Benchmark Discrepancies and Mutual Fund Performance Evaluation

K.J. Martijn Cremers
mcremers@nd.edu

Mendoza College of Business
University of Notre Dame
Notre Dame, IN 46556

Jon A. Fulkerson
jfulkerson1@udayton.edu

School of Business Administration
University of Dayton
Dayton, OH 45469

Timothy B. Riley*
tbriley@uark.edu

Sam M. Walton College of Business
University of Arkansas
Fayetteville, AR 72701

First Draft: October 2017
This Draft: October 2019

Abstract

We introduce a new holdings-based procedure to identify whether a mutual fund has a benchmark discrepancy, which we define as a benchmark other than the prospectus benchmark best matching a fund's investment strategy. Funds with a benchmark discrepancy tend to be riskier than their prospectus benchmarks indicate. As a result, the funds on average outperform their prospectus benchmarks—before further risk-adjusting—despite underperforming the benchmarks that best match their portfolios. High active share funds outperform to a greater degree if there is no benchmark discrepancy, suggesting that active managers with more skill are less likely to have a benchmark discrepancy.

* Contact author. We thank Chris Clifford, Brigitte Dooley, Brad Jordan, Yuekun Liu, Berk Sensoy, Sara Shirley, Jeff Stark, Russ Wermers, Qing Yan, seminar participants at the University of Dayton, members of the CFA Society Arkansas, and participants at the 2018 Financial Management Association Annual Meeting (especially discussant Andrei Simonov); 2019 Southwest Finance Symposium (especially discussant Yifan Liu); 2019 Eastern Finance Association Annual Meeting (especially discussant Fan Chen); and 2019 Financial Management Association European Conference (especially discussant Javier Estrada) for their helpful comments and suggestions. All errors are our own. Disclosure: Martijn Cremers is currently an independent director at Ariel Investments and a consultant to Touchstone Investments and State Street Associates, and Jon Fulkerson and Timothy Riley have both been previously compensated by Natixis Investment Managers and the Investment Adviser Association.

1. Introduction

The evaluation of the performance of an investment product, such as an actively managed mutual fund, generally involves comparing the performance of that product with a benchmark. The benchmark could be an index that follows the same style as the product's portfolio (e.g., the S&P 500); be based on the portfolio's exposure to different factors (e.g., Carhart, 1997); or be based on the characteristics of the portfolio's holdings (e.g., Daniel, Grinblatt, Titman, and Wermers, 1997). In practice, investors appear to allocate capital by focusing on comparatively simple measures of performance. Sensoy (2009) finds that, even after controlling for performance relative to a mutual fund's factor exposures, investors react strongly to performance relative to a fund's prospectus benchmark index—i.e., the benchmark that a fund declares in its prospectus to meet regulatory requirements.[1]

A focus on benchmark-adjusted performance would be reasonable for any mutual fund with a prospectus benchmark that accurately matches the fund's actual investment strategy. For any fund with an inaccurate match, however, the prospectus benchmark will over- or under-estimate performance. For example, a fund with a portfolio that is riskier than the fund's prospectus benchmark could outperform using prospectus-benchmark-adjusted returns, while underperforming after more appropriate risk-adjustment. Given that investor capital flows respond strongly to the prior performance of mutual funds (Chevalier and Ellison, 1997), the fund in that example could attract additional inflows from investors who fail to fully control for risk. This possibility highlights the importance of evaluating the degree to which there is a discrepancy between the prospectus benchmark and a fund's actual investment strategy.

[1] Several other studies also support the idea that investors place considerable weight on simpler measures of performance: Del Guercio and Tkac (2008); Elton, Gruber, and Blake (2014); Barber, Huang, and Odean (2016); Berk and van Binsbergen (2016); Blocher and Molyboga (2017); and Agarwal, Green, and Ren (2018).

In this study, we introduce a novel holdings-based procedure to assess whether a fund has a benchmark discrepancy. Implementing our procedure using actively managed U.S. equity mutual funds, we find that benchmark discrepancies are common and show that the prospectus benchmarks typically understate risk when a discrepancy exists. Funds with benchmark discrepancies on average outperform their prospectus benchmarks, despite underperforming their more risk-appropriate benchmarks. Investors who are allocating capital, nevertheless, give substantial weight to the performance of these funds relative to their prospectus benchmarks. As we show, investors could significantly improve their capital allocations by better accounting for benchmark discrepancies.

Our procedure for identifying benchmark discrepancies is based on fund holdings. We first assess which of several potential benchmarks has the most overlap with a fund's current holdings. We evaluate overlap using the Cremers and Petajisto (2009) active share measure. The benchmark that generates the lowest active share is considered to have the most overlap. This 'AS benchmark' is different from the prospectus benchmark in 67% of our sample.

The differences between the benchmarks is not always economically meaningful. The AS benchmark and the prospectus benchmark can have similar portfolios. To address this issue, our procedure also requires low overlap between the AS and prospectus benchmark. We measure this overlap with *Benchmark Mismatch*, which we define as the active share of the AS benchmark relative to the prospectus benchmark. When *Benchmark Mismatch* is low, the two benchmarks have holdings that largely overlap. Perhaps one benchmark is just a subset of the other (e.g., the S&P 500 and the S&P 500 Growth). But when *Benchmark Mismatch* is high, the two benchmarks are materially different (e.g., the S&P 600 Growth and the Russell 2000).

Our procedure labels a fund as having a benchmark discrepancy if *Benchmark Mismatch* is greater than 60%. The criterion is the same as the active share criterion applied by Cremers and Petajisto (2009) to distinguish truly active managers from closet indexers. It ensures that the AS and prospectus benchmarks have differences that are economically meaningful.[2] Applied to our sample, we find that in 39% of the cases in which the benchmarks differ, *Benchmark Mismatch* is above 60%. As a result, we label 26% of funds in the average month as having a benchmark discrepancy.[3] The funds labeled as such end 2014 managing a combined $284 billion.

We evaluate the importance of benchmark discrepancies on several dimensions, starting with their effect on performance evaluation. We find that, among funds with benchmark discrepancies, the AS and prospectus benchmarks have substantially different returns. The average return on the prospectus benchmark is lower than that of the AS benchmark by about 1.50% per year. Accordingly, funds with a benchmark discrepancy generally need less return to beat the prospectus benchmark compared to the return needed to beat the AS benchmark.

The impact of the differences between the benchmarks is apparent when evaluating benchmark-adjusted fund performance. Funds with a benchmark discrepancy outperform their prospectus benchmarks on average by 0.66% per year, but underperform their AS benchmarks by 0.84% per year. Likewise, using prospectus-benchmark-adjusted returns, funds with a benchmark discrepancy outperform those without a benchmark discrepancy on average by 1.04% per year. If instead the AS benchmarks are used for funds with benchmark discrepancies, the difference in performance between the two groups is negligible. Taken together, a primary conclusion of our

[2] While 60% is a somewhat ad hoc threshold, the average tracking error is lower with respect to the AS benchmark when *Benchmark Mismatch* is greater than 60% (*t*-stat = 4.92), but not when *Benchmark Mismatch* is less than 60%. In subsequent sections, we consider the distribution of *Benchmark Mismatch* and the effect of various cut-offs.

[3] A fund's benchmark discrepancy status can change quarterly, but statuses do not change often. A fund with a benchmark discrepancy today has an 86% chance to have a benchmark discrepancy in one year. We do not find any evidence that benchmark discrepancies are related to market timing or to Brown, Harlow, and Starks (1996) style tournament behavior.

study is that, among funds with a benchmark discrepancy, the prospectus benchmark significantly overstates *ex post* performance.

The difference in returns between the AS and prospectus benchmarks of funds with benchmark discrepancies is the result of differences in risk. When a benchmark discrepancy exists, the AS benchmark tends to be riskier than the prospectus benchmark, as indicated by the AS benchmark's larger factor exposures. Larger traditional factor exposures (i.e., market, size, value, and momentum) explain about a third of the average difference in returns between the benchmarks. Larger non-traditional factor exposures (e.g., liquidity, profitability, and quality-minus-junk) along with the traditional exposures explain about 87% of the average difference in returns. Among the non-traditional factors, the Fama and French (2015) profitability factor (*RMW*) has the most explanatory power.

Further analysis shows that funds with benchmark discrepancies have prospectus-benchmark-adjusted returns with substantial residual factor exposure. The AS-benchmark-adjusted returns of those same funds do not. Put another way, when a fund has a benchmark discrepancy, the AS benchmark is generally an accurate match for the fund's actual investment strategy. Consequently, performance evaluation using those funds' AS benchmark-adjusted returns—but not using their prospectus-benchmark-adjusted returns—results in conclusions similar to those from factor model regressions.

Sensoy (2009) was the first to develop a procedure to identify benchmark discrepancies, but his procedure differs significantly from ours. His procedure begins with funds that have prospectus benchmark styles that do not match the Morningstar (2004) style box. For this group, he regresses each fund's full return history on the returns of the fund's prospectus benchmark and, separately, on the returns of a benchmark corresponding to the fund's style box. If the regression using the style box implied benchmark results in a higher R^2, then the fund is labeled as having a benchmark discrepancy.

We elaborate later, but in brief, our procedure differs from that of Sensoy (2009) because (i) our procedure is based on recent fund holdings rather than the full time-series of returns and style boxes and (ii) our procedure requires that the economic magnitude of the differences between the prospectus and alternative benchmark be meaningful. Applied to our sample, the respective procedures find a similar percentage of funds with a benchmark discrepancy: 26% using our procedure and 23% using Sensoy's. The two procedures, however, frequently disagree on which particular funds have a benchmark discrepancy. Among the funds for which at least one of the two procedures identifies a benchmark discrepancy, about 43% only have a discrepancy according to our procedure and another 35% only have a discrepancy according to Sensoy's procedure.[4]

We compare the impact on performance evaluation of the two procedures by focusing on funds that have a benchmark discrepancy according to one procedure but not the other. On the one hand, funds identified as having a benchmark discrepancy by our procedure, but not Sensoy's, have a higher average benchmark-adjusted return when using the prospectus benchmark instead of the AS benchmark. The difference is 1.33% per year. On the other hand, funds identified as having a benchmark discrepancy by Sensoy's procedure, but not ours, have no significant difference in benchmark-adjusted return between their benchmark choices. Based on that result, we conclude that our procedure's benchmark discrepancies have a larger impact on performance evaluation.

[4] In our sample, 50% of the benchmark discrepancies from following Sensoy's procedure come from funds with an S&P 500 prospectus benchmark. Only 11% of our benchmark discrepancies come from such funds. Our procedure's second most common benchmark discrepancy (Russell 2000 Value and S&P 600 Value) and our third most common benchmark discrepancy (Russell 2000 Growth and S&P 600 Growth) are pairs that fall within the same style box.

Because funds with benchmark discrepancies often have high active share, we specifically consider the relation between benchmark discrepancies, active share, and performance. As documented in Cremers and Petajisto (2009), high active share funds tend to outperform in the future, especially if they have outperformed in the past. We find that result is driven by funds without benchmark discrepancies. High active share funds with a benchmark discrepancy have a Cremers, Petajisto, and Zitzewitz (2012) seven-factor alpha of only 0.07% per year. In comparison, high active share funds without a benchmark discrepancy have an alpha of 1.28% per year. If this latter group is further conditioned on past outperformance, the alpha increases to 3.21% per year. Put differently, having a unique, successful investment strategy consistent with the prospectus benchmark at present predicts strong performance in the future. These results support the predictive power of active share, which has been a subject of on-going debate (see Cremers, Fulkerson, and Riley, 2019).

We conclude our analysis by building on Sensoy (2009) and considering the degree to which investors are aware of benchmark discrepancies. The more investors rely on the simple comparison of fund return relative to prospectus benchmark return—rather than on more accurate performance measures—the greater the effect benchmark discrepancies will have on capital allocation. As in Sensoy (2009), we find that investors respond to fund performance relative to the prospectus benchmark even when a benchmark discrepancy exists, but we novelly capture the magnitude of the response. A prospectus benchmark overstating the performance of a fund by 1.00% has the same impact on net fund flows as an actual increase in performance of 0.54%. We also demonstrate that investors discount performance relative to the prospectus benchmark as *Benchmark Mismatch* increases and as the gap between AS and prospectus benchmark returns increases. In line with recent work showing that investors use readily available measures of performance when allocating capital (Ben-David, Li, Rossi, and Song, 2019; Chakraborty, Kumar, Muhlhofer, and Sastry, 2018; and Evans and Sun, 2018), accounting for a fund's CAPM or multifactor alpha still leaves a large role in explaining fund flows for the prospectus-benchmark-adjusted returns.

Together, our results indicate that investors could substantially improve their capital allocations by identifying whether a fund has a benchmark discrepancy. Using our holdings-based procedure, economically significant benchmark discrepancies can be readily identified. That identification is, in part, valuable because funds with such discrepancies are likely to have prospectus benchmarks that give rise to misleading conclusions. It is further valuable because, by accounting for benchmark discrepancies while also conditioning on past performance and active share, investors can detect funds with large, positive expected alphas.

2. Comparison with prior work

Several studies show that mutual funds' prospectus benchmarks and self-declared styles are often inaccurate, but our study is most comparable to Sensoy (2009).[5] He finds that about 31% of mutual funds have a benchmark discrepancy; that a fund's flows are influenced by performance relative to the prospectus benchmark, even when a benchmark discrepancy exists; and that benchmark discrepancies are likely driven by strategic behavior to increase flows. Our study differs from Sensoy (2009) in several important ways.

[5] For example, Brown and Goetzmann (1997); DiBartolomeo and Witkowski (1997); Kim, Shukla, and Tomas (2000); Chan, Chen, and Lakonishok (2002); Elton, Gruber, and Blake (2003, 2014); Bertsch and Idzorek (2004); Brown, Harlow, and Zhang (2009); Wermers (2012); Hirt, Tolani, and Philips (2015); Bams, Otten, and Ramezanifar (2017); Cao, Iliev, and Velthuis (2017); Mateus, Mateus, and Todorovic (2017); Evans and Sun (2018); and DeWoody (2019) all show evidence of apparent misclassification.

To begin, our procedure for identifying benchmark discrepancies is markedly different from that of Sensoy (2009). As explained above, we directly compare fund and benchmark holdings to identify benchmark discrepancies, while Sensoy uses Morningstar (2004) style boxes and fund returns. Specifically, Sensoy's procedure labels a fund as having a benchmark discrepancy if two conditions are met. First, the mode of the fund's Morningstar style box must not match the fund's style as implied by the prospectus benchmark. Second, the returns on the benchmark that corresponds to the mode of the fund's style box must have a greater correlation with the full sample of the fund's returns.

Our procedure, while building on Sensoy (2009), addresses some of its limitations. The benchmark discrepancies identified by Sensoy's procedure are binary and time invariant. The magnitude of the style box and correlation differences are not considered, and each fund permanently has or does not have a benchmark discrepancy. Further, because Sensoy's procedure uses funds' full return and style box histories, the benchmark discrepancies are identified *ex post*, which can potentially cause look-ahead bias. In comparison, our procedure allows us (i) to measure the magnitude of a benchmark discrepancy, using *Benchmark Mismatch*; (ii) to capture time-variation in the appropriate benchmark, important given that Huang, Sialm, and Zhang (2011) show significant time variation in fund risk; and (iii) to identify the appropriate benchmark *ex ante*, which Sharpe (1992) labels an essential component of a benchmark.

Our procedure for identifying benchmark discrepancies is also relatively 'factor agnostic.' Put another way, our procedure makes weaker assumptions about which factors a fund should match with respect to its benchmark. While all the benchmarks used in analysis are nominally built based on size and growth/value dimensions, our procedure does not explicitly consider those dimensions when determining which benchmark best matches each fund. The usage of the

Morningstar style boxes in Sensoy's procedure, however, means that it can only identify benchmark discrepancies when there is a deviation along either the size or growth/value dimension. For example, using Sensoy's procedure, a fund with a small-cap value prospectus benchmark that is labeled small-cap value by its style box cannot have a benchmark discrepancy. Conversely, using our procedure, a benchmark discrepancy can still be identified. The Russell 2000 Value and the S&P 600 Value are both small-cap value benchmarks, but our procedure will identify a benchmark discrepancy for any fund that declares one their prospectus benchmark while having the other as their AS benchmark. The two benchmarks do ostensibly cover the same factors, but their holdings differ significantly, with an average *Benchmark Mismatch* of 69%.

The result of these methodological differences is that our procedure and Sensoy's procedure frequently disagree about which funds have a benchmark discrepancy. We consider the extent, impact, and importance of this disagreement in detail in Section 7.

Beyond differences in procedure and subsequent outcomes, our study is distinct from Sensoy (2009) in several other ways. First, our analysis focuses on the magnitude of and explanations for the difference in returns between the AS and prospectus benchmarks. We find that, compared to their AS benchmarks, the prospectus benchmarks of funds with benchmark discrepancies have lower average returns, with most of that lower return attributable to smaller factor exposures. Sensoy (2009) does not document or focus on a similar difference in returns. Second, we provide new insights into the responsiveness of investors to performance relative to the prospectus benchmark. For instance, we estimate the marginal impact on fund flows from employing a prospectus benchmark that understates risk and show how fund flows are affected by the magnitude of *Benchmark Mismatch*. Third, while Sensoy (2009) shows that funds with a benchmark discrepancy have higher tracking error and higher fees and that investors in those funds would have been better off in equivalent passive investments, our analysis demonstrates how accounting for benchmark discrepancies helps detect funds with positive performance persistence.

3. Data

3.1. Mutual fund sample

Our sample of actively managed mutual funds comes from the Center for Research in Security Prices (CRSP) Survivor-Bias-Free U.S. Mutual Fund database. We focus on U.S. equity funds, although our analysis could potentially be applied to other styles. To identify actively managed funds that almost exclusively invest in U.S. equities, we first exclude any fund that CRSP identifies as an index fund, ETF, or variable annuity; use only funds with Lipper, Strategic Insight, or Wiesenberger investment objective codes consistent with following a traditional long-only U.S. equity strategy; and require funds to invest at least 80 percent of their assets in common stock. We filter out funds with terms in their name associated with index funds or strategies other than traditional long-only U.S. equity strategies.[6] We address the incubation bias identified by Evans (2010) by excluding a fund from the sample until it is at least two years old and until it first reaches at least $20 million in assets.

All of our analysis is conducted at the fund level. We aggregate information across multiple share classes of a fund using the WFICN variable available from MFLINKS. Fund assets are the sum of the assets across all share classes. All other fund characteristics, including returns and expense ratios, are calculated as the asset-weighted average of the share class values.

