1}(Pre_t^{Mar2016})$ is an indicator variable equal to one if time t falls before the introduction of Morningstar globe ratings in March 2016, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and $\theta_{c,t-1}$ denote fund and lagged category-by-time fixed effects, respectively. Our sample covers the period from March 2015 to February 2017. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Beta		Volatility	
	(1)	(2)	(3)	(4)
$\text{ESG}^{\text{Globes Mar2016}} \times \mathbb{1}(\text{Pre}^{\text{Mar2016}})$	0.017***	0.016***	0.272***	0.265***
	(4.05)	(3.89)	(2.96)	(2.90)
log(Family TNA)		0.0003		−0.030
		(0.07)		(−0.64)
log(Fund TNA)		−0.022**		−0.255*
		(−2.51)		(−1.77)
Expense ratio		0.027		0.513
		(0.89)		(1.25)
Turnover ratio		−0.001		−0.101
		(−0.08)		(−0.89)
Fund fixed effects	Yes	Yes	Yes	Yes
Category-by-time FEs	Yes	Yes	Yes	Yes
Observations	20,236	20,236	20,236	20,236
Adjusted R^2	0.68	0.68	0.94	0.94

Table 12: ESG Labels and Risk-Taking (Placebo Tests)

Panel A of this table reports the results of the following linear regression model:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) + \gamma\Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $ESG_i^{Globes\ Mar2016}$ denotes fund i's Morningstar ESG score used to determine its initial globe rating in March 2016. $\mathbb{1}(Pre_t^{Mar2016})$ is an indicator variable equal to one if time t falls before the introduction of Morningstar globe ratings in March 2016, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and $\theta_{c,t-1}$ denote fund and lagged category-by-time fixed effects, respectively. The regressions are estimated for a subsample of funds with ESG labels. A fund is classified as having an ESG label if its name or the description of its principal investment strategy in the prospectus contains ESG-related keywords. Our sample covers the period from March 2015 to February 2017. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

Panel A

	Beta		Volatility	
	(1)	(2)	(3)	(4)
$\text{ESG}^{\text{Globes Mar2016}} \times \mathbb{1}(\text{Pre}^{\text{Mar2016}})$	−0.016	−0.011	−0.226	−0.167
	(−1.02)	(−0.79)	(−1.02)	(−0.81)
log(Family TNA)		0.024		0.383
		(1.20)		(1.35)
log(Fund TNA)		0.038		0.361
		(1.40)		(1.10)
Expense ratio		0.004		−0.153
		(0.03)		(−0.09)
Turnover ratio		−0.035		−0.523
		(−0.98)		(−0.97)
Fund fixed effects	Yes	Yes	Yes	Yes
Category-by-time FEs	Yes	Yes	Yes	Yes
Observations	908	908	908	908
Adjusted R^2	0.54	0.55	0.95	0.95

Table 12–*Continued*

Panel B of this table reports the results of the following linear regression model:

$$\begin{aligned}\mathbb{E}_{t-1}[Risk_{i,t}] &= \rho ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) \times \mathbb{1}(No\ ESG\ Label_i) \\ &+ \delta_1 ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) + \delta_2 \mathbb{1}(Pre_t^{Mar2016}) \times \mathbb{1}(No\ ESG\ Label_i) \\ &+ \gamma\Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}\end{aligned}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $ESG_i^{Globes\ Mar2016}$ denotes fund i's Morningstar ESG score used to determine its initial globe rating in March 2016. $\mathbb{1}(Pre_t^{Mar2016})$ is an indicator variable equal to one if time t falls before the introduction of Morningstar globe ratings in March 2016, and zero otherwise. $\mathbb{1}(\text{No ESG Label}_i)$ is an indicator variable equal to one if fund i did not have an ESG label at the time when the globe ratings were introduced, and zero otherwise. A fund is classified as having an ESG label if either its name or the description of its principal investment strategy in the prospectus contains ESG-related keywords. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and $\theta_{c,t-1}$ denote fund and lagged category-by-time fixed effects, respectively. Our sample covers the period from March 2015 to February 2017. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

