# Performance Evaluation and Self-Designated Benchmark Indexes in the Mutual Fund Industry

Berk A. Sensoy*

January 31, 2008

*University of Southern California, Marshall School of Business. I am grateful to Steve Kaplan, Gene Fama, Toby Moskowitz, and Josh Rauh for their guidance and support. For helpful comments and discussions, I thank Ola Bengtsson, Daniel Bergstresser, John Chalmers, Linda DeAngelo, Diane DelGuercio, Wayne Ferson, Raife Giovinazzo, Chris Jones, Mark Klebanov, Kevin J. Murphy, Lubos Pastor, Francisco Perez-Gonzalez, Jon Reuter, Morten Sorensen, Laura Starks, Amir Sufi, Peter Tufano, Mark Westerfield, and especially Harry DeAngelo and an anonymous referee. I also appreciate the comments of seminar participants at Arizona State, Chicago, Harvard Business School, Oregon, Rochester, USC, and Washington University. I thank Morningstar, Inc. and Don Phillips and Annette Larson for generously providing data.

Contact address: 3670 Trousdale Pkwy #308, MC-0804, Los Angeles, CA 90089; phone: (213) 740-6497; email: bsensoy@usc.edu; webpage: http://www-rcf.usc.edu/~bsensoy.

### Abstract

Almost one-third of actively managed, diversified U.S. equity mutual funds specify a size and value/growth benchmark index in the fund prospectus that does not match the fund's actual style. Nevertheless, these "mismatched" benchmarks matter to fund investors. Performance relative to the specified benchmark is a significant determinant of a fund's subsequent cash inflows, even controlling for performance measures that better capture the fund's style. These incremental flows appear unlikely to be rational responses to abnormal returns. The evidence is consistent with the notion that mismatched self-designated benchmarks result from strategic fund behavior driven by the incentive to improve flows.

# Introduction

Performance evaluation theory stresses the importance of using good benchmarks (Holmstrom, 1979). For example, when determining an airline CEO's bonus, comparing the firm's performance to that of other airlines can improve efficiency by helping to filter out common shocks that are beyond the CEO's control. It would be less efficient to use railroads as the benchmark instead because shocks to the two industries are not perfectly correlated, yet the CEO has an incentive to encourage the use of a railroad benchmark if he believes that airlines are likely to outperform railroads. Of course, the attempt is unlikely to succeed in this setting because a knowledgable corporate board of directors will realize that railroads are not the best benchmark.

In other settings, however, performance evaluation is undertaken by less sophisticated principals than corporate boards. These principals may have limited ability to distinguish useful benchmarks from less useful ones, which may in turn create incentives for agents to try to strategically influence which benchmark is used. There is little systematic evidence on these issues because it is difficult to observe agents' preferred benchmarks and whether principals pay attention to them.

This paper provides such evidence from the mutual fund industry, in which fund investors take the role of unsophisticated principal. Funds' preferred benchmarks are available as a result of the SEC requirement that each fund's prospectus tabulate the fund's historical returns alongside those of a passive benchmark index. The SEC does not regulate which index is used as the benchmark, instead leaving the choice to the fund.

This institutional setting maps naturally into the general issues mentioned above. Some funds' self-designated benchmarks may not do a very good job capturing their exposures to common factors in returns, and so may not be very helpful in evaluating funds' skill at generating abnormal returns. Moreover, at least some mutual fund investors may not be sophisticated enough to see through this when making decisions about purchases and sales of mutual funds, and thereby may not behave in a manner consistent with theories of optimal performance evaluation such as Holmstrom (1979). If so, such "mismatched" benchmarks might make sense to funds from a strategic perspective (Gibbons and Murphy, 1990).

I use a new database of these self-designated mutual fund benchmark indexes to present evidence consistent with all of these possibilities. While this paper is about performance evaluation in the mutual fund industry, which is important in its own right because of the industry's size and importance to the economy, the evidence contributes more generally to the literature on the efficiency and incentive consequences of performance evaluation schemes (e.g. Ehrenberg and Bognanno, 1990).

The evidence also contributes to three major branches of the mutual fund literature: that on how mutual fund managers are and should be evaluated (e.g. Kothari and Warner, 2001; Cohen, Coval, and Pastor, 2005; Warner and Wu, 2005); that on the determinants of mutual fund flows (e.g. Chevalier and Ellison, 1997; Sirri and Tufano, 1998); and that on strategic behavior by mutual funds (e.g. Brown, Harlow, and Starks, 1996; Chevalier and Ellison, 1997, 1999). Especially relevant is the literature that, like this paper, shows that mutual fund flows appear at times to respond (or fail to respond) in irrational ways. Such papers include Musto (1999), Elton, Gruber, and Busse (2004), Cooper, Gulen, and Rau (2005), and Cronqvist (2006).

I begin by showing that the vast majority of actively managed, diversified U.S. equity funds use an S&P or Russell benchmark index that is defined on size and value/growth dimensions. Because Fama and French (1992) and many others find that size and value/growth are associated with average returns and return covariances, for such a benchmark to be maximally useful in netting out priced common factors in returns, it should match the fund's exposure to size and value/growth factors. Yet this is frequently not the case.

In fact, 31.2% of these funds specify a benchmark index that is "mismatched": alternative S&P or Russell size and value/growth-based benchmarks both better match these funds' size and value/growth characteristics and, more importantly, are more correlated with their returns. I refer to these as funds' "corrected" benchmarks. Among these funds, the average excess return $R^2$ with the actual benchmark is 70.6%, versus 82.6% with the corrected benchmark.

I then ask whether mismatched self-designated benchmarks influence fund flows. Do fund investors respond to performance relative to a mismatched benchmark when making decisions about purchases and sales of mutual funds? For this to happen, at least some investors must pay attention to the information in fund prospectuses. According to a recent survey by the Investment Company Institute, the national association of investment companies, 34% of fund investors consult the fund prospectus before purchasing a mutual fund.[1] This figure seems large enough to plausibly

[1] "Understanding Investor Preferences for Mutual Fund Information", August 2006, available at

have an effect on flows, especially considering that the performance table is prominently displayed in the first few pages of the prospectus. Fund advertising also frequently features a comparison of the fund's performance with that of a benchmark (when the comparison is favorable).

In fact, fund investors do pay attention to mismatched benchmarks when directing flows. A fund's performance relative to its self-designated but mismatched benchmark is a significant determinant of its subsequent cash inflows, even controlling for performance measures that better capture the fund's exposure to size and value/growth factors in returns. This is especially true for funds that beat those mismatched benchmarks. This result is robust to a variety of controls and specifications of functional form intended to capture nonlinearities in the relation between flows and performance (Chevalier and Ellison, 1997; Sirri and Tufano, 1998). In particular, the effect is not due to investors simply comparing performance to the S&P 500 regardless of the actual self-designated benchmark.

How should we interpret these results on flows? Is the response of flows to performance relative to a mismatched self-designated benchmark more likely to reflect rational or irrational behavior on the part of fund investors? From a performance evaluation/contracting perspective, because mutual funds generally receive a fixed percentage of assets under management as a fee, cash inflows and outflows are the mechanism by which fund investors (principals) influence fund companies' (agents) compensation. As such, agency theory (e.g. Holmstrom 1979) predicts that investors ought to direct flows in response to risk-adjusted return. Doing so aligns fund companies' desire for increased compensation, which gives them the incentive to take actions to increase flows, with fund investors' interest, maximizing risk-adjusted return.

Thus, the acid test for the interpretation of these flows is whether mismatched benchmarks have incremental power to explain the cross-section of expected returns, and thereby help measure risk-adjusted returns. While one cannot completely rule out this possibility because the pricing kernel is unobservable to the econometrician, pricing tests, documented in detail in section 3.2, suggest that it is unlikely.

As such, it appears unlikely (but cannot be completely ruled out) that the incremental response of flows to performance relative to a mismatched benchmark is a rational response to abnormal returns. I believe the evidence more likely reflects a behavioral element to the composition of

http://ici.org/statements/res/1rpt_06_inv_prefs_full.pdf.

mutual fund flows, consistent with Musto (1999), Elton, Gruber, and Busse (2004), Cooper, Gulen, and Rau (2005), and Cronqvist (2006).

I estimate that the magnitude of the expected incremental gain in flows to funds with mismatched self-designated benchmarks is 2.3% of assets under management per year, which is 14.6% of the average annual flow to those funds (15.8% of assets).

These incremental flows create strategic incentives for funds to self-designate mismatched benchmarks in the first place: mismatched benchmarks can improve funds' expected flows. Several pieces of evidence suggest that mismatched self-designated benchmarks may reflect funds' strategic incentives. First, mismatched self-designated benchmarks are not typically a result of style drift or changing fund styles and so do not appear incidental. Second, value funds are more likely than growth funds to have self-designated benchmarks that are mismatched on value/growth and small-cap funds are more likely than large-cap funds to have self-designated benchmarks that are mismatched on size. These findings are consistent with fund attempts to improve expected flows by taking advantage of the size and value effects documented by Banz (1981) and Fama and French (1992), among others. Third, mismatched self-designated benchmarks are more common among large and high-fee funds, to which the benefit from a given increase in flows (defined as a percentage of assets under management) is larger. Finally, fund family effects are significant determinants of whether a fund has a mismatched self-designated benchmark, again suggesting that mismatched benchmarks are not incidental or random.

Overall, the evidence in this paper further emphasizes the need, recently stressed by Goetzmann, Ingersoll, Spiegel, and Welch (2007), for the development and dissemination of measures of mutual fund performance that are both well-grounded in economic theory and not subject to gaming.

As mentioned above, this paper is related and contributes to several strands of literature. It is related to the body of work studying (possibly naive) consumer decisions in the mutual fund industry and the determinants of mutual fund flows. In addition to the papers cited above, Ippolito (1992) and Lynch and Musto (2003) are prominent papers in this literature. The strategic interpretation of mismatched benchmarks is consistent with evidence of strategic fund behavior in other contexts documented by, among others, Brown, Harlow, and Starks (1996), Chevalier and Ellison (1997,1999), Brown and Goetzmann (1997), Cooper, Gulen, and Rau (2005), and Cohen and Schmidt (2007).