We collect information on funds' prospectus benchmarks from Morningstar Direct and match that data to CRSP using ticker and CUSIP. A fund is dropped from the sample if we cannot

[6] The list of terms used in this filtering process is available upon request.

match it to Morningstar or if Morningstar does not provide a prospectus benchmark. The data on the prospectus benchmarks is cross-sectional, rather than time-series, but changes in the prospectus benchmark are rare. Evans and Sun (2018) estimate that in a given year only about 2% of funds undergo a change.

Securities and Exchange Commission (SEC) rules required mutual funds to provide a benchmark to investors in certain documents released after July 1, 1993.[7] Since the period used in our analysis is 1991 through 2015, there is the potential for survivor bias in the first couple of years of our sample. However, we find that the probability of survivorship in the 1991-1993 CRSP sample is not related to having a prospectus benchmark in Morningstar Direct. Furthermore, we find economically negligible differences in our results across the pre- and post-regulation sub-periods, and our conclusions are the same regardless of whether we include the pre-regulation data.[8]

3.2. Mutual fund holdings

We use the Thomson Reuters Mutual Fund Holdings database as our source of mutual fund holdings. As shown in Schwarz and Potter (2016), this data is not always consistent with the data filed by mutual funds with the SEC; however, the authors find little evidence of systematic bias. The holdings data only contains information on funds' equity positions, so any non-equity

[7] The original rule specifically required all mutual funds provide "a line graph comparing its performance to that of an appropriate broad-based securities market index" as part of "its prospectus or, alternatively, in its annual report to shareholders." It is common to cite December 1, 1998 as the time mutual funds were first required to provide a benchmark to investors; however, that rule only added the requirement that all mutual funds compare "the fund's average annual returns for 1, 5, and 10 years with that of a broad-based securities market index" to the pre-existing disclosure. See Final Rule: Disclosure of Mutual Fund Performance and Portfolio Managers, page 1, https://www.sec.gov/rules/final/33-6988.pdf, and Final Rule: Registration Form Used by Open-End Management Investment Companies, Section I, https://www.sec.gov/rules/final/33-7512r.htm.

[8] For example, the difference in returns between the prospectus benchmarks and the AS benchmarks of funds with a benchmark discrepancy is about the same pre-regulation as post-regulation. The difference in return differences between the two periods is only 0.01% per year (*t*-stat = 0.01).

positions, including cash, are not reflected. We drop any holdings reports with fewer than 20 equity positions, which is an unusual occurrence and may indicate the holdings report is incomplete.[9]

This data is first merged with the CRSP stock database to obtain price information and to adjust for stock splits. It is then merged with the CRSP fund database using MFLINKS. To verify that match, we drop any funds that (i) have asset values in Thomson Reuters and CRSP that are not approximately the same or (ii) have implied gross fund returns from Thomson Reuters and net fund returns from CRSP that are not highly correlated.

3.3. Benchmark holdings

Our procedure to determine which funds have a benchmark discrepancy involves a comparison between fund holdings and the holdings of a set of benchmark indexes. The set always includes a given fund's prospectus benchmark. We constrain our sample of funds to those with prospectus benchmarks in the set containing the Russell 1000, Russell 2000, Russell 3000, Russell Midcap, S&P 500, S&P 400, and S&P 600, plus the value and growth components of those seven benchmarks. Our primary reason for this constraint is that these 21 benchmarks are well-diversified individually and frequently referenced by funds as a group. The prospectus benchmarks for 97.4% of the funds in our initial sample are in the set, with the most common benchmarks being the S&P 500 (28% of funds), Russell 1000 Growth (12%), Russell 1000 Value (11%), and Russell 2000 (11%).

By considering just these 21 benchmarks when comparing holdings, we (i) do not assign any AS benchmark that is outside of the set normally considered by actual funds and (ii) avoid overfitting the benchmark. If we use a more expanded set of benchmarks, including more concentrated and rarely used indexes, then it is more likely that a significant overlap with a

[9] If not incomplete, these funds would likely have difficulty satisfying the requirements to be considered diversified under the Investment Company Act of 1940.

benchmark other than the prospectus benchmark will be unintentional or caused by active stock-picking. Specifically, a fund that is following the style of its prospectus benchmark while also doing a lot of active stock-picking may, by chance, have holdings similar to a relatively obscure index.

Our data on benchmark holdings comes from multiple sources. Russell provided us the constituent weights for their benchmarks, while the constituent weights for the S&P benchmarks come from Compustat. Monthly return data for the benchmarks (with dividends reinvested) comes from Morningstar Direct. Our final sample consists of 197,643 fund-month observations across 1,216 unique funds. The number of funds varies over time. Our sample has 142 unique funds in 1991, 299 in 1996, 633 in 2001, 901 in 2006, 1,053 in 2011, and 931 in 2015.

4. Benchmark discrepancy methodology

4.1. The active share (AS) benchmark

We classify a fund as having a benchmark discrepancy if two conditions are satisfied. First, a benchmark in our set matches the fund's actual investment style better than the prospectus benchmark. Second, that alternative benchmark is substantially different from the prospectus benchmark, i.e., the differences between the two benchmarks are economically meaningful.

Our method of determining the best match focuses on holdings. We determine the benchmark that best matches a given fund's actual investment style by finding the benchmark whose holdings have the greatest overlap with that fund's holdings. The extent to which a fund's holdings overlap with each benchmark is measured using Cremers and Petajisto's (2009) active share, defined as:

$$Active\ Share = \frac{1}{2}\sum_{i=1}^{N}\left|w_{i,f} - w_{i,b}\right| \tag{1}$$

where $w_{i,f}$ is the weight on stock *i* in the fund's portfolio and $w_{i,b}$ is weight on stock *i* in the benchmark. The measure is calculated over all *N* stocks in the investable universe. An alternative formula for active share is given in Cremers (2017):

$$Active\ Share = 1 - \sum_{i=1}^{N} MIN\left(w_{i,f}, w_{i,b}\right) * d[w_{i,f} > 0] \tag{2}$$

where $d[w_{i,f} > 0]$ is a dummy variable equal to one if stock *i* has a positive weight in the fund's portfolio. The formula to calculate active share in Eq. (2) produces the same active share values as the prior formula in Eq. (1) as long as the fund does not employ leverage or shorting, which is the case for almost all funds in our sample. Eq. (2) emphasizes, however, that active share is only lowered by overlapping weights. That is, active share is equal to 1 minus the sum of the overlapping weights.

As active share increases, the fund and a given benchmark have less overlap. An active share of zero means that the fund and a given benchmark have identical stock holdings, and an active share of 100% means that the fund and a given benchmark have no overlap in stocks held. Accordingly, we consider the benchmark that best matches each fund in our sample to be the benchmark that results in the lowest active share (considered across all 21 benchmarks). We label that benchmark the minimum active share benchmark, or simply the 'AS benchmark.'[10]

The AS benchmark for each fund is re-determined every time our data provides a new holdings report, which is quarterly in most instances. Allowing the benchmark to vary over time is important because fund style and risk are not time invariant. For example, Brown and

[10] Following Cremers and Petajisto (2009), researchers tend to approach active share as a measure of the level of active management, but fundamentally, it is a more broadly applicable measure of overlap.

Goetzmann (1997), Chan, Chen, and Lakonishok (2002), and Cao, Iliev, and Velthuis (2017) show evidence of style drift, while Brown, Harlow, and Starks (1996) and Huang, Sialm, and Zhang (2011) show variation over time in overall fund risk taking. The AS benchmark is always assigned *ex ante*, such that the AS benchmark assigned to a fund at the end of quarter t is used for analyzing the fund in quarter $t + 1$.

It is not uncommon for a fund's AS benchmark to change. The average fund is in our sample for 13.5 years and changes its AS benchmark 12.7 times (using an average of 3.7 different AS benchmarks). However, most of these changes are not economically meaningful. For example, more than half of all changes are between AS benchmarks that both result in *Benchmark Mismatch* less than our 60% cut-off. If we lower the number of AS benchmark changes by using the mode of a fund's AS benchmark over the previous three years, our primary results are unchanged.[11]

The AS benchmark changes suggest the possibility of market timing, but we find little evidence of timing ability for funds related to changes in the AS benchmark.[12] For example, in the month after an AS benchmark change, the average difference in returns between the new and old AS benchmark is only 0.94 basis points (t-stat = 0.38). That difference rises to (i) just 2.64 basis points (t-stat = 0.65) when the sample is limited to changes in which *Benchmark Mismatch* increases and (ii) just 4.54 basis points (t-stat = 1.18) when limited to changes in which *Benchmark Mismatch* increases by at least 30% (i.e., within the highest quartile of *Benchmark Mismatch* changes).

[11] We also considered the possibility that a fund could be approximately equidistant between two dissimilar benchmarks, with one of those benchmarks being the prospectus benchmark. For that fund, a small change in holdings could cause a large change in *Benchmark Mismatch*. We find this situation to be rare and dropping observations in which it occurs or dropping the funds in which it occurs does not change our primary results.

[12] Bollen and Busse (2001); Kaplan and Sensoy (2005); Jiang, Yao, and Yu (2007); Mamaysky, Spiegel, and Zhang (2008); Chen, Ferson, and Peters (2010); Elton, Gruber, and Blake (2012); and Kacperczyk, Nieuwerburgh, and Veldkamp (2014) all show evidence of timing ability for mutual funds.

4.2. Benchmark Mismatch

After identifying the set of funds for which the prospectus benchmark differs from the AS benchmark, we determine whether the different benchmarks are dissimilar to an economically significant degree. This step is important because in many cases the AS and prospectus benchmark are quite similar, simply because many benchmarks in our set of 21 are quite similar. For example, the Russell 1000 and Russell 3000 have an active share of 8.0% relative to each other (averaged across our sample period). Given the similarity in those benchmarks' holdings, a fund with the Russell 1000 as its prospectus benchmark and the Russell 3000 as its AS benchmark may have a difference in benchmarks, but that difference is not economically significant.

We measure the dissimilarity between the prospectus and AS benchmarks by calculating how little their respective holdings overlap. We measure overlap by calculating the active share between the two benchmarks. We label the resulting value *Benchmark Mismatch*:

$$Benchmark\ Mismatch = \frac{1}{2}\sum_{i=1}^{N}\left|w_{i,p} - w_{i,AS}\right| \tag{3}$$

where $w_{i,p}$ is the weight on stock *i* in the fund's prospectus benchmark and $w_{i,AS}$ is weight on stock *i* in the fund's AS benchmark.[13] When the holdings of the two benchmarks largely overlap, the active share of the prospectus benchmark with respect to the AS benchmark is low and thus *Benchmark Mismatch* will be small. Hence, an increase in *Benchmark Mismatch* represents an increase in the difference between the holdings of the two benchmarks (or a decrease in the overlap of holdings).

Since *Benchmark Mismatch* captures how different the holdings of the prospectus benchmark are from the holdings of the AS benchmark, we can directly interpret *Benchmark*

[13] As with the active share of a fund relative to a given benchmark, the active share of the benchmarks relative to each other can also be calculated using the MIN() specification shown in Eq. (2).

Mismatch as a measure of the economic magnitude of the differences between those benchmarks. For the main results in the paper, we classify funds with *Benchmark Mismatch* above 60% as having significant economic differences in their benchmarks and thus having a benchmark discrepancy.

While the 60% cut-off is somewhat arbitrary, we set the threshold there for two reasons. First, a similar 60% cut-off is employed in prior work using active share (e.g., Cremers and Petajisto, 2009, and Cremers and Curtis, 2016). Funds in those studies with active share less than 60% are labeled 'closet indexers.' Second, as shown in Section 6, analysis of the benchmark returns suggests that the economic differences between the prospectus and AS benchmarks are minor on average when *Benchmark Mismatch* is less than 60%.[14]

We also consider the difference between the active share of a fund with respect to its prospectus benchmark and the active share of a fund with respect to its AS benchmark. We label this difference *Active Gap*:

$$Active\ Gap = Active\ Share_{Prospectus} - Active\ Share_{AS} \quad (4)$$

As *Active Gap* increases, the difference between the overlap of fund and AS benchmark holdings and the overlap of fund and prospectus benchmark holdings increases. Unlike *Benchmark Mismatch*, *Active Gap* does not directly measure whether the prospectus benchmark and AS benchmark are economically different. However, conditional on having large *Benchmark Mismatch, Active Gap* does indicate the extent to which the activeness of a fund is overstated by using the prospectus benchmark instead of the AS benchmark.

[14] Among funds with *Benchmark Mismatch* greater than 60%, tracking error is lower on average with respect to the AS benchmark compared to prospectus benchmark (*t*-stat = 4.92). That result does not hold for funds with *Benchmark Mismatch* less than 60%.

4.3. Summary Statistics

Table 1 shows summary statistics for the key measures used in our study. The dummy variable *Any Mismatch*, which is equal to one if *Benchmark Mismatch* is above zero, shows that about 67% of observations have a prospectus benchmark different from the AS benchmark. If we require *Benchmark Mismatch* to be greater than 60%, as motivated above, then only 26% of observations have a difference in benchmarks (as indicated by the dummy variable *Large Mismatch*). While the AS benchmark is re-determined with each new holdings report, funds with AS benchmarks that are different from their prospectus benchmarks tend to maintain that difference. The correlation between the value of *Any Mismatch* (*Large Mismatch*) in month t – 12 and month t is 0.59 (0.75). Further, if the prospectus benchmark and AS benchmark are different in month t – 12, then the probability that they will still be different in month t is 86% (untabulated). That result is about the same if the analysis is limited to funds with *Benchmark Mismatch* greater than 60%.[15]

[Table 1 about here]

As shown in Figure 1, the frequency of differences in benchmarks varies through time. The percentage of funds with a difference of any magnitude is as low as 61% (in 1994, 2005, and 2009) and as high as 78% (in 1995) with no obvious trend. The number of funds with *Benchmark Mismatch* greater than 60% varies within the range of 14% to 37%. The significant jumps in the number of large differences in 1992 and 1998 are a result of new benchmarks entering the set.[16]

---

[15] About 45% of mutual funds in our sample never have a *Benchmark Mismatch* greater than 60%, while about 7% of funds always have a *Benchmark Mismatch* greater than 60%. Within the remaining 48% of funds, about 39% have a *Benchmark Mismatch* greater than 60% for more than half of their life.

[16] Excluding the S&P 500, which is available from the start of our sample, the S&P benchmarks enter into our sample over the period 1992 through 1997. Each of the Russell benchmarks is available from the start of our sample.

After 1998, when all 21 benchmarks in our set are available, the number of funds with *Benchmark Mismatch* above 60% slowly decreases from 30% to 21%.

[Figure 1 about here]

The average active share is 80.7% using the prospectus benchmark, compared to 78.4% using the AS benchmark. As such, that difference (i.e., the *Active Gap*) has a mean of 2.3%. Active share is persistent over time, as the annual autocorrelation is above 0.90 using either the prospectus or AS benchmark.

Using net fund returns, the average fund underperforms both the prospectus and AS benchmark (ignoring the costs of investing in those benchmarks); however, the degree of underperformance differs. Relative to the prospectus benchmark, funds on average underperform by 0.33% per year, which is not statistically distinguishable from zero (t-stat = −1.06). In comparison, relative to their AS benchmark, funds underperform by 0.78% per year, which is statistically significant (t-stat = −2.87). This first glance suggests that the choice of benchmark does affect performance evaluation when using a simple comparison of fund return relative to benchmark return. Prospectus benchmarks have noticeably lower returns compared to AS benchmarks in the full sample.

The average *Benchmark Mismatch* and *Active Gap* are 34.9% and 2.3%, respectively, but those values are biased downward by funds with the same prospectus and AS benchmark. Figure 2 shows the cumulative density function (CDF) of *Active Gap* for funds with different prospectus and AS benchmarks. For most of these funds, *Active Gap* is small. About 38% of funds have *Active Gap* below 2%, and about 78% have *Active Gap* below 5%. There are, however, a meaningful number of funds with large *Active Gap*. About 15% of the full sample of funds—including those funds with the same prospectus and AS benchmarks—have *Active Gap* above 5%, and 3% of the full sample have *Active Gap* greater than 10%.

[Figure 2 about here]

Figure 3 shows the CDF of *Benchmark Mismatch* for funds with different prospectus and AS benchmarks. *Benchmark Mismatch* is below 60% for most of these funds, but 38% have *Benchmark Mismatch* above 60% (26% of the full sample). Among that group, about half of the funds have *Benchmark Mismatch* above 80%. If those funds had exactly the same holdings as their AS benchmark, then an investor using the prospectus benchmark would conclude those funds have an active share of at least 80%—close to the full sample average for that variable.

[Figure 3 about here]

5. Funds with large versus small *Benchmark Mismatch*

Before considering performance, we first analyze the characteristics of funds as a function of *Benchmark Mismatch*. As mentioned before, we separate funds using a *Benchmark Mismatch* cut-off of 60% and refer to funds with *Benchmark Mismatch* greater than 60% as having a benchmark discrepancy.

Comparing the prospectus and AS benchmarks of funds with non-zero *Benchmark Mismatch*, funds with and without a benchmark discrepancy differ in several ways. Table 2 shows the five most common benchmark combinations for each group. Funds with small, but non-zero, *Benchmark Mismatch* (i.e., those without a benchmark discrepancy) have prospectus and AS benchmarks that are quite similar. By our construction, the prospectus and AS benchmarks of these funds are closet indexers of each other. The most common difference, an S&P 500 prospectus benchmark and an S&P 500 Growth AS benchmark, has a *Benchmark Mismatch* of 33.0%. In most of the cases in which *Benchmark Mismatch* is small, the AS benchmark is close to or is a complete subset of the prospectus benchmark (or vice versa).

[Table 2 about here]

Conversely, funds with benchmark discrepancies have meaningful differences between their prospectus and AS benchmarks. The most common benchmark discrepancy is a Russell 2000 prospectus benchmark and an S&P 600 Growth AS benchmark. For these funds, overlap is limited as *Benchmark Mismatch* is 77.1%. To illustrate, the Russell 2000 contains all of the stocks with a market capitalization ranking between 1001 and 3000, whereas the S&P 600 Growth contains the growth stocks within the full set of stocks with a market capitalization ranking between 901 and 1500. As a result, even if all stocks with a ranking between 1001 and 1500 were labeled growth by S&P, the two benchmarks could have at most 500 stocks in common. More importantly, since both of the benchmarks weight by market capitalization, (i) any overlapping stocks should have relatively large weights in the Russell 2000 and relatively small weights in the S&P 600 Growth, and (ii) the growth stocks with a ranking between 901 and 1000 will have the largest weights of any stocks in the S&P 600 Growth, but zero weight in the Russell 2000.