Panel B

	Beta		Volatility	
	(1)	(2)	(3)	(4)
$\text{ESG}^{\text{Globes Mar2016}} \times \mathbb{1}(\text{Pre}^{\text{Mar2016}}) \times \mathbb{1}(\text{No ESG Label})$	0.036**	0.037**	0.514**	0.526**
	(2.16)	(2.18)	(2.24)	(2.27)
$\text{ESG}^{\text{Globes Mar2016}} \times \mathbb{1}(\text{Pre}^{\text{Mar2016}})$	−0.018	−0.020	−0.247	−0.266
	(−1.10)	(−1.17)	(−1.05)	(−1.12)
$\mathbb{1}(\text{Pre}^{\text{Mar2016}}) \times \mathbb{1}(\text{No ESG Label})$	−0.018	−0.017	−0.184	−0.184
	(−1.19)	(−1.15)	(−1.05)	(−1.05)
log(Family TNA)		0.001		−0.020
		(0.23)		(−0.41)
log(Fund TNA)		−0.023**		−0.271*
		(−2.53)		(−1.85)
Expense ratio		0.026		0.508
		(0.84)		(1.22)
Turnover ratio		−0.002		−0.117
		(−0.25)		(−1.01)
Fund fixed effects	Yes	Yes	Yes	Yes
Category-by-time FEs	Yes	Yes	Yes	Yes
Observations	19,596	19,596	19,596	19,596
Adjusted R^2	0.68	0.68	0.94	0.94

Table 13: Time-Varying Gap in Star Ratings: ESG vs. Non-ESG Funds

This table reports the results of the following linear regression model:

$$Stars_{i,t} = \delta ESG_i^{Globes\ Mar2016} \times R_{t-35,t}^{ESG} + \gamma \Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}$$

where i indexes mutual funds, and t indexes time (in months). $Stars_{i,t}$ denotes fund i's Morningstar star rating at time t. $ESG_i^{Globes\ Mar2016}$ denotes fund i's Morningstar ESG score used to determine its initial globe rating in March 2016. The measure is standardized to facilitate the interpretation of results. $R_{t-35,t}^{ESG}$ denotes the 36-month cumulative return on an ESG portfolio long the top and short the bottom tercile of stocks sorted by Sustainalytics ESG scores. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and $\theta_{c,t-1}$ denote fund and lagged category-by-time fixed effects, respectively. Our sample covers the period from March 2015 to February 2017. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Stars	
	(1)	(2)
$\mathrm{ESG}^{\mathrm{Globes\ Mar2016}} \times \mathrm{R}_{t-35,t}^{\mathrm{ESG}}$	1.14***	0.94***
	(15.64)	(13.04)
log(Family TNA)		0.02
		(1.63)
log(Fund TNA)		0.46***
		(30.68)
Expense ratio		0.27***
		(4.03)
Turnover ratio		−0.01
		(−0.77)
Fund fixed effects	Yes	Yes
Category-by-month FEs	Yes	Yes
Observations	27,429	27,429
Adjusted R^2	0.81	0.81

Table 14: Time-Varying Risk-Taking: ESG vs. Non-ESG Funds

This table reports the results of the following linear regression model:

