Should Benchmark Indices Have Alpha?

Revisiting Performance Evaluation[*]

Martijn Cremers[†]

Antti Petajisto[‡]

Eric Zitzewitz[§]

July 31, 2008

Abstract

Standard Fama-French and Carhart models produce economically and statistically significant nonzero alphas even for passive benchmark indices such as the S&P 500 and Russell 2000. We find that these alphas primarily arise from the disproportionate weight the Fama-French factors place on small value stocks which have performed well, and from the CRSP value-weighted market index which is a downward-biased benchmark for U.S. stocks. We explore alternative ways to construct these factors as well as alternative models constructed from common and easily tradable benchmark indices. Such index-based models outperform the standard models both in terms of asset pricing tests and performance evaluation of mutual fund managers.

---

[*] We wish to thank Nick Barberis, Ken French, Juhani Linnainmaa, and Christopher Polk for comments, as well as conference and seminar participants at the Adam Smith Asset Pricing conference, INSEAD, Wharton, and Yale. We also thank Frank Russell Co, Standard and Poor's, Dow Jones Wilshire, and Morningstar for providing us with data. We gratefully acknowledge the financial support of the Q-Group.

[†] Yale School of Management, P.O. Box 208200, New Haven, CT, 06520-8200, tel. +1-203-436-0649, martijn.cremers@yale.edu.

[‡] Corresponding author. Yale School of Management, P.O. Box 208200, New Haven, CT, 06520-8200, tel. +1-203-436-0666, http://www.som.yale.edu/Faculty/petajisto, antti.petajisto@yale.edu.

[§] Dartmouth College, 6016 Rockefeller Hall, Hanover, NH 03755, tel: +1-603-646-2891, http://www.dartmouth.edu/~ericz, ericz@dartmouth.edu

1 Introduction

Practitioners typically evaluate money managers by comparing their returns to benchmark indices, such as the S&P 500 for large-cap stocks and the Russell 2000 for small-cap stocks. In contrast, the academic literature has adopted the Carhart four-factor model and the Fama-French three-factor model as the standard benchmarks for performance evaluation. Interestingly, the two approaches can yield very different results, as evidenced by the large nonzero alphas the benchmark indices themselves get with respect to the academic factor models.

For example, regressing the S&P 500 index on the Carhart four-factor model, we get an annual alpha of 0.82% (t = 2.95) over our sample period from 1980 to 2005. The Russell 2000 has an annual alpha of –2.41% (t = –3.35). A passive portfolio that is long S&P 500 Growth and short Russell 2000 Growth has an impressive annual alpha of 5.24% (t = 3.97). Hence, even pure index funds tracking common benchmark indices would appear to have significant positive or negative "skill." Yet these indices represent broad, well-diversified, and passive portfolios which almost by definition should have zero abnormal returns or alphas. In fact, these nonzero index alphas are symptoms of a deeper issue in the Fama-French-Carhart methodology which produces biased alphas for large segments of the equity market.

In this paper we start by investigating the Fama-French methodology to identify the sources of nonzero index alphas. Using various modifications to their methodology, we develop an improved set of Fama-French factors. Furthermore, we explore alternative factor models based on common benchmark indices. Such index-based models actually perform the best in terms of both pricing and performance evaluation, and thus we propose them as good alternatives to the commonly used academic factor models.

These issues go to the heart of performance evaluation, since most money managers are benchmarked against market segments that exhibit significant positive or negative Carhart alphas. The same issues also arise in other contexts, such as evaluating the profitability of an active trading strategy. For example, portfolios of stocks with a tilt toward large growth, large value or small growth will have a significant bias in their Carhart alphas.

The main source of the nonzero index alphas comes from the Fama-French methodology of constructing the Small-minus-Big (SMB) and High minus Low book-to-market (HML) factors. The Fama-French procedure divides stocks into a 2x3 size-by-book-to-market (BM) matrix using two independent sorts, calculates value-weighted average returns for stocks in the six portfolios, and then constructs its factors using equal-weighted differences between these portfolio returns.[1] There is significantly more market capitalization in the Big size and Low BM portfolios. As a result, the equal-weighted portfolios in the Fama-French factors give more weight to a given unit of capitalization if it is in the Small size and High BM (i.e., value) portfolio. Such biases matter because small value stocks have historically outperformed other stocks by a significant margin.

For the large-cap stocks in the S&P 500, the Fama-French and Carhart models produce a market beta close to one and a negative SMB beta to eliminate the small-stock exposure of the market portfolio. Because SMB places equal weights on large value and large growth portfolios, even when the latter has more than three times the market cap, the model-implied benchmark portfolio will have a substantial overweight on large value and a negative weight on small value. A negative beta on HML offsets the large-cap value tilt but at the cost of significantly adding to the negative weight on small value stocks. The resulting outsized negative exposure to small value stocks drags down the performance of the benchmark portfolio, contributing to a positive alpha on the S&P 500.

For the Russell 2000, these models produce a market beta of about one and a large positive SMB beta to reduce exposure to large-cap stocks. However, the equal-weighting of SMB and value-weighting of the market portfolio again severely distort the allocation within large-caps, generating a tilt towards large-cap growth. This is partly offset by a positive loading on HML, but it simultaneously produces a significant overweight in small-cap value, reinforcing the overweighting of small value stocks due to the equal-weighted SMB factor. As a result, the Russell 2000 is compared against a small-cap value-heavy benchmark which has historically performed well, thereby explaining most of the negative index alpha.

Another source of positive alpha for the S&P 500 comes from the choice of the market portfolio. The Carhart model uses the CRSP value-weighted market return,[2] which includes not only U.S. firms but also non-U.S. firms, closed-end funds, and REITs. These other securities dramatically underperform U.S. stocks, getting an annual Carhart alpha of –4.01%. Since the S&P 500 and other indices typically only include U.S. stocks, using the CRSP market proxy contributes to a positive alpha.

---

[1] Specifically, SMB is defined as (Small-Low + Small-Medium + Small-High)/3 minus (Big-Low + Big-Medium + Big-High)/3, and HML is (Small-High + Big-High)/2 minus (Small-Low + Big-Low)/2.

[2] Fama and French (1993) originally use only U.S. common stocks in the market portfolio, but in subsequent papers they use the CRSP value-weighted index, which is also the "market" return provided on Ken French's website.

To see whether any part of index alphas can arise from stock selection within a style-matched portfolio, we perform attribution analysis at the level of 100 size-BM-sorted Fama-French portfolios. Interestingly, the Fama-French component portfolios themselves are mispriced by the Carhart model: the top size decile has a significant positive alpha while the small-cap deciles have significant negative alphas. For the S&P 500, 90% of its alpha comes simply from its passive exposure to the top size decile, so stock selection by the S&P index committee does not play a meaningful role in the index alpha. For the Russell 2000, over 70% of the alpha can be explained by exposure to Fama-French portfolios, indicating that most of its negative alpha arises simply from the negative Carhart alpha of the small-cap segment in general.

Index reconstitution effects are another possible explanation for the underperformance of the small capitalization indices. Petajisto (2006) points out that this is especially likely for the Russell 2000, which is reconstituted every year at the end of June, due to the combination of relatively large turnover in the index and the large amount of assets indexed and benchmarked to it. In anticipation of the one-time demand shock by index investors at the end of June, stocks being added to the Russell 2000 outperform stocks being deleted in June, and the reverse occurs in July, lowering the returns on the index itself. We find that about one half of the negative alpha of the Russell 2000 occurs during June and July, suggesting it is also associated with a reconstitution effect.

As alternatives to the Carhart and Fama-French models, we consider two different approaches: first, modifying the construction of the factors and second, using the common indices themselves as replacement factors. The indices represent well-diversified portfolios that naturally qualify as proxies for systematic factor risks, and consistent with the Arbitrage Pricing Theory of Ross (1976), they could also be related to expected returns. We consider the most widely followed index in each size category, the S&P 500, Russell Midcap, and Russell 2000, as well as their value and growth components. These indices each represent a broad but disjoint segment of the U.S. equity universe.

Our pricing results focus on the 100 Fama-French size-BM-sorted portfolios as our test assets. The four-factor Carhart model has a cross-sectional R^2 of 29% over our time period. Far from being redundant assets, benchmark indices can improve pricing significantly: adding the S&P 500, Russell Midcap, and Russell 2000 to the Carhart model increases the R^2 to 64%. As an alternative to the non-tradable Fama-French factors, a seven-factor index-based model that includes separate value-minus-growth factors for each of the S&P 500, Russell Midcap, and Russell 2000 indices, along with the usual momentum factor, has a cross-sectional R^2 of 58% with relatively low pricing errors. If the three indices themselves are used with only one value-minus-growth factor, the cross-sectional R^2 equals 48%, which is still a considerable improvement over the Carhart model while using the same number of factors.