Table 3 compares the characteristics of funds with and without a benchmark discrepancy. In this analysis, the group without a benchmark discrepancy includes funds with *Benchmark Mismatch* of zero. Funds with a benchmark discrepancy tend to be more actively managed. The average active share with respect to the prospectus benchmark for those funds is 93.5%, compared to 76.8% for funds without a benchmark discrepancy.[17] Funds with a benchmark discrepancy also have fewer assets, are younger, and charge a greater expense ratio. With respect to style as defined by the prospectus benchmark, funds with a benchmark discrepancy tend to disproportionately have

[17] About 74% of funds in the highest quintile of prospectus active share have a benchmark discrepancy.

a small-cap or mid-cap style. About 80.4% of funds with a benchmark discrepancy have a small- or mid-cap style, while only 23.7% of funds without a benchmark discrepancy have those styles. The differences between the two groups across the value-blend-growth dimension of style are slight in comparison.

[Table 3 about here]

Next, we consider the relation between having a benchmark discrepancy and fund characteristics using the following model:

$$BM > 60\%_{i,t} = \alpha + \beta * Active\ Share_{i,t} + \delta * Chars_{i,t} + \gamma * Style_i + FE + \varepsilon_{i,t} \qquad (5)$$

where $BM > 60\%_{i,t}$ is a dummy variable equal to one if the *Benchmark Mismatch* for fund *i* based on quarter *t* holdings is greater than 60%. $Active\ Share_{i,t}$ is a vector of information about fund *i*'s quarter *t* active share. It includes the prospectus benchmark active share and a dummy variable equal to one if that active share is among the top 20% across all funds in quarter *t*. $Chars_{i,t}$ is a vector of characteristics for fund *i* available as of quarter *t* and includes the natural log of assets, natural log of age, expense ratio, turnover ratio, the number of equity positions, and the percentage of fund assets held within institutional share classes. $Style_i$ is a vector of information about fund *i*'s style (based on its prospectus benchmark). It includes a large-cap dummy, a blend dummy, and a growth dummy. FE represents year-quarter fixed effects. We estimate the model using a logit regression and the full sample of fund-quarters, including funds with *Benchmark Mismatch* of zero.

Table 4 presents the results from this regression. As prospectus active share increases, the probability of having a benchmark discrepancy increases. After controlling for fund characteristics and style, that relation becomes non-linear. Funds in the top 20% of active share are more likely to have a benchmark discrepancy than the linear term indicates. The full model without fixed effects, which is shown in column (5), predicts that a fund has a 6.1% probability of having a benchmark discrepancy if at the 50^{th} percentile of active share and the mean of all other variables. If that same fund instead had an active share at the 85th percentile, that probability would increase to 70.1%.

[Table 4 about here]

It can be difficult to identify what makes the portfolios of high active share funds with and without a benchmark discrepancy different. To demonstrate by hypothetical example, consider two funds—A and B—that both have an S&P 500 prospectus benchmark. Fund A's holdings are 20 equally weighted random Russell 3000 stocks, and Fund B's holdings are 20 equally weighted random S&P 500 stocks. Fund A is likely to be in the top 20% of prospectus benchmark active share (99% chance in simulations) and is likely to have a benchmark discrepancy (73% chance in simulations). In comparison, Fund B is likely to be in the top 20% of prospectus benchmark active share (71% chance in simulations) and is unlikely to have a benchmark discrepancy (20% chance in simulations). By taking a unique investment approach that is still consistent with the prospectus benchmark, Fund B is usually able to be highly active without generating a benchmark discrepancy.

The fund characteristics in Table 4 have either limited economic significance or limited statistical significance when considered in the full model.[18] Fund style, however, contains substantial predictive power. Funds with a large-cap style are less likely to have a benchmark discrepancy compared to funds that have a small- or mid-cap style. Returning to the example fund at the 85th percentile of active share, the full model without fixed effects in column (5) predicts

[18] Sensoy (2009) does not find evidence to support the idea that the likelihood of a benchmark discrepancy increases as a fund gets older. The insignificant impact of age in our full model is consistent with his results.

that the benchmark discrepancy probability would be 61.3% if that fund had a large-cap style. In comparison, that probability would be 81.5% if that fund had a small- or mid-cap style.

Sensoy (2009), using his benchmark discrepancy procedure, finds a substantial impact related to fund family in a similar analysis. About 90% of the fund family effects in his model are statistically significant. That result is consistent with other studies showing that decisions about a fund are often made at the family rather than fund level (e.g., Clifford, Fulkerson, Hong, and Jordan, 2014; Evans, Ferreira, and Prado, 2017; and Ma, Tang, and Gomez, 2019). Nevertheless, in untabulated analysis, we do not find similar results if fund family effects are added to our model. Only about 20% of the fund family effects are statistically significant. This difference suggests that the benchmark discrepancies identified by our procedure are determined at the fund, rather than family, level.[19]

Determination at the fund level suggests the possibility of risking shifting or tournament-style behavior by funds with respect to their AS benchmark: a fund without a benchmark discrepancy that underperforms its peers in the first half of a calendar year could switch to having a benchmark discrepancy for the remainder of the calendar year as part of a strategy to recover from its relatively low prior returns.[20] While this could happen on occasion, we find little evidence in the untabulated results discussed below to support the idea that tournament strategies are related to benchmark discrepancies.

[19] While prior research has shown differences between large and small families (e.g., Almazan, Brown, Carlson, and Chapman, 2004; Pollet and Wilson, 2008; and Bhojraj, Cho, and Yehuda, 2012), we do not find a difference in benchmark discrepancy likelihood between large and small families in our full model.
[20] There is a large literature that considers whether and when risk shifting occurs among mutual funds. See, for example, Brown, Harlow, and Starks (1996); Chevalier and Ellison (1997); Busse (2001); Qiu (2003); Basak, Pavlova, and Shapiro (2007); Kempt, Ruenzi, and Thiele (2009); Elton, Gruber, Blake, Krasny, and Ozelge (2010); Huang, Sialm, and Zhang (2011); Schwarz (2012); Basak and Makarov (2014); Sato (2016); Chen, Hughson, and Stoughton (2018); and Lee, Trzcinka, and Venkatesan (2019).

First, those strategies require funds to switch their benchmark discrepancy status, but switches are uncommon. About 52% of funds never switch their status, and just 4% of funds switch to having a benchmark discrepancy at the mid-point of the average calendar year. Second, those strategies imply that shifts are temporary, but most switches are persistent. About 12% of the funds that switch to having benchmark discrepancy at the mid-point of a calendar year never switch back, and about 51% of the funds that do eventually switch back take more than six months to do so. Third, those strategies suggest that switches should be related to recent past performance, but we find no relation between performance in the first half of a calendar year and switching to having a benchmark discrepancy at the calendar year mid-point. On average, the funds that choose to switch to having a benchmark discrepancy at the mid-point of a calendar year have only underperformed the funds that choose not to switch by an economically and statistically insignificant nine basis points in the first half of the calendar year (t-stat = 0.35). That difference is calculated using prospectus-benchmark-adjusted returns, but results are similar using AS-benchmark-adjusted returns. Considering the evidence in totality, rather than being a temporary measure to boost performance in certain circumstances, having a benchmark discrepancy appears to be a more permanent aspect of a fund.[21]

6. Differences in prospectus and AS benchmark returns and fund performance implications

This section first considers whether the prospectus benchmark gives a fund a 'performance boost' when benchmark-adjusting returns. We answer that question by comparing the returns on the AS and prospectus benchmarks of funds with non-zero *Benchmark Mismatch*. We then

[21] The flow-performance relation that could incentivize a fund to have a benchmark discrepancy—discussed in Section 9—holds for every calendar month of the sample, which suggests that funds should have limited motivation to only switch at certain points in the calendar year.

consider (i) how conclusions about a fund's performance can change depending on the benchmark used and (ii) how accounting for benchmark discrepancies affects an investor's ability to select funds that can be expected to outperform in the future.

6.1. Comparing benchmark returns

Table 5 shows the average difference in annualized return between the AS benchmark and prospectus benchmark—that is, the performance boost—for funds as a function of their *Active Gap* and *Benchmark Mismatch*. Panel A divides funds into five ranges of *Benchmark Mismatch*, and Panel B divides funds based on whether *Benchmark Mismatch* is above or below 60%. The ranges for *Active Gap* are the same in both panels. Funds with *Benchmark Mismatch* of zero are excluded from this analysis.

[Table 5 about here]

Focusing first on Panel A, the average performance boost for funds with non-zero *Benchmark Mismatch* is 0.68% per year (*t*-stat = 2.72). However, the performance boost is considerably larger for funds with greater *Benchmark Mismatch*. Funds with *Benchmark Mismatch* above 80% have an average performance boost of 1.64% per year (*t*-stat = 2.97), compared to −0.12% per year (*t*-stat = −0.75) for funds with *Benchmark Mismatch* less than 20%. *Active Gap* does matter, though on a much more limited basis. Compared to funds with *Active Gap* less than 1.25%, funds with *Active Gap* greater than 5% have an average performance boost 0.42% per year greater (*t*-stat = 1.31).

Once *Benchmark Mismatch* is greater than 60%, the average performance boost is consistently economically large and statistically significant. Funds with *Benchmark Mismatch* between 60% and 80% have an average performance boost of 1.37% per year (*t*-stat = 2.25), and the average performance boost for that group is at least 1% per year in each of the different ranges of *Active Gap*. There is some evidence of a performance boost for funds with *Benchmark Mismatch* between 40% and 60%, but on average, the evidence is economically smaller (0.53%) and statistically weaker (*t*-stat = 1.70). The performance boost for that group also varies without an obvious trend depending on *Active Gap*. Overall, after controlling for *Benchmark Mismatch*, *Active Gap* appears to matter little.[22]

As shown in Panel B, if we group funds based on whether *Benchmark Mismatch* is above or below 60%, the results are similar. When *Benchmark Mismatch* is above 60%, the average performance boost is 1.50% per year (*t*-stat = 3.20). In comparison, when *Benchmark Mismatch* is below 60%, the average performance boost is only 0.18% (*t*-stat = 0.94). *Active Gap* has a negligible impact within these groups. Each *Active Gap* range for funds with *Benchmark Mismatch* above 60% shows an economically large and statistically significant average performance boost. Further, there is no difference in average performance boost between the funds in that group with low and high *Active Gap*. Conversely, regardless of *Active Gap*, funds with *Benchmark Mismatch* less than 60% have average performance boosts that are economically small and statistically insignificant. We conclude based on these results that the prospectus benchmark sets a lower bar for funds to clear only when *Benchmark Mismatch* is large.[23]

We consider the determinants of the performance boost more robustly using the following model:

$$R_{AS,i,t} - R_{P,i,t} = \alpha + \beta * Active\ Share_{i,t} + \delta * Mismatch_{i,t} + \gamma * Chars_{i,t} + FE + \varepsilon_{i,t} \qquad (6)$$

[22] High *Active Gap* appears to be related to a lower performance boost for funds with *Benchmark Mismatch* less than 20%. However, there are very few funds with *Active Gap* greater than 3.75% and *Benchmark Mismatch* less than 20%—below 1% of the tested sample—so we believe caution should be exercised in making any inferences concerning those funds.

[23] Elton, Gruber, and Blake (2014) primarily study separate accounts, but they also briefly consider mutual funds and, similar to our results here, show a difference in returns between mutual funds' prospectus benchmarks and their full sample maximum correlation benchmarks (0.74% per year). However, they report neither separate results for the mutual funds whose benchmarks actually differ nor the number of mutual funds whose benchmarks actually differ, and it is uncertain whether the average difference in performance they document is statistically significant.

where $R_{AS,i,t}$ is the annualized return on fund *i*'s AS benchmark in month t and $R_{P,i,t}$ is the annualized return on fund *i*'s prospectus benchmark in month t. $Active\ Share_{i,t}$ is a vector of information about fund *i*'s active share at the start of month *t*. It includes the fund's prospectus benchmark active share and a dummy variable equal to one if that active share is among the top 20%. $Mismatch_{i,t}$ is a vector of information about fund *i*'s mismatch status at the start of month *t*. It includes *Benchmark Mismatch*, *Active Gap*, and a dummy variable equal to one if *Benchmark Mismatch* is among the top 20%. $Chars_{i,t}$ is the same vector of characteristics used in Eq. (5), but measured for fund *i* as of the start of month *t*. FE represents style and year-month fixed effects. We estimate the model using the sample of funds with non-zero *Benchmark Mismatch*.

Table 6 presents the results of these performance boost regressions. Isolated from each other, active share, *Benchmark Mismatch*, and *Active Gap* each predict the performance boost. When considered simultaneously, however, only *Benchmark Mismatch* and active share continue to have predictive power. A 1% increase in *Active Gap* is associated with a 0.095% per year (*t*-stat = 2.62) increase in the performance boost in column (3), which excludes *Benchmark Mismatch* and active share. But in column (4), which includes *Benchmark Mismatch* and active share, the increase is only 0.040% per year (*t*-stat = 1.23). In comparison, a 1% increase in *Benchmark Mismatch* is associated with a 0.033% per year (*t*-stat = 3.21) increase in the performance boost in column (2), which excludes *Active Gap* and active share, and a 0.023% per year (*t*-stat = 2.54) increase in column (4).[24] The relation between active share and the performance boost is consistently strong, but non-linear. After controlling for *Benchmark Mismatch*, a fund in the

[24] Note that a 1% increase in *Active Gap* is a much bigger change than a 1% increase in *Benchmark Mismatch*. In this sample, the standard deviation of *Active Gap* is 2.7%, compared to 24.7% for *Benchmark Mismatch*.

highest quintile of active share has a performance boost 0.46% per year (t-stat = 2.07) greater than implied by the linear coefficient.

[Table 6 about here]

Overall, these results show that a substantial number of funds have a prospectus benchmark that, on average, is easier to outperform compared to the benchmark implied by fund holdings. Funds likely to have this performance boost can be readily identified using *Benchmark Mismatch* and active share.

6.2. Comparing funds' benchmark-adjusted returns

The previous section shows how the prospectus benchmark can set a lower bar for a fund to clear compared to the AS benchmark, if fund performance is evaluated through a simple comparison with benchmark returns. We now consider how different performance evaluation methods lead to different conclusions about fund performance, depending on whether the fund has a benchmark discrepancy.

An important confounding factor in this analysis is that, as shown in Section 5, high active share funds tend to have much greater *Benchmark Mismatch*. Cremers and Petajisto (2009) show that high active share funds have better average future performance, so it is critical to account for active share when considering the relation between fund performance and *Benchmark Mismatch*. Thus, in our analysis, we independently double sort all funds—including those with *Benchmark Mismatch* of zero—into groups based on prospectus benchmark active share and *Benchmark Mismatch*. Using active share, we sort funds into quintiles. Using *Benchmark Mismatch*, we sort funds into two groups based on the 60% cut-off. It is relatively rare for funds to exceed the 60% cut-off and still have low active share; therefore, to avoid reporting results for groups with very few funds, we collapse the four lowest quintiles of active share into a single group.

We evaluate average net fund performance within each group using three performance evaluation models: the prospectus-benchmark-adjusted return, the modified-benchmark-adjusted return, and the Cremers, Petajisto, and Zitzewitz (2012) seven-factor model (henceforth, the CPZ7 model). The modified-benchmark-adjusted return is the fund return minus the prospectus benchmark return if *Benchmark Mismatch* is less than 60%. It is the fund return minus the AS benchmark return if *Benchmark Mismatch* is greater than 60%. The model represents how the prospectus-benchmark-adjusted returns would appear if funds with a benchmark discrepancy were no longer evaluated relative to their prospectus benchmark, but were instead evaluated relative to their AS benchmark. The alpha from the CPZ7 model represents the abnormal performance after accounting for funds' exposures to market, size, value, and momentum factors. It does not rely on the assignment of a singular benchmark, and it helps confirm whether our inferences using the benchmark-adjusted returns are valid.[25]

Table 7 shows the results from this analysis. Looking at the 'All Funds' portion of Table 7, which is not conditional on active share, funds with large *Benchmark Mismatch* on average outperform funds with small *Benchmark Mismatch* by 1.04% per year (*t*-stat = 3.20) using the prospectus-benchmark-adjusted returns. Using the modified-benchmark-adjusted returns, however, there is no economically or statistically significant difference in performance. The CPZ7 model indicates some outperformance by funds with large *Benchmark Mismatch* relative to those with small *Benchmark Mismatch*, but the economic size of the difference is less than it was with the prospectus benchmark (only 0.54% per year) and the difference is not statistically significant at conventional levels (*t*-stat = 1.46). Importantly, the positive relation between *Benchmark Mismatch* and active share prevents us from drawing any strong conclusions.

[25] The results using the alternative Cremers, Petajisto, and Zitzewitz (2012) four-factor model are similar to those from their seven-factor model.

[Table 7 about here]

Therefore, we turn next to the results conditional on active share. Within the high active share quintile and when using prospectus-benchmark-adjusted returns, funds perform about the same whether they have or do not have a benchmark discrepancy. Both groups show marginal evidence of outperformance (about 0.7% per year) that is almost statistically significant at conventional levels (*t*-stats of about 1.6). If we compare fund performance using the modified-benchmark-adjusted returns instead, the performance evaluation noticeably changes. Among high active share funds with a benchmark discrepancy, the average prospectus-benchmark-adjusted return is 0.72% per year (*t*-stat = 1.55), while the average modified-benchmark-adjusted return is −0.92% per year (*t*-stat = −2.32). In other words, while high active share funds with a benchmark discrepancy outperform their prospectus benchmarks on average—albeit without conventional statistical significance—they clearly underperform their AS benchmarks on average. Using modified-benchmark-adjusted returns, high active share funds without a benchmark discrepancy outperform high active share funds with a benchmark discrepancy by 1.67% per year (*t*-stat = 3.32).

Turning to the factor models, the CPZ7 alpha of high active share funds with a benchmark discrepancy is 0.07% per year (*t*-stat = 0.14). In comparison, high active share funds without a benchmark discrepancy have a CPZ7 alpha of 1.28% per year (*t*-stat = 2.14).[26] The difference in alpha between the groups of 1.21% per year is statistically significant (*t*-stat = −2.09). That result indicates that, even outside the context of benchmark-adjusted returns, high active share funds

[26] Using the Fama-French four-factor model (i.e., the Fama and French, 1993, model augmented with the Carhart, 1997, momentum factor), this group of funds has a positive alpha, but the statistical significance varies depending on the dependent variable. Using prospectus-benchmark-adjusted returns, the four-factor alpha is 1.01% per year (*t*-stat = 2.10); however, using AS-benchmark-adjusted returns and excess returns, the four-factor alphas are 0.46% per year (*t*-stat = 1.03) and 0.63% per year (*t*-stat = 0.78). The decrease in alpha when using the Fama-French four-factor model on these funds is consistent with the model's biases, which are documented in Cremers, Petajisto, and Zitzewitz (2012).

without a benchmark discrepancy outperform high active share funds with a benchmark discrepancy.