$$\begin{aligned}\mathbb{E}_{t-1}[Risk_{i,t}] &= \rho ESG_i^{Globes\ Mar2016} \times \mathbb{E}_{t-1}[R_t^{ESG}] \times \mathbb{1}(Pre_t^{Mar2016}) \\ &+ \delta_1 ESG_i^{Globes\ Mar2016} \times \mathbb{E}_{t-1}[R_t^{ESG}] + \delta_2 ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) \\ &+ \gamma\Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}\end{aligned}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $ESG_i^{Globes\ Mar2016}$ denotes fund i's Morningstar ESG score used to determine its initial globe rating in March 2016. $\mathbb{E}_{t-1}[R_t^{ESG}]$ denotes the expected return at time t on an ESG portfolio long the top and short the bottom tercile of stocks sorted by Sustainalytics ESG scores. The realized return R_t^{ESG} is used as a proxy for this expectation. $\mathbb{1}(Pre_t^{Mar2016})$ is an indicator variable equal to one if time t falls before the introduction of Morningstar globe ratings in March 2016, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and $\theta_{c,t-1}$ denote fund and lagged category-by-time fixed effects, respectively. Our sample covers the period from March 2015 to February 2017. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Beta		Volatility	
	(1)	(2)	(3)	(4)
$\mathrm{ESG}^{\mathrm{Globes\ Mar2016}} \times \mathbb{E}_{t-1}[\mathrm{R}_t^{\mathrm{ESG}}] \times \mathbb{1}(\mathrm{Pre}^{\mathrm{Mar2016}})$	-0.28^{*}	-0.28^{*}	-7.73^{*}	-7.73^{*}
	(-1.95)	(-1.90)	(-1.90)	(-1.87)
$\mathrm{ESG}^{\mathrm{Globes\ Mar2016}} \times \mathbb{E}_{t-1}[\mathrm{R}_t^{\mathrm{ESG}}]$	0.16	0.16	5.19	5.21
	(1.25)	(1.23)	(1.34)	(1.33)
$\mathrm{ESG}^{\mathrm{Globes\ Mar2016}} \times \mathbb{1}(\mathrm{Pre}^{\mathrm{Mar2016}})$	0.02^{***}	0.02^{***}	0.25^{***}	0.25^{***}
	(4.08)	(3.91)	(3.11)	(3.03)
log(Family TNA)		0.0003		-0.03
		(0.09)		(-0.60)
log(Fund TNA)		-0.02^{**}		-0.26^{*}
		(-2.51)		(-1.78)
Expense ratio		0.03		0.50
		(0.87)		(1.22)
Turnover ratio		-0.001		-0.10
		(-0.09)		(-0.90)
Fund fixed effects	Yes	Yes	Yes	Yes
Category-by-month FEs	Yes	Yes	Yes	Yes
Observations	20,236	20,236	20,236	20,236
Adjusted R^2	0.68	0.68	0.94	0.94

Table 15: ESG Tilts

This table reports the results of the following linear regression model:

$$ESG_{i,t}^{Sustainalytics} = \delta ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Post_t^{Mar2016}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $ESG_{i,t}^{Sustainalytics}$ is the value-weighted average of Sustainalytics ESG scores of fund i's stock holdings at time t. $ESG_i^{Globes\ Mar2016}$ denotes fund i's Morningstar ESG score used to determine its initial globe rating in March 2016. $\mathbb{1}(Post_t^{Mar2016})$ is an indicator variable that takes a value of one if time t is on or after the introduction of Morningstar globe ratings in March 2016, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and $\theta_{c,t-1}$ denote fund and lagged category-by-time fixed effects, respectively. Our sample covers the period from March 2015 to February 2017. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	$ESG^{Sustainalytics}$	
	(1)	(2)
$ESG^{Globes\ Mar2016} \times \mathbb{1}(Post^{Mar2016})$	0.195**	0.197**
	(2.27)	(2.29)
log(Family TNA)		0.008
		(0.20)
log(Fund TNA)		−0.041
		(−0.74)
Expense ratio		−0.113
		(−0.29)
Turnover ratio		0.032
		(0.37)
Fund fixed effects	Yes	Yes
Category-by-time FEs	Yes	Yes
Observations	17,014	17,014
Adjusted R^2	0.94	0.94

Internet Appendix for "Tilting at Windmills: Biased Benchmarks and the Risk-Taking Response of Mutual Funds"

Table A1: Risk-Taking: Value vs. Growth Funds (Robustness Checks)