In performance evaluation applications, we verify that the benchmark alphas can indeed have a significant impact. When mutual funds are sorted into size and value groups, the Carhart model indicates that small-cap funds underperformed large-cap funds by 2.13% per year from 1996 to 2005. However, this result arises from the fact that the four-factor Carhart alphas of the small-cap benchmark indices are on average an astounding 5.07% per year less than that of the large-cap indices. If instead we control for the benchmark index of a fund, the results are completely reversed, and we find that small-cap funds outperformed large-cap funds by 2.94%. These numbers are economically very large, especially given that mutual fund alphas are generally so close to zero – e.g. Wermers (2000) reports that the average mutual fund manager outperformed the market by about 1.3% per year before expenses and underperformed by about 1% after expenses. The best benchmark model, producing reasonable alpha estimates across all fund groups and regardless of the benchmark index, is the seven-factor index-based model.

In addition to eliminating index-specific biases in alpha estimates, we would also like to see a benchmark model closely track the time series of returns for an individual fund, as this would reduce noise in the alpha estimate of the fund. To investigate this, we compute the out-of-sample tracking error volatility for all mutual funds in our sample, take the average across all funds, and compare it between various benchmark models. We find that the tracking error volatility of the Carhart model can be reduced by subtracting the benchmark index or by using pure index-based factor models – a seven-factor index model decreases tracking error volatility by about 10% on average, and more for larger or less active funds.

The general conclusion from our analysis is that benchmark indices matter both for pricing and performance evaluation. The Fama-French and Carhart models can be particularly misleading in performance evaluation due to the large alphas they assign to passive benchmark indices, and they generate unnecessarily noisy alpha estimates. In addition, we can improve cross-sectional explanatory power in standard asset pricing tests by replacing the SMB and HML factors with index factors. Overall, the best model, both in our pricing and benchmarking tests, is a seven-factor index model, consisting of the S&P 500, Russell Midcap, Russell 2000, a separate value-minus-growth factor for each index, and a momentum factor. If we want to keep the number of factors smaller, a four-factor index model with momentum still dominates the Carhart four-factor model.

Our contribution is methodological as well as conceptual and related to the benchmarking and pricing models of Fama and French (1993), Carhart (1997), and Sharpe (1992). Sharpe's style analysis is one of the few academic studies using benchmark indices for performance evaluation, but he does not analyze model construction in any detail or evaluate alternative model specifications. Daniel, Grinblatt, Titman, and Wermers (1997) present a nonlinear benchmarking methodology based on characteristics-matched portfolios that avoid many of the issues we document, albeit at the cost of requiring knowledge of portfolio holdings and a nontrivial amount of computation. In this paper we want to focus on refining factor models that do not require holdings data, given that this approach remains quite popular among researchers and practitioners.

Chan, Dimmock, and Lakonishok (2006) investigate a similar broad question regarding the robustness of various benchmarking methodologies and the implications for performance evaluation. Their comparison is between academic benchmark models, primarily concentrating on characteristics-based models. In contrast, we focus on all the benchmark indices defined and used by practitioners, including the S&P 500 and Russell 2000, document the positive and negative long-term index alphas under the common academic factor models, and importantly, identify the sources of these alphas. Furthermore, we propose step-by-step improvements to the common academic factor models, eventually ending up with alternative benchmarking and pricing factors that are based on the common benchmark indices used by practitioners and are therefore convenient for anyone to implement. Another related paper is Huij and Verbeek (2007), who argue that transaction costs, trading impact and trading restrictions systematically bias the academic size and BM factor premia and advocate using proxies based on actual mutual fund returns as benchmarks.

This paper proceeds as follows. Section 2 discusses the underlying theoretical concepts, including the criteria for judging pricing and benchmarking models. Section 3 explains the basics of the most common benchmark indices. Section 4 presents the evidence on benchmark index alphas under the Carhart model and investigates the reasons for those alphas. Section 5 compares the pricing ability of the Carhart factors with alternative index-based factors when applied to the usual test portfolios of stocks. Section 6 examines the performance of the Carhart model relative to index-based models in the context of mutual fund performance evaluation. We present our conclusions in Section 7. All tables and figures are in the appendix.

2 Benchmarks for Asset Pricing and Performance Evaluation

2.1 Defining a Good Benchmark Model

How should we define a "good" benchmark model for portfolio performance evaluation? These criteria are not identical to those of a good pricing model, even though pricing models can also be used as benchmark models.

A pricing model should be the simplest possible model that explains the cross-section of expected stock returns. Asset pricing theory suggests that expected returns should be a linear function of betas of the portfolio with respect to one or more priced risk factors. Empirically motivated factors could in principle be derived from any stock characteristic that predicts returns.

A benchmark model should provide the most accurate estimate of a portfolio manager's value added relative to a passive strategy. This implies that a benchmark model should include the pricing model, so that the manager does not get credit for exploiting well-known cross-sectional patterns in stock returns. However, a benchmark model may also include non-priced factors to reduce noise in alpha estimates.[3] For example, even if value and size were not priced, they should still be included in a benchmark model simply because there are extended periods of time when one size-value segment significantly outperforms or underperforms the rest of the market.

The difference between a pricing model and a benchmark model may be clearest in the context of an event study conducted around a single calendar-time event. For example, in July 2002 all remaining foreign firms were removed from the S&P 500 index. Including industry returns over the event period in the benchmark model would help to more accurately estimate the short-term price impact on these firms, even when the industry exposures are not priced ex ante. Most event studies take this approach. A pricing model would ignore such industry returns and thus produce a noisier and ex post biased result.[4] Eventually the return differential due to nonpriced factors will by definition converge to zero, so including them in the benchmark model does not help over very long periods of time. But in practice the distinction can be extremely

---

[3] A benchmark model should not control for characteristics of fund managers (such as SAT scores) even if they predict fund returns, because identifying such skill is the very purpose of the model.

[4] Furthermore, even over a period of ten years, one market segment such as large-cap growth stocks may still outperform other market segments by an economically large amount, even when there was no obvious ex ante risk premium.

useful, as individual money managers often need to be evaluated over relatively short periods of time.

For the performance evaluation of a large pool of money managers, it is convenient to apply a generally applicable and relatively parsimonious model rather than deal with ad hoc adjustments due to a manager's average exposure to e.g. industry risk. Most academic literature therefore has chosen to use the most popular pricing models also as benchmark models, leading to the prevalent use of the Fama-French three-factor model and Carhart four-factor model.

In contrast to the academic literature, practitioners generally compare money managers against their self-declared benchmark indices such as the S&P 500. While the mere subtraction of the benchmark index return may oversimplify performance evaluation, a set of multiple benchmark indices may be convenient factors for pricing and benchmarking purposes. They certainly satisfy the criteria of well-diversified portfolios called for by the Arbitrage Pricing Theory of Ross (1976) to be used as proxies for systematic factor risk.[5] Therefore the factors could even be priced, and at the very least they help reduce noise in alpha estimates.

2.2 Improving Benchmark and Pricing Models

We can improve a benchmark model by improving the alpha estimates in one of two ways: by reducing noise or by reducing biases. To reduce noise in the alpha estimates for a manager, we should select a benchmark portfolio that better mimics the manager's actual portfolio. This may require including non-priced factors, as discussed earlier, or perhaps the manager's self-declared benchmark index.

In addition to random short-term noise, a manager's alpha may also have a more systematic long-term bias due to the manager's investment universe as indicated by his benchmark index. After all, a manager randomly selecting stocks within the universe of a benchmark index will on average earn the same alpha as the benchmark index. This, as we discussed in the introduction, would also lead to the odd conclusion that pure index funds would display significant positive or negative skill. To eliminate this bias when evaluating managers, we should either adjust for the alpha of the benchmark index or use a benchmark model that does not produce such biases across common benchmarks.

[5] Lehmann and Modest (1987) employ a variety of actual APT benchmarks to investigate the sensitivity of abnormal mutual fund performance to the benchmark chosen and find that rankings are quite sensitive.