The results for the lower quintiles of active share are similar to those from the full sample. Although, like the full sample results, they should be considered cautiously because of the positive correlation between *Benchmark Mismatch* and active share. Within the bottom four quintiles of active share, the funds with a benchmark discrepancy tend to have a much higher active share than the funds without a benchmark discrepancy.

6.3. Accounting for active share, past performance, and benchmark discrepancies

The outperformance of high active share funds is concentrated among those without a benchmark discrepancy. If that group is limited to just those funds also in the top 20% of prior year CPZ7 alpha, outperformance further increases. In untabulated tests, we find that the funds in that subgroup have a CPZ7 alpha of 2.18% per year (*t*-stat = 2.51).

Here, we expand upon that finding by re-examining a key result from Cremers and Petajisto (2009): funds in the highest quintiles of both active share and benchmark-adjusted return during the previous year significantly outperform in the future. Our goal is to test the impact on that result from accounting for benchmark discrepancies. To accomplish that goal, we first conditionally sort funds by prospectus benchmark active share, by prospectus-benchmark-adjusted return over the previous year, and by whether the fund has a benchmark discrepancy (in that order). Using the resulting groups, we then form equal weight portfolios and estimate annualized alphas using the CPZ7 model. We use active share and past performance relative to the prospectus benchmark in the sorting process to capture the portfolios that investors would form if taking the prospectus benchmark at face value; however, we evaluate the portfolios using the CPZ7 model to generate a more accurate, less benchmark-dependent measure of subsequent performance.

[Table 8 about here]

Table 8 shows the performance of the portfolios at each level of sorting. An investor focusing on the prospectus benchmark and choosing to buy only the funds in the highest quintiles of past performance and active share obtains a portfolio with an alpha of 2.31% per year (*t*-stat = 3.01). If that investor drops the funds with a benchmark discrepancy from that portfolio, the alpha increases to 3.21% per year (t-stat = 2.37).[27] In comparison, if that investor drops the funds without a benchmark discrepancy, the alpha decreases to 1.72% per year (*t*-stat = 1.97). While the difference in alpha between the prior two portfolios is economically large, it is not statistically significant at conventional levels (*t*-stat = 1.13). That result is attributable, at least in part, to the relatively small number of funds within each portfolio after sorting on three dimensions.

7. Comparison with the procedure in Sensoy (2009)

In Sections 1 and 2, we discussed the details of Sensoy's (2009) benchmark discrepancy identification procedure, as well as the main differences between his procedure and ours. In this section, we compare the results from the two procedures.[28] We first compare the procedures in terms of which funds are labeled as having a benchmark discrepancy. We then consider the extent to which the choice of procedure affects performance evaluation.

Sensoy finds that 31.2% of his sample has a benchmark discrepancy over the period 1994 to 2004. Among the funds he finds to have a benchmark discrepancy, the average R^2 from regressing fund returns on the returns of the alternative benchmark is 82.6%, whereas using the

[27] Using the AS benchmark instead of the prospectus benchmark in the sorting process does not meaningfully change the alpha for this portfolio (3.15% per year, *t*-stat = 2.70). Further, if the benchmark in the sorting process is switched, the alpha for this portfolio is still the largest among those tested.

[28] While we do not exactly replicate Sensoy's results, we aim to follow his procedure closely. The only difference between Sensoy's procedure and our version of his procedure is the set of benchmarks. Sensoy uses 12 benchmarks, but we use a larger set of 21 benchmarks as motivated in Section 3.3. Our set of benchmarks includes all of those in Sensoy's set.

returns of the prospectus benchmark results in an R^2 of 70.6%. Applying Sensoy's procedure in our sample, which covers the period 1991 to 2015, we find that 23.1% of funds have a benchmark discrepancy. Within that group, the average R^2 is 87.6% with respect to the alternative benchmark versus 80.2% with respect to the prospectus benchmark.

Figure 4 shows a broad comparison between Sensoy's procedure and our procedure. The pie chart shows the commonality in fund-month observations identified as having a benchmark discrepancy. In this analysis and in subsequent comparisons, we only consider a benchmark discrepancy to exist by our procedure if *Benchmark Mismatch* is greater than 60%. The observations are at the fund-month level because whether a fund has a benchmark discrepancy is time-varying in our procedure.

[Figure 4 about here]

About 60% of fund-month observations do not have a benchmark discrepancy using either procedure; about 17% only have a benchmark discrepancy according to our procedure; and about 14% only have a benchmark discrepancy according to Sensoy's procedure. Just 2% of observations have a benchmark discrepancy according to both procedures that results in the same alternative benchmark, with another 7% having both procedures identify a benchmark discrepancy but with different alternative benchmarks. All considered, the two procedures generate notably different conclusions about which funds have a benchmark discrepancy.

A clear difference between the procedures can be recognized in their evaluation of funds with an S&P 500 prospectus benchmark. About 63% of the benchmark discrepancies in Sensoy's original analysis come from funds with an S&P 500 prospectus benchmark (50% in our sample period). In comparison, only 11% of the benchmark discrepancies identified using our procedure come from funds with an S&P 500 prospectus benchmark. Furthermore, in Sensoy's original analysis, about 73% of the funds with an S&P 500 prospectus benchmark and a benchmark discrepancy have either a Russell 1000 Value or a Russell 1000 Growth alternative benchmark. Both of those benchmark combinations have an average *Benchmark Mismatch* near 45%, well below our 60% cut-off.[29] Put another way, the most frequent benchmark discrepancies by Sensoy's procedure are not considered benchmark discrepancies by our procedure.

Given these differences, we next consider fund performance relative to the prospectus, AS, and Sensoy benchmarks. Table 9 shows the average annualized benchmark-adjusted performance of funds conditional on whether each procedure identifies a benchmark discrepancy. Considered separately, both procedures find evidence of a performance boost—i.e., the AS or Sensoy benchmark return minus the prospectus benchmark return—for funds with a benchmark discrepancy. Our procedure, however, finds a significantly larger performance boost compared to Sensoy's. Funds with a benchmark discrepancy according to our procedure have an average performance boost of 1.52% per year (t-stat = 3.14), while funds with a benchmark discrepancy according to Sensoy's procedure have an average performance boost of 0.76% per year (t-stat = 1.64).

[Table 9 about here]

More importantly, when our procedure identifies a benchmark discrepancy and Sensoy's does not, there is a large performance boost, but the reverse is not true. Funds identified by our procedure, but not by Sensoy's, have an average performance boost of 1.33% per year (t-stat = 2.25). Funds identified by Sensoy's procedure, but not by ours, have an average performance boost

[29] The average performance boost for funds with an S&P 500 prospectus benchmark and a Russell 1000 Value or a Russell 1000 Growth AS benchmark is 0.46% per year (t-stat = 1.37).

of only 0.34% per year (t-stat = 0.82).[30] When both procedures agree there is a benchmark discrepancy, the funds have a large average performance boost relative to both alternative benchmarks. These results, as a whole, indicate that our procedure's benchmark discrepancies have a greater impact on fund performance evaluation.[31]

8. Factor exposures of the prospectus and AS benchmarks

The expected return on a passively managed index should be driven solely by factor exposures, as passive indexes by construction have no alpha (arguably, see Cremers, Petajisto and Zitzewitz, 2012). Therefore, the significant difference in average return between the prospectus and AS benchmarks of funds with benchmark discrepancies must logically arise out of differences in factor exposures. Since the average return on the AS benchmarks is greater than the average return on the prospectus benchmarks, the AS benchmarks should have greater net factor exposure than the prospectus benchmarks. In this section, we first analyze factor exposure differences between the two benchmarks for funds with a benchmark discrepancy. We then test whether the AS benchmark or the prospectus benchmark more accurately reflects those funds' actual factor exposures.

[30] Elton, Gruber, and Blake (2014) use full sample fund-benchmark correlations alone to identify benchmarks for mutual funds. If we replicate their procedure, the results are similar to those presented in this section. Funds identified by our procedure as having a benchmark discrepancy, but not by theirs, have an average performance boost of 1.29% per year (t-stat = 1.78), while funds identified by their procedure, but not by ours, have an average performance boost of only 0.38% per year (t-stat = 1.37).

[31] A potential alternative explanation for these results is that the AS benchmarks overstate the risk of funds with a benchmark discrepancy according to our procedure. However, as we show in Section 8, the AS benchmarks for those funds accurately reflect both traditional and non-traditional factor exposures. In untabulated tests, we also find no evidence that the Sensoy benchmarks of funds with a benchmark discrepancy according to Sensoy's procedure systematically overstate or understate net factor exposure. The difference in the performance results in this section arises from the particular benchmark discrepancies identified by each procedure, not from the accuracy of the alternative benchmarks selected.

We model the difference between the AS benchmark returns and the prospectus benchmark returns as:

$$Return_{AS,t} - Return_{pro,t} = \beta * Factor_t + \varepsilon_t \quad (7)$$

where $Return_{AS,t}$ is the annualized return on the AS benchmark averaged across all tested funds in month t, and $Return_{pro,t}$ is the annualized return on the prospectus benchmark averaged across all tested funds in month t. $Factor_t$ is a vector of factor returns in month t. The base model includes all the factors in the Cremers, Petajisto, and Zitzewitz (2012) seven-factor model.[32] We intentionally exclude a constant from the model because the return differences between the two benchmarks should be explainable by just differences in factor exposures—there should not be any alpha.[33] The model is estimated using all funds with *Benchmark Mismatch* greater than 60% and across various investment style subgroups as implied by the prospectus benchmark.

Table 10 shows the exposures to the CPZ7 factors. In this test, we consider the extent to which the traditional market, size, value, and momentum factors, which are included in the CPZ7 model, explain the average difference in returns between the prospectus and AS benchmarks. As a reference, the first set of rows in the table reports average annualized benchmark-adjusted returns for each group of funds using both benchmarks. This indicates the economic magnitude of the performance boost within each group. The second set of rows then reports estimated coefficients associated with the model in Eq. (7). In the third set of rows, we report the R^2 from the model and the sum of the products of the estimated factor exposures and the annualized factor returns. The

[32] In a few instances, the CPZ7 model explains all the variation in returns between a fund's prospectus and AS benchmarks because the model's index-based factors correspond to those two benchmarks. For example, a fund with an S&P 500 prospectus benchmark and a Russell Midcap AS benchmark will have a difference in returns that is fully explained by the *RMS5* factor. The small number of funds whose two benchmarks correspond to a CPZ7 factor are dropped in these tests.

[33] Our conclusions are the same if a constant is included in the model. Sensoy (2009, page 35) uses the constant from a similar model to test whether "the mismatched benchmarks have incremental explanatory power for the cross-section of expected returns" and finds they do not.

‘Total Factor Return’ row can be compared to the ‘Difference’ row to determine how much of the average difference in the returns between the benchmarks can be explained by differences in traditional factor exposures.

[Table 10 about here]

Using all funds with a benchmark discrepancy, the factors explain 0.57% (*t*-stat = 2.65) of the 1.50% per year difference in average returns between the prospectus and AS benchmarks. The primary differences relate to the two size factors. The coefficient associated with *RMS5* (the difference in returns between the Russell Midcap index and the S&P 500 index) is positive, which indicates the prospectus benchmark has a lower exposure to the factor than the AS benchmark. However, the reverse is true of the coefficient associated with *R2RM* (the difference in returns between the Russell 2000 index and the Russell Midcap index). Overall, only 38% (=0.57%/1.50%) of the average difference in returns between the benchmarks is explained by the traditional factors included in the CPZ7 model.

Looking across the style subgroups, there is substantial variation. For large-cap funds, differences in traditional factor exposures explain 1.60% (*t*-stat = 1.64) of the 2.38% per year difference in returns. Most of the explanatory power comes from the prospectus benchmark having lower small- and mid-cap exposure compared to the AS benchmark. In other words, large-cap funds with a benchmark discrepancy tend to own stocks with lower market capitalization than is implied by their prospectus benchmarks.

In comparison, funds with a small- or mid-cap style have a smaller difference in benchmark returns at 1.12% per year, with differences in traditional factor exposures explaining only 0.41% (*t*-stat = 1.10). The AS benchmarks of small- and mid-cap funds have less small-cap exposure than their prospectus benchmarks imply. Both large-cap and small-/mid-cap funds have AS benchmarks that lean more towards the center of the market capitalization distribution than their prospectus benchmarks, which increases the average difference in returns for large-cap funds and decreases the difference for small-/mid-cap funds. A similar lean towards the center of the value-blend-growth distribution can be seen among growth and value funds (e.g.., the coefficients on *S5VS5G*, *RMVRMG*, and *R2VR2G*). As a result, among value funds with a benchmark discrepancy, there is no statistically significant difference in returns between the prospectus and AS benchmarks.

The CPZ7 model explains a limited portion of the difference in returns between the benchmarks. This indicates that the prospectus and AS benchmarks should differ from each other across dimensions that are unrelated to traditional factors. Hence, we next consider some 'non-traditional' factors. Cochrane (2011, page 1047) remarks that there is now "a zoo of new factors" in the academic literature that purport to explain the cross-section of returns. Rather than attempt to test all potential factors, we consider the explanatory power of a subset of non-traditional factors that have either received a particularly large amount of attention or have been shown to be particularly robust (in, e.g., Feng, Giglio, and Xiu, 2017).

The non-traditional factors we consider are: the Fama and French (2015) profitability (*RMW*) and investment (*CMA*) factors; the Stambaugh and Yuan (2017) management (*MGMT*) and performance (*PERF*) factors; the Frazzini and Pedersen (2014) betting against beta (*BAB*) factor; the Asness, Frazzini, and Pedersen (2019) quality-minus-junk (*QMJ*) factor; and the Pastor and Stambaugh (2003) traded liquidity (*LIQ*) factor. These factors are added as a group to our base CPZ7 model, and the previous analysis is repeated.

Table 11 shows that once these non-traditional factors are included, the difference in benchmark returns that can be explained considerably increases (and is statistically significant for each group of funds considered). The total factor return now captures most of the average difference in benchmark returns. Considering all funds with a benchmark discrepancy, 1.31% (*t*-stat = 4.48) of the 1.50% per year difference in benchmark returns can be explained by the expanded model. After including the non-traditional factors, the R^2 almost doubles from 27.7% to 52.4%.

[Table 11 about here]

The factor that most consistently adds new explanatory power is the profitability factor *RMW*. Across all groups, the prospectus benchmark has a lower *RMW* exposure than the AS benchmark. This result indicates that, on average, funds with a benchmark discrepancy tend to have AS benchmarks that invest more in profitable companies than their prospectus benchmarks. The economic impact of this particular difference in exposure is large. The difference in *RMW* exposure adds 0.57% per year to the total factor return when considering all funds with a benchmark discrepancy. Considering style subgroups, the quality-minus-junk factor *QMJ* and the management factor *MGMT* also have some economically large and statistically significant explanatory power.

These results, as a whole, show that funds with benchmark discrepancies have AS benchmarks with greater average net factor exposure than those same funds' prospectus benchmarks. This difference could occur because the AS benchmarks overstate funds' net exposure, or it could occur because the prospectus benchmarks understate net exposure (or both). To evaluate these possibilities, we note that if a benchmark effectively captures a fund's net factor exposure, then the benchmark-adjusted return should be the same as the estimated alpha from regressing that benchmark-adjusted return against various factors. Conversely, if a benchmark understates (overstates) net exposure, then the benchmark-adjusted return should be greater (less) than the estimated alpha. We consider whether the prospectus benchmarks or AS benchmarks of funds with benchmark discrepancies better reflect net factor exposure using the following model:

$$Return_{fund,t} - Return_{bench,t} = \alpha + \beta * Factor_t + \varepsilon_t \quad (8)$$

where $Return_{fund,t}$ is the average annualized net return across all tested funds in month t and $Return_{bench,t}$ is the average annualized return on those funds' benchmarks in month t. We consider both the prospectus and AS benchmarks in our analysis. $Factor_t$ is a vector of factor returns in month t. Depending on the specification, it includes either no factors, the CPZ7 factors, or the CPZ7 factors and the non-traditional factors. We include a constant in this model because the difference between a fund's return and its benchmark's return should not necessarily be explained by factor exposures alone. The model is estimated using all funds with *Benchmark Mismatch* greater than 60% and for the subgroup of those funds with a large-cap prospectus benchmark style. That subgroup is of particular interest because our tests in Tables 10 and 11 indicate that those funds have the largest difference in average returns between their prospectus and AS benchmarks.

[Table 12 about here]

The results from this model are presented in Table 12. Using all funds with a benchmark discrepancy in Panel A, the prospectus-benchmark-adjusted returns are unaffected by the traditional factors (CPZ7). Adjusting for those factors reduces the abnormal return by only 0.15% per year (t-stat = −0.33). When considering non-traditional factors, however, the prospectus benchmark significantly understates net factor exposure. The abnormal return based on prospectus-benchmark-adjusted returns decreases from 0.65% per year using no factors to −0.29% per year after adding both the traditional and non-traditional factors to the model (CPZ7+). This large change in performance of −0.94% per year is statistically significant (t-stat = −2.37) and indicates that the prospectus benchmark sets a lower bar for the funds than is appropriate given the funds' net factor exposure.

The factors matter less when using the AS-benchmark-adjusted returns. Even after including the non-traditional factors in the model, the abnormal return does not significantly change. The change when switching from no factors to all factors is 0.57% per year, which is not statistically significant at conventional levels (*t*-stat = 1.44). That is, if fund performance is adjusted using the AS benchmark, only relatively minor net factor exposure remains.[34]

The above outcomes are magnified if we focus on just large-cap funds with a benchmark discrepancy.[35] The abnormal return based on prospectus-benchmark-adjusted returns decreases by 1.18% per year (*t*-stat = −1.92) after adjusting for the CPZ7 factors, which shows those funds' prospectus benchmarks significantly understate net exposure to traditional factors. After including the non-traditional factors in the model, that decrease becomes 1.94% per year (*t*-stat = −3.05), indicating that those funds' prospectus benchmarks also understate net exposure to non-traditional factors. In comparison, the AS-benchmark-adjusted returns are unaffected by both the traditional and non-traditional factors. Specifically, among funds with a benchmark discrepancy and a prospectus benchmark implying a large-cap style, adjusting performance using the AS benchmark leaves little net factor exposure, regardless of whether an investor considers non-traditional factors.