The frequency of mismatched self-designated benchmarks is consistent with Elton, Gruber, and Blake (2003), who study a sample of 108 funds and find that funds have substantial exposures to size and value/growth factors in returns that are not captured by their benchmarks, and with Cremers and Petajisto (2007), who find that funds typically have a high proportion of holdings that differ from those of the fund's theoretically correct benchmark index.

Mismatched benchmarks that complicate the task of identifying a fund's true factor exposures and lead to excess flows are consistent with theoretical work such as Carlin (2006), who argues that financial service providers have incentives to strategically complicate their pricing schedules because they can earn economic rents in equilibrium from doing so.

Finally, from the perspective of performance evaluation theory, mismatched self-designated benchmarks among mutual funds may be viewed as an example of the phenomenon described by Gibbons and Murphy (1990), in which an agent subject to relative performance evaluation chooses a reference group other than the one preferred by the principal.[2] In a corporate setting, Murphy (1995) reports strategic choices of peer groups used for performance comparisons in company annual reports, though he is not able to test whether these peer groups matter to consumers of annual reports.

This paper proceeds as follows. The next section describes the data. Section 2 compares funds to their benchmarks, remaining agnostic regarding whether benchmark choices appear strategic. Section 3 analyzes investor reaction to mismatched benchmarks in terms of flows, and discusses the interpretation and magnitude of these flows. Section 4 considers whether mismatched benchmarks are plausibly attributed to strategic behavior. Section 5 concludes.

# 1 Data

The main database comes from Morningstar and contains self-designated benchmark information for 1,981 actively managed, diversified U.S. equity mutual funds. The database also contains holdings and monthly return data for the vast majority of these funds over 1994-2004. Holdings

[2] To the extent that benchmark changes are rare in practice, the mapping to theory is especially strong if one thinks of this phenomenon as encompassing ex post behavior that makes the reference group, chosen ex ante, not the one preferred by the principal.

data are typically semiannual, the statutory reporting requirement during the sample period, and are available through June, 2004. Returns data are monthly and run through July, 2004. I also obtain monthly returns on the Fama-French factor portfolios Rm-Rf, SMB, HML, and Rf, as well as the momentum factor UMD from Ken French's website. Finally, I obtain monthly returns on Russell and S&P/Barra indexes from the websites of their parent companies.

The benchmark data have two limitations. First, the benchmark information is cross-sectional, not panel. This is unlikely to be a problem for this study. Conversations with industry insiders, including members of the boards of directors of two mutual fund families as well as data providers Morningstar and Thomson Financial, indicate that in practice benchmark changes are rare, perhaps because the SEC frowns on benchmark changes. Sections 2 and 4 demonstrate that the differences between fund benchmarks and portfolios and the frequency of mismatched benchmarks that I document are not driven by unoberserved benchmark changes.

Second, although the SEC only began requiring funds to report their historical returns alongside those of a passive benchmark index in 1999, some funds voluntarily did so earlier. My data does not identify these. I report evidence for the full 1994-2004 sample period. Results are similar when restricting attention to 1999-2004.

An additional complication is that some funds have two benchmarks - designated primary and secondary by Morningstar. The Morningstar data provide incomplete coverage of secondary benchmarks, so I supplement the benchmark data using fund prospectuses from 1999 or the earliest available, if later. Because the purpose of this paper is to identify and explore the properties of funds with mismatched benchmarks, I want to be conservative. For each fund, I use the primary benchmark as that fund's benchmark unless it is mismatched according to the methodology of section 2. If the primary benchmark is mismatched, I use the prospectus to check and correct if necessary the primary benchmark information, and, if the prospectus indicates a second benchmark that is not mismatched, I use the second benchmark as the fund's benchmark. This procedure ensures that a fund's benchmark as defined in this paper is not mismatched as long as at least one of its primary and secondary benchmarks is not mismatched.

Table 1 provides summary information on benchmark use in the resulting sample. Of the 1,981 funds, 1,815 (91.6%) have one of the following benchmarks: S&P 500, Russell 1000, Russell 1000 Value, Russell 1000 Growth, S&P Midcap 400, Russell Midcap, Russell Midcap Value, Russell Midcap Growth, S&P Smallcap 600, Russell 2000, Russell 2000 Value, or Russell 2000 Growth. Thus, the vast majority of funds use S&P or Russell indexes that are defined on size and value/growth. These twelve benchmarks include the top 10 most commonly used benchmarks and 94.6% of total net assets. I restrict my attention to funds with these benchmarks because in these cases it is straightforward to compare the fund's exposure to size and value/growth factors in returns to that of the benchmark.

The funds collectively held $985.1 billion in assets on June 30, 2004. The S&P 500 is by far the most popular benchmark choice, representing 44.4% of funds and 61.3% of June 2004 assets.

# 2 Comparison of funds and their benchmarks

Fama and French (1992) and many others find that size and value/growth are associated with average returns and return covariances. Therefore, for a fund's self-designated size and value/growth benchmark index to be maximally useful in netting out priced common factors in returns, it should match the fund's exposure to those factors. In this section, I compare funds to their self-designated benchmarks in order to investigate how useful these benchmarks actually are to fund investors. I also investigate whether alternative benchmarks better capture funds' factor exposures.

## *2.1 Covariance comparisons*

I begin with fund-level Fama-French 3-factor model regressions to explain funds' monthly benchmark-adjusted returns:

$$R_{i,t} - R_{Bench,i,t} = \alpha_i + \beta_i \left(R_{M,t} - R_{f,t}\right) + s_i SMB_t + h_i HML_t + e_{i,t}, \tag{1}$$

where $R_{i,t}$ is fund $i$'s return in month $t$ and $R_{Bench,i,t}$ is the return of fund $i$'s self-designated benchmark in month $t$. The factor loadings in each of these regressions identify differences between the fund's and the benchmark's average exposures to the factors. To be included in the regressions, a fund must have at least 24 monthly return observations.

Table 2 displays statistics from the distribution of regression coefficients to assess how frequently and on which factors the differences are significant. A substantial fraction (40.0%) of funds have significantly different market exposure than their self-designated benchmarks. This figure ranges from 21.5% for funds benchmarked to the Russell 1000 Growth to 50.0% for funds benchmarked to the Russell 1000. Most differences are negative: of the 40.0%, 27.6% have less market exposure and 12.4% more. This pattern is fairly uniform across benchmarks and is at least partially due to fund cash holdings.

A still greater fraction of funds - 57.9% - display significant loadings on SMB relative to their benchmarks. These tend to work in the direction of positive expected benchmark-adjusted returns: 41.5% of funds have significantly positive SMB loadings relative to their benchmarks and only 16.5% have negative. Similarly, 61.6% of funds have significant HML loadings relative to their benchmarks. There is a tendency to value: 33.3% of funds have significantly positive HML loadings relative to their benchmarks and 28.3% have negative.

Overall, the evidence shows that funds frequently have significantly different factor loadings than their self-designated benchmarks. The direction of the differences tends to be toward higher expected returns (small stocks and value stocks), consistent with attempts to beat the benchmark.

The deviations are economically as well as statistically significant. The magnitude of the typical SMB (HML) coefficient is about 0.25 (0.35). The SMB (HML) premium in my sample period is 0.15% (0.40%) per month. So the typical SMB deviation results in a performance difference relative to the benchmark of 0.0375% per month, or roughly 0.45% per year. For HML, the figures are 0.14% per month, or 1.68% per year.

The differences documented in table 2 are not due to unobserved self-designated benchmark changes. The 1994, 1999, and 2003 cross-sections all show meaningful differences between the factor loadings of funds and those of their benchmarks (though statistical significance is less common in any one cross-section due to a paucity of monthly observations).

### *2.2 Characteristics comparisons*

Another way to compare fund investments to their self-designated benchmarks is to compare characteristics. A simple and intuitive, albeit coarse, way to do this is to use Morningstar's stylebox classifications, which categorize each fund portfolio using the size and value/growth characteristics of the fund's stock holdings. The categories are large value, large blend, large growth, medium value, medium blend, medium growth, small value, small blend, and small growth. The mode of the time-series distribution of a fund's portfolio styleboxes is a measure of the fund's typical size and value/growth characteristics. Table 3 compares this measure of fund characteristics to that implied by the fund's benchmark.

Table 3's main message confirms that of table 2: there are frequent differences between funds' actual size and value/growth characteristics, as measured by their Morningstar styleboxes, and those of their self-designated benchmarks. As in table 2, these differences are more common on value/growth rather than size.

The most common difference is funds with a value or growth orientation using size-appropriate but value/growth neutral benchmarks. For example, the S&P 500 is the most popular benchmark among large-cap funds, even those with a value or growth style. The S&P 500 and the Russell 1000, another large blend index, together account for 63.2% of large value funds and 69.4% of large growth funds, as well as 96.6% of large blend funds. The medium blend S&P Midcap 400 and Russell Midcap account for 13.2% of medium value funds and 34.6% of medium growth funds, as well as 40.2% of medium blend funds. Similarly, the small blend S&P Smallcap 600 and Russell 2000 account for 58.8% of small value funds and 54.7% of small growth funds, as well as 77.0% of small blend funds.

Differences on size are less common. However, the S&P 500 is quite popular with medium cap funds, accounting for 47.0%, 42.3%, and 23.7% of medium value, medium blend, and medium growth funds, respectively.

The differences documented in table 3 are not due to unobserved benchmark changes. The 1994, 1999, and 2003 cross-sections all display qualitatively similar patterns to those reported in table 3 for the full sample.