2.3 Our Test Design for Benchmark and Pricing Models

To test how well a model can do as a benchmark for money managers, we test for all of the aforementioned properties. First, a new model should track the time series of returns better than the old models, producing lower tracking error volatility. Second, a model should not generate significant benchmark-specific biases in alphas for any of the common benchmark indices. Furthermore, if funds are sorted into simple styles according to value and size dimensions, the model should produce only modest alphas for all groups, and not extreme positive values for some and extreme negative for others (unless we really consider it plausible that the *average* managerial skill varies from large positive to large negative values across market segments). Third, the model should also do a satisfactory job pricing assets, so that it can explain a significant fraction of the cross-section of returns on test assets such as size and book-to-market-sorted portfolios of stocks.

Could a model do well in some of these tests but poorly in others? Certainly. The tracking error analysis only explains time-series fluctuation in fund returns but not their average level. If the average level of fund returns is explained by a characteristic or a time-varying factor exposure, instead of a constant factor exposure, it does not matter for tracking error. Yet it may matter a lot for pricing assets and also for average returns across fund style groups. Momentum could potentially play such a role: prior literature has found that it matters for pricing, but it is not a systematic risk factor in the sense that any risk exposure in a momentum portfolio is likely to consist of time-varying exposures to industry portfolios, and thus it would not necessarily matter for tracking error volatility. Alternatively, a pricing model may do well when explaining the average returns on stock portfolios over a long period of time, but if it does not include any nonpriced factors which nevertheless significantly influence stock returns, the model will produce noisy estimates of alpha over any (short) horizon practical for performance evaluation.

3 Benchmark Index Data

We include all the US equity benchmark indices that are most commonly used by practitioners. This covers a total of 23 indices from three index families: Standard and Poor’s, Frank Russell, and Dow Jones Wilshire. We have data directly from all three index providers, covering monthly and daily index returns as well as month-end index constituents.

The main S&P indices are the S&P 500, S&P MidCap 400, and S&P SmallCap 600. The S&P 500 is the most common large-cap benchmark index, consisting of approximately the largest 500 stocks. It is further divided into a growth and value style, with equal market capitalization in each, forming the S&P500 Growth and Value indices which together sum up to the S&P 500. The S&P 400 and S&P 600 consist of 400 mid-cap and 600 small-cap stocks, respectively, and they are also further divided into separate value and growth indices.

From the Russell family we have 12 indices: the Russell 1000, Russell 2000, Russell 3000 and Russell Midcap indices, plus the value and growth components of each. The Russell 3000 covers the largest 3,000 stocks in the U.S. and the Russell 1000 covers the largest 1,000 stocks. Russell 2000 is the most common small-cap benchmark, consisting of the smallest 2,000 stocks in the Russell 3000. The Russell Midcap index contains the smallest 800 stocks in the Russell 1000.

Finally, we include the two most popular Wilshire indices, namely the Wilshire 5000 and Wilshire 4500. The Wilshire 5000 covers essentially the entire U.S. equity market, with about 5,000 stocks in 2004 and peaking at over 7,500 stocks in 1998. The Wilshire 4500 is equal to the Wilshire 5000 minus the 500 stocks in the S&P 500 index, which makes it a mid-cap to small-cap index.

Since 1998, all mutual funds have had to report a benchmark index to the SEC. The popularity of each index can be seen in Table 1, which shows the self-reported benchmark indices for US all-equity mutual funds. The benchmark data is from Morningstar Direct, representing a snapshot of live funds in January 2007.[6] The most common benchmark index is the S&P500. Russell 2000 is the second-most popular benchmark, and its value and growth components are also relatively popular, while the most common general mid-cap index is the S&P400, although the Russell Midcap group of indices is collectively more popular.[7] Wilshire indices are less common in terms of the number of funds, but they each have a significant amount of assets benchmarked to them.

Figure 1 shows the fraction of ordinary common stocks of U.S. firms covered by the most common indices as a function of market capitalization. Each month and for each market cap rank, we compute the fraction of the neighboring 20 stocks (market cap ranks) that are in the index. The figure reports the average index membership density from 1997 to 2005.[8] For S&P indices in

---

[6] Some funds still have missing or unspecified benchmarks in the database, so the total number of US equity funds in the database is slightly higher.

[7] Overall, the Russell style indices have begun to dominate S&P style indices recently, whereas the S&P500 style indices used to be more popular in the 1990s. Boyer (2006) provides more details on the S&P500 style indices.

[8] The S&P 600 Growth and Value index data do not start until 1/1997.

Panel A, two features stand out: First, the indices do not cover all stocks, which arises from S&P's relatively tight selection criteria on profitability and other firm characteristics. Second, the market cap boundaries of each index are very flexible, as market cap is only one of S&P's selection criteria. In contrast, Russell indices in Panel B cover virtually their entire target universe, and they have strict market cap cutoffs.[9]

Figure 2 shows the fraction of all stocks covered by the most common indices along two dimensions: small vs. large and growth vs. value. We divide the CRSP universe into 10x10 portfolios by market cap and book-to-market as defined by Fama and French, as well as two additional groups: "N" for common stocks of U.S. firms not included in the Fama-French portfolios (such as new listings), and "O" for all other share codes (i.e., other than common stocks) in the CRSP market index. The color of each cell indicates the fraction of market cap that a particular index covers among those stocks. The common benchmark indices divide their market cap roughly equally between the growth and value components, but since the Fama-French benchmarks contain a much greater share of market cap in the growth deciles, the index coverage figures seem oddly tilted toward growth, especially among large-cap stocks in the S&P 500. The other CRSP share codes are generally not included in the U.S. equity indices, but many of the new listings are included in the Russell and Wilshire indices before they qualify for the Fama-French 10x10 portfolios.

4 Alphas of Benchmark Indices

4.1 Baseline Results

In this section, we present the evidence on non-zero benchmark index alphas that provide the main motivation for the paper. Table 2 presents estimates of Carhart alphas for the major Russell, S&P, and Wilshire indices from 1980 to 2005.[10] Alphas are positive and statistically

[9] The only reason we do not see discrete steps at 1,000 and 3,000 is that we have averaged across market cap rankings throughout the year, whereas Russell updates its indices only once a year.

[10] We use a sample period back to January 1980 when possible. For some indices (see the footnote to Table 2 for a list), the first available return data are from a later month, so for these indices our sample period is shorter. The Russell 1000, 2000, and 3000 indices were introduced in January 1984, and returns from 1980-1983 were calculated by Russell based on a back-casting of their index construction rule. One might worry that including these years creates the possibility of a bias, but given the nature of the Russell indices, which include every stock within a given capitalization range, it seems unlikely that the index construction rules were biased towards raising prior performance. We do obtain similar results (i.e. positive alphas for the Russell 1000 and a negative alpha for the Russell 2000) for 1980-1983, and dropping these 4 years does not meaningfully affect our results.

significant for the general and Growth versions of the large-cap indices (the Russell 1000 and S&P 500) and are negative and statistically significant for the general and Growth versions of the small-cap indices (the Russell 2000 and S&P 600). The alpha for the Wilshire 5000 is very close to zero as expected, given that it approximates the CRSP value-weighted index (which is included as a factor in the Carhart model).

In unreported results we examine the robustness of non-zero benchmark alphas across subperiods and models. Alphas for the general and Growth versions of the large-cap indices are positive in almost every five-year period examined, with the exception being for the general indices in 2001-2005. Likewise, they are negative in almost every period for the general and Growth versions of the small-cap indices, with the exception of the Russell 2000 in 1986-1990. Benchmark index alphas are similar for the Fama-French and Carhart models, reflecting generally minor loadings on the momentum factor. In contrast, results for the CAPM are quite different, indicating that the CAPM does not control for the outperformance of small and value stocks during our time period.

Following most of the recent literature, we calculate our alphas in-sample, estimating factor weights over our entire sample period and calculating the alpha in a given subperiod as the regression residual plus the constant. A potential weakness of this approach is that it assumes that the factor loading of the benchmarks are constant throughout the entire sample period. While this assumption is more justified for a benchmark index than for an actively managed portfolio, one might worry that, e.g., the increasing number of stocks caused the exposure of the Russell 2000 to various decile portfolios to change and that this change might be correlated with the performance of the decile portfolios in a way that biases our results. In unreported results, we have estimated three and four-factor alphas using betas estimated from a trailing 60-month window and calculated benchmark alphas that were qualitatively similar.

4.2 Sources of Benchmark Alphas

4.2.1 Construction of the Market and Fama-French Factors

The standard Fama-French model makes a number of methodological choices. In this section we reexamine these choices and consider whether they contribute to the benchmark alphas. Fama and French (1993, p. 9) note that the choices made in constructing their factors “are arbitrary ... and we have not searched over alternatives.” Presumably, they avoided searching over alternatives to avoid the temptation to data mine. This is an important concern for us as well; in proposing or recommending alternative choices, we are always guided by an effort to mimic the choices made in the construction of the actual benchmark indices and real-world portfolios.