In the big picture, funds with benchmark discrepancies perform better relative to their prospectus benchmarks because their prospectus benchmarks tend to understate net factor

[34] While investors may have different opinions on different factors (e.g., anomalies, systematic risk, or data mining), the AS benchmarks—all of which are commonly known and can be tracked at low cost through passive vehicles—account well for these funds' exposures to all tested factors. Hence, no investor should be willing to pay large active management fees to obtain these funds' factor exposure deviations from their prospectus benchmarks. Put another way, managers of funds with benchmark discrepancies are not adding value by providing exposure to otherwise unobtainable factors.

[35] If we repeat the analysis in Panel A with large-cap funds excluded, the results are similar. The abnormal return based on prospectus-benchmark-adjusted returns decreases by 0.76% per year (*t*-stat = −2.13) after adding both the traditional and non-traditional factors to the model.

exposure. Given that, on average, the AS benchmarks of those same funds neither overstate nor understate net factor exposure, using the AS benchmark whenever a fund has a benchmark discrepancy can be expected to result in a more accurate measure of fund performance compared to using the prospectus benchmark.

9. Funds flows and the prospectus benchmark

The importance of benchmark discrepancies depends, in part, on the extent to which investors actually rely on comparing the fund's return to the prospectus benchmark's return when evaluating performance. On the one hand, if investors can identify benchmark discrepancies and ignore those prospectus benchmarks, then the performance boost should have no impact on the competition for capital between funds. On the other hand, if investors cannot identify benchmark discrepancies or fail to fully discount the prospectus benchmark when they do identify benchmark discrepancies, then the competition for capital between funds will be affected.

In this section, we examine net flows to determine how performance relative to the prospectus benchmark affects investors' capital allocations. We model the relation between a fund's net flows and past performance as follows:

$$Flow_{i,t} = \theta + \beta * Performance_{i,t} + \gamma * Mismatch_{i,t} + \delta * Chars_{i,t} + FE + \varepsilon_{i,t} \qquad (9)$$

where $Flow_{i,t}$ is the percentage implied net flow for fund i in month t.[36] $Performance_{i,t}$ is a vector of information about fund i's performance over the year ending at the start of month t. It includes the difference between fund i's return and fund i's AS benchmark return, the difference

[36] The calculation of implied net flows is based on fund returns and assets. It assumes that all inflows and outflows occur at the end of the month. That assumption is obviously incorrect. However, Clifford, Fulkerson, Jordan, and Waldman (2013) find that implied net flows have a correlation of 0.996 with the actual net flows calculated from funds' SEC Form N-SAR filings.

between fund i's AS benchmark return and prospectus benchmark return, and fund i's annualized CAPM alpha.[37] In some instances, we use the actual numeric fund returns. In other instances, the three returns measures are (i) ranked across all funds at the start of each month and (ii) evenly scaled from zero to one based on that ranking (with one being the highest). Using the rankings allows for a more natural test of a potential non-linear relation between fund performance and net flows (shown in, e.g., Chevalier and Ellison, 1997; Sirri and Tufano, 1998; Fant and O'Neal, 2000; and Huang, Wei, and Yan, 2007).

$Mismatch_{i,t}$ is the *Benchmark Mismatch* for fund *i* as of the start of month *t*, and $Chars_{i,t}$ is a vector of characteristics for fund *i* available as of the start of month *t*. It contains the same characteristics as in Eq. (5). FE represents style and year-month fixed effects. The model is estimated using the sample of fund-months with different prospectus and AS benchmarks (i.e., *Benchmark Mismatch* > 0).[38]

Table 13 contains estimates of this model. In the first three columns, we consider whether fund flows depend on fund performance relative to the prospectus benchmark. If such a relation exists, then an increase in the performance boost—the difference in returns between the AS benchmark and the prospectus benchmark—should increase flows. After controlling for performance relative to the AS benchmark, a 1% increase in the performance boost increases flows by 0.07% per month (0.84% annualized, *t*-stat = 14.03). That effect is about half that of a 1% increase in performance relative to the AS benchmark, and thus seems economically meaningful. Controlling for a fund's CAPM alpha lessens, but does not eliminate, the effect of the performance boost on flows. While investors are influenced by multiple measures of performance, these results

[37] Our conclusions are the same if alternative measures of alpha (e.g., Fama-French four-factor or CPZ7) are used or if performance relative to the S&P 500 is used.
[38] Results are similar using the full sample of fund-months.

indicate that performance relative to the prospectus benchmark is an important determinant of investors' capital allocations.[39]

[Table 13 about here]

As the magnitude of *Benchmark Mismatch* increases, we would expect the importance of the prospectus benchmark to decrease. In the fourth column, we interact our performance measures with *Benchmark Mismatch* and find that as *Benchmark Mismatch* increases flows are less sensitive to the performance boost. A 10% increase in *Benchmark Mismatch* reduces the impact on flows of a 1% performance boost by 0.01% per month (t-stat = −5.64). Conversely, increases in *Benchmark Mismatch* have no impact on the weight investors give to performance relative to the AS benchmark. While these results indicate some sophistication on the part of investors, the performance boost still has an economically large and statistically significant impact on flows when *Benchmark Mismatch* is 100%.

The final columns consider non-linearity in the relation between fund flows and performance. We expect benchmark discrepancies to be particularly salient to investors when the performance boost is relatively large. A fund that outperforms its prospectus benchmark by 10% should receive more scrutiny than a fund that outperforms its prospectus benchmark by 1%. As indicated by the squared performance terms, flows have a convex relation with AS-benchmark-adjusted returns but a concave relation with the performance boost, which confirms our hypothesis. Further, consistent with Clifford, Jordan, and Riley (2014), the flow-performance relation among the largest funds (i.e., the top 20% by total assets) is linear with respect to AS-benchmark-adjusted

[39] While Del Guercio and Reuter (2014) and Barber, Huang, and Odean (2016) both suggest that the flows of more sophisticated investors are less responsive to simple measures of performance, we find similar results regardless of a fund's level of institutional ownership or likelihood to be broker-sold.

returns; however, regardless of fund size, the flow-performance relation is concave with respect to the performance boost.

10. Conclusion

Risk-adjustment is central to performance evaluation. To facilitate that process, SEC regulations require that mutual funds provide a benchmark to investors in the fund prospectus. Given that funds rarely change their prospectus benchmark and market themselves in a competitive environment to investors that often have limited sophistication, we expect that some funds will act strategically when constructing their portfolios. Building on Sensoy (2009), we find that while most funds appear to have risk-appropriate prospectus benchmarks, a substantial portion of prospectus benchmarks understate risk and, consequently, overstate relative performance. Funds benefit from that performance overstatement because investors' flows respond to performance relative to the prospectus benchmark even when a benchmark discrepancy exists. In general, researchers and investors should exercise significant caution when using prospectus benchmarks to evaluate fund performance.

Our results contribute to several related topics in the literature. First, they suggest researchers should be careful when choosing benchmarks for performance evaluation. Benchmark-adjusted returns are common in studies of mutual funds (e.g., Angelidis, Giamouridis, and Tessaromatis, 2013; Berk and van Binsbergen, 2015; Cremers, Ferreira, Matos, and Starks, 2016; and Pastor, Stambaugh, and Taylor, 2017), and if studies use prospectus benchmarks, there could be substantial noise and biases in the results. In the majority of current studies, researchers using benchmark-adjusted returns assign their own benchmark or rely on benchmark providers such as Morningstar, so we do not expect the current bias in the literature to be large.[40]

Second, despite the fact that they often fail to match the fund's actual investment strategy, our paper demonstrates the importance of prospectus benchmarks for funds and investors. Barber, Huang, and Odean (2016) and Berk and van Binsbergen (2016) both indicate that performance relative to the Sharpe (1964) and Lintner (1965) capital asset pricing model (CAPM) best explains the flow-performance relation, but neither study considers benchmark-adjusted returns (prospectus or AS). We show that the impact on flows of benchmark-adjusted returns in general, and prospectus-benchmark-adjusted returns specifically, are economically meaningful even after accounting for CAPM alpha.

Third, our fund flow results add to a growing number of studies that show how fund investors respond to information of questionable economic value. Cooper, Gulen, and Rau (2005) find that funds who change their name to align with popular investment styles receive larger flows, regardless of whether the name change reflects a change in the fund's portfolio. Kaniel and Parham (2017) demonstrate that funds just making the cut-off to qualify for lists in the *Wall Street Journal* receive larger flows in periods when those lists are labeled 'Category Kings.' Hartzmark and Solomon (2018) show that fund flows are heavily influenced by price-only relative performance—with investors overlooking the impact of dividends—while Harris, Hartzmark, and Solomon (2015) show that funds who artificially increase their dividends attract larger flows, despite such behavior increasing turnover and taxes.[41] In a similar fashion, we show that a fund overstating its

[40] Angelidis, Giamouridis, and Tessaromatis (2013) conclude that skill can be misstated when funds' self-declared benchmarks are ignored; however, they use inferred self-declared benchmarks, rather than actual prospectus benchmarks, in their analysis. Hence, their conclusions are more about the value of accounting for self-declared style than the biases of prospectus benchmarks.

[41] Other examples of investors responding to information of questionable economic value include Jain and Wu (2000), Solomon, Soltes, and Sosyura (2014), and Niessen-Ruenzi and Ruenzi (2018).

performance by using an inaccurate prospectus benchmark will receive larger flows even when the magnitude of the inaccuracy is economically significant.

Fourth, our comparison of different benchmark-adjusted returns adds to the debate on the appropriate factor structure for evaluating fund performance. It is common to control for market risk and the size and value factors (e.g., the Fama and French, 1993, model and the Cremers, Petajisto, and Zitzewitz, 2012, model), but Harvey, Liu, and Zhu (2016) and Hou, Xue, and Zhang (2019) find that there are hundreds of apparent pricing anomalies that could be used to form pricing factors. Whether the use of any or all of those additional factors is appropriate in the evaluation of mutual fund performance remains unclear. If these non-traditional factors are not directly investable, then providing exposure to them through active management could represent a value-added activity. Our results make it evident that mutual funds often have exposures to non-traditional factors that (i) are not indicated by their prospectus benchmarks and (ii) impact performance. However, funds and their AS benchmarks, which are all directly investable at low cost, generally have similar exposures to the non-traditional factors; therefore, the additional performance arising from funds' factor exposure deviations relative to their prospectus benchmarks should not represent alpha from an investor's prospective.

Finally, our results contribute to the debate on mutual fund manager skill. On the one hand, studies including Carhart (1997), French (2008), and Fama and French (2010) find little evidence of skill. On the other hand, studies such as Kosowski, Timmermann, Wermers, and White (2006), Barras, Scaillet, and Wermers (2010), and Berk and van Binsbergen (2015) find material evidence of skill. Complicating the matter are recent works raising concerns related to methodology (Andrikogiannopoulou and Papakonstantinou, 2019; Barras, Scaillet, and Wermers, 2019; and Harvey and Liu, 2019) and out-of-sample replicability (Jones and Mo, 2019, and Riley, 2019).

Consistent with the logic of Kacperczyk, Sialm, and Zheng (2008), Amihud and Goyenko (2013), and Hoberg, Kumar, and Prabhala (2018), our results support the existence of meaningful investment skill for at least some funds. Like Cremers and Petajisto (2009), we show that funds with high active share outperform, but like Petajisto (2013) and Cremers and Pareek (2016), we also show that high active share funds' outperformance is concentrated within a subgroup. In our case, high active share funds without a benchmark discrepancy.

Thinking more broadly, the motivation for fund sponsors and fund managers to allow persistent benchmark discrepancies is clear. Fund sponsors' profits increase with assets under management, and Ma, Tang, and Gomez (2019) show that 79% of fund managers have pay directly linked to fund performance. It is less clear why investors accept persistent benchmark discrepancies. Perhaps the broadness of the SEC's benchmark requirement ("appropriate broad-based securities market index") makes complaints difficult. Or, perhaps the reliance on traditional size and value/growth categories makes it difficult for investors to identify benchmark discrepancies when so many are driven by non-traditional factors. Regardless of the reason, since investors appear to rely heavily on the benchmark in the prospectus, benchmark discrepancies should be a primary concern. The wider dissemination of manipulation-proof measures, such as that of Goetzmann, Ingersoll, Spiegel, and Welch (2007), may help alleviate this issue.

References

Agarwal, Vikas, Clifton Green, and Honglin Ren. 2018. Alpha or beta in the eye of the beholder: What drives hedge fund flows?. *Journal of Financial Economics* 127, 417-434.

Almazan, Andres, Keith Brown, Murray Carlson, and David Chapman. 2004. Why constrain your mutual fund manager?. *Journal of Financial Economics* 73, 289-321.

Amihud, Yakov, and Ruslan Goyenko. 2013. Mutual fund's R^2 as predictor of performance. *Review of Financial Studies* 26, 667-694.

Andrikogiannopoulou, Angie, and Filippos Papakonstantinou. 2019. Reassessing false discoveries in mutual fund performance: Skill, luck, or lack of power?. *Journal of Finance*, forthcoming.

Angelidis, Timotheos, Daniel Giamouridis, and Nikolaos Tessaromatis. 2013. Revisiting mutual fund performance evaluation. *Journal of Banking & Finance* 37, 1759-1776.

Asness, Clifford, Andrea Frazzini, and Lasse Pedersen. 2019. Quality minus junk. *Review of Accounting Studies* 24, 34-112.

Bams, Dennis, Roger Otten, and Ehsan Ramezanifar. 2017. Investment style misclassification and mutual fund performance. Working paper.

Barber, Brad, Xing Huang, and Terrance Odean. 2016. Which factors matter to investors? Evidence from mutual fund flows. *Review of Financial Studies* 29, 2600-2642.

Barras, Laurent, Olivier Scaillet, and Russ Wermers. 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *Journal of Finance* 65, 179-216.

Barras, Laurent, Olivier Scaillet, and Russ Wermers. 2019. Reassessing false discoveries in mutual fund performance: Skill, luck, or lack of power? A reply. Working paper.

Basak, Suleyman, and Dmitry Makarov. 2014. Strategic asset allocation in money management. *Journal of Finance* 69, 179-217.

Basak, Suleyman, Anna Pavlova, and Alexander Shapiro. 2007. Optimal asset allocation and risk shifting in money management. *Review of Financial Studies* 20, 1583-1621.

Ben-David, Itzhak, Jiacui Li, Andrea Rossi, and Yang Song. 2019. What do mutual fund investors really care about?. Working paper.

Berk, Jonathan, and Jules van Binsbergen. 2015. Measuring skill in the mutual fund industry. *Journal of Financial Economics* 118, 1-20.

Berk, Jonathan, and Jules van Binsbergen. 2016. Assessing asset pricing models using revealed preference. *Journal of Financial Economics* 119, 1-23.

Bertsch, Fred, and Thomas Idzorek. 2004. The style drift score. *Journal of Portfolio Management* 31, 76-83.

Bhojraj, Sanjeev, Young Jun Cho, and Nir Yehuda. 2012. Mutual fund family size and mutual fund performance: The role of regulatory changes. *Journal of Accounting Research* 50, 647-684.

Blocher, Jesse, and Marat Molyboga. 2017. The revealed preference of sophisticated investors. *European Financial Management* 23, 839-872.

Bollen, Nicolas, and Jeffrey Busse. 2001. On the timing ability of mutual fund managers. *Journal of Finance* 56, 1075-1094.

Brown, Keith, W.V. Harlow, and Hanjiang Zhang. 2009. Staying the course: The role of investment style consistency in the performance of mutual funds. Working paper.

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

Brown, Stephen, and William Goetzmann. 1997. Mutual fund styles. *Journal of Financial Economics* 43, 373-399.

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

Cao, Charles, Peter Iliev, and Raisi Velthuis. 2017. Style drift: Evidence from small-cap mutual funds. *Journal of Banking & Finance* 78, 42-57.

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

Chakraborty, Indraneel, Alok Kumar, Tobias Muhlhofer, and Ravi Sastry. 2018. If you show it, they will use it: The "CAPM heuristic" and mutual fund allocation decisions. Working paper.

Chan, Louis, Hsiu-Lang Chen, and Josef Lakonishok. 2002. On mutual fund investment styles. *Review of Financial Studies* 15, 1407-1437.

Chen, Joseph, Eric Hughson, and Neal Stoughton. 2018. Strategic mutual fund tournaments. Working paper.

Chen, Yong, Wayne Ferson, and Helen Peters. 2010. Measuring the timing ability and performance of bond mutual funds. *Journal of Financial Economics* 98, 72-89.

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

Clifford, Christopher, Bradford Jordan, and Timothy Riley. 2014. Average fund versus average dollars: Implications for mutual fund research. *Journal of Empirical Finance* 28, 249-260.

Clifford, Christopher, Jon Fulkerson, Bradford Jordan, and Steve Waldman. 2013. Risk and fund flows. Working paper.

Clifford, Christopher, Jon Fulkerson, Xin Hong, and Bradford Jordan. 2014. Investment restrictions and fund performance. Working paper.

Cochrane, John. 2011. Presidential address: Discount rates. *Journal of Finance* 66, 1047-1108.

Cooper, Michael, Huseyin Gulen, and Raghavendra Rau. 2005. Changing names with style: Mutual fund name changes and their effects on fund flows. *Journal of Finance* 60, 2825-2858.

Cremers, Martijn, and Ankur Pareek. 2016. Patient capital outperformance: The investment skill of high active share managers who trade infrequently. *Journal of Financial Economics* 122, 288-306.

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

Cremers, Martijn, and Quinn Curtis. 2016. Do mutual fund investors get what they pay for? The legal consequences of closet index funds. *Virginia Law & Business Review* 11, 31-93.

Cremers, Martijn, Antti Petajisto, and Eric Zitzewitz. 2012. Should benchmark indices have alpha? Revisiting performance evaluation. *Critical Finance Review* 2, 1-48.

Cremers, Martijn, Jon Fulkerson, and Timothy Riley. 2019. Challenging the conventional wisdom on active management: A review of the past 20 years of academic literature on actively managed mutual funds. *Financial Analysts Journal*, forthcoming.

Cremers, Martijn, Miguel Ferreira, Pedro Matos, and Laura Starks. 2016. Indexing and active fund management: International evidence. *Journal of Financial Economics* 120, 539-560.

Cremers, Martijn. 2017. Active share and the three pillars of active management: Skill, conviction, and opportunity. *Financial Analysts Journal* 73, 61-79.

Daniel, Kent, Mark Grinblatt, Sheridan Titman, and Russ Wermers. 1997. Measuring mutual fund performance with characteristic-based benchmarks. *Journal of Finance* 52, 1035-1058.

Del Guercio, Diane, and Jonathan Reuter. 2014. Mutual fund performance and the incentive to generate alpha. *Journal of Finance* 69, 1673-1704.