### *2.3 Mismatched and corrected benchmarks*

The differences documented in tables 2 and 3 suggest that some funds' self-designated benchmarks do not do a very good job capturing the funds' exposures to size and value/growth factors in returns. One must be careful, however. That a fund's benchmark-adjusted returns load significantly on the Fama-French factors does not by itself imply that the specified benchmark is suboptimal. It may be the closest possible match despite the significant differences. For example, a fund benchmarked to the S&P 500 may hold enough small-cap stocks to generate a positive loading on SMB, but not enough to make a small or even mid cap benchmark a better overall match. Similarly, a fund's typical characteristics as measured by Morningstar styleboxes may differ from those of the benchmark even if the benchmark does an excellent job capturing the fund's exposure to common factors in returns.[3] Thus we cannot conclude that some funds have suboptimal benchmarks based solely on tables 2 and 3.

All the same, the frequency and magnitude of the differences documented in tables 2 and 3 suggest that for at least some funds an alternative benchmark among the twelve I consider may better capture the fund's exposure to common factors in returns. To investigate this possibility, I define a fund's candidate corrected benchmark as the benchmark whose style matches the fund's Morningstar style as described in section 2.2.[4] For the 49.6% of funds whose candidate corrected benchmark is not the same as its actual benchmark, the key question is whether the candidate corrected benchmark does a better job than the actual benchmark in capturing the fund's exposure to common factors in returns. If the corrected benchmark is more correlated with the fund's returns than the actual benchmark, it is picking up more of the variation in the fund's returns (including that associated with common factors in returns) and so is a better match for the fund.

Table 4 shows that this is true for 31.2% of funds. I call these funds' self-designated benchmarks "mismatched". They are mismatched because the corrected benchmarks better match the funds' typical Morningstar styles and, more importantly, are more correlated with their returns. Among funds with mismatched benchmarks, the average $R^2$ with the actual benchmark is 70.6%, versus 82.6% with the corrected benchmark.

I emphasize that I use Morningstar styleboxes only to identify funds whose self-designated benchmarks *may* be mismatched, and rely on correlation analysis to determine whether the candidate corrected benchmark is in fact better matched. This procedure has two main advantages. First, it avoids the reliance on the idiosyncracies of Morningstar's classification system that would result if Morningstar styleboxes alone were used to determine mismatched benchmarks. Second, it is conservative in that it avoids the multiple comparisons problem of searching all possible benchmarks

[3] There are at least two reasons why this may be the case. The characteristics-based methodology used in the construction of styleboxes is related but far from identical to that used by Fama and French (1992) to construct their size (S,B) and value/growth (H,L) factor portfolios. Moreover, even if there were a perfect match between the Morningstar and Fama-French characteristics-based procedures, there is variation in returns that is related to covariances, controlling for characteristics (Davis, Fama, and French,1997).

[4] The candidate corrected benchmarks are the Russell 1000 Value, S&P 500, Russell 1000 Growth, Russell Midcap Value, S&P Midcap 400, Russell Midcap Growth, Russell 2000 Value, Russell 2000, and Russell 2000 Growth. They correspond, respectively, to large value, large blend, large growth, medium value, medium blend, medium growth, small value, small blend, and small growth.

for the one with the highest correlation with the fund. Compared to either of these alternative methodologies, I am less likely to spuriously label a fund as having a mismatched self-designated benchmark.[5]

### *2.4 Discussion*

The analysis so far identifies funds whose self-designated benchmarks are mismatched in the sense that the corrected benchmark better captures the fund's average exposure to common factors in returns. I have not yet analyzed the time-series profile of differences from the self-designated benchmark. This is important for the interpretation of mismatched benchmarks, in particular whether mismatched benchmarks are likely to reflect strategic fund behavior. For example, if a fund started off with a good benchmark but then drifted away from that benchmark over time, we would not necessarily want to call this behavior strategic even if the benchmark ends up being mismatched. However, whether fund styles change over time is important only to the *interpretation* of mismatched benchmarks, and is completely irrevelent to the question of whether benchmarks *are* (on average) mismatched. This is because I label a benchmark mismatched if and only if the candidate corrected benchmark is more correlated with the fund's returns, and not with respect to changes or lack thereof in fund styles.

The analysis so far also does not consider the types of funds that have mismatched benchmarks and how they differ from the benchmark. This is also important for the interpretation of mismatched benchmarks.

I analyze these issues in detail in section 4.

[5] In an untabulated test, I confirm that the identified corrected benchmarks do a better job than mismatched self-designated benchmarks of capturing funds' exposures to size and value/growth factors in returns. For funds with mismatched benchmarks, I collapse the cross-section by taking cross-sectional averages of actual benchmark-adjusted return and corrected benchmark-adjusted return in each month. I then run a stacked time-series regression of these average returns on the Fama-French factors. Compared to funds' actual benchmark-adjusted returns, their corrected benchmark-adjusted returns produce SMB and HML coefficients that are significantly closer to zero.

# 3 Fund investor reaction to mismatched self-designated benchmarks

In this section, I analyze the response of mutual fund investors to mismatched self-designated benchmarks. The next subsection tests whether inflows of new investment to mutual funds respond to performance relative to a mismatched self-designated benchmark. Subsections 3.2 and 3.3 discuss the interpretation and magnitude of these results.

## *3.1 Does performance relative to a mismatched self-designated benchmark influence fund flows?*

I now examine whether fund investors respond to performance relative to a self-designated but mismatched benchmark when making decisions about purchases and sales of mutual funds. For this to happen, at least some investors must pay attention to the information in fund prospectuses. There is some evidence that fund investors might do this. According to a recent survey by the Investment Company Institute, the national association of investment companies, 34% of fund investors consult the fund prospectus before purchasing a mutual fund. This figure seems large enough to plausibly have an effect on flows, especially considering that the performance table is prominently displayed in the first few pages of the prospectus. Fund advertising also frequently features a comparison of the fund's performance with that of a benchmark (when the comparison is favorable).

My tests are based on regressions of flows to funds with mismatched self-designated benchmarks on different measures of fund performance as well as various controls. Following most of the existing literature on fund flows (e.g. Chevalier and Ellison 1997, Sirri and Tufano 1998), I calculate flows for fund $i$ in year $t+1$ as the percentage growth of new assets, assuming that all flows take place at the end of the year:

$$Flow_{i,t+1} = \frac{TNA_{i,t+1}}{TNA_{i,t}} - (1 + R_{i,t+1})\,, \tag{2}$$

where $TNA_{i,t}$ is the total net assets under management of fund $i$ at the end of year $t$, and $R_{i,t+1}$ is the total return of fund $i$ in year $t+1$.

Again following the existing literature, I investigate the relation between flows in year $t+1$

and performance in year $t$. I estimate regressions of the form:

$$Flow_{i,t+1} = \alpha + r_{i,t} * \beta + X_{i,t} * \gamma + \delta_t + \varepsilon_{i,t}, \tag{3}$$

where $r_{i,t}$ is a vector of performance measures for fund $i$ in year $t$, including the main variable of interest, self-designated benchmark-adjusted return. The other performance variables are return relative to the corrected benchmark, Fama-French 3-factor alpha, and return relative to the market. Return relative to the market (the Fama-French market factor) is included as a control only for funds not benchmarked to the S&P 500 because the market and the S&P 500 are almost perfectly correlated. I set this variable equal to zero in the regressions if the fund is benchmarked to the S&P 500. $X_{i,t}$ is a vector of controls that might influence fund inflows for reasons unrelated to performance. These include fund age, fund expense ratio, and the natural logarithm of fund assets (in billions of dollars). Chevalier and Ellison (1997) and Sirri and Tufano (1998) find that these variables do affect fund flows.

In all specifications, I follow Bergstresser and Poterba (2002) and include a set of year-specific indicator variables for each of the nine different fund investment styles. These style classifications are based on Morningstar's stylebox classifications at the end of each year. These year*style indicator variables allow for changing investor tastes for funds with different investment styles and also capture all year and style fixed effects. They also absorb residual variation and thus reduce concerns about the impact of spurious factors on the regression results.

Previous research (Sirri and Tufano 1998, Chevlier and Ellison 1997) documents significant nonlinearities in the relation between flows and historical returns. Specifically, Chevalier and Ellison (1997) observe that fund investors are quicker to reward good performance relative to the market than to punish bad. That is, flows are more strongly related to positive market-adjusted performance than negative. Failing to capture these nonlinearities can cause spurious inference because of an omitted variables problem. To be precise, because the different performance measures are correlated, spurious significance may be attributed to a performance measure in a linear model if it picks up part of the nonlinearity in the (true) relation between flows and another performance measure. In light of Chevalier and Ellison (1997), I parsimoniously attempt to account for nonlinearities by allowing positive performance relative to a reference index to affect flows differently than negative performance.

Following Chevalier and Ellison (1997) and Bergstresser and Poterba (2002), the regressions include only those funds that were at least two years old at the end of year $t$, had at least \$10 million in assets at the end of year $t$, and had year $t+1$ flows of less than 10 (1,000%). To reduce the effect of outliers, I drop the 46 fund-years in which benchmark-adjusted return is more than 3 standard deviations above the mean. This improves the model fit and does not affect the conclusions.

Table 5 displays the results of these regressions. To conserve space, I do not report in the table that fund size and age have a reliably negative effect on fund flows, which is consistent with prior research. However, in minor contrast with prior work, there is no reliable effect of a fund's expense ratio on flows.

The main message of table 5 is that performance relative to a mismatched self-designated benchmark has substantial explanatory power for fund flows. Beating the actual benchmark is always associated with higher flows, even when controlling for the effects on flows of the other performance variables. Trailing the benchmark carries little, if any, penalty. These results are consistent with relatively naive investors paying attention to these self-designated benchmarks. This interpretation is strengthened by the results in table 5 that it is mainly negative performance relative to the corrected benchmark and negative Fama-French alpha that reduces flows. Thus it appears to be mainly relatively sophisticated investors who penalize funds for poor performance.