Specifically, we examine four choices: 1) the universe of assets included in the market factor, 2) the weighting of component portfolios when constructing factors, 3) the imposition of a common value factor for small and large stocks, and 4) the boundaries between size and book-to-market (BM) categories. In each case, we propose alternative choices that are more consistent with the construction of the benchmark indices and real-world portfolios. We find that these alternative choices lead the factor models to more closely approximate the mix of stocks held by the index or portfolio in question, and individually and collectively reduce benchmark alphas and their variance.

For their market proxy, Fama and French (1993) use a value-weighted portfolio of the stocks they use in their Size and BM portfolios, plus stocks with negative book equity. Specifically, they include common stocks of U.S.-headquartered and listed firms (CRSP share codes 10 and 11) that have a sufficiently long history,[11] thus excluding new issues. Carhart (1997) and most of the subsequent literature instead use the CRSP value-weighted index, which includes all U.S.-headquartered and listed common stocks, as well as closed-end funds, REITs, foreign firms with primary listings in the U.S., and other asset types such as certificates, shares of beneficial interest, and units.[12] This is also the market return researchers commonly obtain from Ken French's website.

It turns out that the choice of which securities to include in the market proxy significantly affects risk-adjusted returns. Table 3 reports Carhart alphas for the different components of the CRSP value-weighted index, which has an alpha of exactly zero by construction since it is included as a factor in the model. U.S. common stocks (share codes 10 and 11) collectively have an alpha of 23 basis points per year over our 1980-2005 period, while the stocks included in the Fama-French size-BM-sorted portfolios have an alpha of 51 basis points per year. These differences are explained respectively by the underperformance of other assets such as closed-end

---

[11] This means that Compustat and CRSP data for the firm must have started 3.5-4.5 years and 0.5-1.5 years earlier, respectively, depending on the month.

[12] American Depository Receipts (ADRs) are the only securities included in the CRSP dataset but excluded from the CRSP value-weighted index.

funds and of stocks with insufficient data or negative book value. The underperformance of the latter group is also consistent with the general long-term underperformance of IPOs.[13]

Given that the Carhart model is most often used as a benchmark for domestic non-specialized equity mutual fund portfolios, we can use the holdings of these portfolios or their self-declared benchmark indices as a guideline for what to include in the market factor (Table 3). New issues are included in these portfolios, while closed-end funds, foreign firms, and assets such as shares of beneficial interest are excluded from the indices and are held at much lower rates by funds, if at all. Foreign firms are less likely to be included in indices or funds. REITs are the closest call; they are held by the benchmark indices and by some equity mutual funds, but less so by the domestic non-specialized equity funds that the Carhart model is typically applied to. For this reason, we exclude them from the market factor, but as their inclusion affects the average return of the market proxy by less than one basis point per year, results are very similar if they are included.

The second choice involves the weighting of stocks in constructing factors. In their seminal paper, Fama and French (1993) construct factors capturing the relative performance of small and value stocks using the following procedure. They sort U.S. common stocks into six value-weighted portfolios based on whether a stock's market capitalization is "Big" (above the NYSE median) or "Small" (below the median) and whether its book-to-market (BM) ratio is "High" (top 3 deciles), "Medium" (middle 4 deciles), or "Low" (bottom 3 deciles). Fama and French then equal-weight across these six portfolios in constructing their factors; their small-minus-big (SMB) factor is (Small-Low + Small-Medium + Small-High)/3 – (Big-Low + Big-Medium + Big-High)/3 and their high-minus-low BM (HML) factor is (Small-High + Big-High)/2 – (Small-Low + Big-Low)/2. Given that the total market capitalization differs significantly across the six portfolios (Table 4, Panel B), this implies that a given unit of market capitalization receives a different weight in the Fama-French factors depending on which 2x3 portfolio it is part of (e.g., more than 3 times more weight if it is in Big-High as opposed to Big-Low). As mentioned above, Fama and French exclude stocks with negative book equity or with no book equity data available for the fiscal year ending in the prior calendar year from the six portfolios, and so these stocks receive zero weight in their factors.

---

[13] See Ritter (1991) for the long-term IPO performance, and see also Barber and Lyon (1997), who discuss the associated reverse problem of the 'new listing bias, which arises because … sampled firms generally have a long post-event history of returns, while firms that constitute the index typically include new firms that begin trading subsequent to the event month.'

Table 4 examines the effect of equal-weighting and the exclusion of stocks with no or negative book equity (which we call "No BM" or “None” for short) on the average returns of the SMB factor during our sample period. Panel A reports average returns for Fama and French’s 2x3 portfolios, along with two portfolios of "No BM" stocks (Big-None and Small-None). Panel B reports the average share of the CRSP VW index represented by these portfolios. It is apparent that: 1) the outperformance of Small stocks is present mainly among Medium and High BM stocks, and 2) within the Small and Big stocks, the Low BM stock portfolios represent more market capitalization than the Medium and High BM portfolios. Relative to using value-weighted Small and Big portfolios, equal-weighting overweights the Medium and High BM stocks where the outperformance of small stocks is largest. In the first two columns of Panel C, we calculate the Small minus Big factor using equal and value-weighted Small and Big portfolios and find that value-weighting reduces the return spread by about 10 basis points per year.

It is also apparent from Table 4 that: 3) the No BM stocks underperform, and 4) they compose a much larger share of the Small portfolio than of the Big portfolio. Excluding them therefore increases average SMB returns by 24 basis points per year. Taken together, value weighting and adding the No BM stocks reduces average SMB returns from 1.47 percent per year to 1.13 percent per year. Since small and large-cap benchmark indices and portfolios include new stocks from the No BM portfolios and do not exhibit the overweighting of High BM stocks of the Fama-French factors, we examine how the latter version of SMB affects benchmark alphas in the following section.

The third choice made by Fama and French and followed by the subsequent literature is to apply a single value factor (HML) that equal-weights the outperformance of value stocks among Small and Big stocks. As the returns in Table 4, Panel A illustrate, the outperformance of High BM stocks over Low BM stocks is much more pronounced among Small stocks (13.21 – 4.85 = 8.36 percent per year) than Big stocks (9.20 – 7.61 = 1.59 percent per year). Using a model that forces the large-cap and small-cap value effects to be equal is likely to generate positive alphas for Small-Value and Large-Growth portfolios and negative alphas for Small-Growth and Large-Value portfolios, and this is indeed what we find in Table 2. While the Arbitrage Pricing Theory of Ross (1976) predicts that returns should be linearly related to factors, APT does not rule out separate value factors for large and small stocks. Indeed, the industry practice of focusing portfolios on a particular capitalization range makes a decoupling of the large and small-cap value effects seem less surprising. As a result, we will experiment with models that allow for separate Big and Small stock HML factors (BHML and SHML, respectively; see also Moor and Sercu (2006)).

The fourth choice we will revisit is the partition of stocks into two size categories (Big and Small) and three or four BM categories (Low, Medium, or High BM, and None). In contrast, the industry practice has historically been to partition stocks into three or four size categories (Large, Mid, Small, and Micro) but only two BM categories (Growth and Value, with some indices and portfolios including both), and this practice is reflected in the Russell and S&P family of indices.

Figure 2 shows how the holdings of the benchmark indices map into the Fama-French 10x10 portfolios. Given that size deciles are defined using NYSE breakpoints, the mapping of indices like the Russell 1000 (which includes approximately the 1,000 largest capitalization stocks) has not changed as much over our sample period as one might expect. The S&P 500 primarily includes stocks from NYSE deciles 9 and 10, while midcap stocks are drawn mostly from deciles 6-8. The Russell 2000 includes stocks from deciles 2-5, while the microcaps (included only in the Wilshire 5000) are primarily in decile 1. The Growth components of the benchmark indices include stocks from only the 2-3 lowest-BM deciles, while stocks in the other 7-8 deciles are usually in the Value component. This is because the indices construct the Growth and Value components so that they evenly divide the market-cap of the index, and this leads the Value component to include many more stocks.

Table 6 reports the SMB and HML betas of the Fama-French 10x10 Size-BM portfolios, along with an eleventh column of stocks with No or Negative book equity data. Three observations can be made. First, only the largest cap decile is clearly negatively correlated with SMB; the Midcaps (deciles 6-8) are positively correlated with SMB despite being included among Big stocks, which should mechanically induce a negative correlation. Second, BM deciles 4-9 (Medium and High in the Fama-French scheme) are all positively correlated with HML. Third, the None (No BM) column has a modest negative correlation with HML. One could argue, based on these correlations, that Midcaps should be included with Small rather than Large cap stocks; Medium BM stocks should be included with High BM stocks, and the None portfolio of stocks should be included with Low BM stocks.