This table reports the results of the following linear regression model:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta \mathbb{1}(Growth_i) \times \mathbb{1}(Pre_t^{Jun2002}) + \gamma \Gamma_{i,t-1} + \theta_i + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $\mathbb{1}(Growth_i)$ is an indicator variable equal to one if fund i belongs to one of the Morningstar growth categories (Large Growth, Mid-cap Growth, or Small Growth), and zero if it belongs to one of the Morningstar value categories (Large Value, Mid-cap Value, or Small Value). $\mathbb{1}(Pre_t^{Jun2002})$ is an indicator variable equal to one if time t falls before the change in the Morningstar peer groups used for the star ratings in June 2002, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and θ_t denote fund and time fixed effects, respectively. The regressions are estimated for the periods from June 2001 to May 2003 and from June 2000 to May 2004. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Jun 2001 – May 2003		Jun 2000 – May 2004	
	Beta	Volatility	Beta	Volatility
	(1)	(2)	(3)	(4)
$\mathbb{1}$(Growth) × $\mathbb{1}$(Pre$^{\text{Jun2002}}$)	0.23***	4.59***	0.28***	7.79***
	(5.03)	(4.14)	(5.79)	(5.91)
log(Family TNA)	0.002	0.03	−0.001	−0.04
	(0.41)	(0.26)	(−0.13)	(−0.32)
log(Fund TNA)	0.10***	2.48***	0.09***	2.56***
	(4.47)	(3.27)	(8.69)	(6.61)
Expense ratio	−0.07*	−2.47**	−0.09**	−2.72**
	(−1.87)	(−2.15)	(−2.64)	(−2.56)
Turnover ratio	−0.02**	−0.99**	−0.002	−0.32
	(−2.19)	(−2.66)	(−0.17)	(−1.09)
Fund fixed effects	Yes	Yes	Yes	Yes
Time fixed effects	Yes	Yes	Yes	Yes
Observations	11,700	11,700	23,394	23,394
Adjusted R^2	0.74	0.86	0.70	0.82

Table A2: Style Drift: Role of Investment Mandates

This table reports the results of the following linear regression model:

$$Drift_{i,t} = \mathbb{1}(Mandate_{i,t-1}) + \gamma\Gamma_{i,t-1} + \theta_t + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $Drift_{i,t}$ denotes fund i's *% Outside Style Box* or *Value/Growth Dispersion* at time t. *% Outside Style Box*$_{i,t}$ is the percentage of fund i's stock holdings whose Morningstar equity style assignments (based on value–growth scores) fall outside the fund's designated value/growth style box at time t. *Value/Growth Dispersion*$_{i,t}$ is the weighted variance of the value–growth scores of fund i's stock holdings at time t. $\mathbb{1}(Mandate_{i,t-1})$ is an indicator variable equal to one if fund i has a value or growth mandate at time $t-1$. A fund is classified as having such a mandate if its name or its prospectus benchmark name contains a value or growth designation. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_t denotes time fixed effects. Our sample covers the period from 1988 to 2022. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	% Outside Style Box		Value/Growth Dispersion	
	(1)	(2)	(3)	(4)
$\mathbb{1}$(Mandate)	-12.63^{***}	-12.43^{***}	-13.80^{***}	-13.87^{***}
	(-23.91)	(-23.59)	(-20.26)	(-21.17)
log(Family TNA)		-0.10		0.67^{***}
		(-0.75)		(4.35)
log(Fund TNA)		-0.53^{***}		-0.45^{**}
		(-2.87)		(-2.10)
Expense ratio		-1.85^{**}		-3.09^{***}
		(-2.49)		(-3.55)
Turnover ratio		-1.55^{***}		2.91^{***}
		(-4.13)		(6.57)
Time fixed effects	Yes	Yes	Yes	Yes
Observations	247,876	247,876	247,876	247,876
Adjusted R^2	0.14	0.15	0.28	0.30

Table A3: Validity of the Instrument: Distance to Rating Thresholds

This table reports the average realized distance to a rating threshold for quintiles sorted by our proxy for expected distance to rating thresholds, as well as the difference between the first and last quintiles. For each category–month, funds with the same rating at time $t-1$ are sorted into deciles based on returns from 36 months earlier. The most extreme deciles (1 and 10) are assigned to the first expected-distance quintile (close), deciles 2 and 9 to the second quintile, and so on. The realized distance to a rating threshold is estimated using percentile rankings of Morningstar risk-adjusted returns, following the Morningstar star ratings methodology as closely as possible. We define $\mathbb{1}(Close_{i,t})$ to equal one if fund i belongs to the first distance quintile in month t, based on its return from month $t-36$ (and thus known at $t-1$). Although our proxy for expected distance to rating thresholds is based on both active and index funds, the validity test sample consists only of index funds, which are unlikely to make active investment decisions to influence star ratings at time t. The sample begins after the change in the star ratings methodology in June 2002—when Morningstar risk-adjusted returns became more straightforward to compute—and ends in 2022.

