JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 57, No. 2, Mar. 2022, pp. 543–571
© THE AUTHOR(S), 2021. PUBLISHED BY CAMBRIDGE UNIVERSITY PRESS ON BEHALF OF THE MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON
doi:10.1017/S0022109021000119

Benchmark Discrepancies and Mutual Fund Performance Evaluation

K. J. Martijn Cremers
University of Notre Dame Mendoza College of Business
mcremers@nd.edu

Jon A. Fulkerson
University of Dayton School of Business Administration
jfulkerson1@udayton.edu

Timothy B. Riley
University of Arkansas Walton College of Business
tbriley@uark.edu (corresponding author)

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. We find that 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 adjustments, despite underperforming the benchmarks that best match their portfolios.

I. 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

We thank an anonymous referee, Chris Clifford, Jennifer Conrad (the editor), 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 Chartered Financial Analyst (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: Cremers is currently an independent director at Ariel Investments and a consultant to Touchstone Investments and State Street Associates, and Fulkerson and Riley have both been previously compensated by Natixis Investment Managers and the Investment Adviser Association.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

to performance relative to a fund's prospectus benchmark index, that is, the benchmark that a fund declares in its prospectus to meet regulatory requirements.[1]

Focusing on benchmark-adjusted performance would be reasonable for mutual funds with prospectus benchmarks that accurately match the funds' actual investment strategy. For any fund with an inaccurate match, however, the prospectus benchmark could over- or underestimate performance. Consider a fund with a portfolio that is substantially riskier than the fund's prospectus benchmark. That fund could outperform using prospectus-benchmark-adjusted returns but underperform if compared to a more appropriate, and thus riskier, benchmark. 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.

In this article, we introduce a novel 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 benchmark typically understates 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 a fund relative to its prospectus benchmark, even if the prospectus benchmark is inaccurate.

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 (AS) measure. The benchmark that generates the lowest AS is considered to have the most overlap. This "AS benchmark" is different from the prospectus benchmark in 65% of our sample.

Differences between the benchmarks are 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 benchmarks. 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: One benchmark is often 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 AS criterion applied by Cremers and Petajisto (2009) to distinguish truly active

[1]Several other studies also support the idea that investors place considerable weight on simpler measures of performance (i.e., 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)).

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

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 37% of the cases in which the benchmarks differ, BENCHMARK_MISMATCH is above 60%. As a result, we label 24% of our fund observations as having a benchmark discrepancy. The funds labeled as such end 2014 managing a combined $234 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 approximately 1.46% per year (t-statistic = 3.08). Accordingly, funds with a benchmark discrepancy generally need less return to beat the prospectus benchmark compared with 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 0.17% per year, but they underperform their AS benchmarks by 1.26% per year. Likewise, using prospectus-benchmark-adjusted returns, we find that funds with a benchmark discrepancy outperform those without a discrepancy on average by 0.96% per year. If the AS benchmarks are instead used for funds with benchmark discrepancies, funds without a benchmark discrepancy outperform those with a discrepancy on average by 0.47% per year. Taken together, a primary conclusion of our article 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, nontraditional factor exposures (e.g., liquidity) in conjunction with the traditional exposures explain approximately 89% of the average difference in returns.

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, will result 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

[2] Although 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-statistic = 3.68) but not when BENCHMARK_MISMATCH is less than 60%.

begins with funds that have prospectus benchmark styles that do not match the Morningstar 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: 24% using our procedure and 22% 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, approximately 43% only have a discrepancy according to our procedure, and another 37% only have a discrepancy according to Sensoy's procedure.

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 1.31% per year (t-statistic = 2.20). On the other hand, funds identified as having a benchmark discrepancy by Sensoy's procedure, but not ours, have no significant difference in average benchmark-adjusted return between their benchmark choices. The difference is just 0.18% per year (t-statistic = 0.40). Based on that result, we conclude that our procedure's benchmark discrepancies have a larger impact on performance evaluation.

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, we find that investors respond to fund performance relative to the prospectus benchmark even when a benchmark discrepancy exists, but our new procedure can capture the magnitude of that response. For instance, 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)), accounting for a fund's capital asset pricing model (CAPM) or multifactor alpha still leaves a large role in explaining fund flows for the prospectus-benchmark-adjusted returns.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

II. Comparison with Prior Work

Previous work has shown that common mutual fund style classifications are often inaccurate (e.g., Brown and Goetzmann (1997), DiBartolomeo and Witkowski (1997)), that mutual funds' styles tend to drift over time (e.g., Wermers (2012), Cao, Iliev, and Velthuis (2017)), and that asset managers frequently supply erroneous benchmarks (e.g., Sensoy (2009), Elton et al. (2014)). Given our focus on mutual funds' prospectus benchmarks, our article is most comparable to Sensoy. He finds that approximately 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 article differs from Sensoy's in several important ways.

To begin, our procedure for identifying benchmark discrepancies is markedly different from that of Sensoy (2009). As explained previously, we directly compare fund and benchmark holdings to identify benchmark discrepancies, whereas Sensoy uses Morningstar style boxes and fund returns. Our procedure addresses some of the limitations of Sensoy's procedure. First, 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. Second, 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 to i) measure the magnitude of a benchmark discrepancy, using BENCHMARK_MISMATCH; ii) capture time variation in the appropriate benchmark, important given that Huang, Sialm, and Zhang (2011) show significant time variation in fund risk; and iii) 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. Although all the benchmarks used in our 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 use of the Morningstar style boxes in Sensoy's (2009) 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 its prospectus benchmark while having the other as its AS benchmark. The two benchmarks do ostensibly cover the same factors, but their holdings differ significantly, with an average BENCHMARK_MISMATCH of 70%.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

The result of these methodological differences is that our procedure and Sensoy's (2009) procedure frequently disagree about which funds have a benchmark discrepancy. We previewed these results earlier and provide a more detailed discussion in Section VIII.

Beyond differences in procedure, our article is distinct from Sensoy's (2009) in several other ways. Of greatest importance, our analysis focuses on the magnitude of and explanations for the difference in returns between the AS and prospectus benchmarks. We find that compared with their AS benchmarks, the prospectus benchmarks of funds with benchmark discrepancies have lower average returns, with most of that lower return attributable to lower factor exposures. Sensoy does not document or focus on a similar difference in returns. Moreover, we provide new insights into the responsiveness of investors to performance relative to the prospectus benchmark. We estimate, for example, 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.

III. Data

A. Mutual Fund Sample

Our sample of actively managed mutual funds comes from the 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 exclude any fund that CRSP identifies as an index fund, exchange-traded 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% of their assets in common stock. We also filter out funds with terms in their names that are associated with indexing or strategies other than traditional long-only U.S. equity (available from the authors). To address incubation bias (Evans (2010)), we exclude a fund from the sample until it is at least 2 years old and until it first reaches at least $20 million in assets.

The entirety of our analysis is conducted at the fund level. We aggregate information across multiple share classes of a fund using the WFICN variable available in Mutual Fund Links (MFLINKS). Fund assets are the sum of the assets across all share classes. All other fund characteristics, including returns, are calculated as the asset-weighted average of the share-class values.

We collect information on funds' prospectus benchmarks from Morningstar Direct and match those data to CRSP using ticker, CUSIP, and assets.[3] A fund is dropped from the sample if we cannot match it to Morningstar or if Morningstar does not provide a prospectus benchmark. The data on the prospectus benchmarks are 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 approximately 2%

[3]We use the primary prospectus benchmark reported in Morningstar in all our tabulated analyses; however, for approximately 32% of our final sample of funds, Morningstar also reports a secondary prospectus benchmark. If we exclude funds with a reported secondary prospectus benchmark, our primary results hold.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

of funds undergo a change. U.S. Securities and Exchange Commission (SEC) rules first required mutual funds to provide a benchmark to investors in certain documents released after July 1, 1993.[4] Because the time period used in our analysis is 1991–2015, there is the potential for survivor bias in the first few years of our sample; however, our conclusions are the same regardless of whether we include the preregulation data.[5]

B. Mutual Fund Holdings

We use the Thomson Reuters database as our source of mutual fund holdings. Those data only contain information on funds' equity positions, so any nonequity positions, including cash, are not reflected. These holdings are merged with the CRSP fund database using MFLINKS. To verify the merge, 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.