Given these results, we suggest modifications to make the academic partitions more similar to the industry approach. The first is to divide Big stocks (NYSE deciles 6-10) into Large (deciles 9-10) and Mid-cap stocks (deciles 6-8). The second modification is to include Medium BM stocks with High BM stocks. We do not include the None portfolios with the Low BM stocks, since some of these stocks can even be characterized as extreme value stocks (e.g., those in financial distress with negative book equity), although including them makes little difference to the results that follow.

4.2.2 Construction of the Factors: An Illustration

To illustrate the aforementioned arguments and to clarify the intuition behind the nonzero index alphas, let us consider two "target portfolios:" Fama-French size decile 10, which contains the typical large stocks in the S&P 500 index, and size decile 4, which contains the typical small stocks in the Russell 2000 index. A regression of either portfolio on the Fama-French factors determines an appropriate three-factor benchmark portfolio, where the alpha is the difference in return between the target and the benchmark portfolio. Ideally, the benchmark portfolio should have the same broad category exposures as the target portfolio; if the two differ significantly, this may be a source of nonzero alpha. We conduct the analysis for the Fama-French three-factor model to keep it more transparent, but the mechanism is virtually identical for the Carhart model with the added momentum factor.

Panels A and B in Table 5 show the portfolio weights of the Fama-French factor portfolios using the 2x3 size-BM grid extended with a fourth column for all other stocks. The market portfolio weights are like in Table 4, except that now that "None" column includes all the securities included in the CRSP market portfolio (and not just U.S. stocks).

The left-hand side of Panel C shows the weights that the size decile 10 (large stocks) has on the 2x4 grid. The right-hand side of the panel shows the regression coefficients when the return on this portfolio is regressed on the returns on the Fama-French factors: the negative beta on SMB was expected, but the nonzero beta on HML may be surprising. Below the factor betas, we see the 2x4 portfolio weights implied by the three-factor model.

The 2x4 weights of the target portfolio differ from the benchmark weights particularly in small caps, where the target portfolio has a zero weight and the benchmark portfolio has an aggressive –19.1% weight; furthermore, a 13% difference comes from small value stocks alone. Since the benchmark is so heavily underweighted in small value which has performed very well (see Panel A), it has suffered from poor performance over long time periods, contributing to a positive alpha on the target portfolio.

Why does the benchmark portfolio get such a large underweight on small value? Since the market beta is about one, we start the benchmark portfolio with essentially the market weights in Panel A. As previously discussed, SMB places equal weights on all six component portfolios (Panel B), so it will reduce the weight on small value stocks (market weight 2.0%) too much compared to small growth stocks (market weight 3.5%). Furthermore, a large negative beta on SMB will increase too much the weight on large value while not increasing enough the weight on large growth. To reduce this overweight on large value, we get a negative beta on HML. But this comes at the cost of reducing the weight on small value even more, producing a 13% underweight.

The small stocks in size decile 4 exhibit largely the opposite effect. When regressed on the three-factor model, market beta is again about one, but SMB and HML betas are positive. The equal-weighting of SMB implies that the large positive SMB beta produces an overweight in small value and underweight in small growth. Furthermore, the SMB weights would generate a considerable growth bias in large stocks: about +18% weight in large growth and –15% weight in large value. A positive HML is needed to offset this growth tilt, but it comes with the cost of increasing the small-cap value bias even more. As a result, the benchmark portfolio has a 40% weight on small value while the target portfolio has only 19% on it, with the opposite weights on small growth. Given the performance record of small value relative to small growth (Panel A), this value tilt in the three-factor benchmark makes a significant contribution to a negative alpha on the target portfolio.

4.2.3 Benchmark Alphas from Alternative Models

In this subsection, we examine how alternative choices in constructing factors affect the loading of these factors on Size-BM portfolios and how they affect benchmark alphas. Panel A of Table 7 contains the results for the S&P 500 and Panel B for the Russell 2000.[14] Each panel estimates several alternative models and calculates the weights implied by the resulting betas on a 3x4 set of Size-BM portfolio (Large, Mid, and Small size; Low, Medium, High, and No BM).[15] These implied weights are then compared with both the weights estimated from flexible models (which include each of the 12 portfolios as a factor) and with the actual percentage of the index accounted for by each portfolio as calculated from holdings data. This comparison helps identify instances in which the structure of the factor model leads to a mismatch between the model-implied loadings on the 3x4 portfolios and the index's actual loadings.

The first column in Table 7A estimates the standard Carhart four-factor model for the S&P 500. It has a beta on the CRSP VW index of about 1.0 and a beta on SMB of -0.21, with

[14] The full table (available upon request) contains 9 panels, one each for the combined, Growth, and Value versions of the S&P 500, Russell 2000, and Russell Midcap.

[15] Each model implies a benchmark portfolio, given by the sum of product of the Fama-FrenchCarhart factor portfolios and the estimated betas. This particular benchmark portfolio (i.e., the 'fitted' or explained return) in turn implies specific weights on the portfolios in the 3x4 size-by-BM space, which can be quite different from the actual average weights of the benchmark on these portfolios (based on the flexible model including all 12 factors or the holdings).

very small betas on HML and UMD. As a result of the negative SMB beta, the model loads more heavily than the market factor on the Large and Midcap Low, Medium, and High BM portfolios. If we compare the model's loadings to the actual holdings of the index, we see that the Carhart model overweights the Midcap portfolios and produces a value tilt in large-cap stocks. It also has significant negative weights on small stocks, particularly small value which has performed well.

Subsequent columns modify the Carhart model as described above. The second column replaces with CRSP-VW with a value-weighted average of only U.S. common stocks (share codes 10 and 11). The third column replaces the equal-weighted SMB of Fama-French with a version that value-weights the High, Medium, and Low BM portfolios; the fourth column also includes the No BM stocks in SMB. The fifth column replaces HML with BHML and SHML, while the sixth column replaces SMB with SMM (Small minus Mid) and MML (Mid minus Large). The seventh column adds a Midcap HML factor (i.e., Mid-High minus Mid-Low) and changes BHML to include only the top two size deciles (i.e., the true large-caps). The eighth column includes Medium BM stocks with High BM stocks when constructing the HML factors.

As the models become more flexible, one can observe the fit improve between the implied and actual weights on the 3x4 portfolios. For the S&P 500, the most significant improvement comes from splitting the SMB factor into SMM and MML; this change keeps the model from being forced to include Midcaps in the benchmark. The second most significant improvement comes from including the No BM portfolios in SMB; this allows Large-None and Small-None to have weights above and below their weight in the market portfolio, respectively.

The alpha of the S&P 500, which is 82 basis points per year in the Carhart model, also declines as the models become more flexible. Replacing the CRSP-VW index with U.S. common stocks (column 2) reduces the alpha by 23 basis points, or roughly the difference in the average returns of these two indices. Value-weighting SMB (column 3) decreases the alpha by another 26 basis points to 33 basis points per year, which is no longer statistically significant. Replacing HML with BHML and SHML (column 5) further decreases the alpha to 11 basis points per year, whereas more elaborate models (columns 6-8) marginally increase the alpha to about 20 basis points. Overall, the first two steps (up to column 3) are the most important in terms of reducing the alpha and bringing the model-implied 3x4 portfolio weights closer to the actual index weights.

Table 7B conducts the same exercise for the Russell 2000. The Carhart model estimates loadings of approximately 1.1, 0.8, and 0.2 on the market, SMB, and HML, respectively. The positive HML loading may be puzzling, as a comparison of the actual loadings of the Russell

2000 and those of the market proxy on the Small portfolios suggest that the market proxy has relative loadings that are approximately correct, and the equal-weighted SMB factor adds an overweighting of small value. The problem is that the market beta of 1.1, combined with a 0.8 beta on the equal-weighted SMB, produces a significant negative loading on Big-High and positive loading on Big-Low. The positive HML beta partially corrects, but at the cost of adding to the overweighting of small value already present due to the equal-weighting of SMB.

By giving the Russell 2000 a benchmark that overweights Big-Low and Small-High and underweights Big-High and Small-Low, the Carhart model depresses the estimated alpha in time period such as ours, when the value effect is larger among small stocks. Switching from an equal to a value-weighted SMB in column 3 increases the estimated alpha by full percentage point per year, from -2.6 to -1.6 percent. Even in the more flexible models, the negative alpha of the Russell 2000 remains significant, although as we show below, the remaining alpha is concentrated in June and July, suggesting that it is related to the annual reconstitution of the index on June 30.