In the model using self-designated benchmark-adjusted return as the only explanatory return variable, beating (trailing) the benchmark by one percentage point predicts a 3.13 (1.20) percentage point increase (decrease) in inflows in the following year. Of course, the most interesting specifications are those in which performance relative to the corrected benchmark and Fama-French 3-factor alpha are included as controls. These performance variables generally also have significant explanatory power for flows. Their inclusion reduces, but does not eliminate, the importance of self-designated benchmark-adjusted return. Even controlling for these performance measures, beating the benchmark by one percentage point predicts between a 1.71 and a 1.80 percentage point increase in inflows in the following year. Controlling for the penalties earned by trailing the corrected benchmark and having negative Fama-French alpha, there is no additional penalty for trailing the self-designated benchmark.

The main result concerning the significance of self-designated benchmark-adjusted return survives several robustness checks, which for brevity I do not tabulate. First, another way to account for nonlinearities in the flow-performance relations is suggested by Berk and Green (2004), whose model of fund flows in a fully rational world with symmetric information implies that the optimal relation between flows and risk-adjusted returns is quadratic over a wide range. They argue that a quadratic form fits Chevalier and Ellison's (1997) empirically determined flow-performance relation fairly well. The main result is robust to entering each performance measure quadratically instead of forcing the inflection point, if there is one, to be at zero.

Second, models 3, 4, and 5 refute concerns that flows are simply responding to performance relative to the market (or S&P 500, which is almost perfectly correlated with the market) regardless of funds' specified benchmarks or their corrected benchmarks. If that were so, the coefficient on positive market-adjusted return would be significantly positive and the coefficient on positive self-designated benchmark-adjusted return would be reduced relative to model 2. As a further check, I estimate all models using only funds whose benchmark is not the S&P 500 and find that benchmark-adjusted return is still highly significant and of similar magnitude (even when controlling for market-adjusted return).

Third, the results are robust to substituting year dummies, style dummies, and a control for the yearly flow to all funds of a given style instead of the full set of year*style dummies. Bergstresser and Poterba (2002) use this alternative specification in some of their models.

Fourth, the results are are robust to Ippolito's (1992) correction to account for the fact that flows happen throughout the year:

$$Flow_{i,t+1}^{Ipp} = Flow_{i,t+1} / \left(1 + R_{i,t+1}/2\right). \tag{4}$$

The results in table 5 show that fund investors at least partially rely on funds' self-designated but mismatched benchmarks as a reference group for the determination of fund flows, and so suggest that future work on the relation between fund flows and performance should consider carefully the measures of performance to use. In particular, tests based solely on raw or market-adjusted returns, a common practice, may suffer from an omitted variable bias.

### *3.2 Interpreting the results on flows*

Is the response of flows to performance relative to a mismatched self-designated benchmark more likely to reflect rational or irrational behavior on the part of fund investors? From a performance evaluation/contracting perspective, because mutual funds generally receive a fixed percentage of assets under management as a fee, cash inflows and outflows are the mechanism by which fund investors (principals) influence fund companies' (agents) compensation. As such, agency theory (e.g. Holmstrom 1979) predicts that investors ought to direct flows in response to risk-adjusted return. Doing so aligns fund companies' desire for increased compensation, which gives them the incentive to take actions to increase flows, with fund investors' interest, maximizing risk-adjusted return.

Thus, the acid test for the interpretation of these flows is whether mismatched benchmarks have incremental power to explain the cross-section of expected returns, and thereby help measure risk-adjusted returns. Controlling for the Fama-French factors Rm-Rf, SMB, and HML and the corrected benchmark (which are controlled for in table 5), does the mismatched benchmark price anything? I investigate this question in two sets of tests, and document them in table 6.[6]

In the first set of tests, I use the 25 Fama-French size and book-to-market portfolios as the cross-section of assets to be explained. For each of the 36 mismatched-corrected benchmark pairs in my sample, I do the following (the sample period is that of my main sample, January 1994-July 2004). First, I regress the excess returns of the 25 Fama-French portfolios on the Fama-French factors and the excess returns of the corrected benchmark (call this model 1). I then add the excess returns of the mismatched benchmark to the right hand side and estimate the regressions again (call this model 2). For each model, I compute the average absolute pricing error across the 25 portfolios and the Gibbons, Ross, and Shanken (GRS, 1989) statistic for the hypothesis that all of the 25 intercepts are equal to 0.0.

The null hypothesis is that, for each mismatched-corrected benchmark pair, the two models do an equally good job explaining the cross-section of expected returns represented by the 25 Fama-French portfolios. Inspection of the pricing errors and GRS statistics presented in table 6 is comforting from the perspective of this null. Specifically, the average absolute pricing errors across

[6] I thank an anonymous referee for suggesting these tests.

the two models do not differ much (the difference is less than 3 basis points per month for most mismatched-corrected benchmark pairs and is less than 10 basis points per month for all pairs). Consistent with this, the GRS statistics are similar across the two models as well.

I also conduct a more formal test of this null hypothesis for each mismatched-corrected benchmark pair. The test statistic is $\left(\alpha_1' var(\alpha_1)^{-1}\alpha_1 - \alpha_2' var(\alpha_2)^{-1}\alpha_2\right)$, where $\alpha_1$ is a vector of the 25 estimated intercepts from model 1 and $var(\alpha_1)$ is the corresponding estimated variance-covariance matrix. I bootstrap this test statistic to obtain critical values. Table 6 shows that for only two of the 36 mismatched-corrected benchmark pairs is the hypothesis that the test statistic is equal to zero rejected at the 5% level. This is close to the 1.8 rejections (36*0.05) we would expect by pure chance.

Overall, the results of this first set of tests are consistent with the null hypothesis that the mismatched benchmark does not help the Fama-French factors and the corrected benchmark price the 25 Fama-French portfolios. In this framework, it is difficult to construct a powerful test of the null hypothesis for all the mismatched-corrected benchmark pairs simultaneously because the set of benchmarks that are the mismatched benchmark for at least one fund and the set of benchmarks that are the corrected benchmark for at least one fund mostly overlap.

The second test takes a different approach in order to circumvent this difficulty. This test considers whether the Fama-French 3-factor model can jointly explain differences in the expected returns of each pair of mismatched and corrected benchmarks. I construct zero-investment spread portfolios that are long the mismatched benchmark and short the corrected benchmark. I then regress each spread portfolio on the Fama-French factors and test whether the intercepts are jointly equal to zero using the GRS test of Gibbons, Ross, and Shanken (1989). As shown at the bottom of table 6, the GRS test fails to reject (at the 5% level) the hypothesis that the the intercepts are all equal to 0.0.[7] The p-value is 0.089 based on an F distribution and 0.143 when bootstrapped, suggesting that the residuals may not be normally distributed. The last two columns of table 6 show that the individual intercepts are statistically indistinguishable (at the 5% level) from zero for 26 of the 36 spread portfolios.

The GRS test result is consistent with the hypothesis that the Fama-French 3-factor model can

[7] This test can, of course, be based only on linearly independent combinations of the spread portfolios, of which there are 11. That is, 25 of the 36 spread portfolios can be formed from linear combinations of the other 11.

*jointly* price the spread portfolios. Nevertheless, one might wonder whether the overall incremental response of flows to performance relative to a mismatched self-designated benchmark is driven by precisely those funds for which the Fama-French 3-factor model does not *individually* price the corresponding spread portfolio. Such flows would make sense as a rational response to abnormal returns. To address this issue, I drop from the flow regressions in table 5 all of the 32 funds whose mismatched and corrected benchmarks correspond to the 10 spread portfolios whose alphas are significantly nonzero in the Fama-French 3-factor model. The results in table 5 are virtually unchanged, so these funds are not driving the main flow results.

Overall, these two sets of tests suggest that it is unlikely that the incremental responses of flows to performance relative to mismatched self-designated benchmarks are rational responses to abnormal returns. Rather, they suggest that the evidence in table 5 more likely reflects a behavioral element to the composition of mutual fund flows, consistent with Del Guercio and Tkac (2002), Musto (1999), Elton, Gruber, and Busse (2004), Cooper, Gulen, and Rau (2005), and Cronqvist (2006). That being said, because the pricing kernel is unobservable to the econometrician, one cannot completely rule out the possibility that mismatched benchmarks do have incremental pricing power and that these flows are in fact a rational response to abnormal returns. My interpretation of the tests presented in this subsection is that this possibility is unlikely, but certainly not impossible.

### *3.3 Magnitude of incremental flows*

What is the magnitude of the expected incremental gain in flows to funds with mismatched benchmarks? The flow-performance relations estimated in table 5 allow us to estimate this. The coefficients on self-designated benchmark-adjusted return in the regressions imply that, conditional on the other performance measures, funds are rewarded for beating a mismatched benchmark, but are barely, if at all, penalized for trailing it. For each model in the table, the expected flows due to the mismatched self-designated benchmark are consistently estimated by multiplying a fund's self-designated benchmark-adjusted return by the estimated coefficient. From this we must subtract the estimated expected flows under the counterfactual that the fund has the corrected benchmark instead of its actual (mismatched) benchmark. This results in an estimate of the gain in expected flows from having a mismatched self-designated benchmark. The estimate from the model with full controls is a statistically and economically significant 2.3% of fund assets per year. To put this number in perspective, it is 14.6% of the average annual flow to funds with mismatched self-designated benchmarks (15.8% of assets per year).

It is also useful to investigate where these incremental flows are going. From the perspective of a fund investor who believes that the self-designated benchmark reflects the risk profile of the fund, purchasing a fund with a mismatched self-designated benchmark offers a worse risk-return tradeoff on average than purchasing a fund whose benchmark is not mismatched. For funds whose self-designated benchmarks are not mismatched, the average benchmark-adjusted return is -0.018% per month and the average fund's standard deviation of benchmark-adjusted return is 2.074% per month. For funds with mismatched self-designated benchmarks, the corresponding figures are -0.024% and 3.435%. Funds with mismatched benchmarks charge higher fees, averaging an expense ratio of 1.453%, versus 1.378% for funds whose benchmarks are not mismatched. A difference of this magnitude is economically significant (Bris et. al., 2007). These facts suggest that directing incremental flows to funds with mismatched self-designated benchmarks is unlikely to be good for fund investors.