C. Benchmark Holdings

Our benchmark-discrepancy procedure involves a comparison between fund holdings and the holdings of a set of benchmark indexes. The set of benchmarks includes 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 7 benchmarks. Our primary reason for this limitation is that these 21 benchmarks are individually well diversified and, as a group, frequently referenced by funds. We constrain our sample of funds to include only those with a prospectus benchmark within this set. The prospectus benchmarks for 93% of the funds in our initial sample are in the set, with the most common benchmarks being the S&P 500 (31% of funds), Russell 1000 Growth (11%), Russell 1000 Value (10%), and Russell 2000 (8%).

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 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 come from multiple sources. Russell provided us the constituent weights for its benchmarks, whereas the constituent weights for the S&P benchmarks come from Compustat. Monthly return data for the benchmarks (with dividends reinvested) come from Morningstar Direct.

[4]See "Final Rule: Disclosure of Mutual Fund Performance and Portfolio Managers" at https://www.sec.gov/rules/final/33-6988.pdf.

[5]On a similar note, if we split our time period into halves, Jan. 1991 to June 2003 and July 2003 to Dec. 2015, we find similar results in each half.

Our final sample consists of 245,512 fund-month observations across 1,745 unique funds. The number of funds varies over time, with a minimum of 148 in Nov. 1992 and a maximum of 1,262 in Feb. 2007.

IV. Benchmark-Discrepancy Methodology

A. The Active Share 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 identifying 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) AS, defined as follows:

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

where $w_{i,f}$ is the weight on stock i in the fund's portfolio, and $w_{i,b}$ is the weight on stock i in the benchmark. The measure is calculated over all N stocks in the investable universe.

As AS increases, the fund and a given benchmark have less overlap. An AS of 0 means that the fund and a given benchmark have identical stock holdings, and an AS 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 AS benchmark* or simply the *AS benchmark.*

The AS benchmark for each fund is redetermined every time our data provide a new holdings report, which is quarterly in most instances. 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 $t+1$. Allowing the benchmark to vary over time is important because fund style and risk are not time invariant. Regardless, if we stabilize the AS benchmark by using the mode of a fund's AS benchmark over the previous 3 years, our primary results are unchanged. On the same subject, the fact that the AS benchmark can change suggests the possibility of market timing (e.g., Bollen and Busse (2001)), but we find little evidence of timing ability related to changes in the AS benchmark. For example, in the month after a change in the AS benchmark, the average difference in returns between the new and old AS benchmark is only 1.6 basis points (bps) (t-statistic = 0.64).

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

B. Benchmark Mismatch

After identifying the set of funds for which the prospectus benchmark differs from the AS benchmark, we then 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. For example, the Russell 1000 and Russell 3000 have an AS 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 how dissimilar prospectus and AS benchmarks are 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 as BENCHMARK_MISMATCH:

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

where $w_{i,p}$ is the weight on stock i in the fund's prospectus benchmark, and $w_{i,AS}$ is the weight on stock i in the fund's AS benchmark. When the holdings of the two benchmarks largely overlap, the AS of the prospectus benchmark with respect to the AS benchmark will be low, and thus BENCHMARK_MISMATCH will be small. Hence, an increase in BENCHMARK_MISMATCH represents a decrease in the overlap between the holdings of the two benchmarks.

Because BENCHMARK_MISMATCH captures the overlap in holdings between the prospectus benchmark and the AS benchmark, we can directly interpret BENCHMARK_MISMATCH as a measure of the economic magnitude of the dissimilarity between those benchmarks. For the main results in the article, we classify funds with BENCHMARK_MISMATCH above 60% as having significant economic dissimilarity in their benchmarks and thus having a benchmark discrepancy.

Although the 60% cutoff is somewhat arbitrary, we set the threshold there for two reasons. First, a similar 60% cutoff is employed in prior work using AS (e.g., Cremers and Petajisto (2009), Cremers and Curtis (2016)). Funds in those articles with an AS of less than 60% are labeled *closet indexers.* Second, as shown in Section VII, analysis of benchmark returns suggests that the economic differences between the prospectus and AS benchmarks are minor, on average, when BENCHMARK_MISMATCH is less than 60%.

We also consider the difference between the AS of a fund with respect to its prospectus benchmark and the AS of a fund with respect to its AS benchmark. We label this difference as ACTIVE_GAP:

$$\text{ACTIVE_GAP} = \text{ACTIVE_SHARE}_{\text{prospectus}} - \text{ACTIVE_SHARE}_{\text{AS}}. \tag{3}$$

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 dissimilar. However, conditional on having a 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.

V. Key Measures and Fund Characteristics

Table 1 shows summary statistics for the key measures used in our article (Panel A) and average characteristics for funds with BENCHMARK_MISMATCH greater than and less than 60% (Panel B). Starting with Panel A, the dummy variable ANY_MISMATCH, which is equal to 1 if BENCHMARK_MISMATCH is above 0, shows that approximately 65% of observations have a prospectus benchmark different from the AS benchmark. If we require BENCHMARK_MISMATCH to be greater than 60%, as motivated previously, then only 24% of observations have a difference in benchmarks (as indicated by the dummy variable LARGE_MISMATCH). Funds with an AS benchmark different from their prospectus benchmark tend to maintain that difference. The correlation between the value

TABLE 1

Summary Statistics

Table 1 shows summary statistics for our full sample and for funds with low and high BENCHMARK_MISMATCH. Panel A shows statistics for key measures using our full sample of funds. ANY_MISMATCH is a dummy variable equal to 1 if the prospectus and AS benchmarks are different. LARGE_MISMATCH is a dummy variable equal to 1 if the prospectus and AS benchmarks are different and BENCHMARK_MISMATCH is greater than 60%. 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 $t-12$. Panel B compares the mean 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 0). LARGE_CAP through VALUE are dummy variables for style and show 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 independently double-clustered by fund and year-month.

Panel A. Full-Sample Statistics

	Mean	Std. Dev.	P25	P50	P75	$\rho_{t,t-12}$
ANY_MISMATCH	0.65	0.48	0.00	1.00	1.00	0.57
LARGE_MISMATCH	0.24	0.43	0.00	0.00	0.00	0.73
PROSPECTUS_ACTIVE_SHARE	79.8%	13.9%	70.2%	81.9%	91.4%	0.92
MINIMUM_ACTIVE_SHARE	77.4%	13.7%	67.8%	79.7%	88.8%	0.91
BENCHMARK_MISMATCH	33.3%	31.4%	0.0%	32.5%	50.1%	0.69
ACTIVE_GAP	2.3%	3.2%	0.0%	1.2%	3.5%	0.74
PROSPECTUS_ADJUSTED_RETURN	−0.55%	19.24%	−10.76%	−0.73%	9.36%	0.01
AS_ADJUSTED_RETURN	−0.94%	18.93%	−11.23%	−0.98%	9.16%	0.01

Panel B. Characteristics of Funds with High and Low BENCHMARK_MISMATCH

	Full Sample	BM > 60%	BM ≤ 60%	Difference	*t*-Stat.
PROSPECTUS_ACTIVE_SHARE	80.0%	92.9%	76.1%	16.8%	48.15
ASSETS ($billions)	1.14	0.78	1.25	−0.48	−6.05
AGE (years)	15.8	13.0	16.6	−3.6	−7.29
EXPENSE_RATIO	1.22%	1.32%	1.19%	0.14%	10.19
TURNOVER_RATIO	80.1%	82.9%	79.3%	3.6%	1.66
NUMBER_OF_HOLDINGS	89.3	96.3	87.2	9.1	3.39
INSTITUTIONAL_OWNERSHIP (% of assets)	26.8%	24.5%	27.5%	−3.0%	−2.05
LARGE_CAP	62.7%	19.0%	76.3%	−57.3%	−34.36
SMALL/MID_CAP	37.3%	81.0%	23.7%	57.3%	34.36
GROWTH	31.4%	29.1%	32.2%	−3.1%	−1.51
BLEND	43.9%	42.5%	44.3%	−1.8%	−0.77
VALUE	24.7%	28.4%	23.5%	4.9%	2.27

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

of ANY_MISMATCH (LARGE_MISMATCH) in month t and month $t - 12$ is 0.57 (0.73).