Panels for the other indices (available on request) reveal significant improvements in the fit between the model-implied and actual weights on the 3x4 portfolios where one would expect. For the Russell Midcap index, the most significant improvement comes from splitting SMB into SMM and MML. For the Growth and Value components of the S&P 500 and Russell 2000, splitting HML into BHML and SHML yields the biggest improvement. For the Growth and Value components of the Russell Midcap index, improvements come from splitting SMB and adding MidHML.

For most indices, the most flexible model (8), which contains 7 factors, yields nearly as close a fit as the fully flexible model, which includes 13 (UMD and the 12 3x4 portfolios). This is partly due to the flexible model estimating negative betas on certain 3x4 portfolios. A non-negative least squares (NNLS) version of this model, which restricts betas on the 12 portfolios to be non-negative, yields a closer fit with the actual index holdings but qualitatively similar alphas to the fully flexible model.

Table 8 presents an overview of the results for the nine indices. The absolute value of average index alphas and the sum of their squares clearly decline as one moves from left to right and the methodological gap between the academic model and portfolio and index construction in the financial industry narrows. The fit between the models' implied loadings on the 3x4 portfolios and the actual holdings also improves. It is important to note that while each improvement from model (4) to (7) comes at the expense of adding a factor to the model, the more sizeable reduction in alphas from model (1) to (4) does not. While the performance of model (4) appears to clearly dominate model (1), further improvements in performance come at the expense of adding factors, and thus involve tradeoffs in terms of the amount of data required to obtain accurate beta estimates.

4.2.4 Attribution Analysis

Is there an upper bound on how much of the index alphas we can hope to explain with factor models based on size- and value-sorted portfolios? To answer this question, we can trace the index alphas to two possible sources: 1) exposure to passive size- and value-sorted portfolios, and 2) stock selection within these broader portfolios. The decomposition between the two sources of alpha tells us whether the index stocks have different returns relative to other stocks with similar characteristics – for example, whether S&P tends to select higher-alpha stocks for its indices. The stock selection alpha is unlikely to be explained with any factor model, but the rest of the alpha in principle could be explained as it comes from passive and broad-based portfolios of stocks. Furthermore, whichever the source of alpha turns out to be, we would like to identify what subset of stocks it comes from.

As benchmark portfolios for this attribution analysis, we pick the 10x10 Fama-French portfolios which are also the basis for creating the common Fama-French factors. To cover the full universe of the CRSP market index, we again add 10x2 portfolios to include the remaining U.S. firms as well as the other share codes, as discussed earlier.

Panel A in Table 9 shows that the four-factor alpha of the S&P 500, about 81 bp per year, comes almost exclusively from just two of the 100 Fama-French portfolios. The two most extreme growth portfolios within the top size decile have large positive four-factor Carhart alphas of 371 bp and 296 bp per year (Panel A in Table 6), and they contain about 35% of the value of the S&P 500 index.[16] The more value-oriented large-cap portfolios have negative alphas, but given the smaller weight of the index in these portfolios, the overall effect is not enough to offset

[16] This analysis is based on the holdings of Fama-French portfolios and benchmark indices. Because we do not perfectly replicate the 10x10 Fama-French component portfolios, some small discrepancies arise when compared to the 10x10 portfolio returns from Ken French's web site. Nevertheless, the match is economically very close and does not seem to affect our results. Furthermore, the index alphas in this analysis also differ from the official results by a 1-3bp per year because the attribution analysis requires that we compute index returns from month-end holdings.

the two growth portfolios. Overall, 73 bp out of 81 bp of the S&P 500 alpha comes from the top market cap decile.[17]

We also compare the alphas of the S&P 500 index stocks with the alphas of all Fama-French constituent stocks in each portfolio. We construct a mimicking portfolio that has the same weights on the 10x10 Fama-French portfolios as the S&P 500, except that the mimicking portfolio holds all stocks in each 10x10 portfolio. The attribution analysis for this portfolio looks almost exactly like that for the S&P 500; the differences between the two are shown in Panel A. All the differences are only a few basis points per year, and the total difference is 11 bp per year, indicating that almost 90% of the S&P 500 alpha comes from its exposures to particular Fama-French portfolios and not from any well-informed stock selection by the S&P index selection committee.

Panel B repeats the same analysis for the Russell 2000. In contrast to the S&P 500, we see that the alpha is spread rather uniformly across all Fama-French component portfolios. Compared to a portfolio that holds the same weights in all Fama-French component portfolios, the Russell 2000 exhibits some negative "stock selection," amounting to 69 bp per year. Almost all of it comes from the upper and lower boundaries of the index (size deciles 2 and 5-6, while size deciles 3-4 show very little selection alpha). This suggests that index reconstitution may be creating a slight drag on returns, consistent with Petajisto (2006). However, about 70% of the Russell 2000 negative alpha, 169 bp out of 238 bp per year, still comes simply from its exposure to Fama-French portfolios.

If we were to repeat this exercise for other indices, Panel A in Table 6 would allow us to predict the results when combined with the index membership density from Figure 2. For example, Figure 2 indicates that S&P 500 Growth has virtually all of its weight in the three large-growth corner portfolios, all of which have large positive alphas in Table 6, so the index alpha will be the highest among the group. The largest negative alphas occur for small growth stocks, the cells targeted by the Russell 2000 Growth and S&P 600 Growth, so these indices must have the largest negative Carhart alphas, purely because they happen to cover that segment of the equity market.

---

[17] The alpha contributions of individual cells do not add up exactly to the marginal portfolio alphas because each cell alpha is estimated separately, and due to time-variation in weights across cells this is not the same as estimating the value-weighted marginal portfolio alpha (without time-variation in weights, the numbers would add up exactly). Because portfolio weights across the 100 Fama-French portfolios are more stable across size than across value deciles, the alpha contributions add up better across size deciles.

4.2.5 Index Reconstitution

As mentioned above, an additional possible explanation for the negative alpha of the small-cap indices is given by the index reconstitution effects discussed by Petajisto (2006). Additions to and deletions from the Russell indices are determined once per year based on closing market capitalizations on May 31 and are implemented at the end of June.[18] Stocks being added to the Russell 2000 outperform those being deleted in June, as one might expect if arbitrageurs were purchasing added stocks and selling deleted stocks in anticipation of index tracking portfolios being forced to trade them at the end of June. Some of the excess performance of the added stocks in June reverts in July.

These return patterns should depress the returns of the Russell 2000 relative to non-Russell 2000 stocks and may contribute to the negative alpha we find. One would expect these effects to be concentrated in June and July, and thus a simple test of whether the index reconstitution effect is an important source of the negative alpha of the Russell 2000 is to compare the June and July alphas with those from other months. In Table 10, we estimate for the Russell 2000 and its growth component three models: the Carhart model, model (4) (Carhart with a market factor that includes only U.S. common stocks and a value-weighted SMB factor that includes the No BM portfolios), and model (8) (model 4, with SMB split into SMM and MML, HML replaced by BHML, MidHML, and SHML, and the Medium BM stocks included with the High BM stocks in the HML factor). We add to each model an indicator variable for June and July; the constant in the model captures the average alpha from August to May, while the June-July coefficient captures any extra alpha in these months, which could be due to reconstitution.

We find that the alphas for June and July are negative and significant and collectively explain at least half of the negative alphas for these indices. The proportion that is not explained by the June-July coefficient drops by about half from model (1) to model (8). For models (4) and (8), the August-to-May alpha is no longer statistically significant at even the 10 percent level, while the June-July coefficient remains highly significant. In unreported versions of these regressions that include an indicator variable for each month, the June and July coefficients are both significant and of roughly equal size. The only other months with non-zero alphas are December (positive) and January (negative), consistent with the well-known January effect.

[18] Historically the Russell reconstitution has taken place at the close on the last trading day in June. In 2004, Russell changed this to the Friday that falls between June 21 and June 27.

Index reconstitution for the S&P indices occurs on a more ad hoc basis at multiple times throughout the year, with decisions being made by a committee rather than a publicized rule. Thus we cannot apply the same test design for the S&P indices.

5 Explaining Stock Returns with Factor Models

The failure of the Carhart model to correctly price standard equity benchmark indices suggests that incorporating the indices directly in such a model could affect cross-sectional pricing results. That is explored in this section. Our motivation is the Arbitrage Pricing Theory of Ross (1976), which shows that expected returns can only arise from exposure to systematic risk factors. These well-diversified benchmark portfolios are natural proxies for such systematic risk factors, and thus they might work in cross-sectional pricing.