Distance to a rating threshold					
Q1	Q2	Q3	Q4	Q5	Q5–Q1
6.68	6.90	7.17	7.27	7.24	-0.56^{***}
					(-3.49)

Table A4: Risk-Taking: ESG vs. Non-ESG Funds (Robustness Checks I)

This table reports the results of the following linear regression model:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta ESG_i^{Globes\ Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) + \gamma\Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $ESG_i^{Globes\ Mar2016}$ denotes fund i's Morningstar ESG score used to determine its initial globe rating in March 2016. The measure is standardized to facilitate the interpretation of results. $\mathbb{1}(Pre_t^{Mar2016})$ is an indicator variable equal to one if time t falls before the introduction of Morningstar globe ratings in March 2016, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and $\theta_{c,t-1}$ denote fund and lagged category-by-time fixed effects, respectively. The regressions are estimated for the periods from September 2014 to August 2017 and from March 2014 to February 2018. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Sep2014 – Aug2017		Mar2014 – Feb2018	
	Beta	Volatility	Beta	Volatility
	(1)	(2)	(3)	(4)
ESG score$^{\text{Mar2016}}$ × $\mathbb{1}$(Pre Mar2016)	0.014***	0.200***	0.007**	0.098*
	(3.71)	(2.73)	(2.00)	(1.79)
log(Family TNA)	0.001	−0.026	−0.002	−0.043
	(0.26)	(−0.65)	(−0.70)	(−1.18)
log(Fund TNA)	−0.011*	−0.032	−0.003	0.051
	(−1.93)	(−0.42)	(−0.64)	(0.98)
Expense ratio	0.020	0.468*	0.017	0.436**
	(0.90)	(1.73)	(0.93)	(2.04)
Turnover ratio	0.011	0.079	0.014**	0.118
	(1.60)	(0.90)	(2.12)	(1.49)
Fund fixed effects	Yes	Yes	Yes	Yes
Category-by-time FEs	Yes	Yes	Yes	Yes
Observations	30,133	30,133	39,793	39,793
Adjusted R^2	0.69	0.95	0.71	0.95

Table A5: Risk-Taking: ESG vs. Non-ESG Funds (Robustness Checks II)

This table reports the results of the following linear regression model:

$$\mathbb{E}_{t-1}[Risk_{i,t}] = \delta\, Globes_i^{Mar2016} \times \mathbb{1}(Pre_t^{Mar2016}) + \gamma\Gamma_{i,t-1} + \theta_i + \theta_{c,t-1} + \varepsilon_{i,t}$$

where i indexes mutual funds and t indexes time (in months). $\mathbb{E}_{t-1}[Risk_{i,t}]$ denotes fund i's expected market beta or total volatility at time t, estimated from stock holdings at the end of time $t-1$. Volatility is annualized and reported as a percentage. $Globes_i^{Mar2016}$ denotes fund i's Morningstar globe rating in March 2016. $\mathbb{1}(Pre_t^{Mar2016})$ is an indicator variable equal to one if time t falls before the introduction of Morningstar globe ratings in March 2016, and zero otherwise. $\Gamma_{i,t-1}$ is a vector of lagged fund characteristics: log(Family TNA), log(Fund TNA), expense ratio, and turnover ratio. θ_i and $\theta_{c,t-1}$ denote fund and lagged category-by-time fixed effects, respectively. Our sample covers the period from March 2015 to February 2017. Standard errors are double-clustered by fund and time, and t-statistics are reported in parentheses, with statistical significance at the 10%, 5%, and 1% levels indicated by *, **, and ***, respectively.

	Beta		Volatility	
	(1)	(2)	(3)	(4)
$\text{Globes}^{\text{Mar2016}} \times \mathbb{1}(\text{Pre}^{\text{Mar2016}})$	0.011***	0.010***	0.161**	0.158**
	(3.46)	(3.36)	(2.55)	(2.51)
log(Family TNA)		0.00002		−0.035
		(0.01)		(−0.71)
log(Fund TNA)		−0.023**		−0.277*
		(−2.59)		(−1.88)
Expense ratio		0.025		0.479
		(0.80)		(1.12)
Turnover ratio		−0.002		−0.106
		(−0.35)		(−1.07)
Fund fixed effects	Yes	Yes	Yes	Yes
Category-by-time FEs	Yes	Yes	Yes	Yes
Observations	20,153	20,153	20,153	20,153
Adjusted R^2	0.68	0.68	0.94	0.94