In the Supplementary Material, we show the cumulative density function (CDF) of BENCHMARK_MISMATCH for funds with different prospectus and AS benchmarks. Among those funds, approximately 17% have BENCHMARK_MISMATCH above 80%. To put that in context, if those funds had exactly the same holdings as their AS benchmark, then an investor using the prospectus benchmark would conclude that those funds have an active share of at least 80%, close to AS's full-sample average.

The frequency of benchmark discrepancies varies over time. As shown in the Supplementary Material, the number of funds with BENCHMARK_MISMATCH greater than 60% varies within the range of 12% to 42%, with significant jumps in the number of such funds in 1992 and 1998 because of new benchmarks entering the set.[6] After 1998, the number of funds with BENCHMARK_MISMATCH greater than 60% slowly decreases from 27% to 18%.

The average AS is 79.8% using the prospectus benchmark, compared with 77.4% using the AS benchmark. As such, that difference (i.e., the ACTIVE_GAP) has a mean of approximately 2.3%. The Supplementary Material shows the CDF of ACTIVE_GAP for funds with different prospectus and AS benchmarks. It indicates that approximately 23% of those funds have ACTIVE_GAP above 5%.

Using net returns, the average fund underperforms both the prospectus and AS benchmarks (ignoring the costs of investing in those benchmarks); however, the degree of underperformance differs. Relative to the prospectus benchmark, funds on average underperform by 0.55% per year. In comparison, relative to the AS benchmark, funds underperform by 0.94% per year. 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 with AS benchmarks in the full sample.

Panel B of Table 1 compares the characteristics of funds with a benchmark discrepancy (BENCHMARK_MISMATCH > 60%) to those of funds without a benchmark discrepancy.[7] In this analysis, the group without a benchmark discrepancy includes funds with BENCHMARK_MISMATCH of 0. Funds with a benchmark discrepancy tend to be more actively managed. The average AS with respect to the prospectus benchmark for those funds is 92.9%, compared with 76.1% for funds without a benchmark discrepancy. 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 a small- or mid-cap style. Approximately 81.0% of

[6]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 1992–1997. Each of the Russell benchmarks is available from the start of our sample.

[7]The Supplementary Material shows the most common benchmark combinations for funds with and without a benchmark discrepancy. The most common combination for funds without a benchmark discrepancy is an S&P 500 prospectus benchmark and an S&P 500 Growth AS benchmark, which has a BENCHMARK_MISMATCH of 32.9%. The most common combination for funds with a benchmark discrepancy is a Russell 2000 Value prospectus benchmark and an S&P 600 Value AS benchmark, which has a BENCHMARK_MISMATCH of 69.1%.

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

VI. Benchmark Discrepancies and Fund Characteristics

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

$$\text{BM} > 60\%_{i,t} = \alpha + \beta \times \text{ACTIVE_SHARE}_{i,t} + \delta \times \text{CHARS}_{i,t} + \gamma \times \text{STYLE}_i + \text{FE} + \varepsilon_{i,t}, \tag{4}$$

where $\text{BM} > 60\%_{i,t}$ is a dummy variable equal to 1 if the BENCHMARK_MISMATCH for fund i based on quarter t holdings is greater than 60%, and 0 otherwise. $\text{ACTIVE_SHARE}_{i,t}$ is a vector of information about fund i's quarter t AS. It includes the prospectus benchmark AS and a dummy variable equal to 1 if that AS is among the top 20% across all funds in quarter t, and 0 otherwise. $\text{CHARS}_{i,t}$ is a vector of characteristics for fund i available as of quarter t. It includes the natural logarithm of assets, natural logarithm of age, expense ratio, turnover ratio, number of equity positions, and 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 0.

Table 2 presents the results from this regression. As prospectus AS increases, the probability of having a benchmark discrepancy increases. After controlling for fund characteristics and style, that relation becomes nonlinear. Funds in the top 20% of AS are more likely to have a benchmark discrepancy than the linear term indicates. The fund characteristics have either limited economic significance or limited statistical significance when considered in the full model. Fund style, however, contains substantial predictive power. Consistent with the summary statistics, funds with a large-cap style are less likely to have a benchmark discrepancy compared with funds that have a small- or mid-cap style.

That benchmark discrepancies are relatively uncommon among large-cap funds is likely driven by differences between the universes of large- and small-/mid-cap stocks. Because the benchmarks are weighted by market capitalization and because market capitalizations are highly skewed, the effective number of stocks in each benchmark, defined as the number of equally weighted stocks necessary to produce the same Herfindahl index as the actual benchmark, is much smaller for large-cap benchmarks than for small- and mid-cap benchmarks (e.g., 163 for the Russell 1000 compared with 1,102 for the Russell 2000). Hence, to the extent that funds invest in assets inside their benchmarks, large-cap funds face a universe with fewer effective stocks, which means that they effectively have fewer options to diverge from their benchmarks and potentially create a benchmark discrepancy. The ASs of large-cap funds tend to be lower for the same reason.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

TABLE 2

Probability of Having a Benchmark Discrepancy

Table 2 shows estimates from the following logit model:

$$BM > 60\%_{i,t} = \alpha + \beta \times ACTIVE_SHARE_{i,t} + \delta \times CHARS_{i,t} + \gamma * STYLE_i + FE + \varepsilon_{i,t},$$

where $BM > 60\%_{i,t}$ is a dummy variable equal to 1 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, $CHARS_{i,t}$ is a vector of characteristics for fund i available as of quarter t, $STYLE_i$ is a vector of information about fund i's style based on its prospectus benchmark, and FE represents year-quarter fixed effects. The model is estimated using the full sample (including fund-quarters with BENCHMARK_MISMATCH equal to 0). t-statistics are reported in parentheses below each coefficient and are calculated using standard errors independently double-clustered by fund and year-quarter.

	1	2	3	4	5	6
PROSPECTUS_ACTIVE_SHARE	0.25	0.24			0.29	0.29
	(29.89)	(26.37)			(20.15)	(19.46)
TOP_20%_AS_DUMMY		0.13			0.31	0.33
		(1.37)			(2.93)	(3.04)
ASSETS			−0.02		0.04	0.05
			(−0.79)		(1.44)	(1.46)
AGE			−0.29		−0.02	0.03
			(−4.99)		(−0.23)	(0.43)
EXPENSE_RATIO			1.01		−0.12	−0.25
			(7.83)		(−0.84)	(−1.72)
TURNOVER_RATIO			0.00		0.00	0.00
			(−1.24)		(1.32)	(0.62)
NUMBER_OF_HOLDINGS			0.00		0.01	0.01
			(5.57)		(12.15)	(12.05)
INSTITUTIONAL_OWNERSHIP			0.00		0.00	0.00
			(−0.07)		(0.05)	(−0.14)
LARGE_CAP_DUMMY				−2.86	−1.00	−0.94
				(−24.72)	(−9.06)	(−8.01)
BLEND_DUMMY				0.23	−0.51	−0.54
				(1.92)	(−4.03)	(−4.18)
GROWTH_DUMMY				−0.70	−0.20	−0.17
				(−5.41)	(−1.48)	(−1.24)
Fixed effects	No	No	No	No	No	Yes
No. of obs.	66,029	66,029	66,029	66,029	66,029	66,029

Large-cap funds could respond to the smaller effective number of stocks in their benchmarks by investing more outside their benchmarks. Empirically, however, large-cap funds invest substantially less outside their prospectus benchmarks, on average, compared with small- and mid-cap funds (17.2% vs. 36.7%). This tendency can also be explained by the skewed distribution of market capitalizations. That skewness creates a substantial difference in the relative ambiguity of the cap universes: The universe of large-cap stocks is relatively well defined, with the typical stocks inside, say, the Russell 1000 having very different market capitalization than the stocks outside the Russell 1000. In comparison, the universe of small- and mid-cap stocks is much less clear, and thus funds following small- and mid-cap benchmarks are more likely to invest outside their benchmarks, which increases the probability of a benchmark discrepancy.