We consider the three most common indices, the S&P 500, Russell Midcap, and Russell 2000, as well as their value and growth components. As these indices are the most widely followed and each representative of broad and separate sections of the U.S. equity universe, we argue that the indices are particularly good candidate factors in the Arbitrage Pricing Theory framework.

Our cross-sectional pricing analysis consists of three parts. First, we add these benchmark indices (plus the value and growth component of the Russell 3000) directly to the four-factor Carhart model. Second, we consider the modified Fama-French factors. Third, we construct benchmark index-only pricing models. In all three cases, we compare the pricing ability of the models in terms of cross-sectional R^2 and pricing errors as measured by the Hansen-Jagannathan distance, and for a variety of test portfolios.

We consider three different sets of test portfolios in order to investigate the robustness of our results: (i) 100 value-weighted portfolios based on a 10x10 sort on size and book-to-market, (ii) 90 value-weighted portfolios based on a 10x10 sort on size and book-to-market, where the 10 portfolios from the smallest size decile (i.e., the microcaps) are excluded, and (iii) 25 value-weighted portfolios based on a 5x5 sort on size and book-to-market.[19] For cross-sectional pricing tests, 25 portfolios may be a very small sample, so our main analysis focuses on the set of 100

[19] These portfolio returns are provided on Ken French's website, for which we are grateful. We also considered the 49 value-weighted industry portfolios, but found few significant differences across models there. Generally, both the cross-sectional R^2 and the pricing errors are low for all models.

size-BM portfolios. To save space, only results for this set are reported (the other results are available upon request).

Panel A of Table 11 reports the time series correlations of the Fama-French factors together with their alternative versions. The correlation of the Fama-French market and HML factors equals -49%, which is primarily driven by small stocks (see SHML). The correlation between their market and SMB factors equals 19%, which is largely due to the difference between midcap and large stocks (see MML) rather than small cap versus midcap stocks (see SMM). The correlation between the Fama-French factors HML and SMB equals -43%. While this is largely due to the correlation of SMB with SHML, the correlation between SMB and BHML is still surprisingly large and negative (about -24%). However, changing the SMB factor to be value-weighted with a 50^{th}-percentile size cutoff significantly reduces the relation with BHML (to a correlation of about -8%).

The correlations of the Fama-French factors with the benchmark index factors are reported in Panel B of Table 11. We consider seven different benchmark index-based factors. The first three are the S&P 500 (S5) and two market cap-spread portfolios, (i) the difference between the Russell Midcap index and the S&P 500 (RM-S5) and (ii) the difference between the Russell 2000 and the Russell Midcap index (R2-RM). Second, we employ four index-based value-growth spread portfolios, namely the difference between the value and growth components of the S&P 500 (S5V–S5G), the Russell Midcap index (RMV–RMG), the Russell 2000 index (R2V–R2G) and finally the Russell 3000 index (R3V–R3G). The correlation of the Fama-French market (i.e. CRSP VW) portfolio is obviously highest with the S&P 500 (98%), and we again find that the market is most strongly negatively correlated with the value-growth spread portfolio for small stocks (R2V-R2G). SMB has a correlation of 2% with the S&P 500 and of 92% with R2-RM. The correlation of HML with the value-growth spread portfolios is about 90% for R3V-R3G, R2V-R2G, and RMV-RMG, but only 72% for S5V-S5G.

Panel A of Table 12 presents the results for various cross-sectional OLS regressions of mean excess returns of the 100 value-weighted size-BM-sorted test portfolios regressed on their factor betas using 239 monthly returns from 2/1986 to 12/2005. We start in February 1986 because the Russell Midcap value and growth components first become available then, and we want all cross-sectional models to be directly comparable with an identical time period. However, our main conclusions for models not using the Russell Midcap value and growth components would remain unchanged if we used the longer time period of 1980-2005 instead.

The econometric approach used is the two-stage cross-sectional regression. In the first stage, the multivariate betas are estimated using OLS. The second stage is a single cross-sectional regression of average excess returns on betas, estimated again with OLS. Following Shanken (1992), the second stage standard errors are corrected for the bias induced by sampling errors in the first-stage betas. In addition, we test our econometric specification using the Hansen-Jagannathan (HJ) distance. Hansen and Jagannathan (1997) demonstrate how to measure the distance between a true stochastic discount factor that prices all assets and the one implied by the asset pricing model. If the model is correct, the HJ distance should not be significantly different from zero, which is evaluated by calculating the asymptotic p-values using the test developed in Jagannathan and Wang (1996).[20]

As a reference, the Carhart four-factor model in column 1 has a cross-sectional R^2 equal to 28.6%, with a HJ distance of 0.69 and a low p-value of only 9.4% (i.e., pricing errors as large or larger than these would be unlikely if the model held perfectly). Subsequently adding the S&P 500 (S5), RM-S5, R2-RM, and R3V-R3G increases the R^2 to 34.4%, 59.2%, 63.5%, and 63.5%, respectively (columns 2-5).

These very significant increases in the cross-sectional R^2 indicate that the four Carhart factors fail to capture significant size-related systematic factors in the cross-section of stock returns. In particular the exposure to midcap stocks is missing, as adding RM-S5 results in a significant jump in the cross-sectional R^2. The addition of RM-S5 in column 3 further lowers the cross-sectional coefficients on HML and UMD by more than half and makes them insignificant. Finally, adding the index factors decreases the HJ-statistic from 0.69 to 0.65, but the p-value remains low at 22%.

Our finding that the four Fama-French and Carhart factors do not fully capture significant size-related systematic factors in the cross-section of stocks can only partially be remedied by the alternative Fama-French factors discussed in the previous section. Columns 6 in Panel A reports the pricing results for the same seven-factor model as in column 6 of Table 7, with a cross-sectional R^2 of 47.5%, falling clearly short of the R^2 of 63.5% for the six-factor model including S5 and RM-S5 in column 3.[21] Further, the alternative construction of the Fama-French market,

---

[20] We also computed the empirical p-values assuming normality as in Hodrick and Zhang (2000) using Monte Carlo simulations under each model holding exactly. Ahn and Gadarowski (2003) indicate that the small sample properties of the HJ-distance can be quite far from the asymptotic distribution and depend on the number of assets and the number of time periods. These p-values indicate a very similar pattern as the asymptotic p-values.

[21] Adding the four benchmark-based factors (not reported) to column 8 further increases the R^2 to 74%.

SMB and HML factors (as described in the previous section, not reported) makes no difference for cross-sectional pricing of these test portfolios.

In Panel B of Table 12, we consider the pricing performance of purely index-based factor models using the 100 value-weighted size-BM-sorted test portfolios. As an alternative to the non-tradable Fama-French factors, we consider the following index factors: S5 rather than the CRSP market portfolio, R2-S5 as an alternative to SMB, and R3V-R3G as an alternative to HML. We further add RM-S5 and R2-RM to capture the importance of midcap stocks. Finally, we consider the value and growth components of the S&P 500, Russell Midcap, and Russell 2000 separately (i.e., S5V-S5G, RMV-RMG and R2V-R2G, respectively). The models in columns 1-4 include the momentum factor UMD, but since this is not an actual benchmark followed in practice, we also consider the same models without UMD in columns 5-8.

In general, the index-based models easily improve upon the cross-sectional R^2 of the four-factor Carhart model of 29.5%. For example, the four-factor models in columns 1 and 6 have an R^2 of 32.6% and 48.3%, respectively, with comparable HJ distances. Interestingly, the models without UMD in columns 6-7 have an almost identical R^2 to the corresponding models with UMD in columns 2-3, with again comparable HJ distances. This indicates that UMD hardly matters for the cross-sectional pricing of these test assets once exposure to the size and value-growth benchmarks is accounted for (even though UMD's coefficient remains statistically significant for all models in columns 1-4).

In unreported results, we consider the cross-sectional pricing results for two other sets of test portfolios: 90 value-weighted size-BM-sorted test portfolios (i.e., excluding from the 10x10 sort the smallest market cap decile which consists of microcap stocks) and the 25 value-weighted size-BM-sorted test portfolios (5x5 sort).

For the 90 value-weighted size-BM test portfolios, the main result of excluding the microcaps is that pricing errors go down. As the microcaps are in none of the benchmarks considered here, it seems logical that the improvement is the largest there. For example, the p-value of the HJ-distance of the seven-factor model equals 82.1%, while the corresponding p-value for the same model using the 100 size-BM portfolios was 43.5% (see column 6 of Panel A).