A calculation that sheds further light on the disposition of these incremental flows assumes that they remain in those funds (with mismatched self-designated benchmarks) for the life of the sample, and compares the excess return and Sharpe ratio they earn to what they would have earned if invested in the corrected benchmark or in the actual (mismatched) benchmark.[8] I use the coefficients in the last column of table 5 as well as funds' actual benchmark-adjusted returns to estimate the incremental flows. Over the sample period, the average dollar of incremental flows earns an excess return of 0.482%, 0.561%, and 0.619% per month if invested in, respectively, a fund with a mismatched self-designated benchmark, the corresponding corrected benchmark index, or the corresponding mismatched benchmark index itself. The corresponding Sharpe ratios are 0.090, 0.105, and 0.134. These results suggest that the incremental flows might have been better directed to a low-cost index fund for either the corrected or the actual (mismatched) benchmark index.

[8] I thank an anonymous referee for suggesting this calculation.

# 4 Are mismatched self-designated benchmarks strategic?

So far we have seen that funds' self-designated benchmarks are frequently mismatched and that fund investors nevertheless pay attention to them, directing incremental flows to funds that beat those benchmarks. I now consider whether mismatched self-designated benchmarks represent strategic behavior by funds. The response of flows to performance relative to mismatched benchmarks creates a strategic incentive for funds to have mismatched self-designated benchmarks in an attempt to improve inflows. In this section, I develop and and present evidence supporting five hypotheses consistent with the idea that mismatched self-designated benchmarks are strategic.

*H1: Mismatched self-designated benchmarks are not a result of style drift or changing fund styles.*

If a fund started off with a good benchmark but then drifted away from that benchmark over time, we would not necessarily want to call this behavior strategic even if the benchmark ends up being mismatched. On the other hand, if a fund's benchmark were mismatched from the beginning, that would be evidence that it might be strategically chosen. I investigate this idea by repeating the methodology of section 2.3 using only data from 1999, the first year of the SEC requirement to tabulate returns alongside those of a benchmark.

I find that 31.5% of funds have self-designated benchmarks that are mismatched in 1999, compared to 31.2% of funds whose self-designated benchmarks are mismatched overall. Moreover, 65.0% of funds whose self-designated benchmarks are mismatched overall also have mismatched benchmarks when considering 1999 alone.

Thus, self-designated benchmarks are just as frequently mismatched in 1999 as in the overall sample and most funds with mismatched self-designated benchmarks overall did not match those benchmarks in 1999. This evidence does not support the idea that mismatched benchmarks are primarily due to style drift or changing fund styles: funds with mismatched benchmarks did not generally start off with good benchmarks but drift away from them over time. This evidence is consistent with Chen, Chan, and Lakonishok (2002), who find that fund styles are generally stable over time.

*H2: Value funds are more likely than growth funds to have mismatched benchmarks on value/growth.*

*H3: Small-cap funds are more likely than large-cap funds to have mismatched benchmarks on size.*

*H4: Funds with higher fees and/or more assets under management are more likely to have mismatched benchmarks.*

*H5: There are fund family effects in the pattern of mismatched benchmarks.*

The evidence in table 5 implies that if fund self-designated benchmarks and portfolios differ in their risk attributes in such a way that the fund manager believes that expected benchmark-adjusted return is positive, the manager expects the fund to gain in terms of flows. Thus if fund portfolios systematically differ from mismatched self-designated benchmarks in the direction of positive expected returns, it would be evidence that mismatched benchmarks might be strategic. Unfortunately, the key concept here is manager beliefs about positive expected returns, and any pattern of differences between fund risk attributes and those of their mismatched self-designated benchmarks is consistent with fund managers believing the fund has positive expected benchmark-adjusted return.

What I can test is whether funds' actual systematic risk differences from their mismatched self-designated benchmarks are (on average) consistent with the size and value effects documented by Banz (1981), Fama and French (1992), and others. This work shows that on average small stocks outperform large stocks and value stocks outperform growth stocks. In light of this evidence, funds with a value orientation ought to be more likely to have a mismatched self-designated benchmark on value/growth (i.e. a neutral or growth benchmark) than funds with a growth orientation (for which a value/growth mismatch would mean a neutral or value benchmark). This is hypothesis H2. Similarly, small-cap funds ought to be more likely to have a mismatched self-designated benchmark on size (i.e. a mid-cap or large-cap benchmark) than large-cap funds (for which a size mismatch would mean a mid-cap or small-cap benchmark). This is hypothesis H3.

The interpretation that mismatched self-designated benchmarks are consistent with strategic fund incentives to attract flows suggests that they should be more common among funds that benefit more from a given increase in flows. Because flows improve fund profits through fees, and because a given change in flows as a percentage of assets is a larger dollar amount for funds with more assets under management, mismatched benchmarks should be more common among high-asset, high-fee funds. This is hypothesis H4.

Finally, there may be fund family effects. If mismatched self-designated benchmarks are strategic, then family affilation may be a significant determinant of whether a fund has a mismatched benchmark, for two reasons. Profits generated by a mutual fund accrue to the parent company (family) and individual fund benchmarks may be set at the family level. This is hypothesis H5.

I test hypotheses H2, H3, H4, and H5 in table 7. The table displays the results of fund-level logit models of the probability of having a mismatched self-designated benchmark, either on value/growth (column 1), size (column 2), or overall (column 3). The independent variables are indicator variables for whether the fund is primarily a value fund, growth fund, small-cap fund, or large-cap fund, as determined by the mode of the fund's portfolio styleboxes (the missing category is mid-cap blend), the time-series average of the fund's expense ratio, and the time-series average of the natural logarithm of the fund's assets under management. I also include fund family fixed effects for each of the 44 fund families that have at least 10 funds in the sample. This regression setup allows me to test each of the four hypotheses while controlling for the effects of the variables suggested by the other three, and thereby helps avoid spurious inference. In particular, there is no inconsistency between the prediction of H2, that small-cap funds are more likely to have mismatched self-designated benchmarks on size, and H4, that funds with greater assets under management are more likely to have mismatched self-designated benchmarks. The regression setup essentially estimates the within category (small-cap, etc.) effect of greater assets under management on the probability of having a mismatched self-designated benchmark.

The first column of table 7 presents evidence consistent with H2: value funds are more likely than growth funds to have mismatched self-designated benchmarks on value/growth. The difference between the two coefficients is statistically significant at the 10% level (p-value 0.077). This result is consistent with the patterns documented in section 2.

The second column of table 7 presents evidence consistent with H3: small-cap funds are more likely than large-cap funds to have mismatched self-designated benchmarks on size (though both are less likely than mid-cap funds). The difference between the two coefficients is statistically significant at the 1% level. This result is also consistent with the patterns documented in section 2.

The evidence in all three columns is consistent with H4. The coefficient on expense ratio is positive and significant at the 1% level in all three models. The coefficient on log assets is positive in all three models and significant at the 1% level in two of the three.

Similarly, the evidence in all three columns is consistent with H5. In the three models, 90.9%, 90.9%, and 86.4% of the 44 family fixed effects are significantly different from zero. Interestingly, more family fixed effects are negative than positive: 54.5%, 54.5%, and 54.5% versus 36.4%, 36.4%, and 31.8%. In these figures, I count those families for which all observations are dropped because the familiy fixed effect predicts failure (not having a mismatched benchmark) perfectly as having a significantly negative family fixed effect. In the second column, there are fewer observations in the regression because more family fixed effects predict failure perfectly. Not a single family fixed effect in any model predicts success (having a mismatched benchmark) perfectly. This is evidence that some families are more likely than others to have funds with mismatched benchmarks, and that there are families that eschew mismatched self-designated benchmarks.

Specifically, the fund families that have no funds with mismatched self-designated benchmarks (among those with at least 10 funds in the sample) are GE Funds, Pioneer Funds, and Northern Trust Funds. Conversely, the fund families that have positive fixed effects in the model in the third column of table 7 are Columbia Funds, Delaware Investments, Eaton Vance, AXA Enterprise, First American Funds, Franklin Templeton, Gartmore, Hartford, Janus, Oppenheimer, Pimco, Strong, TA Idex, and Van Kampen.

Overall, the evidence supports the interpretation that mismatched self-designated benchmarks are consistent with strategic fund incentives to attract flows. Each of the hypotheses developed above are supported by the evidence.

From the perspective of performance evaluation theory, strategically mismatched self-designated benchmarks among mutual funds may be viewed as an example of the phenomenon described by Gibbons and Murphy (1990), in which an agent subject to relative performance evaluation strategically chooses a reference group other than the one preferred by the principal. To the extent that benchmark changes are rare in practice, the mapping to theory is especially strong if one thinks of the phenomenon of undesirable reference group choice as encompassing ex post behavior that makes the reference group, chosen ex ante, not the one preferred by the principal. In a corporate setting, Murphy (1995) reports strategic choices of peer groups used for performance comparisons in company annual reports: companies apparently chose their peer groups strategically to make their performance appear more favorable. In other applications, Dye (1992) argues that relative performance evaluation provides incentives for companies to operate in industries with inept rivals, and Carmichael (1988) argues that untenured faculty have an incentive to recruit inferior colleagues.

# 5 Conclusions

This paper shows that 31.2% of equity mutual funds specify an S&P or Russell size and value/growth-based benchmark index in the fund prospectus that is “mismatched": alternative S&P or Russell size and value/growth-based benchmarks both better match these funds’ size and value/growth characteristics and are more correlated with their returns. Nevertheless, these mismatched self-designated benchmarks matter to fund investors. Performance relative to the specified benchmark, especially above the benchmark, is a significant determinant of a fund’s subsequent cash inflows, even controlling for performance measures that better capture the fund’s style.

These incremental flows appear unlikely to be rational responses to abnormal returns, and provide a strategic incentive for funds to have benchmarks and portfolios that systematically differ in their risk attributes. I document several pieces of evidence suggesting that mismatched self-designated benchmarks may be due to funds’ strategic incentives to improve inflows.