Unrelated to market-cap differences, Sensoy (2009), using his benchmark discrepancy procedure, finds a substantial impact related to fund family in a similar analysis. Nevertheless, in untabulated analysis, we do not find similar results if fund family effects are added to our model. This difference suggests that the benchmark discrepancies identified by our procedure are determined at the fund, rather than family, level. This fund-level determination, in turn, suggests the possibility of

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

risking shifting or tournament-style behavior (e.g., Brown, Harlow, and Starks (1996)) by funds with respect to their AS benchmarks: 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. In untabulated results, however, we find that changes in benchmark-discrepancy status are uncommon; just 5.6% of funds that do not have a benchmark discrepancy switch to having a benchmark discrepancy at mid-year, which suggests that the strategy, if used at all, is used rarely.

VII. Differences in Prospectus and AS Benchmark Returns

This section first considers whether the prospectus benchmark gives funds a "performance boost" when benchmark-adjusting returns. We answer that question by comparing the returns on the AS and prospectus benchmarks for funds with nonzero BENCHMARK_MISMATCH. We then consider how conclusions about a fund's performance can change depending on the benchmark.

A. Comparing Benchmark Returns

Table 3 shows the average difference in annualized return between the AS benchmark and the prospectus benchmark (i.e., the performance boost) for funds as a function of their ACTIVE_GAP and BENCHMARK_MISMATCH. We divide funds into five ranges of BENCHMARK_MISMATCH and five ranges of ACTIVE_GAP. Funds with BENCHMARK_MISMATCH of 0 are excluded from this analysis.

TABLE 3

Difference in Benchmark Returns as a Function of ACTIVE _GAP and BENCHMARK_MISMATCH

Table 3 shows the average difference in annualized return between the active share (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 preset cutoffs, and the average difference is reported for each of the resulting groups. The "High – Low" column reports the difference in the results between the BM > 80 and 0 < BM ≤ 20 groups, and the "High – Low" row reports the difference in results between the AG > 5 and 0 < AG ≤ 1.25 groups. *t*-statistics are reported in parentheses below each coefficient and are calculated using standard errors independently double clustered by fund and year-month.

	All	0 < BM ≤ 20	20 < BM ≤ 40	40 < BM ≤ 60	60 < BM ≤ 80	BM > 80	High – Low
All	0.59% (2.37)	−0.11% (−0.72)	−0.06% (−0.18)	0.49% (1.56)	1.27% (2.05)	1.68% (2.98)	1.79% (3.00)
0 < AG ≤ 1.25	0.36% (1.69)	−0.02% (−0.15)	0.01% (0.03)	0.22% (0.76)	0.93% (1.48)	1.48% (2.38)	1.50% (2.54)
1.25 < AG ≤ 2.5	0.69% (2.64)	−0.02% (−0.16)	−0.05% (−0.16)	0.84% (2.28)	1.53% (2.35)	1.93% (2.75)	1.95% (2.82)
2.5 < AG ≤ 3.75	0.71% (2.35)	−0.20% (−1.00)	−0.03% (−0.10)	0.89% (2.01)	1.67% (2.32)	1.24% (1.82)	1.44% (2.02)
3.75 < AG ≤ 5	0.60% (1.80)	−0.35% (−1.41)	−0.07% (−0.17)	0.80% (1.54)	1.22% (1.65)	1.28% (1.71)	1.63% (2.02)
AG > 5	0.65% (1.81)	−0.87% (−2.06)	−0.20% (−0.40)	0.08% (0.20)	0.99% (1.47)	1.99% (2.77)	2.86% (2.87)
High – Low	0.29% (0.92)	−0.85% (−2.25)	−0.21% (−0.53)	−0.14% (−0.34)	0.07% (0.13)	0.51% (0.67)	1.36% (1.43)

The average performance boost for funds with nonzero BENCHMARK_MISMATCH is 0.59% per year (t-statistic = 2.37).[8] 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 1.68% per year (t-statistic = 2.98), compared with −0.11% per year (t-statistic = −0.72) for funds with BENCHMARK_MISMATCH less than 20%. Once BENCHMARK_MISMATCH is greater than 60%, the average performance boost is economically large and statistically significant. Funds with BENCHMARK_MISMATCH between 60% and 80% have an average performance boost of 1.27% per year (t-statistic = 2.05). There is some evidence of a performance boost for funds with BENCHMARK_MISMATCH between 40% and 60%, but on average, the performance boost is economically smaller (0.49%) and statistically weaker (t-statistic = 1.56). The level of ACTIVE_GAP does not appear to affect the performance boost: Funds with ACTIVE_GAP less than 1.25% and funds with ACTIVE_GAP greater than 5% have a similar performance boost.[9]

We consider several robustness tests related to these results, with each available in the Supplementary Materials. First, we group funds more coarsely based on just the 60% BENCHMARK_MISMATCH cutoff, and the results are similar. When BENCHMARK_MISMATCH is greater than 60%, the average performance boost is 1.46% per year (t-statistic = 3.08). In comparison, when BENCHMARK_MISMATCH is less than 60%, the average performance boost is only 0.11% (t-statistic = 0.51). Second, our results from a fixed-effects panel regression indicate that BENCHMARK_MISMATCH, not ACTIVE_GAP, is the primary driver of the performance boost and that the effect of BENCHMARK_MISMATCH holds after accounting for AS. Third, we find that across all styles, except for value funds, the performance boost is economically large and statistically significant when BENCHMARK_MISMATCH is greater than 60%. Fourth, if we switch to using only funds with a Russell index as their prospectus benchmark and using only Russell indexes as potential AS benchmarks, the performance boost when BENCHMARK_MISMATCH is greater than 60% is still large and statistically significant. Fifth, the effectiveness of BENCHMARK_MISMATCH could potentially be negatively affected by problems in the holdings data (e.g., window dressing, nonequity holdings, and within-quarter trading), but we find that conditioning on past tracking error, which uses only returns and should avoid those issues, does not substantially influence the effect of BENCHMARK_MISMATCH. We conclude, based on the totality of these results, that when BENCHMARK_MISMATCH is large, the prospectus benchmark sets a lower bar for funds to clear. Given that a large

[8]Our t-statistics in this table are independently double-clustered by fund and year-month. We find this method of clustering to be more conservative (i.e., lower t-statistics) than independently double-clustering by benchmark-pair and year-month or single-clustering by benchmark-pair-year-month. To eliminate the clustering issue, we also, in untabulated analysis, formed long–short portfolios using paired benchmarks (weighted by frequency of occurrence) and found similar results.

[9]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 5% and BENCHMARK_MISMATCH less than 20% (0.17% of the tested sample), so we believe caution should be exercised in making inferences concerning those funds.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

BENCHMARK_MISMATCH is a common occurrence, a substantial number of funds operate using that lower bar.

B. Comparing Funds' Benchmark-Adjusted Returns

The previous section shows how, compared with the AS benchmark, the prospectus benchmark can set a lower bar for a fund if 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 Sections V and VI, high-AS funds, in comparison to low-AS funds, tend to have much greater BENCHMARK_MISMATCH. Cremers and Petajisto (2009) show that high-AS funds also have better average future performance, so it is critical to account for AS 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 0, into groups based on the prospectus benchmark AS and BENCHMARK_MISMATCH. Using AS, we sort funds into quintiles. Using BENCHMARK_MISMATCH, we sort funds into two groups based on the 60% cutoff. It is relatively rare for funds to exceed the 60% cutoff and still have low AS; therefore, to avoid reporting results for groups with very few funds, we collapse the four lowest quintiles of AS into a single group.

We evaluate average net fund performance within each group using two performance evaluation models: The prospectus-benchmark-adjusted return and the modified-benchmark-adjusted return. 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%. This 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.

Table 4 shows the results from this analysis. We look first at the "All Funds" portion of Table 4, which is not conditional on AS. Within this portion, funds with large BENCHMARK_MISMATCH on average outperform funds with small BENCHMARK_MISMATCH by 0.96% per year (t-statistic = 2.69) using the prospectus-benchmark-adjusted returns. Using the modified-benchmark-adjusted returns, however, there is no statistically significant difference in performance, with the funds with small BENCHMARK_MISMATCH actually outperforming by 0.47% per year (t-statistic = 1.38). Importantly, the positive relation between BENCHMARK_MISMATCH and AS prevents us from drawing any strong conclusions.