Del Guercio, Diane, and Paula Tkac. 2008. Star power: The effect of Morningstar ratings on mutual fund flow. *Journal of Financial and Quantitative Analysis* 43, 907-936.

DeWoody, Madeline. 2019. A tale of two benchmarks: Performance benchmarking in the mutual fund industry. Working paper.

DiBartolomeo, Dan, and Erik Witkowski. 1997. Mutual fund misclassification: Evidence based on style analysis. *Financial Analysts Journal* 53, 32-43.

Elton, Edwin, Martin Gruber, and Christopher Blake. 2003. Incentive fees and mutual funds. *Journal of Finance* 58, 779-804.

Elton, Edwin, Martin Gruber, and Christopher Blake. 2012. An examination of mutual fund timing ability using monthly holdings data. *Review of Finance* 16, 619-645.

Elton, Edwin, Martin Gruber, and Christopher Blake. 2014. The performance of separate accounts and collective investment trusts. *Review of Finance* 18, 1717-1742.

Elton, Edwin, Martin Gruber, Christopher Blake, Yoel Krasny, and Sadi Ozelge. 2010. The effect of holdings data frequency on conclusions about mutual fund behavior. *Journal of Banking & Finance* 34, 912-922.

Evans, Richard, Miguel Ferreira, and Melissa Porras Prado. 2017. Fund performance and equity lending: Why lend what you can sell?. *Review of Finance* 21, 1093-1121.

Evans, Richard, and Yang Sun. 2018. Models or stars: The role of asset pricing models and heuristics in investor risk adjustment. Working paper.

Evans, Richard. 2010. Mutual fund incubation. *Journal of Finance* 65, 1581-1611.

Fama, Eugene, and Kenneth French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33, 3-56.

Fama, Eugene, and Kenneth French. 2010. Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance* 65, 1915-1947.

Fama, Eugene, and Kenneth French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116, 1-22.

Fant, Franklin, and Edward O'Neal. 2000. Temporal changes in the determinants of mutual fund flows. *Journal of Financial Research* 23, 353-371.

Feng, Guanhao, Stefano Giglio, and Dacheng Xiu. 2017. Taming the factor zoo. Working paper.

Frazzini, Andrea, and Lasse Pedersen. 2014. Betting against beta. *Journal of Financial Economics* 111, 1-25.

French, Kenneth. 2008. Presidential address: The cost of active investing. *Journal of Finance* 63, 1537-1573.

Goetzmann, William, Jonathan Ingersoll, Matthew Spiegel, and Ivo Welch. 2007. Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies* 20, 1503-1546.

Harris, Lawrence, Samuel Hartzmark, and David Solomon. 2015. Juicing the dividend yield: Mutual funds and the demand for dividends. *Journal of Financial Economics* 116, 433-451.

Hartzmark, Samuel, and David Solomon. 2018. Reconsidering returns. Working paper.

Harvey, Campbell, and Yan Liu. 2019. False (and missed) discoveries in financial economics. Working paper.

Harvey, Campbell, Yan Liu, and Heqing Zhu. 2016. …And the cross-section of expected returns. *Review of Financial Studies* 29, 5-68.

Hirt, Joshua, Ravi Tolani, and Christopher Philips. 2015. Global equity benchmarks: Are prospectus benchmarks the correct barometer?. Vanguard Research.

Hoberg, Gerard, Nitin Kumar, and Nagpurnanand Prabhala. 2018. Mutual fund competition, managerial skill, and alpha persistence. *Review of Financial Studies* 31, 1896-1929.

Hou, Kewei, Chen Xue, and Lu Zhang. 2019. Replicating anomalies. *Review of Financial Studies*, forthcoming.

Huang, Jennifer, Clemens Sialm, and Hanjiang Zhang. 2011. Risk shifting and mutual fund performance. *Review of Financial Studies* 24, 2575-2616.

Huang, Jennifer, Kelsey Wei, and Hong Yan. 2007. Participation costs and the sensitivity of fund flows to past performance. *Journal of Finance* 62, 1273-1311.

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

Jiang, George, Tong Yao, and Tong Yu. 2007. Do mutual funds time the market?. Evidence from portfolio holdings. *Journal of Financial Economics* 86, 724-758.

Jones, Christopher, and Haitao Mo. 2019. Out-of-sample performance of mutual fund predictors. *Review of Financial Studies*, forthcoming.

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng. 2008. Unobserved actions of mutual funds. *Review of Financial Studies* 21, 2379-2416.

Kacperczyk, Marcin, Stijn van Nieuwerburgh, and Laura Veldkamp. 2014. Time-varying fund manager skill. *Journal of Finance* 69, 1455-1484.

Kaniel, Ron, and Robert Parham. 2017. WSJ Category Kings – The impact of media attention on consumer and mutual fund investment decisions. *Journal of Financial Economics* 123, 337-356.

Kaplan, Steven, and Berk Sensoy. 2005. Do mutual funds time their benchmarks?. Working paper.

Kempf, Alexander, Stefan Ruenzi, and Tanja Thiele. 2009. Employment risk, compensation incentives, and managerial risk taking: Evidence from the mutual fund industry. *Journal of Financial Economics* 92, 92-108.

Kim, Moon, Ravi Shukla, and Michael Tomas. 2000. Mutual fund objective misclassification. *Journal of Economics and Business* 52, 309-323.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White. 2006. Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *Journal of Finance* 61, 2551-2596.

Lee, Jung, Charles Trzcinka, and Shyam Venkatesan. 2019. Do portfolio manager contracts contract portfolio management?. *Journal of Finance,* forthcoming.

Lintner, John. 1965. The valuation of risk assets and the selection of risk investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47, 13-37.

Ma, Linlin, Yuehua Tang, and Juan-Pedro Gomez. 2019. Portfolio manager compensation in the U.S. mutual fund industry. *Journal of Finance* 74, 587-638.

Mamaysky, Harry, Matthew Spiegel, and Hong Zhang. 2008. Estimating the dynamics of mutual fund alphas and betas. *Review of Financial Studies* 21, 233-264.

Mateus, Irina, Cesario Mateus, and Natasa Todorovic. 2017. The impact of benchmark choice on US mutual fund benchmark-adjusted performance and ranking. Working paper.

Morningstar. 2004. Fact Sheet: The Morningstar Style Box. White paper available at https://corporate.morningstar.com/US/documents/MethodologyDocuments/FactSheets/MorningstarStyleBox_FactSheet_.pdf

Niessen-Ruenzi, Alexandra, and Stefan Ruenzi. 2018. Sex matters: Gender bias in the mutual fund industry. *Management Science,* forthcoming.

Pastor, Lubos, and Robert Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111, 642-685.

Pastor, Lubos, Robert Stambaugh, and Lucian Taylor. 2017. Do funds make more when they trade more?. *Journal of Finance* 72, 1483-1528.

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

Pollet, Joshua, and Mungo Wilson. 2008. How does size affect mutual fund behavior?. *Journal of Finance* 63, 2941-2969.

Qiu, Jiaping. 2003. Termination risk, multiple managers, and mutual fund tournaments. *Review of Finance* 7, 161-190.

Riley, Timothy. 2019. Can mutual fund stars still pick stocks?: A replication and extension of Kosowski, Timmermann, Wermers, and White (2006). *Critical Finance Review*, forthcoming.

Sato, Yuki. 2016. Fund tournaments and asset bubbles. *Review of Finance* 20, 1383-1426.

Schwarz, Christopher, and Mark Potter. 2016. Revisiting mutual fund portfolio disclosure. *Review of Financial Studies* 29, 3519-3544.

Schwarz, Christopher. 2012. Mutual fund tournaments: The sorting bias and new evidence. *Review of Financial Studies* 25, 913-936.

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

Sharpe, William. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19, 425-442.

Sharpe, William. 1992. Asset allocation: Management style and performance measurement. *Journal of Portfolio Management* 18, 7-19.

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

Solomon, David, Eugene Soltes, and Denis Sosyura. 2014. Winners in the spotlight: Media coverage of fund holdings as a driver of flows. *Journal of Financial Economics* 113, 53-72.

Stambaugh, Robert, and Yu Yuan. 2017. Mispricing factors. *Review of Financial Studies* 30, 1270-1315.

Wermers, Russ. 2012. Matter of style: The causes and consequences of style drift in institutional portfolios. Working paper.

Figure 1: Percentage of funds with different prospectus and AS benchmarks by quarter

This figure shows (i) the percentage of funds in each quarter with different prospectus and AS benchmarks and (ii) the percentage of funds in each quarter with different prospectus and AS benchmarks and *Benchmark Mismatch* greater than 60%.

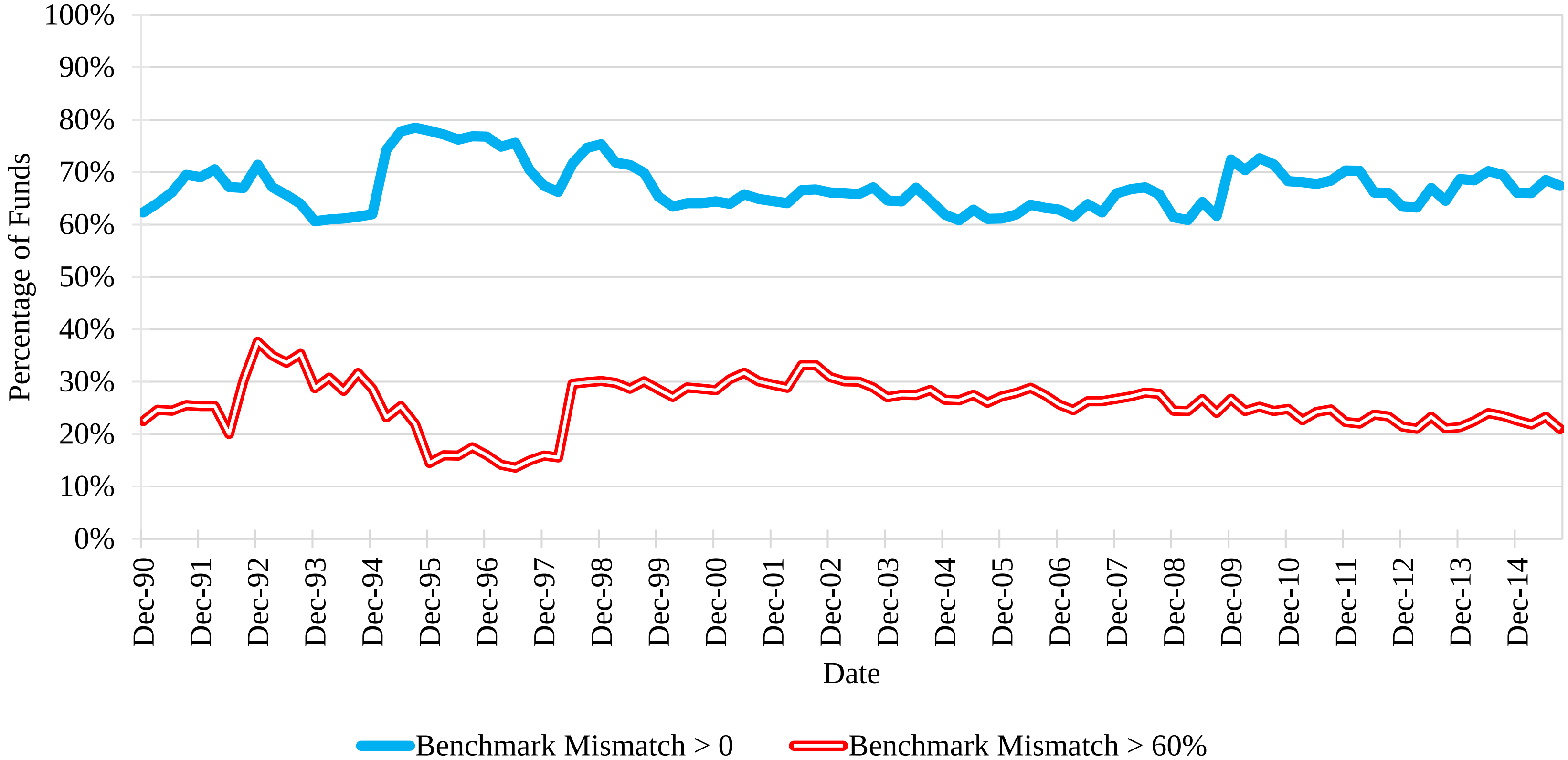

Figure 2: Distribution of *Active Gap*

This figure shows a cumulative density function (CDF) of *Active Gap* for all fund-month observations where the prospectus benchmark does not match the AS benchmark.

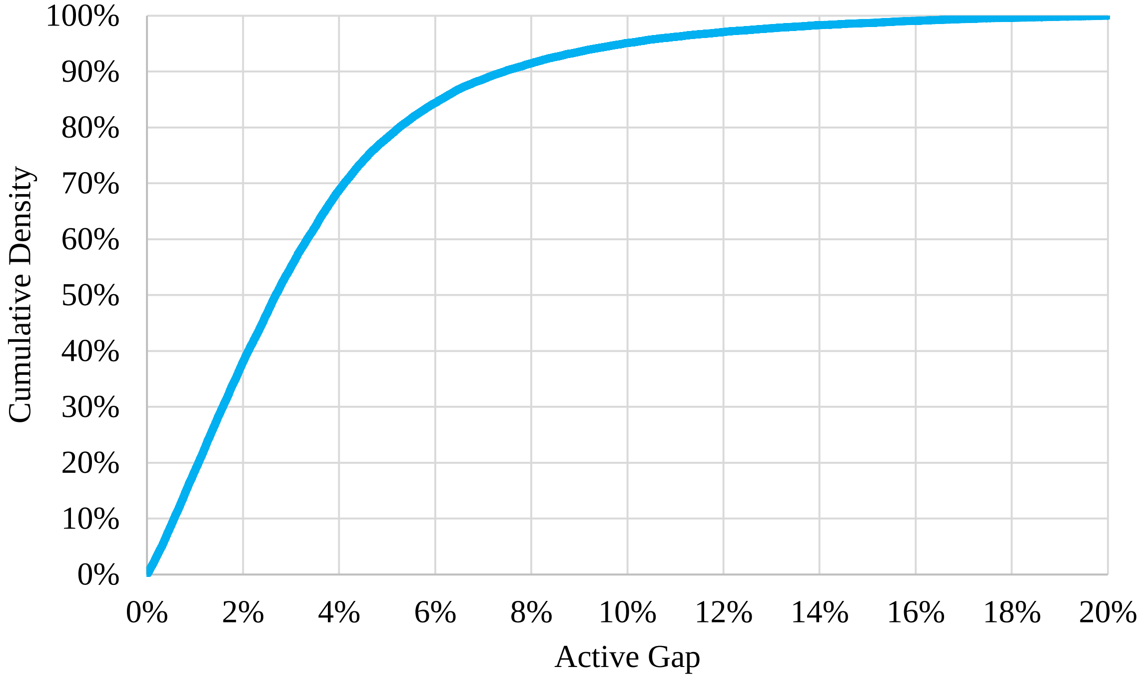

Figure 3: Distribution of *Benchmark Mismatch*

This figure shows a cumulative density function (CDF) of *Benchmark Mismatch* for all fund-month observations where the prospectus benchmark does not match the AS benchmark.

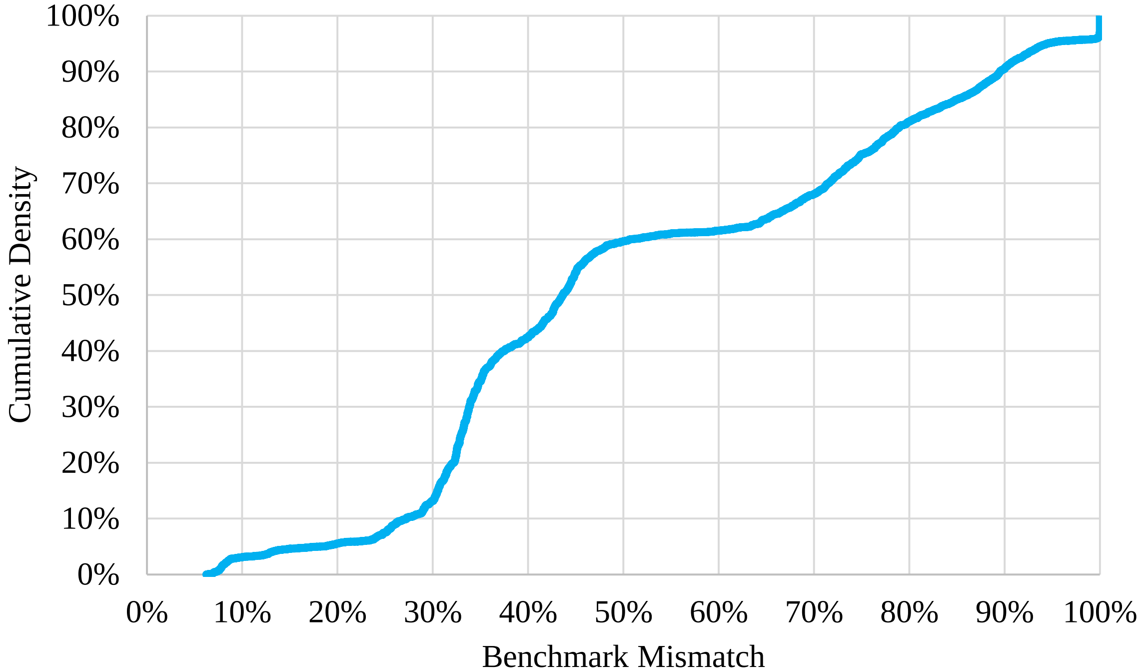

Figure 4: Overlap between Sensoy and *Benchmark Mismatch* (BM) identification procedures

This figure shows the percentage of the full sample of fund-months identified as having a benchmark discrepancy using two different procedures. In the BM procedure, a fund is considered to have a benchmark discrepancy if *Benchmark Mismatch* is greater than 60%. In the Sensoy procedure, a fund is considered to have a benchmark discrepancy if Morningstar style boxes and fund-benchmark correlations indicate a more appropriate benchmark. If the figure legend reads 'BM = Yes', then there is a benchmark discrepancy using the BM procedure. If the figure legend reads 'Sensoy = Yes', then there is a benchmark discrepancy using the Sensoy procedure. When both procedures identify a benchmark discrepancy, the alternative benchmark identified as the more appropriate benchmark is sometimes the same across procedures and sometimes different. We report those two groups separately.