Overall, the evidence in this paper further emphasizes the need, recently stressed by Goetzmann, Ingersoll, Spiegel, and Welch (2007), for the development and dissemination of measures of mutual fund performance that are both well-grounded in economic theory and not subject to gaming.

## References


Banz, Rolf W. 1981, “The relationship between return and market value of common stocks”, *Journal of Financial Economics* 9: 3-18.

Bergstresser, Daniel and James Poterba 2002, “Do After-Tax Returns Affect Mutual Fund Inflows?”, *Journal of Financial Economics* 63: 381-414.

Berk, Jonathan B. and Richard C. Green 2004, “Mutual Fund Flows and Performance in Rational Markets”, *Journal of Political Economy* 112: 1269-1295.

Bertrand, Marianne and Sedhil Mullainathan 2001, “Are CEOs Rewarded for Luck? The Ones Without Principals Are”, *Quarterly Journal of Economics* 116: 901-932.

Bris, Arturo, Huseyin Gulen, Padma Kadiyala, and P. Raghavendra Rau 2007, "Good Stewards, Cheap Talkers, or Family Men? The Impact of Mutual Fund Closures on Fund Managers, Flows, Fees, and Performance", *Review of Financial Studies* 20:953-982.

Brown, Stephen J. and William N. Goetzmann 1997, "Mutual Fund Styles", *Journal of Financial Economics* 43:373-399.

Brown, Keith C., W. Harlow, and Laura Starks 1996, “Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry”, *Journal of Finance* 51: 85-110.

Carhart, Mark M. 1997, “On Persistence in Mutual Fund Performance”, *Journal of Finance* 52: 57-82.

Carlin, Bruce 2006, "Strategic Price Complexity in Retail Financial Markets", working paper.

Carmichael, H. Lorne 1988, “Incentives in Academia: Why is there Tenure?” *Journal of Political Economy* 96: 453-472.

Chan, Louis, Hsiu-Lang Chen, and Josef Lakonishok 2002, "On Mutual Fund Investment Styles", *Review of Financial Studies* 15:1407-1437.

Chevalier, Judith and Glenn Ellison 1997, “Risk-Taking by Mutual Funds as a Response to Incentives”, *Journal of Political Economy* 105: 1167-1200.

Chevalier, Judith and Glenn Ellison 1999, "Career Concerns of Mutual Fund Managers", *Quarterly Journal of Economics* 114: 389-432.

Cohen, Lauren and Breno Schmidt 2007, "Attracting Flows by Attracting Big Clients: Conflicts of Interest and Mutual Fund Portfolio Choice", working paper.

Cohen, Randolph, Joshua Coval, and Lubos Pastor 2005, "Judging Fund Managers by the Company they Keep", *Journal of Finance* 60:1057-1096.

Cooper, Michael, Gulen, Huseyin and P. Raghavendra Rau 2005, "Changing Names with Style: Mutual Fund Name Changes and their Effects on Fund Flows", *Journal of Finance* 60: 2825-2858.

Cremers, Martijn and Antti Petajisto 2007, "How Active is Your Fund Manager? A New Measure that Predicts Performance", working paper.

Cronqvist, Henrik 2006, "Advertising and Portfolio Choice", working paper.

Del Guercio, Diane and Paula A. Tkac 2002, "The Determinants of the Flow of Funds of Managed Portfolios: Mutual Funds vs. Pension Funds", *Journal of Financial and Quantitative Analysis* 37: 523-557.

Dye, Ronald A. 1992, "Relative Performance Evaluation and Project Selection", *Journal of Accounting Research* 30: 27-52.

Ehrenberg, Ronald and Michael Bognanno 1990, "Do Tournaments Provide Incentive Effects?", *Journal of Political Economy* 98: 1307-1324.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake 2003, "Incentive Fees and Mutual Funds", *Journal of Finance* 58: 779-804.

Elton, Edwin J., Martin J. Gruber, and Jeffrey A. Busse 2004, "Are Investors Rational? Choices Among Index Funds", *Journal of Finance* 59: 261-288.

Eun, Cheol and Bruce Resnick, *International Financial Management*, McGraw-Hill Irwin.

Fama, Eugene F. and Kenneth R. French 1992, "The Cross-Section of Expected Stock Returns", *Journal of Finance* 47: 427-465.

Gibbons, Robert and Kevin J. Murphy 1990, “Relative Performance Evaluation for Chief Executive Officers”, *Industrial and Labor Relations Review* 43: 30S-51S.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken 1989, "A Test of the Efficiency of a Given Portfolio", *Econometrica* 57:1121-1152.

Goetzmann, William, Jonathan Ingersoll, Matthew Spiegel, and Ivo Welch 2007, "Portfolio Performance Manipulation and Manipulation-Proof Performance Measures", *Review of Financial Studies*, forthcoming.

Holmstrom, Bengt 1979, “Moral Hazard and Observability”, *Bell Journal of Economics* 10:74-91.

Ippolito, Richard 1992, "Consumer Reaction to Measures of Poor Quality: Evidence from the Mutual Fund Industry", *Journal of Law and Economics* 35:45-70.

Kothari, S.P. and Jerold Warner, 2001, "Evaluating Mutual Fund Performance", *Journal of Finance* 56:1985-2010.

Lynch, Anthony and David Musto, 2003, "How Investors Interpret Past Fund Returns", *Journal of Finance* 58: 2033-2058.

Murphy, Kevin J. 1995, "Politics, Economics, and Executive Compensation", *University of Cincinnati Law Review* 63: 713-748.

Musto, David, 1999, "Investment Decisions Depend on Portfolio Disclosures", *Journal of Finance* 54: 935-952.

Pastor, Lubos and Robert F. Stambaugh, 2000, "Comparing Asset Pricing Models: An Investment Perpective", *Journal of Financial Economics* 56:335-381.

Sirri, Erik R. and Peter Tufano 1998, “Costly Search and Mutual Fund Flows”, *Journal of Finance* 53: 1589-1622.

Treynor, Jack and Fischer Black, "How to Use Security Analysis to Improve Portfolio Selection", *Journal of Business* 46:66-86.

Warner, Jerold and Joanna Wu 2005, "Changes in Mutual Fund Advisory Contracts", working paper.

**Table 1**
**Benchmark usage summary**

Summary statistics on mutual fund self-designated benchmarks. The sample period is 1994-2004. The sample consists of actively managed, diversified U.S. equity mutual funds. "Benchmark style" refers to the size and value/growth style of the benchmark, not necessarily the fund. Rank is relative to all benchmarks in the database. Percentages are based on all funds with benchmark data. "Assets" are total net assets, in billions of dollars, on June 30, 2004.

| Benchmark | Benchmark Style | # funds rank | # funds | % funds | Assets | % assets |
|---|---|---|---|---|---|---|
| S&P 500 | Large blend | 1 | 879 | 44.4 | 637.4 | 61.3 |
| Russell 2000 | Small blend | 2 | 263 | 13.3 | 105.1 | 10.1 |
| Russell 1000 Growth | Large growth | 3 | 116 | 5.9 | 35.8 | 3.4 |
| Russell 1000 Value | Large value | 4 | 111 | 5.6 | 36.5 | 3.5 |
| Russell 2000 Growth | Small growth | 5 | 108 | 5.5 | 29.9 | 2.9 |
| S&P Midcap 400 | Medium blend | 6 | 89 | 4.5 | 29.8 | 2.9 |
| Russell 2000 Value | Small value | 7 | 56 | 2.8 | 18.8 | 1.8 |
| Russell Midcap Growth | Medium growth | 8 | 54 | 2.7 | 32.4 | 3.1 |
| Russell Midcap Value | Medium value | 9 | 40 | 2.0 | 17.9 | 1.7 |
| Russell Midcap | Medium blend | 10 | 37 | 1.9 | 24.1 | 2.3 |
| Russell 1000 | Large blend | 11 | 33 | 1.7 | 13.3 | 1.2 |
| S&P Smallcap 600 | Small blend | 13 | 29 | 1.5 | 4.1 | 0.4 |
| | | Total: | 1,815 | 91.8 | 985.1 | 94.6 |

**Table 2**
**Covariance differences between funds and their self-designated benchmarks**

Coefficient distributions from fund-level Fama-French 3-factor regressions to explain monthly benchmark-adjusted returns: $R_{i,t} - R_{Bench,i,t} = \alpha_i + \beta_i (R_{M,t} - R_{f,t}) + s_i SMB_t + h_i HML_t + e_{i,t}$. For each benchmark and coefficient, the columns display the percentage of funds for which the coefficient is significantly different from, greater than, and less than 0. In the latter two columns, the numbers in square brackets are the cross-sectional average coefficient values among funds with significant coefficients.