Therefore, we turn next to the results conditional on AS. Within the high-AS quintile and when using prospectus-benchmark-adjusted returns, average fund performance is statistically the same whether or not the funds have a benchmark discrepancy. The difference, which is to the advantage of the funds with a benchmark discrepancy, is 0.53% per year (t-statistic = 1.25). But if we compare

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

TABLE 4

Performance of Funds as a Function of BENCHMARK_MISMATCH and Active Share

Table 4 shows the performance of various groups of funds using two different models. To form the groups, the full sample of fund-months, including fund-months with matching prospectus and active share (AS) benchmarks, is sorted independently on prospectus benchmark AS and BENCHMARK_MISMATCH (BM). With respect to active share, funds are sorted into quintiles at the beginning of each month. Funds in the fifth quintile (highest AS) are tested separately from the other four quintiles. The difference in results between those groups is considered in the "Q5−Q1234" rows. With respect to BENCHMARK _MISMATCH, funds are sorted based on whether BENCHMARK_MISMATCH is greater than 60%. The difference in results between those groups is considered in the "Difference" column. The "Prospectus" model reports the annualized average of the differences between the fund net return and the prospectus benchmark return. The "Modified" model reports the annualized average of the 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 two models. *t*-statistics are reported in parentheses below each coefficient and are calculated using standard errors independently double-clustered by fund and year-quarter.

	Model	All	BM > 60%	BM ≤ 60%	Difference
All funds	Prospectus	−0.57%	0.17%	−0.79%	0.96%
		(−1.87)	(0.38)	(−2.65)	(2.69)
	Modified	−0.90%	−1.26%	−0.79%	−0.47%
		(−3.34)	(−3.88)	(−2.65)	(−1.38)
	Difference	0.33%	1.43%	0.00%	1.43%
		(3.08)	(3.09)	—	(3.09)
Prospectus AS quintile 5	Prospectus	0.25%	0.40%	−0.13%	0.53%
		(0.57)	(0.85)	(−0.27)	(1.25)
	Modified	−0.92%	−1.24%	−0.13%	−1.12%
		(−2.72)	(−3.52)	(−0.27)	(−2.52)
	Difference	1.16%	1.65%	0.00%	1.65%
		(3.39)	(3.39)	—	(3.39)
Prospectus AS quintiles 1, 2, 3, and 4	Prospectus	−0.77%	−0.20%	−0.85%	0.65%
		(−2.61)	(−0.42)	(−2.84)	(1.64)
	Modified	−0.89%	−1.28%	−0.85%	−0.44%
		(−3.16)	(−3.53)	(−2.84)	(−1.19)
	Difference	0.12%	1.09%	0.00%	1.09%
		(2.28)	(2.28)	—	(2.28)
Q5−Q1234	Prospectus	1.02%	0.60%	0.72%	−0.12%
		(3.33)	(2.14)	(1.87)	(−0.30)
	Modified	−0.02%	0.04%	0.72%	−0.68%
		(−0.07)	(0.13)	(1.87)	(−1.64)
	Difference	1.04%	0.56%	0.00%	0.56%
		(3.47)	(2.07)	—	(2.07)

performance using the modified-benchmark-adjusted returns instead, the relative performance of the funds without a benchmark discrepancy noticeably improves. The performance of the funds with a benchmark discrepancy decreases by 1.65% per year (*t*-statistic = 3.39) after switching from prospectus- to modified-benchmark-adjusted returns. As a result, when using the modified-benchmark-adjusted returns, the funds without a benchmark discrepancy outperform those with a benchmark discrepancy by 1.12% per year (*t*-statistic = 2.52). The results for the lower quintiles of AS are similar to those from the full sample: Switching from prospectus- to modified-benchmark-adjusted returns significantly changes the impression of the relative performance of funds with and without a benchmark discrepancy.

Notably, after switching to modified-benchmark-adjusted returns, the relation between AS and fund performance documented by Cremers and Petajisto (2009) only holds for funds without a benchmark discrepancy. Among funds with a benchmark discrepancy, the difference in modified-benchmark-adjusted returns between the top quintile and the other 4 quintiles is only 0.04% per year (*t*-statistic = 0.13), whereas among funds without a benchmark discrepancy, that difference is

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

0.72% per year (t-statistic = 1.87). This result suggests that funds without a benchmark discrepancy drive the relation between AS and fund performance. More generally, this result shows that when measuring fund activeness and its relations (e.g., Kacperczyk, Sialm, Zheng (2005), Amihud and Goyenko (2013), and Pastor, Stambaugh, and Taylor (2020)), researchers should, for correct inferences, rely on either their own evaluations or independent evaluations of a fund's style and portfolio characteristics rather than those supplied by the fund.

VIII. Comparison with Sensoy's Procedure

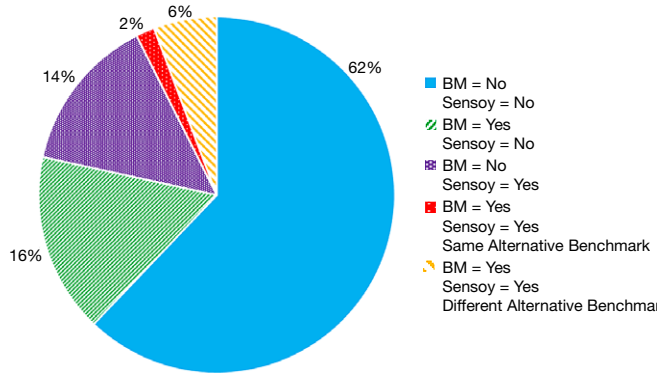
In Sections I and II, we discussed the details of Sensoy's (2009) benchmark-discrepancy-identification procedure, as well as the main differences between his procedure and ours. In this section, we compare the results from the two procedures.[10] We first compare the procedures in terms of which funds are labeled as having a benchmark discrepancy. We then consider the extent to which the choice of procedure affects performance evaluation.

Sensoy (2009) finds that 31.2% of his sample has a benchmark discrepancy over the period 1994–2004. Applying Sensoy's procedure in our sample, which covers the period 1991–2015, we find that 21.6% of funds have a benchmark discrepancy. Figure 1 shows a more specific comparison between Sensoy's procedure and our procedure. The pie chart shows the commonality in fund-month

FIGURE 1

Overlap between Sensoy and BENCHMARK_MISMATCH Identification Procedures

Figure 1 shows the percentage of the full sample of fund-months identified as having a benchmark discrepancy using two different procedures. In the BENCHMARK_MISMATCH (BM) procedure, a fund is considered to have a benchmark discrepancy if BENCHMARK_MISMATCH is greater than 60%. In the Sensoy (2009) 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.



[10]Although 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 III. Our set of benchmarks includes all of those in Sensoy's set.

observations identified as having a benchmark discrepancy. In this analysis and in subsequent comparisons, we consider a benchmark discrepancy to exist by our procedure only if BENCHMARK_MISMATCH is greater than 60%.[11] The observations are at the fund-month level because whether a fund has a benchmark discrepancy is a time-varying factor in our procedure.

Approximately 62% of fund-month observations do not have a benchmark discrepancy using either procedure, approximately 16% have a benchmark discrepancy only according to our procedure, and approximately 14% have a benchmark discrepancy only according to Sensoy's (2009) procedure. Just 2% of observations have a benchmark discrepancy according to both procedures that results in the same alternative benchmark, with another 6% 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. Approximately 63% of the benchmark discrepancies in Sensoy's (2009) original analysis come from funds with an S&P 500 prospectus benchmark. In comparison, only 10% of the benchmark discrepancies identified using our procedure come from funds with an S&P 500 prospectus benchmark. Furthermore, in Sensoy's original analysis, approximately 73% of the funds with both an S&P 500 prospectus benchmark and a benchmark discrepancy have either a Russell 1000 Value or a Russell 1000 Growth alternative benchmark. Those benchmark combinations each have an average BENCHMARK_MISMATCH near 45%, well below our 60% cutoff. 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 (2009) benchmarks. Table 5 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., a positive difference between the AS or Sensoy benchmark return and the prospectus benchmark return) for funds with a benchmark discrepancy. Our procedure, however, finds a significantly larger performance boost compared with Sensoy's (and one that is statistically significant). Funds with a benchmark discrepancy according to our procedure have an average performance boost of 1.48% per year (t-statistic = 3.01), whereas funds with a benchmark discrepancy according to Sensoy's procedure have an average performance boost of 0.68% per year (t-statistic = 1.37).