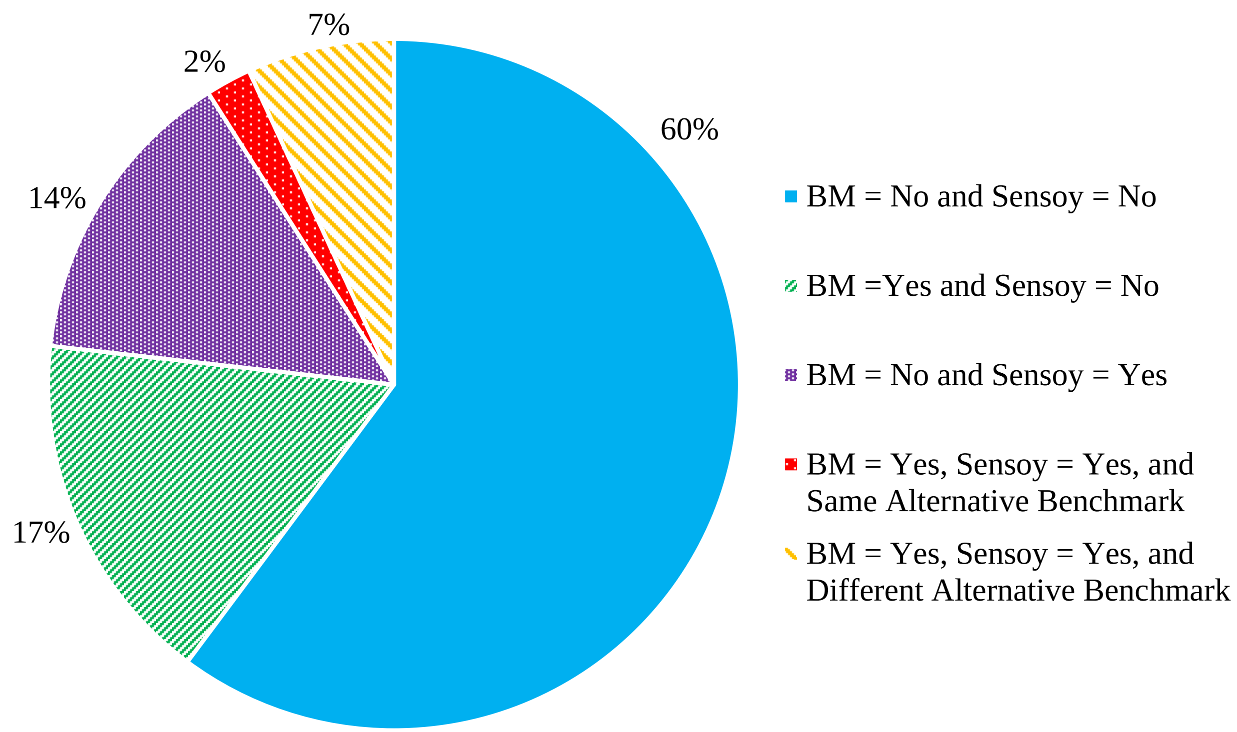

Table 1: Full sample summary statistics

This table shows full sample summary statistics. Any Mismatch is dummy variable equal to one if the prospectus and AS benchmarks are different. Large Mismatch is a dummy variable equal to one if the prospectus and AS benchmarks are different and *Benchmark Mismatch* is greater than 60%. Prospectus Active Share is the active share of the fund relative to the benchmark listed in the fund's prospectus. Minimum Active Share is the lowest active share of the fund across all tested benchmarks. *Benchmark Mismatch* is the active share of the fund's prospectus benchmark relative to its AS benchmark. *Active Gap* is the difference between the fund's prospectus active share and minimum active share. Prospectus-Adjusted Return is the fund's annualized net return less the annualized return on the fund's prospectus benchmark. AS-Adjusted Return is the fund's annualized net return less the annualized return on the fund's AS benchmark. P25, P50, and P75 are the 25th, 50th, and 75th percentiles, respectively. $\rho_{t,t-12}$ is the correlation between the fund's value in month t and month $t - 12$.

	Mean	Standard Deviation	P25	P50	P75	$\rho_{t,t-12}$
Any Mismatch	0.67	0.47	0.00	1.00	1.00	0.59
Large Mismatch	0.26	0.43	0.00	0.00	1.00	0.75
Prospectus Active Share	80.7%	14.1%	71.1%	83.5%	92.5%	0.93
Minimum Active Share	78.4%	13.8%	68.9%	81.0%	89.7%	0.92
Benchmark Mismatch	34.9%	32.1%	0.0%	33.0%	63.5%	0.72
Active Gap	2.3%	3.1%	0.0%	1.3%	3.5%	0.75
Prospectus-Adjusted Return	-0.33%	18.90%	-10.50%	-0.57%	9.52%	0.02
AS-Adjusted Return	-0.78%	18.48%	-10.99%	-0.88%	9.20%	0.01

Table 2: Most common differences between the prospectus and AS benchmarks

This table shows the five most common differences between the prospectus benchmark and the AS benchmark. Panel A shows the most common benchmark differences for all fund-months with *Benchmark Mismatch* greater than zero and less than or equal to 60%. Panel B shows the most common benchmark differences for all fund-months with *Benchmark Mismatch* greater than 60%. For each benchmark difference listed, the percentage of that sample with that benchmark difference is reported. The median *Active Gap* and the mean *Benchmark Mismatch* are also provided for each benchmark difference.

Panel A: 0 < Benchmark Mismatch ≤ 60%

Prospectus Benchmark	AS Benchmark	Percentage of Differences	Active Gap	Benchmark Mismatch
S&P 500	S&P 500 Growth	19.5%	3.1%	33.0%
Russell 1000 Growth	S&P 500 Growth	14.4%	1.8%	30.2%
Russell 1000 Value	S&P 500 Value	9.6%	2.1%	32.7%
S&P 500	Russell 1000 Growth	8.0%	3.4%	43.4%
S&P 500	S&P 500 Value	7.3%	1.9%	35.8%

Panel B: Benchmark Mismatch > 60%

Prospectus Benchmark	AS Benchmark	Percentage of Differences	Active Gap	Benchmark Mismatch
Russell 2000	S&P 600 Growth	11.6%	3.9%	77.1%
Russell 2000 Value	S&P 600 Value	9.6%	2.0%	68.6%
Russell 2000 Growth	S&P 600 Growth	9.0%	1.7%	69.0%
Russell 2000	S&P 600 Value	5.7%	2.6%	75.6%
S&P 500	Russell Midcap Growth	4.4%	6.7%	90.3%

Table 3: Characteristics of funds conditional on *Benchmark Mismatch*

This table compares the characteristics of funds with *Benchmark Mismatch* (BM) greater than 60% to funds with *Benchmark Mismatch* less than or equal to 60% (including funds with *Benchmark Mismatch* of zero). Panel A reports basic fund characteristics. Prospectus Active Share is the active share of the fund relative to the benchmark listed in the fund's prospectus. Assets is the net assets of the fund reported in billions of dollars. Age is the age of the oldest share class of the fund and is reported in years. Expense Ratio and Turnover Ratio are the annual expense and turnover ratios as reported by the fund. Number of Holdings is the number of equity positions held by the fund. Institutional is the percentage of the fund's net assets held within institutional share classes. Panel B reports the percentage of funds within each group that have a particular style. The styles are determined based on the prospectus benchmark. Each fund is identified as either large-cap or small-/mid-cap and either growth, blend, or value. The *t*-statistics for the differences between the groups are calculated using standard errors clustered by fund and year-month.

Panel A: Fund Characteristics

	Full Sample	BM > 60%	BM ≤ 60%	Difference	*t*-stat
Prospectus Active Share	80.7%	93.5%	76.8%	16.7%	40.84
Assets (billions of $)	1.16	0.88	1.26	-0.38	-3.71
Age (years)	16.1	13.8	16.9	-3.0	-5.14
Expense Ratio	1.20%	1.29%	1.16%	0.13%	8.91
Turnover Ratio	77.2%	77.4%	77.1%	0.3%	0.11
Number of Holdings	89.4	92.3	88.3	4.0	1.31
Institutional (% of assets)	26.1%	23.5%	27.0%	-3.5%	-2.08

Panel B: Prospectus Benchmark Style

	Full Sample	BM > 60%	BM ≤ 60%	Difference	*t*-stat
Large Cap	61.8%	19.6%	76.3%	-56.7%	-27.87
Small/Mid Cap	38.2%	80.4%	23.7%	56.7%	27.87
Growth	29.7%	24.4%	31.5%	-7.1%	-3.25
Blend	46.7%	49.6%	45.7%	3.9%	1.45
Value	23.6%	26.0%	22.8%	3.1%	1.34

Table 4: Probability of having a benchmark discrepancy

This table shows estimates from the following logit model:

$$BM > 60\%_{i,t} = \alpha + \beta * Active\ Share_{i,t} + \delta * Chars_{i,t} + \gamma * Style_i + FE + \varepsilon_{i,t}$$

where $BM > 60\%_{i,t}$ is a dummy variable equal to one if *Benchmark Mismatch* for fund *i* in quarter *t* is greater than 60%. $Active\ Share_{i,t}$ is a vector of information about fund i's quarter *t* active share. It includes active share relative to the prospectus benchmark and a dummy variable equal to one if the prospectus benchmark active share is among the top 20% in the quarter. $Chars_{i,t}$ is a vector of characteristics for fund i available as of quarter *t*. It includes the natural log of assets, natural log of age, expense ratio, turnover ratio, the number of equity positions, and the percentage of fund assets held within institutional share classes. $Style_i$ is a vector of information about fund i's style based on its prospectus benchmark. It includes a large-cap dummy, a blend dummy, and a growth dummy. FE represents year-quarter fixed effects. The model is estimated using the full sample (including fund-quarters with *Benchmark Mismatch* equal to zero). *t*-statistics are reported in brackets below each coefficient and are calculated using standard errors clustered by fund and year-quarter.

	(1)	(2)	(3)	(4)	(5)	(6)
Prospectus Active Share	0.25	0.25			0.29	0.31
	[25.69]	[22.70]			[15.71]	[15.73]
Top 20% AS Dummy		0.07			0.27	0.25
		[0.63]			[2.21]	[1.91]
Assets			-0.03		0.05	0.04
			[-0.81]		[1.33]	[1.12]
Age			-0.24		0.02	0.06
			[-3.54]		[0.22]	[0.63]
Expense Ratio			1.09		-0.17	-0.34
			[7.11]		[-0.98]	[-1.93]
Turnover Ratio			-0.00		0.00	0.00
			[-2.11]		[0.64]	[0.05]
Number of Holdings			0.00		0.01	0.01
			[3.92]		[7.99]	[8.19]
Institutional Ownership			-0.00		0.00	-0.00
			[-0.60]		[0.07]	[-0.28]
Large Cap Dummy				-2.88	-1.02	-0.94
				[-21.31]	[-7.77]	[-6.72]
Blend Dummy				0.40	-0.48	-0.52
				[2.84]	[-3.12]	[-3.28]
Growth Dummy				-0.81	-0.30	-0.25
				[-5.43]	[-1.87]	[-1.53]
Fixed Effects	No	No	No	No	No	Yes
Observations	53,316	53,316	53,316	53,316	53,316	53,316

Table 5: Difference in benchmark returns as a function of *Active Gap* and *Benchmark Mismatch*

This table shows the average difference in annualized return between the AS benchmark and the prospectus benchmark for fund-months in which those benchmarks are different. Fund-months are sorted unconditionally on *Active Gap* (AG) and *Benchmark Mismatch* (BM) based on pre-set cut-offs, and the average difference is reported for each of the resulting groups. Panel A sorts funds into five groups based on *Benchmark Mismatch*, and Panel B sorts funds into two groups based on *Benchmark Mismatch*. The 'High – Low' column reports the difference in the results between the '0 < BM ≤ 20' and 'BM > 80' groups in Panel A and difference in results between the 'BM ≤ 60' and 'BM > 60' groups in Panel B. In both panels, the 'High – Low' row reports the difference in results between the '0 < AG ≤ 1.25' and 'AG > 5' groups. *t*-statistics are reported in brackets below each coefficient and are calculated using standard errors clustered by fund and year-month.

Panel A: Five Ranges for Benchmark Mismatch

	All	0 < BM ≤ 20	20 < BM ≤ 40	40 < BM ≤ 60	60 < BM ≤ 80	BM > 80	High − Low
All	0.68%	-0.12%	0.04%	0.53%	1.37%	1.64%	1.75%
	[2.72]	[-0.75]	[0.18]	[1.70]	[2.25]	[2.97]	[3.04]
0 < AG ≤ 1.25	0.43%	0.02%	0.15%	0.24%	1.00%	1.26%	1.24%
	[2.15]	[0.12]	[0.71]	[0.76]	[1.64]	[1.99]	[2.05]
1.25 < AG ≤ 2.5	0.68%	-0.03%	-0.12%	0.84%	1.61%	1.87%	1.90%
	[2.81]	[-0.18]	[-0.47]	[2.19]	[2.52]	[2.84]	[2.92]
2.5 < AG ≤ 3.75	0.75%	-0.29%	-0.03%	0.71%	1.62%	1.67%	1.96%
	[2.53]	[-1.40]	[-0.09]	[1.56]	[2.33]	[2.57]	[2.85]
3.75 < AG ≤ 5	0.73%	-0.45%	0.07%	0.75%	1.12%	1.69%	2.14%
	[2.19]	[-1.60]	[0.19]	[1.68]	[1.58]	[2.26]	[2.62]
AG > 5	0.84%	-0.75%	0.18%	0.26%	1.45%	1.62%	2.37%
	[2.31]	[-1.78]	[0.46]	[0.67]	[2.10]	[2.54]	[2.68]
High − Low	0.42%	-0.77%	0.03%	0.02%	0.45%	0.36%	1.13%
	[1.31]	[-2.02]	[0.10]	[0.05]	[0.86]	[0.55]	[1.34]

Panel B: Two Ranges for Benchmark Mismatch

	0 < BM ≤ 60	BM > 60	High − Low
All	0.18%	1.50%	1.32%
	[0.94]	[3.20]	[3.07]
0 < AG ≤ 1.25	0.15%	1.09%	0.94%
	[1.02]	[2.13]	[1.85]
1.25 < AG ≤ 2.5	0.14%	1.72%	1.58%
	[0.78]	[3.32]	[3.15]
2.5 < AG ≤ 3.75	0.18%	1.64%	1.46%
	[0.79]	[3.03]	[2.92]
3.75 < AG ≤ 5	0.26%	1.40%	1.14%
	[0.97]	[2.39]	[2.04]
AG > 5	0.20%	1.56%	1.36%
	[0.61]	[2.90]	[2.78]
High − Low	0.05%	0.47%	0.42%
	[0.17]	[0.87]	[0.83]

Table 6: Predictors of the difference between AS and prospectus benchmark returns

This table shows results from the following model:

$$R_{AS,i,t} - R_{P,i,t} = \alpha + \beta * Active\ Share_{i,t} + \delta * Mismatch_{i,t} + \gamma * Chars_{i,t} + FE + \varepsilon_{i,t}$$

where $R_{AS,i,t}$ is the annualized return on fund i's AS benchmark in month t and $R_{P,i,t}$ is the annualized return on fund i's prospectus benchmark in month t. $Active\ Share_{i,t}$ is a vector of information about fund i's active share at the start of month t. It includes the fund's prospectus benchmark active share and a dummy variable equal to one if the prospectus benchmark active share is among the top 20% at the start of the month. $Mismatch_{i,t}$ is a vector of information about fund i's mismatch status at the start of month t. It includes *Benchmark Mismatch*, *Active Gap*, and a dummy variable equal to one if *Benchmark Mismatch* is among the top 20% at the start of the month. $Chars_{i,t}$ is a vector of characteristics for fund i available as of the start of month t. It includes the natural log of assets, natural log of age, expense ratio, turnover ratio, the number of equity positions, and the percentage of fund assets held within institutional share classes. The characteristics are included in all presented models, but the coefficients associated with the characteristics are suppressed in the table. FE represents style and year-month fixed effects. The model is estimated using the sample of fund-months with different prospectus and AS benchmarks. t-statistics are reported in brackets below each coefficient and are calculated using standard errors clustered by fund and year-month.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Prospectus Active Share	0.062			0.039	0.052		0.033
	[3.21]			[2.72]	[2.95]		[2.38]
Benchmark Mismatch		0.033		0.023		0.039	0.029
		[3.21]		[2.54]		[3.00]	[2.68]
Active Gap			0.095	0.040			
			[2.62]	[1.23]			
Top 20% AS Dummy					0.773		0.461
					[3.19]		[2.07]
Top 20% BM Dummy						-0.432	-0.456
						[-0.65]	[-0.67]
Characteristic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	125,352	125,352	125,352	125,352	125,352	125,352	125,352

Table 7: Performance of funds as a function of *Benchmark Mismatch* and active share

This table shows the performance of different groups of funds using multiple models. To form the groups, the full sample of fund-months—including fund-months with matching prospectus and AS benchmarks—is sorted independently on prospectus benchmark active share and *Benchmark Mismatch* (BM). With respect to active share, funds are sorted into quintiles at the beginning of each month. Funds in fifth quintile (those with highest active share) are tested separately from funds in the other four quintiles. The difference in results between those groups is considered in the 'Q5 – Q1234' portion of the table. With respect to *Benchmark Mismatch*, funds are sorted based on whether *Benchmark Mismatch* is greater than or less than 60%. The difference in results between those groups is considered in the 'Diff' column. We adjust the returns using three different models. The 'Prospectus' model reports the annualized average of the monthly average differences between the fund net return and the prospectus benchmark return. The 'Modified' model reports the annualized average of the monthly average differences between the fund net return and the prospectus benchmark return if *Benchmark Mismatch* is less than 60%. If *Benchmark Mismatch* is greater than 60%, then the AS benchmark is used instead of the prospectus benchmark. The 'Difference' row reports the differences between the 'Prospectus' and 'Modified' models. The 'CPZ7' model regresses the time-series of monthly average excess fund net returns against the Cremers, Petajisto, and Zitzewitz (2012) seven factors. We report the annualized intercept (i.e., the alpha) from that regression. *t*-statistics are reported in brackets below each value.