| | $\beta_i$ | | | $s_i$ | | | $h_i$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Benchmark | % ≠ 0 | % > 0 | % < 0 | % ≠ 0 | % > 0 | % < 0 | % ≠ 0 | % > 0 | % < 0 | # funds |
| S&P 500 | 45.8 | 11.1 [0.21] | 34.7 [-0.18] | 67.3 | 64.2 [0.27] | 3.1 [-0.12] | 66.2 | 41.3 [0.33] | 24.9 [-0.33] | 799 |
| Russell 1000 | 50.0 | 28.6 [0.09] | 21.4 [-0.30] | 53.6 | 42.9 [0.16] | 10.7 [-0.09] | 78.6 | 60.7 [0.25] | 17.9 [-0.27] | 28 |
| Russell 1000 Value | 39.8 | 19.4 [0.17] | 20.4 [-0.12] | 30.6 | 27.6 [0.15] | 3.1 [-0.13] | 41.8 | 15.3 [0.16] | 26.5 [-0.15] | 98 |
| Russell 1000 Growth | 21.5 | 10.3 [0.14] | 11.2 [-0.18] | 53.3 | 48.6 [0.25] | 4.7 [-0.10] | 50.5 | 47.7 [0.24] | 2.8 [-0.18] | 107 |
| S&P Midcap 400 | 39.0 | 12.2 [0.22] | 26.8 [-0.16] | 35.4 | 25.6 [0.27] | 9.8 [-0.14] | 69.5 | 8.5 [0.22] | 61.0 [-0.51] | 82 |
| Russell Midcap | 30.3 | 12.1 [0.27] | 18.2 [-0.12] | 42.4 | 21.2 [0.38] | 21.2 [-0.13] | 75.8 | 45.5 [0.33] | 30.3 [-0.45] | 33 |
| Russell Midcap Value | 45.7 | 25.7 [0.31] | 20.0 [-0.17] | 42.9 | 28.6 [0.23] | 14.3 [-0.10] | 48.6 | 2.9 [0.21] | 45.7 [-0.25] | 35 |
| Russell Midcap Growth | 27.7 | 2.1 [0.19] | 25.5 [-0.26] | 46.8 | 36.2 [0.23] | 10.6 [-0.19] | 44.7 | 21.3 [0.38] | 23.4 [-0.31] | 47 |
| S&P Smallcap 600 | 31.8 | 0.0 | 31.8 [-0.11] | 63.6 | 0.0 | 63.6 [-0.14] | 50.0 | 9.1 [0.09] | 40.9 [-0.35] | 22 |
| Russell 2000 | 37.6 | 13.1 [0.22] | 24.5 [-0.20] | 61.2 | 6.9 [0.24] | 54.3 [-0.31] | 70.6 | 35.9 [0.33] | 34.7 [-0.50] | 245 |
| Russell 2000 Value | 40.4 | 29.8 [0.26] | 10.6 [-0.20] | 63.8 | 2.1 [0.17] | 61.7 [-0.23] | 31.9 | 4.3 [0.16] | 27.7 [-0.26] | 47 |
| Russell 2000 Growth | 25.8 | 7.2 [0.31] | 18.6 [-0.23] | 37.1 | 3.1 [0.25] | 34.0 [-0.25] | 46.4 | 8.2 [0.36] | 38.1 [-0.41] | 97 |
| Overall | 40.0 | 12.4 [0.21] | 27.6 [-0.18] | 57.9 | 41.5 [0.26] | 16.5 [-0.24] | 61.6 | 33.3 [0.31] | 28.3 [-0.37] | 1640 |

**Table 3**
**Characteristics differences between funds and their self-designated benchmarks**

Morningstar classifies fund portfolios into one of nine size and value/growth "styleboxes" based on portfolio holdings. The mode of the time-series distribution of a fund's portfolio styleboxes is a measure of the fund's typical size and value/growth characteristics. This table displays the distribution of fund self-designated benchmarks and this measure of chracteristics. Each column displays the percentage of funds with a given benchmark index, conditional on the fund's characteristics.

| | Fund characteristics | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Benchmark | Large Value | Large Blend | Large Growth | Medium Value | Medium Blend | Medium Growth | Small Value | Small Blend | Small Growth | # funds |
| S&P 500 | 60.2 | 92.1 | 67.0 | 47.0 | 42.3 | 23.7 | 1.2 | 2.4 | 1.0 | 873 |
| Russell 1000 | 3.0 | 4.5 | 2.4 | | | | | | | 33 |
| Russell 1000 Value | 36.0 | 2.2 | | 7.2 | 1.0 | | 1.2 | | | 111 |
| Russell 1000 Growth | | 0.6 | 30.0 | | | 0.9 | | | | 116 |
| S&P Midcap 400 | 0.4 | | | 6.0 | 21.6 | 28.9 | | | 0.5 | 89 |
| Russell Midcap | | | | 7.2 | 18.6 | 5.7 | | | 0.5 | 37 |
| Russell Midcap Value | 0.4 | 0.3 | | 31.3 | 9.3 | | | 2.4 | | 40 |
| Russell Midcap Growth | | | 0.3 | | | 25.1 | | | | 54 |
| S&P Smallcap 600 | | | | | | | 3.6 | 13.5 | 4.2 | 29 |
| Russell 2000 | | 0.3 | 0.3 | 1.2 | 7.2 | 10.0 | 54.2 | 63.5 | 49.5 | 263 |
| Russell 2000 Value | | | | | | | 39.8 | 17.5 | 0.5 | 56 |
| Russell 2000 Growth | | | | | | 5.7 | | 0.8 | 44.0 | 108 |
| # funds | 264 | 356 | 373 | 83 | 97 | 211 | 83 | 126 | 216 | 1,809 |

**Table 4**
**Mismatched self-designated benchmarks**

The first column displays the percentage of funds with a given self-designated benchmark whose characteristics do not match those of the benchmark (from table 3). The second, most important, column displays the percentage of funds that are more correlated with the fund's candidate corrected benchmark than with the actual benchmark. These funds' actual benchmarks are mismatched because the corrected benchmark does a better job capturing the fund's exposure to common factors in returns. The last column displays, for funds with mismatched benchmarks, the average difference in $R^2$ of regressions of fund returns on those of the corrected benchmark and those of the actual (mismatched) benchmark.

| Benchmark | Characteristics difference (% funds) | **Benchmark mismatched (% funds)** | # funds | Average $R^2$ difference if mismatched |
|---|---|---|---|---|
| S&P 500 | 62.0 | **40.7** | 879 | 0.13 |
| Russell 1000 | 51.5 | **24.2** | 33 | 0.10 |
| Russell 1000 Value | 14.4 | **7.2** | 111 | 0.08 |
| Russell 1000 Growth | 3.4 | **3.4** | 116 | 0.07 |
| S&P Midcap 400 | 76.4 | **44.9** | 89 | 0.14 |
| Russell Midcap | 51.4 | **32.4** | 37 | 0.14 |
| Russell Midcap Value | 35.0 | **17.5** | 40 | 0.06 |
| Russell Midcap Growth | 1.9 | **0.0** | 54 | |
| S&P Smallcap 600 | 41.4 | **17.2** | 29 | 0.08 |
| Russell 2000 | 69.6 | **44.1** | 263 | 0.09 |
| Russell 2000 Value | 41.1 | **5.4** | 56 | 0.12 |
| Russell 2000 Growth | 12.0 | **4.6** | 108 | 0.04 |
| Overall | 50.4 | **31.2** | 1,815 | 0.12 |

**Table 5**
**Determinants of flows to funds with mismatched self-designated benchmarks**

Panel regressions of the form $Flow_{i,t+1} = r_{i,t} * \beta + X_{i,t} * \gamma + \delta_t + \varepsilon_{i,t}$ for funds with mismatched self-designated benchmarks, where $r_{i,t}$ is a vector of performance measures for fund $i$ in year $t$. Specifically, $R_{i,t}$ is fund $i$'s (raw) return in year $t$, $\alpha_{i,t}$ is fund $i$'s Fama-French 3-factor alpha in year $t$, $R_{Mismatched,i,t}$ is fund $i$'s self-designated benchmark's return in year $t$, $R_{Corrected,i,t}$ is fund $i$'s corrected benchmark's return in year $t$, and $R_{Mkt,t}$ is the market return in year $t$. The control variables include the natural logarithm of a fund's total net assets, the fund's age, the fund's expense ratio, and year*style interaction dummies. Heteroskedasticity-robust standard errors (in parentheses) are clustered by year. *, **, and *** denote significance at the 10%, 5%, and 1% levels respectively.

| Independent variable | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| $R_{i,t} - R_{Mismatchedi,t}$ (above 0) | 3.13*** | 1.80*** | 1.85*** | 2.37*** | 1.71*** |
| | (0.36) | (0.31) | (0.29) | (0.28) | (0.24) |
| $R_{i,t} - R_{Mismatched,i,t}$ (below 0) | 1.20*** | 0.08 | -0.01 | 0.47 | -0.08 |
| | (0.29) | (0.46) | (0.46) | (0.29) | (0.41) |
| $R_{i,t} - R_{Corrected,i,t}$ (above 0) | | 1.91*** | 1.55** | | 1.15** |
| | | (0.45) | (0.51) | | (0.37) |
| $R_{i,t} - R_{Corrected,i,t}$ (below 0) | | 1.17** | 0.89** | | 0.76* |
| | | (0.41) | (0.36) | | (0.35) |
| $R_{i,t} - R_{Mkt,t}$ (non-S&P 500, above 0) | | | 0.63 | 0.92 | 0.56 |
| | | | (0.61) | (0.65) | (0.63) |
| $R_{i,t} - R_{Mkt,t}$ (non-S&P 500, below 0) | | | 0.78* | 0.85 | 0.67* |
| | | | (0.37) | (0.46) | (0.36) |
| $\alpha_{i,t}$ (above 0) | | | | 0.85* | 0.75* |
| | | | | (0.45) | (0.38) |
| $\alpha_{i,t}$ (below 0) | | | | 0.65** | 0.52** |
| | | | | (0.20) | (0.23) |
| Additional controls: | $\ln(TNA)_{i,t}$, $Age_{i,t}$, $ExpRatio_{i,t}$, $year * style$ dummies. | | | | |
| $R^2$ | 0.28 | 0.30 | 0.30 | 0.31 | 0.31 |
| Num. Obs. | 2,113 | 2,113 | 2,113 | 2,113 | 2,113 |