More importantly, when our procedure identifies a benchmark discrepancy and Sensoy's (2009) 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 1.31% per year (t-statistic = 2.20). Funds identified by Sensoy's procedure, but not by ours, have an average performance boost of only 0.18% per year (t-statistic = 0.40). These results, as a whole, indicate that our

[11] We use BENCHMARK_MISMATCH as a continuous measure in the Supplementary Material and find results consistent with those presented in this section.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

TABLE 5

Comparison of the Sensoy and BENCHMARK_MISMATCH Identification Procedures with Respect to the Difference in Benchmark Returns

Table 5 shows the average annualized benchmark-adjusted net fund return for fund-months identified as having a benchmark discrepancy using two different procedures. In the BENCHMARK_MISMATCH (BM) procedure, a fund is considered to have a benchmark discrepancy if BENCHMARK_MISMATCH is greater than 60%. In the Sensoy (2009) 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 active share (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. *t*-statistics are reported in parentheses below each coefficient and are calculated using standard errors independently double-clustered by fund and year-month.

Benchmark Discrepancy?	BM / Sensoy	Yes / —	— / Yes	Yes / No	No / Yes	Yes / Yes
Benchmark-adjusted return	Prospectus	0.24% (0.50)	−0.44% (−0.78)	0.03% (0.05)	−0.98% (−1.96)	0.69% (0.90)
	AS	−1.24% (−3.55)	—	−1.28% (−3.47)	—	−1.11% (−2.91)
	Sensoy	—	−1.12% (−3.98)	—	−1.16% (−4.20)	−1.05% (−2.80)
Differences	Prospectus − AS	1.48% (3.01)	—	1.31% (2.20)	—	1.80% (2.65)
	Prospectus − Sensoy	—	0.68% (1.37)	—	0.18% (0.40)	1.73% (2.56)

procedure's benchmark discrepancies have a greater impact on fund performance evaluation.[12]

As an alternative procedure based exclusively on fund and benchmark returns, we also consider using funds' tracking errors to identify benchmark discrepancies. In particular, we compare funds' prospectus-benchmark tracking errors over the previous 36 months with funds' lowest tracking errors over that period among our benchmark set. We identify benchmark discrepancies by applying different thresholds for the minimum difference between those values. Unlike Sensoy's (2009) procedure, this procedure avoids look-ahead bias and dynamically updates whether a fund has a benchmark discrepancy. Relative to our BENCHMARK_MISMATCH procedure, the main advantage of this exclusively return-based procedure is that it is not subject to concerns related to holdings (e.g., window dressing, nonequity positions, and trading between quarters). As shown in the Supplementary Material, the funds identified as having a benchmark discrepancy by our BENCHMARK_MISMATCH procedure, but not by the tracking-error procedures, tend to have economically large and statistically significant performance boosts, whereas those identified by the tracking-error procedures, but not by our BENCHMARK_MISMATCH procedure, do not. Consequently, as in the Sensoy-procedure comparison, we conclude that our BENCHMARK_MISMATCH procedure is more useful in identifying funds for which benchmark discrepancies have a substantial impact on performance evaluation.

[12]Elton et al. (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 1.31% per year (*t*-statistic = 1.77), whereas funds identified by their procedure, but not by ours, have an average performance boost of only 0.38% per year (*t*-statistic = 1.34).

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

IX. Factor Exposures of the Prospectus and AS benchmarks

A. Comparing Benchmark Returns

The expected return on a passively managed index should be driven solely by factor exposures because passive indexes, by construction, have no alpha (arguably, because index-reconstitution frictions exist; 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. Because 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 analyze factor-exposure differences between the two benchmarks for funds with a benchmark discrepancy.

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

$$\text{RETURN}_{\text{AS},t} - \text{RETURN}_{\text{pro},t} = \beta \times \text{FACTOR}_t + \varepsilon_t, \tag{5}$$

where $\text{RETURN}_{\text{AS},t}$ is the annualized return on the AS benchmark averaged across all tested funds in month t, and $\text{RETURN}_{\text{pro},t}$ 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 et al. (2012) 7-factor model (henceforth, the CPZ7 model). That model focuses on the traditional market, size, value, and momentum exposures.[13] The expanded model (henceforth, the CPZ7+ model) also includes the Fama and French (2015) profitability and investment factors; the Stambaugh and Yuan (2017) management and performance factors; the Frazzini and Pedersen (2014) betting-against-beta factor; the Asness, Frazzini, and Pedersen (2019) quality-minus-junk factor; and the Pastor and Stambaugh (2003) traded liquidity factor.[14] We intentionally exclude a constant from the model because the model should explain the return differences between the two benchmarks by just the differences in factor exposures (i.e., there should not be any alpha), but our results are similar if a constant is included (see the Supplementary Material).[15] The model is estimated using all funds with BENCHMARK_MISMATCH greater than 60% and across various investment style subgroups.

[13]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. The small number of funds whose two benchmarks correspond to a CPZ7 factor are dropped in these tests.

[14]Cochrane (2011, page 1047) remarks that there is now "a zoo of new factors." Rather than attempting to test all potential new factors, we focus on the explanatory power of this particular subset because each has either received a particularly large amount of attention or been shown to be particularly robust (in, e.g., Feng, Giglio, and Xiu (2020)).

[15]The resulting alphas from including a constant in the CPZ7 model are, across all styles, economically large and statistically significant; conversely, the resulting alphas from including a constant in the CPZ7+ model are, with the exception of the large-cap style, economically small and statistically insignificant.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

TABLE 6

Factor Exposure Differences Between the AS and Prospectus Benchmarks

Table 6 shows results from the following model:

$$\text{RETURN}_{\text{AS},t} - \text{RETURN}_{\text{pro},t} = \beta * \text{FACTOR}_t + \varepsilon_t,$$

where $\text{RETURN}_{\text{AS},t}$ is the average annualized return on the active share (AS) benchmarks in month t across all funds with BENCHMARK_MISMATCH greater than 60%. $\text{RETURN}_{\text{pro},t}$ 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; see the text for details. 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) and using different combinations of factors. In Panel A, the "Prospectus" ("AS") row reports the average of the annualized monthly average differences between the fund net returns and the prospectus (AS) benchmark returns. The "Difference" row reports the differences between the "Prospectus" and "AS" rows. In Panel B, we only use the Cremers et al. (2012) 7 factors in the model and report the total factor return (i.e., the sum of the products of the estimated factor exposures and the annualized factor returns) and the R^2 for the model. Panel C repeats Panel B, but it uses the full set of factors (CPZ7+). t-statistics associated with tests of whether the values are different from 0 are reported in parentheses below the values.

Style	All	Large	Small/Mid	Growth	Value	Blend
Panel A. Benchmark Return Difference by Style						
Prospectus	0.75%	1.28%	0.30%	0.87%	−0.41%	0.66%
	(1.37)	(0.90)	(0.60)	(1.25)	(−0.60)	(1.09)
AS	−0.66%	−1.21%	−0.71%	−0.50%	−1.18%	−1.05%
	(−1.34)	(−2.00)	(−1.32)	(−0.61)	(−2.29)	(−2.08)
Difference	1.41%	2.49%	1.01%	1.37%	0.77%	1.71%
	(3.55)	(2.17)	(1.87)	(1.93)	(1.24)	(3.36)
Panel B. Explanatory Power: CPZ7 Model						
R^2	23.0%	81.4%	46.0%	54.6%	39.7%	39.1%
Total factor return	0.48%	1.41%	0.31%	0.56%	−0.12%	0.74%
	(2.49)	(1.36)	(0.85)	(1.05)	(−0.30)	(2.30)
Panel C. Explanatory Power: CPZ7+ Model						
R^2	49.9%	85.6%	55.1%	64.0%	65.7%	55.8%
Total factor return	1.25%	1.90%	1.09%	1.27%	0.80%	1.49%
	(4.49)	(1.79)	(2.72)	(2.23)	(1.59)	(3.95)

Panel A of Table 6 reports average annualized benchmark-adjusted returns using both benchmarks. The difference in those returns indicates the economic magnitude of the performance boost. Panels B and C report the R^2 from the base model and expanded model, respectively, and the sum of the products of the estimated factor exposures and the annualized factor returns. The resulting "Total Factor Return" can be compared to the average difference in benchmarks returns to determine how much of the return difference is explained by factor-exposure differences. The coefficients associated with the individual factors are available in the Supplementary Material.