	Method	All	BM > 60%	BM ≤ 60%	Diff
All Funds	Prospectus	-0.25%	0.53%	-0.51%	1.04%
		[-0.78]	[1.23]	[-1.67]	[3.20]
	Modified	-0.59%	-0.86%	-0.51%	-0.35%
		[-2.07]	[-2.38]	[-1.67]	[-1.11]
	Difference	0.35%	1.39%	0.00%	1.39%
		[3.28]	[3.30]	-	[3.30]
	CPZ7	-0.53%	-0.11%	-0.65%	0.54%
		[-1.59]	[-0.23]	[-2.06]	[1.46]
Prospectus Active Share Quintile 5	Prospectus	0.75%	0.72%	0.75%	-0.02%
		[1.83]	[1.55]	[1.60]	[-0.04]
	Modified	-0.42%	-0.92%	0.75%	-1.67%
		[-1.21]	[-2.32]	[1.60]	[-3.32]
	Difference	1.16%	1.65%	0.00%	1.65%
		[3.62]	[3.64]	-	[3.64]
	CPZ7	0.50%	0.07%	1.28%	-1.21%
		[1.05]	[0.14]	[2.14]	[-2.09]
Prospectus Active Share Quintiles 1, 2, 3, and 4	Prospectus	-0.49%	0.27%	-0.62%	0.89%
		[-1.55]	[0.59]	[-1.97]	[2.50]
	Modified	-0.64%	-0.72%	-0.62%	-0.10%
		[-2.11]	[-1.82]	[-1.97]	[-0.30]
	Difference	0.14%	0.99%	0.00%	0.99%
		[2.36]	[2.14]	-	[2.14]
	CPZ7	-0.79%	-0.31%	-0.82%	0.51%
		[-2.47]	[-0.68]	[-2.67]	[1.12]
Q5 - Q1234	Prospectus	1.24%	0.46%	1.37%	-0.91%
		[4.18]	[1.36]	[3.09]	[-1.69]
	Modified	0.22%	-0.20%	1.37%	-1.57%
		[0.72]	[-0.60]	[3.09]	[-2.98]
	Difference	1.02%	0.66%	0.00%	0.66%
		[3.66]	[1.99]	-	[1.99]
	CPZ7	1.29%	0.38%	2.10%	-1.72%
		[4.18]	[0.95]	[4.26]	[-2.94]

Table 8: Performance of funds as a function of active share, past performance, and having a benchmark discrepancy

This table shows the Cremers, Petajisto, and Zitzewitz (2012) seven-factor alpha for different groups of funds. The alpha for a given group is estimated using the time-series of the monthly net returns on an equal weight portfolio of the group's funds. The reported alpha is annualized. To form the groups, the full sample of fund-months—including fund-months with matching prospectus and AS benchmarks—is first sorted based on prospectus benchmark active share. The 'Bottom 80%' group contains the funds within the lowest 80% of active share at the beginning of each month. The 'Top 20%' group contains the funds within the highest 20%. Next, funds within each of those active share groups are sorted each month based on their prospectus-benchmark-adjusted return during the previous year. The 'Bottom 80%' group contains the funds within the lowest 80% of prospectus-benchmark-adjusted return. The 'Top 20%' group contains the funds within the highest 20%. Finally, funds within each active share/past performance group are sorted based on whether the fund has a benchmark discrepancy. If a fund's *Benchmark Mismatch* is greater than 60% at the beginning of the month, then it is placed in the 'Yes' group. Otherwise, it is placed in the 'No' group. *t*-statistics are reported in brackets below each measurement.

Active Share	Bottom 80%	Top 20%
CPZ7 Alpha	-0.64%	0.71%
	[-2.25]	[1.37]

Past Performance	Bottom 80%	Top 20%	Bottom 80%	Top 20%
CPZ7 Alpha	-0.94%	0.53%	0.32%	2.31%
	[-2.65]	[0.85]	[0.57]	[3.01]

Benchmark Discrepancy	Yes	No	Yes	No	Yes	No	Yes	No
CPZ7 Alpha	-0.82%	-0.93%	0.72%	0.42%	-0.03%	1.03%	1.72%	3.21%
	[-1.44]	[-2.69]	[1.00]	[0.65]	[-0.05]	[1.41]	[1.97]	[2.37]

Table 9: Comparison of the Sensoy (2009) and *Benchmark Mismatch* identification procedures with respect to the difference in benchmark returns

This table shows the average benchmark-adjusted net fund return for fund-months identified as having a benchmark discrepancy using two different procedures. In the BM procedure, a fund is considered to have a benchmark discrepancy if *Benchmark Mismatch* is greater than 60%. In the Sensoy procedure, a fund is considered to have a benchmark discrepancy if Morningstar style boxes and fund-benchmark correlations indicate a more appropriate benchmark. The reported returns are adjusted using three benchmarks. The 'Prospectus' row reports the fund returns less the prospectus benchmark returns. The 'AS' row reports the fund returns less the AS benchmark returns. The 'Sensoy' row reports the fund returns less the returns on the appropriate benchmarks identified by the Sensoy procedure. The 'Pro – AS' row reports the difference between the 'Prospectus' and 'AS' results, and the 'Pro – Sensoy' row reports the difference between the 'Prospectus' and 'Sensoy' results. All returns are annualized. *t*-statistics are reported in brackets below each coefficient and are calculated using standard errors clustered by fund and year-month.

Benchmark Discrepancy?						
	BM	Yes	-	Yes	No	Yes
	Sensoy	-	Yes	No	Yes	Yes
Benchmark-Adjusted Return	Prospectus	0.37%	-0.16%	0.27%	-0.52%	0.52%
		[0.74]	[-0.31]	[0.49]	[-1.04]	[0.82]
	AS	-1.16%	-	-1.06%	-	-1.29%
		[-3.44]		[-2.92]		[-3.54]
	Sensoy	-	-0.91%	-	-0.86%	-0.97%
			[-3.17]		[-3.05]	[-2.44]
Differences	Prospectus − AS	1.52%	-	1.33%	-	1.81%
		[3.14]		[2.25]		[3.20]
	Prospectus − Sensoy	-	0.76%	-	0.34%	1.49%
			[1.64]		[0.82]	[2.59]

Table 10: Factor exposure differences between the AS and prospectus benchmarks

This table shows results from the following model:

$$Return_{AS,t} - Return_{pro,t} = \beta * Factor_t + \varepsilon_t$$

where $Return_{AS,t}$ is the average annualized return on the AS benchmarks in month *t* across all funds with *Benchmark Mismatch* greater than 60%. $Return_{pro,t}$ is the average annualized return on the prospectus benchmarks in month *t* across the same funds. $Factor_t$ is a vector of factor returns in month *t*. The factors included are all of those in the Cremers, Petajisto, and Zitzewitz (2012) seven-factor model. The model is estimated using the full sample of fund-months with *Benchmark Mismatch* greater than 60%. It is also estimated for subgroups with different styles (as indicated by funds' prospectus benchmarks). The 'Prospectus' row reports the average of the monthly average differences between the fund net returns and the prospectus benchmark returns. The 'AS' row reports the average of the monthly average differences between the fund net returns and the AS benchmark returns. Both of those values are annualized. The 'Difference' row reports the differences between the 'Prospectus' and 'AS' rows. It is equal to the average of $Return_{AS,t}$ less the average of $Return_{pro,t}$. Rows 'S5RF' through 'UMD' report the estimated coefficients of the model. The 'Total Factor Return' row reports the sum of the products of the estimated factor exposures and the annualized factor returns. *t*-statistics associated with tests of whether the values in the table are different from zero are reported in brackets below each value.

	(1)	(2)	(3)	(4)	(5)	(6)
Style	All	Large	Small/Mid	Growth	Value	Blend
Prospectus	0.66%	1.05%	0.18%	0.97%	-0.61%	0.70%
	[1.16]	[0.78]	[0.31]	[1.51]	[-0.92]	[0.94]
AS	-0.84%	-1.33%	-0.94%	-0.80%	-1.01%	-1.08%
	[-1.70]	[-2.38]	[-1.72]	[-0.99]	[-1.93]	[-1.75]
Difference	1.50%	2.38%	1.12%	1.77%	0.40%	1.78%
	[3.67]	[2.13]	[1.94]	[2.19]	[0.71]	[3.21]
S5RF	-0.01	0.02	-0.01	-0.02	-0.01	-0.01
	[-1.14]	[1.22]	[-0.77]	[-1.18]	[-0.59]	[-1.46]
RMS5	0.12	0.69	-0.05	0.02	0.01	0.26
	[4.99]	[11.89]	[-1.49]	[0.42]	[0.49]	[9.25]
R2RM	-0.12	0.16	-0.22	-0.13	-0.11	-0.08
	[-6.83]	[3.75]	[-9.01]	[-5.20]	[-3.94]	[-3.54]
S5VS5G	-0.04	-0.08	-0.03	-0.04	-0.08	-0.05
	[-2.08]	[-2.02]	[-1.09]	[-1.22]	[-2.55]	[-1.98]
RMVRMG	-0.02	0.14	-0.08	-0.02	0.03	-0.06
	[-0.72]	[1.67]	[-2.83]	[-0.69]	[0.76]	[-2.08]
R2VR2G	0.04	-0.15	0.10	0.25	-0.15	0.03
	[1.49]	[-2.52]	[3.78]	[7.28]	[-3.87]	[1.03]
UMD	0.02	0.01	0.03	0.06	-0.02	0.02
	[2.39]	[0.95]	[2.08]	[3.50]	[-1.76]	[1.99]
R^2	27.7%	75.7%	49.9%	63.3%	36.9%	34.5%
Total Factor Return	0.57%	1.60%	0.41%	0.87%	-0.18%	0.73%
	[2.65]	[1.64]	[1.01]	[1.35]	[-0.52]	[2.24]

Table 11: Non-traditional factor exposure differences between the AS and prospectus benchmarks

This table shows results from the following model:

$$Return_{AS,t} - Return_{pro,t} = \beta * Factor_t + \varepsilon_t$$

where $Return_{AS,t}$ is the average annualized return on the AS benchmarks in month *t* across all funds with *Benchmark Mismatch* greater than 60%. $Return_{pro,t}$ is the average annualized return on the prospectus benchmarks in month *t* across the same funds. $Factor_t$ is a vector of factor returns in month *t*. The factors include all of those in the Cremers, Petajisto, and Zitzewitz (2012) seven-factor model; the Fama and French (2015) profitability (RMW) and investment (CMA) factors; the Stambaugh and Yuan (2017) management (MGMT) and performance (PERF) factors; the Frazzini and Pedersen (2014) betting against beta (BAB) factor; the Asness, Frazzini, and Pedersen (2019) quality-minus-junk (QMJ) factor; and the Pastor and Stambaugh (2004) traded liquidity (LIQ) factor. The model is estimated using the full sample of fund-months with *Benchmark Mismatch* greater than 60%. It is also estimated for subgroups with different styles (as indicated by funds' prospectus benchmarks). Rows 'S5RF' through 'LIQ' report the estimated coefficients of the model. The 'Total Factor Return' row reports the sum of the products of the estimated factor exposures and the annualized factor returns. *t*-statistics associated with tests of whether the values in the table are different from zero are reported in brackets below each value.

	(1)	(2)	(3)	(4)	(5)	(6)
Prospectus Style	All	Large	Small/Mid	Growth	Value	Blend
S5RF	0.00	0.05	0.00	-0.02	0.02	0.01
	[0.26]	[2.95]	[0.07]	[-1.28]	[2.05]	[0.53]
RMS5	0.13	0.68	-0.01	0.04	0.02	0.28
	[6.27]	[13.34]	[-0.39]	[0.82]	[0.48]	[10.50]
R2RM	-0.08	0.21	-0.19	-0.10	-0.05	-0.03
	[-4.52]	[6.02]	[-7.52]	[-3.86]	[-2.21]	[-1.52]
S5VS5G	0.02	0.03	0.00	-0.01	0.03	0.05
	[1.16]	[0.70]	[0.13]	[-0.43]	[1.22]	[1.73]
RMVRMG	-0.10	0.07	-0.13	-0.08	-0.06	-0.15
	[-4.95]	[1.02]	[-4.50]	[-2.13]	[-2.03]	[-6.02]
R2VR2G	0.01	-0.12	0.06	0.19	-0.15	0.02
	[0.52]	[-2.04]	[1.76]	[4.66]	[-5.46]	[0.57]
UMD	0.01	-0.01	0.02	0.06	-0.05	-0.00
	[0.97]	[-0.26]	[1.48]	[3.24]	[-3.05]	[-0.30]
RMW	0.14	0.08	0.09	0.15	0.12	0.14
	[4.47]	[1.34]	[2.31]	[3.17]	[3.37]	[3.85]
CMA	-0.02	0.03	-0.04	-0.02	0.03	-0.01
	[-0.63]	[0.64]	[-1.41]	[-0.38]	[0.90]	[-0.35]
MGMT	-0.00	-0.10	0.06	0.05	-0.06	-0.02
	[-0.07]	[-2.59]	[2.07]	[1.35]	[-2.28]	[-0.70]
PERF	-0.00	0.03	-0.02	-0.03	0.03	0.02
	[-0.13]	[0.89]	[-0.95]	[-1.47]	[1.68]	[0.98]
QMJ	0.04	0.10	0.05	-0.00	0.09	0.06
	[1.33]	[2.61]	[1.35]	[-0.09]	[2.52]	[1.59]
BAB	0.02	0.00	0.01	0.02	0.01	0.02
	[2.46]	[0.23]	[0.55]	[1.48]	[0.83]	[1.76]
LIQ	-0.00	0.02	0.02	0.00	0.01	-0.02
	[-0.15]	[1.63]	[2.31]	[0.27]	[1.37]	[-2.51]
R^2	52.4%	81.1%	58.2%	67.3%	59.4%	53.9%
Total Factor Return	1.31%	2.18%	1.23%	1.50%	0.74%	1.50%
	[4.48]	[2.17]	[2.80]	[2.26]	[1.71]	[3.70]

Table 12: Prospectus- and AS-benchmark-adjusted returns evaluated with factor models

This table shows results from the following model:

$$Return_{fund,t} - Return_{bench,t} = \alpha + \beta * Factor_t + \varepsilon_t$$

where $Return_{fund,t}$ is the average annualized net return in month *t* across all tested funds, and $Return_{bench,t}$ is the average annualized return on those same funds' benchmarks in month *t*. In the 'Prospectus Adjusted' row, the fund's prospectus benchmark is used. In the 'AS Adjusted' row, the AS benchmark is used. $Factor_t$ is a vector of month *t* factor returns. In the column labeled 'Return', no factors are included in the model. In the columns with the heading 'CPZ7', all of the factors in the Cremers, Petajisto, and Zitzewitz (2012) seven-factor model are included. In the columns with the heading 'CPZ7+', all of the non-traditional factors discussed in Table 11 are also included. Within the 'CPZ7' and 'CPZ7+' headings, the 'Alpha' column reports the α from the model, and the 'Change' column reports the difference between that α and the value from the 'Return' column in the same row. The 'Difference' row reports the differences between the values in the 'Prospectus Adjusted' and 'AS Adjusted' rows. Panel A shows results using the full sample of fund-months with *Benchmark Mismatch* greater than 60%. Panel B shows results using just the funds within that group that also have a large-cap style (based on their prospectus benchmark). *t*-statistics are reported in brackets below each value.

Panel A: Benchmark Mismatch > 60%

		CPZ7		CPZ7+	
	Return	Alpha	Change	Alpha	Change
Prospectus Adjusted	0.65%	0.51%	-0.15%	-0.29%	-0.94%
	[1.36]	[1.16]	[-0.33]	[-0.72]	[-2.37]
AS Adjusted	-0.87%	-0.40%	0.47%	-0.30%	0.57%
	[-2.23]	[-1.08]	[1.27]	[-0.75]	[1.44]
Difference	1.52%	0.91%	-0.61%	0.01%	-1.51%
	[3.45]	[2.28]	[-1.54]	[0.03]	[-4.24]

Panel B: Benchmark Mismatch > 60% and Large Cap Style

		CPZ7		CPZ7+	
	Return	Alpha	Change	Alpha	Change
Prospectus Adjusted	1.05%	-0.14%	-1.18%	-0.89%	-1.94%
	[0.87]	[-0.22]	[-1.92]	[-1.40]	[-3.05]
AS Adjusted	-1.36%	-0.95%	0.41%	-0.91%	0.45%
	[-3.11]	[-2.21]	[0.96]	[-2.05]	[1.00]
Difference	2.41%	0.81%	-1.60%	0.02%	-2.38%
	[2.10]	[1.52]	[-3.00]	[0.04]	[-4.52]

Table 13: Response of investor flows to different measures of performance

This table shows results from the following model:

$$Flow_{i,t} = \theta + \beta * Performance_{i,t} + \gamma * Mismatch_{i,t} + \delta * Chars_{i,t} + FE + \varepsilon_{i,t}$$

where $Flow_{i,t}$ is the percentage implied net flow for fund *i* in month *t*. $Performance_{i,t}$ is a vector of information about fund *i*'s performance over the year ending at the start of month *t*. It includes the difference between fund *i*'s net return and the return on fund *i*'s AS benchmark, the difference between the returns on fund *i*'s AS and prospectus benchmarks, and fund *i*'s annualized CAPM alpha. In columns (1) through (4), the actual numeric return values are used. In columns (5) through (7), each of the three return measures in $Performance_{i,t}$ is ranked at the start of each month and then evenly scaled from zero to one (with one being the highest). $Mismatch_{i,t}$ is the *Benchmark Mismatch* (BM) for fund *i* as of the start of month *t*. $Chars_{i,t}$ is a vector of characteristics for fund *i* available as of the start of month *t*. It includes the natural log of assets, natural log of age, expense ratio, turnover ratio, the number of equity positions, and the percentage of fund assets held within institutional share classes. The characteristics are included in all of the presented models, but the coefficients are suppressed. FE represents style and year-month fixed effects. The model is estimated using the sample of fund-months with different prospectus and AS benchmarks. In column (7), only the funds with different benchmarks that are also in the top 20% of total assets at the start of month *t* are used to estimate the model. *t*-statistics are reported in brackets below each coefficient and are calculated using standard errors clustered by fund and year-month.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fund Ret − AS Ret	0.12	0.13	0.09	0.129	1.97	1.26	2.24
	[23.27]	[23.86]	[15.35]	[13.38]	[18.90]	[5.46]	[4.78]
AS Ret − Prospectus Ret		0.07	0.04	0.140	0.69	1.83	1.83
		[14.03]	[8.76]	[9.98]	[9.16]	[8.30]	[5.50]
CAPM Alpha			0.07		1.78		
			[14.05]		[16.93]		
Benchmark Mismatch				-0.002			
				[-1.52]			
(Fund Ret − AS Ret) * BM				0.000			
				[0.63]			
(AS Ret − Prospectus Ret) * BM				-0.001			
				[-5.64]			
(Fund Ret − AS Ret)^2						1.83	0.57
						[7.48]	[1.25]
(AS Ret − Prospectus Ret)^2						-0.73	-0.83
						[-3.39]	[-2.46]
Returns	Actual	Actual	Actual	Actual	Ranking	Ranking	Ranking
Sample	BM > 0	BM > 0	BM > 0	BM > 0	BM > 0	BM > 0	BM > 0 Size Q5
Characteristic Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	122,411	122,411	122,411	122,411	122,411	122,411	24,363