**Table 6**
**Mismatched and corrected benchmarks: pricing tests**

Pricing tests for each mismatched-corrected benchmark pair. The first three columns show the mismatched benchmark, the corrected benchmark, and the number of corresponding funds ("Ru" abbreviates "Russell"). For example, the S&P 500 is the corrected benchmark for 2 funds whose self-designated benchmark is the Russell 1000 Value and is mismatched. Model 1 refers to excess-return time series regressions in which the 25 Fama-French size and book-to-market portfolios are the dependent portfolios and the explanatory portfolios are the Fama-French factors (Rm-Rf, SMB, and HML) and the corrected benchmark. Av $|\alpha|$ is the average absolute value of the 25 estimated intercepts (alphas). GRS is the F-statistic of Gibbons, Ross, and Shanken (1989), testing the hypothesis that all 25 intercepts are equal to 0.0. Model 2 is the same as model 1 but also adds the mismatched benchmark as an additional explanatory portfolio. Av $|\alpha|$ Diff. is the difference between Av $|\alpha|$ for model 1 and model 2 (i.e. the difference in average absolute pricing errors). $p\left(\Delta\right)$ is the p-value for the hypothesis that the intercepts in model 1 are no larger than those in model 2. The test statistic is $\left(\alpha_1' var\left(\alpha_1\right)^{-1}\alpha_1 - \alpha_2' var\left(\alpha_2\right)^{-1}\alpha_2\right)$, where $\alpha_1$ is a vector of the 25 estimated intercepts from model 1 and $var\left(\alpha_1\right)$ is the corresponding estimated variance-covariance matrix. The reported p-values are obtained by bootstrapping this test statistic. "Spread port." refers to a zero-cost portfolio long the mismatched benchmark and short the corrected benchmark. $\alpha$ and $t\left(\alpha\right)$ are the intercept and t-statistic of the intercept from a time-series regression of the spread portfolio on the Fama-French factors. Spread portfolios GRS is the GRS statistic for the hypothesis that all of the spread portfolio intercepts are equal to 0.0. All regressions use 127 monthly observations from January 1994-July 2004.

| Mismatched Benchmark | Corrected Benchmark | N | Model 1 Av $\|\alpha\|$ | Model 1 GRS | Model 2 Av $\|\alpha\|$ | Model 2 GRS | Av $\|\alpha\|$ Diff. | $p\,(\Delta)$ | Spread port. $\alpha$ | Spread port. $t\,(\alpha)$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Ru 1000 Value | S&P 500 | 2 | 0.25 | 2.71 | 0.21 | 2.62 | 0.04 | 0.250 | -0.16 | -1.74 |
| Ru 1000 Value | Ru Mid Value | 4 | 0.21 | 2.73 | 0.20 | 2.73 | 0.00 | 0.644 | 0.04 | 0.47 |
| Ru 1000 Value | S&P 400 | 1 | 0.26 | 2.84 | 0.23 | 2.78 | 0.03 | 0.283 | -0.13 | -0.78 |
| Ru 1000 Value | Ru 2000 Value | 1 | 0.19 | 2.75 | 0.19 | 2.84 | 0.00 | 0.838 | 0.14 | 1.31 |
| Ru 1000 | Ru 1000 Value | 6 | 0.22 | 2.80 | 0.24 | 2.50 | -0.02 | 0.043 | 0.15 | 1.73 |
| Ru 1000 | Ru 1000 Growth | 2 | 0.23 | 2.64 | 0.25 | 2.55 | -0.02 | 0.193 | -0.11 | -1.39 |
| S&P 500 | Ru 1000 Value | 108 | 0.22 | 2.80 | 0.21 | 2.62 | 0.01 | 0.114 | 0.16 | 1.74 |
| S&P 500 | Ru 1000 Growth | 153 | 0.23 | 2.64 | 0.23 | 2.56 | 0.00 | 0.202 | -0.10 | -1.16 |
| S&P 500 | Ru Mid Value | 27 | 0.21 | 2.73 | 0.20 | 2.60 | 0.00 | 0.176 | 0.21 | 1.72 |
| S&P 500 | S&P 400 | 26 | 0.26 | 2.84 | 0.28 | 2.68 | -0.03 | 0.128 | 0.03 | 0.19 |
| S&P 500 | Ru Mid Growth | 39 | 0.25 | 3.06 | 0.28 | 2.85 | -0.02 | 0.117 | 0.00 | 0.01 |
| S&P 500 | Ru 2000 Value | 1 | 0.19 | 2.75 | 0.18 | 2.56 | 0.00 | 0.081 | 0.30 | 2.66 |
| S&P 500 | Ru 2000 | 2 | 0.18 | 2.69 | 0.18 | 2.64 | 0.00 | 0.267 | 0.37 | 3.09 |
| S&P 500 | Ru 2000 Growth | 2 | 0.17 | 2.47 | 0.17 | 2.46 | 0.00 | 0.426 | 0.47 | 2.90 |
| Ru 1000 Growth | S&P 500 | 2 | 0.25 | 2.71 | 0.23 | 2.56 | 0.02 | 0.132 | 0.10 | 1.16 |
| Ru 1000 Growth | Ru Mid Growth | 2 | 0.25 | 3.06 | 0.23 | 2.88 | 0.02 | 0.097 | 0.10 | 0.56 |
| Ru Mid Value | Ru 1000 Value | 1 | 0.22 | 2.80 | 0.20 | 2.73 | 0.02 | 0.163 | -0.04 | -0.47 |
| Ru Mid Value | S&P 400 | 4 | 0.26 | 2.84 | 0.21 | 2.64 | 0.04 | 0.094 | -0.17 | -1.35 |
| Ru Mid Value | Ru 2000 | 2 | 0.18 | 2.69 | 0.17 | 2.66 | 0.01 | 0.346 | 0.16 | 1.67 |
| Ru Mid | Ru Mid Value | 5 | 0.21 | 2.73 | 0.21 | 2.74 | 0.00 | 0.594 | 0.09 | 1.15 |
| Ru Mid | Ru Mid Growth | 6 | 0.25 | 3.06 | 0.22 | 2.97 | 0.04 | 0.142 | -0.12 | -0.96 |
| Ru Mid | Ru 2000 Growth | 1 | 0.17 | 2.47 | 0.16 | 2.42 | 0.01 | 0.439 | 0.35 | 3.17 |
| S&P 400 | Ru 1000 Value | 1 | 0.22 | 2.80 | 0.23 | 2.78 | -0.01 | 0.469 | 0.13 | 0.78 |
| S&P 400 | Ru Mid Value | 2 | 0.21 | 2.73 | 0.21 | 2.64 | -0.01 | 0.212 | 0.17 | 1.35 |
| S&P 400 | Ru Mid Growth | 36 | 0.25 | 3.06 | 0.25 | 3.03 | 0.00 | 0.282 | -0.03 | -0.21 |
| S&P 400 | Ru 2000 Growth | 1 | 0.17 | 2.47 | 0.18 | 2.30 | -0.01 | 0.151 | 0.43 | 2.92 |
| Ru 2000 Value | Ru 2000 | 3 | 0.18 | 2.69 | 0.18 | 2.66 | 0.00 | 0.328 | 0.06 | 0.84 |
| Ru 2000 | S&P 500 | 1 | 0.25 | 2.71 | 0.18 | 2.64 | 0.07 | 0.266 | -0.37 | -3.09 |
| Ru 2000 | Ru 1000 Growth | 1 | 0.23 | 2.64 | 0.19 | 2.62 | 0.04 | 0.391 | -0.46 | -3.15 |
| Ru 2000 | Ru Mid Value | 1 | 0.21 | 2.73 | 0.17 | 2.66 | 0.03 | 0.320 | -0.16 | -1.67 |
| Ru 2000 | S&P 400 | 4 | 0.26 | 2.84 | 0.16 | 2.47 | 0.10 | 0.064 | -0.33 | -2.43 |
| Ru 2000 | Ru Mid Growth | 14 | 0.25 | 3.06 | 0.18 | 2.76 | 0.08 | 0.089 | -0.36 | -2.17 |
| Ru 2000 | Ru 2000 Value | 34 | 0.19 | 2.75 | 0.18 | 2.66 | 0.01 | 0.207 | -0.06 | -0.84 |
| Ru 2000 | Ru 2000 Growth | 61 | 0.17 | 2.47 | 0.18 | 2.67 | -0.01 | 0.926 | 0.10 | 1.45 |
| S&P 600 | Ru 2000 Growth | 5 | 0.17 | 2.47 | 0.17 | 2.49 | 0.00 | 0.643 | 0.24 | 1.82 |
| Ru 2000 Growth | Ru Mid Growth | 5 | 0.25 | 3.06 | 0.17 | 2.55 | 0.09 | 0.038 | -0.47 | -3.01 |
| | | | | | | | | Spread portfolios GRS | 1.67 | |
| | | | | | | | | F-distribution p-value | 0.089 | |
| | | | | | | | | Bootstrap p-value | 0.143 | |

**Table 7**
**Probability of having a mismatched self-designated benchmark**

Fund-level logit models for the probability of having a mismatched self-designated benchmark. The dependent variables are indicator variables for whether the fund has a mismatched benchmark on size, value/growth, or either. The independent variables are the time-series average of the fund's expense ratio, the time-series average of the natural logarithm of the fund's assets under management, both measured relative to all funds with the same style, and dummy variables for whether the fund is a value fund, growth fund, small-cap fund, or large-cap fund. The models also include fund family fixed effects for each of the 44 fund families that have at least 10 funds in the sample. Heteroskedasticity-robust standard errors (in parentheses) are clustered by fund family. *, **, and *** denote significance at the 10%, 5%, and 1% levels respectively.

| | Benchmark mismatched on value/growth | Benchmark mismatched on size | Benchmark mismatched |
|---|---|---|---|
| Value fund | 3.74*** | 0.31* | 2.48*** |
| | (0.24) | (0.17) | (0.15) |
| Growth fund | 3.56*** | -0.23 | 2.21*** |
| | (0.27) | (0.15) | (0.18) |
| Small-cap fund | -0.49*** | -3.13*** | -0.99*** |
| | (0.13) | (0.32) | (0.16) |
| Large-cap fund | -0.08 | -4.95*** | -0.71*** |
| | 0.12 | (0.37) | (0.15) |
| Expense ratio | 0.52*** | 0.20*** | 0.49*** |
| | (0.11) | (0.05) | (0.12) |
| ln (Assets) | 0.24*** | 0.09 | 0.23** |
| | (0.08) | (0.08) | (0.09) |
| % family dummies significantly > 0 | 36.4 | 36.4 | 31.8 |
| % family dummies significantly < 0 | 54.5 | 54.5 | 54.5 |
| Pseudo $R^2$ | 0.25 | 0.39 | 0.19 |
| # funds | 1,563 | 1,402 | 1,563 |