Using all funds with a benchmark discrepancy, the traditional factors in the base model explain 0.48% (t-statistic = 2.49) of the 1.41% per year difference in average returns between the prospectus and AS benchmarks. Put another way, exposures to the traditional factors explain just 34% (= 0.48/1.41) of the difference in average returns between the benchmarks.

That result indicates that the prospectus and AS benchmarks differ from each other across dimensions that are unrelated to traditional factors. Hence, we next consider the expanded model (CPZ7+) that contains some "nontraditional" factors. Once these nontraditional factors are included, the difference in average returns that can be explained considerably increases. Using all funds with a benchmark discrepancy, 1.25% (t-statistic = 4.49) of the 1.41% per year difference in average returns can be explained by the expanded model (i.e., 89% of the difference is

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

explained). The R^2 more than doubles when switching from the CPZ7 model to the CPZ7+ model, increasing from 23.0% to 49.9%. Untabulated analysis indicates that the profitability factor adds the largest new explanatory power. Within each of the style subgroups, the proportion of the difference in average returns that is explained by the model increases substantially when the nontraditional factors are included. On average, approximately 94% of the difference in average returns is explained.[16]

B. Comparing Funds' Benchmark-Adjusted Returns

We next test whether, when a benchmark discrepancy exists, the AS benchmark or the prospectus benchmark more accurately reflects funds' actual factor exposures. The previous results show that compared with prospectus benchmarks, funds with benchmark discrepancies have AS benchmarks with greater average net factor exposure. 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:

$$\text{RETURN}_{\text{fund},t} - \text{RETURN}_{\text{bench},t} = \alpha + \beta \times \text{FACTOR}_t + \varepsilon_t, \tag{6}$$

where $\text{RETURN}_{\text{fund},t}$ is the average annualized net return across all tested funds in month t, and $\text{RETURN}_{\text{bench},t}$ is the average annualized return on those funds' benchmarks in month t. We consider both the prospectus benchmark and the AS benchmark 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 nontraditional 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% (Panel A of Table 7) and for the subgroup of those funds with a large-cap style (Panel B). That subgroup is of particular interest because our tests in Table 6 indicate that large-cap funds have the largest difference in average returns between their prospectus and AS benchmarks.

The results from this model are presented in Table 7. Using all funds with a benchmark discrepancy, the prospectus-benchmark-adjusted returns are unaffected by the traditional factors. Adjusting for those factors reduces the abnormal return by only 0.18% per year (t-statistic = −0.39). When considering nontraditional factors,

[16]We replicate this analysis in the Supplementary Material using the Carhart (1997) factors in the base model instead of the CPZ7 factors. Overall, we find weaker results, which is consistent with that model having relatively poor power with respect to common benchmarks (Cremers et al. (2012)); however, after we include the nontraditional factors, we still find a substantial increase in the proportion of the difference in average returns explained.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

TABLE 7

Prospectus- and AS-Benchmark-Adjusted Returns Evaluated with Factor Models

Table 7 shows results from the following model:

$$\text{RETURN}_{\text{fund},t} - \text{RETURN}_{\text{bench},t} = \alpha + \beta \times \text{FACTOR}_t + \varepsilon_t,$$

where $\text{RETURN}_{\text{fund},t}$ is the average annualized net return in month t across all tested funds, and $\text{RETURN}_{\text{bench},t}$ 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 et al. (2012) 7-factor model are included. In the columns with the heading "CPZ7+," all of the nontraditional factors 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 parentheses below each value.

		CPZ7		CPZ7+	
	RETURN	Alpha	Change	Alpha	Change
Panel A. BENCHMARK_MISMATCH > 60%					
Prospectus adjusted	0.63%	0.45%	−0.18%	−0.28%	−0.91%
	(1.27)	(0.97)	(−0.39)	(−0.67)	(−2.19)
AS adjusted	−0.76%	−0.43%	0.32%	−0.24%	0.52%
	(−1.86)	(−1.14)	(0.85)	(−0.55)	(1.22)
Difference	1.39%	0.88%	−0.50%	−0.04%	−1.43%
	(3.27)	(2.15)	(−1.23)	(−0.12)	(−3.88)
Panel B. BENCHMARK MISMATCH > 60% and Large-Cap Style					
Prospectus adjusted	1.44%	0.11%	−1.33%	−0.38%	−1.82%
	(1.14)	(0.18)	(−2.12)	(−0.61)	(−2.95)
AS adjusted	−1.32%	−1.16%	0.16%	−0.90%	0.42%
	(−2.74)	(−2.44)	(0.33)	(−1.82)	(0.86)
Difference	2.76%	1.27%	−1.49%	0.52%	−2.24%
	(2.35)	(2.84)	(−3.32)	(1.18)	(−5.11)

however, the prospectus benchmark significantly understates net factor exposure. The abnormal return based on prospectus-benchmark-adjusted returns decreases from 0.63% per year using no factors to −0.28% per year after using both the traditional and nontraditional factors. This large change in performance, −0.91% per year (t-statistic = −2.19), 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 nontraditional factors in the model, the abnormal return is not significantly altered. The change when switching from no factors to all factors is 0.52% per year, which is not statistically significant at conventional levels (t-statistic = 1.22). That is, if fund performance is adjusted using the AS benchmark, only relatively minor net factor exposure remains.[17]

These outcomes are magnified if we focus on just large-cap funds with a benchmark discrepancy. In that case, the abnormal return based on prospectus-benchmark-adjusted returns decreases by 1.33% per year (t-statistic = −2.12) after

[17]Although 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.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

adjusting for the CPZ7 factors and 1.82% per year (t-statistic = −2.95) after further adjusting for the nontraditional factors. These results indicate that large-cap funds with benchmark discrepancies have prospectus benchmarks that understate net exposure to both traditional and nontraditional factors.[18] In comparison, the AS-benchmark-adjusted returns are generally unaffected by both the traditional and nontraditional 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 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 than using the prospectus benchmark.

X. Fund 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 a benchmark discrepancy, 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:

$$\text{FLOW}_{i,t} = \theta + \beta \times \text{PERFORMANCE}_{i,t} + \gamma \times \text{MISMATCH}_{i,t} + \delta \times \text{CHARS}_{i,t} + \text{FE} + \varepsilon_{i,t}, \tag{7}$$

where $\text{FLOW}_{i,t}$ is the percentage implied net flow for fund i in month t. $\text{PERFORMANCE}_{i,t}$ 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 between fund i's AS benchmark return and prospectus benchmark return, and fund i's annualized capital asset pricing model (CAPM) alpha (results are robust to alternative alphas, e.g., Carhart (1997)). 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 0 to 1 based on that ranking (with 1 being the best performing fund). Using the rankings allows for a more natural test of a potential nonlinear relation between fund performance and net flows (e.g., Chevalier and Ellison (1997), Sirri and Tufano (1998)).

[18] If we perform this same analysis using all styles except large cap, the abnormal return based on prospectus-benchmark-adjusted returns decreases by 0.71% per year (t-statistic = −1.93) after adding both the traditional and nontraditional factors to the model.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

TABLE 8

Response of Investor Flows to Different Measures of Performance

Table 8 shows results from the following model:

$$FLOW_{i,t} = \theta + \beta \times PERFORMANCE_{i,t} + \gamma \times MISMATCH_{i,t} + \delta \times CHARS_{i,t} + FE + \varepsilon_{i,t},$$

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. $MISMATCH_{i,t}$ 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. FE represents style and year-month fixed effects. The model is estimated using the sample of fund-months with different prospectus and active share (AS) benchmarks. t-statistics are reported in parentheses below each coefficient and are calculated using standard errors independently double-clustered by fund and year-month.

	1	2	3	4	5	6
FUND_RETURN – AS_RETURN	0.11	0.13	0.08	0.13	1.89	1.33
	(24.80)	(25.65)	(16.47)	(14.85)	(20.41)	(6.11)
AS_RETURN – PROSPECTUS_RETURN		0.07	0.04	0.13	0.70	1.90
		(15.20)	(9.37)	(11.06)	(10.88)	(9.17)
CAPM_ALPHA			0.07		1.86	
			(14.84)		(18.07)	
BENCHMARK_MISMATCH				−0.003		
				(−2.43)		
(FUND_RETURN – AS_RETURN) × BM				0.000		
				(0.17)		
(AS_RETURN – PROSPECTUS_RETURN) × BM				−0.001		
				(−6.02)		
(FUND_RETURN – AS_RETURN)^2						1.73
						(7.64)
(AS_RETURN – PROSPECTUS_RETURN)^2						−0.79
						(−3.95)
Returns	Actual	Actual	Actual	Actual	Ranking	Ranking
Characteristic controls	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
No. of obs.	147,883	147,883	147,883	147,883	147,883	147,883

$MISMATCH_{i,t}$ 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 equation (4). 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), but the results are similar using all fund-months.

Table 8 contains the 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 (i.e., 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 (t-statistic = 15.20). 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. Although investors are influenced by multiple measures of performance, these results indicate that performance relative to the prospectus benchmark is an important determinant of investors' capital allocations.

As the magnitude of BENCHMARK_MISMATCH increases, we would expect the importance of the prospectus benchmark to decrease. In the fourth

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

column, we interact our performance measures with BENCHMARK_MISMATCH and find that as BENCHMARK_MISMATCH increases, flows are less sensitive to the performance boost. An increase in BENCHMARK_MISMATCH of 10% reduces the base impact of the performance boost by 8% (= 0.001 × (10)/0.13). Conversely, increases in BENCHMARK_MISMATCH have no impact on the weight investors give to performance relative to the AS benchmark. Although these results indicate some sophistication on the part of investors, even when BENCHMARK_MISMATCH is 100%, the performance boost still has an economically large and statistically significant impact on flows.

The final columns consider nonlinearity 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.

XI. 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 although 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.

More broadly, fund sponsors and fund managers could have incentives to allow persistent benchmark discrepancies. 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 nontraditional factors. Regardless of the reason, because investors appear to rely heavily on the benchmark in the prospectus, benchmark discrepancies should be a significant concern.

Supplementary Material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0022109021000119.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

References

Agarwal, V.; C. Green; and H. Ren. "Alpha or Beta in the Eye of the Beholder: What Drives Hedge Fund Flows?" *Journal of Financial Economics*, 127 (2018), 417–434.

Amihud, Y., and R. Goyenko. "Mutual Fund's R^2 as Predictor of Performance." *Review of Financial Studies*, 26 (2013), 667–694.

Asness, C.; A. Frazzini; and L. Pedersen. "Quality Minus Junk." *Review of Accounting Studies*, 24 (2019), 34–112.

Barber, B.; X. Huang; and T. Odean. "Which Factors Matter to Investors? Evidence from Mutual Fund Flows." *Review of Financial Studies*, 29 (2016), 2600–2642.

Ben-David, I.; J. Li; A. Rossi; and Y. Song. "What Do Mutual Fund Investors Really Care About?" Working Paper, Ohio State University (2019).

Berk, J., and J. van Binsbergen. "Assessing Asset Pricing Models Using Revealed Preference." *Journal of Financial Economics*, 119 (2016), 1–23.

Blocher, J., and M. Molyboga. "The Revealed Preference of Sophisticated Investors." *European Financial Management*, 23 (2017), 839–872.

Bollen, N., and J. Busse. "On the Timing Ability of Mutual Fund Managers." *Journal of Finance*, 56 (2001), 1075–1094.

Brown, K.; W. V. Harlow; and L. Starks. "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry." *Journal of Finance*, 51 (1996), 85–110.

Brown, S., and W. Goetzmann. "Mutual Fund Styles." *Journal of Financial Economics*, 43 (1997), 373–399.

Cao, C.; P. Iliev; and R. Velthuis. "Style Drift: Evidence from Small-Cap Mutual Funds." *Journal of Banking and Finance*, 78 (2017), 42–57.

Carhart, M. "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52 (1997), 57–82.

Chevalier, J., and G. Ellison. "Risk Taking by Mutual Funds as a Response to Incentives." *Journal of Political Economy*, 105 (1997), 1167–1200.

Cochrane, J. "Presidential Address: Discount Rates." *Journal of Finance*, 66 (2011), 1047–1108.

Cremers, M., and Q. Curtis. "Do Mutual Fund Investors Get What They Pay For? The Legal Consequences of Closet Index Funds." *Virginia Law and Business Review*, 11 (2016), 31–93.

Cremers, M., and A. Petajisto. "How Active Is Your Fund Manager? A New Measure That Predicts Performance." *Review of Financial Studies*, 22 (2009), 3329–3365.

Cremers, M.; A. Petajisto; and E. Zitzewitz. "Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation." *Critical Finance Review*, 2 (2012), 1–48.

Daniel, K.; M. Grinblatt; S. Titman; and R. Wermers. "Measuring Mutual Fund Performance with Characteristic-Based Benchmarks." *Journal of Finance*, 52 (1997), 1035–1058.

Del Guercio, D., and P. Tkac. "Star Power: The Effect of Morningstar Ratings on Mutual Fund Flow." *Journal of Financial and Quantitative Analysis*, 43 (2008), 907–936.

DiBartolomeo, D., and E. Witkowski. "Mutual Fund Misclassification: Evidence Based on Style Analysis." *Financial Analysts Journal*, 53 (1997), 32–43.

Elton, E.; M. Gruber; and C. Blake. "The Performance of Separate Accounts and Collective Investment Trusts." *Review of Finance*, 18 (2014), 1717–1742.

Evans, R. "Mutual Fund Incubation." *Journal of Finance*, 65 (2010), 1581–1611.

Evans, R., and Y. Sun. "Models or Stars: The Role of Asset Pricing Models and Heuristics in Investor Risk Adjustment." Working Paper, University of Virginia (2018).

Fama, E., and K. French. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics*, 116 (2015), 1–22.

Feng, G.; S. Giglio; and D. Xiu. "Taming the Factor Zoo: A Test of New Factors." *Journal of Finance*, 75 (2020), 1327–1370.

Frazzini, A., and L. Pedersen. "Betting against Beta." *Journal of Financial Economics*, 111 (2014), 1–25.

Huang, J.; C. Sialm; and H. Zhang. "Risk Shifting and Mutual Fund Performance." *Review of Financial Studies*, 24 (2011), 2575–2616.

Kacperczyk, M.; C. Sialm; and L. Zheng. "On the Industry Concentration of Actively Managed Equity Mutual Funds." *Journal of Finance*, 60 (2005), 1983–2011.

Ma, L.; Y. Tang; and J. Gomez. "Portfolio Manager Compensation in the U.S. Mutual Fund Industry." *Journal of Finance*, 74 (2019), 587–638.

Pastor, L., and R. Stambaugh. "Liquidity Risk and Expected Stock Returns." *Journal of Political Economy*, 111 (2003), 642–685.

Pastor, L.; R. Stambaugh; and L. Taylor. "Fund Tradeoffs." *Journal of Financial Economics*, 138 (2020), 614–634.

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press

Sensoy, B. "Performance Evaluation and Self-Designated Benchmark Indexes in the Mutual Fund Industry." *Journal of Financial Economics*, 92 (2009), 25–39.
Sharpe, W. "Asset Allocation: Management Style and Performance Measurement." *Journal of Portfolio Management*, 18 (1992), 7–19.
Sirri, E., and P. Tufano. "Costly Search and Mutual Fund Flows." *Journal of Finance*, 53 (1998), 1589–1622.
Stambaugh, R., and Y. Yuan. "Mispricing Factors." *Review of Financial Studies*, 30 (2017), 1270–1315.
Wermers, R. "Matter of Style: The Causes and Consequences of Style Drift in Institutional Portfolios." Working Paper, University of Maryland (2012).

https://doi.org/10.1017/S0022109021000119 Published online by Cambridge University Press