For the 25 value-weighted size-BM test portfolios, the cross-sectional R^2 of the standard four-factor model equals 48%, with a p-value of the HJ-distance of 7.4%. This low p-value does not increase as alternative Fama-French factors or index-based factors are added, and thus it remains extremely low for the pure index models. The advantages of the index models are least pronounced here, with a cross-sectional R^2 of 43.4% and 53.1% for the four-factor and seven-factor models (corresponding to the models in columns 1 and 4, respectively, of Panel B in Table 12).

Overall, we conclude that adding the index-based factors to the four-factor Carhart model can improve asset pricing by producing large increases in the cross-sectional R^2, with the biggest impact coming from a midcap factor. Replacing the Carhart model entirely with index-based factors also improves the cross-sectional R^2 for the 100 size-BM test portfolios. Separate value-minus-growth factors for different size groups, whether based on indices or Fama-French component portfolios, can further improve the pricing performance of a model.

6 Explaining Mutual Fund Returns with Factor Models

The failure of the Carhart four-factor model to explain the returns of the most common benchmark indices can also have important consequences for performance evaluation. We assess the impact of benchmark models on inferences about the skill of a money manager by turning to mutual fund returns. This allows us to search over alternative benchmark models and focus on the ones that seem most appropriate for benchmarking. Our sample consists of all-equity mutual funds investing in the US market, so non-stock holdings or timing across bonds and stocks do not play a role in our analysis. We only consider funds with at least \$10M in total net assets.

6.1 Tracking Error Volatility

6.1.1 Methodology

To accurately estimate a fund manager's alpha, we would like to use a benchmark that closely tracks his portfolio return over time, thus producing tighter standard errors on alpha. This is measured by tracking error volatility (sometimes just called "tracking error" for simplicity), which is the time-series standard deviation of the difference between a fund return and a benchmark return. We evaluate various benchmark models on this criterion by computing and comparing the average tracking error volatility produced by each model.

We start with the standard models in the literature: the CAPM, the Fama-French three-factor model, and the Carhart four-factor model. As before, we also try a six-factor model, where we add the S&P500 and Russell 2000 to the Carhart model, as well as modified Fama-French models using a value-weighted SMB that includes "no BM" stocks and a market return on U.S. stocks only (MOD4, or column 4 in Table 7) and a seven-factor model with separate value factors for large-caps, mid-caps, and small-caps (MOD7, or column 6 in Table 7).

Alternatively, we can construct benchmark models purely from the most common benchmark indices. IDX4 refers to a simple four-factor model consisting of the S&P500, Russell 2000, Russell 3000 Value minus Russell 3000 Growth, and a momentum factor. This model roughly captures the main dimensions we are interested in: the market (S5 and R2 together), value vs. growth, small-cap vs. large-cap, and momentum. We further refine the basic model by splitting the small-cap index into separate value and growth components (IDX5, containing S5V-S5G and R2V-R2G), adding the Russell Midcap index (IDX6a), and adding a midcap value-minus-growth factor (IDX7). We also test the last model without momentum (IDX6b).

In addition to various benchmark models on the right-hand side, we also try two different return specifications on the left-hand side. One is the excess return on a fund relative to the risk-free rate. The other is the benchmark-adjusted return on a fund, which means the return in excess of a fund's benchmark index. The benchmark index of a fund is estimated separately each time the fund reports its portfolio holdings; we follow the methodology of Cremers and Petajisto (2007) in selecting the index that produces the lowest Active Share, i.e., the index that has the greatest overlap with the fund's portfolio holdings. The rationale behind the benchmark-adjustment is simple: if the benchmark index already captures most of the style differences across funds, then we may not even need a complicated model to account for the residual style differences.

To estimate tracking errors for each model, we first need to estimate betas of funds with respect to each model. We estimate betas based on twelve months of daily data on fund returns and index returns. We repeat the beta estimation each time a fund reports its portfolio holdings in the Thomson database, which usually occurs quarterly or semiannually, using the twelve months prior to the report date. Tracking error is then computed for each fund using monthly out-of-sample returns.

We focus on the time period 1996-2005. If we were to start the period earlier, we would have to include years when some indices had not been officially launched and were not known to investors. This does not have to be a concern for asset pricing tests, but it probably had an impact on fund manager behavior. Also starting in 1998, the SEC required all mutual funds to disclose a benchmark index in their prospectuses, so it is likely that managers have been more benchmark-aware in the years after that change.

6.1.2 Results

Panel A of Table 13 shows the equal-weighted annualized tracking error across all of our benchmark models using excess returns or benchmark-adjusted return as the dependent variable.

In terms of excess return, the average fund has experienced volatility of 17.35% per year. Controlling for the market portfolio reduces it by about a half to 8.28%, and the Fama-French three-factor model reduces it further to 6.50% per year. Adding the Carhart momentum factor makes little difference for tracking error. When we add the S&P500 and Russell 2000, tracking error declines to 6.10%. The methodological changes in the factor construction of the Carhart model have a very small effect on tracking error, reducing it to 6.40%, but the more elaborate seven-factor model reduces tracking error to 6.15%.

The pure index models produce a generally lower tracking error. A four-factor model with S&P500, Russell 2000, R3V-R3G, and UMD produces about 30 bp lower tracking error than the Carhart four-factor model. Adding a midcap index together with midcap and smallcap value factors further reduces tracking error to 5.80%. This is 64 bp, or 10%, lower than with the Carhart model, indicating an economically meaningful improvement in tracking error when using the seven-factor index model. The six-factor index model without momentum performs essentially just as well.

Alternatively, if we simply subtract the benchmark index return from fund return, tracking error already decreases to 6.91%, which is much closer to the four-factor tracking error than the CAPM tracking error. Regressing the benchmark-adjusted return on the Fama-French or Carhart models produces tracking errors 30-32 bp lower than with excess return, which indicates that a fund's official benchmark can capture significant risk exposures beyond the standard three or four factors. However, with the four or seven-factor index models the benchmark-adjustment no longer makes a difference. This has an important practical implication: we can simply apply the four or seven-factor index models for all funds without having to determine their benchmark indices first.

Panel B repeats the same exercise but using only relatively passive funds which are therefore easiest to explain with factor models. We compute each fund's Active Share as in Cremers and Petajisto (2007), and we select funds in the bottom 50% of Active Share within each benchmark index. We find that all tracking errors go down by about 120-140bp per year. In particular, tracking error for the Carhart model decreases from 6.44% to 5.20%, while the index models improve slightly more, reaching tracking errors of 4.73% for the four-factor model and 4.47% for the seven-factor model.

6.1.3 Robustness of Beta Estimation

When estimating betas, it is not obvious what the time horizon or the sampling frequency should be. We try four different methods: monthly data over five or three years, and daily data over twelve or six months. Monthly data is convenient to use, but it requires a longer history of returns and it may mismeasure betas if they vary over time. Daily data allows for a large number of data points while keeping the beta estimates current, but it may introduce problems due to stale prices for some stocks. Panel C of Table 13 shows the average out-of-sample tracking errors across the four estimation methods. The main conclusion from the results is that daily data produces superior estimates to monthly data.

With monthly data, even a simple benchmark-adjustment performs as well out-of-sample as the Fama-French and Carhart models on excess returns. The four-factor index model performs best, while adding more factors slightly increases out-of-sample tracking error. Whether we use three or five years of data does not matter much for models with only a few factors, but models with at least five factors are clearly better estimated from a longer dataset.

With daily data, it does not matter whether we use six or twelve months of data. In general, the twelve-month estimates perform slightly better, except for the CAPM where we need to estimate only one parameter. Tracking error improves monotonically as we add new factors, at least up to the seven-factor model.

Because daily beta estimates perform so much better out-of-sample than monthly beta estimates, it appears that any staleness in prices does not interfere much with beta estimation. Stale prices would undoubtedly be more important for individual stocks, but mutual funds hold broad portfolios of stocks, so the average staleness in fund return is likely to be close to the average staleness in benchmark index return. Nevertheless, we investigated daily beta estimates further to see whether including leads and lags would improve our estimates; we find that it does not.[22]

6.2 Cross-section of Fund Returns

To find out if there are spurious cross-sectional patterns in fund alphas induced by the choice of benchmark model, we next analyze the level of fund alphas across a variety of benchmark models. In order to form groups among similar funds and to maximize cross-sectional differences across groups, we create nine portfolios of funds from a two-dimensional sort on size and value. In particular, we determine the fund groups from their benchmark indices: the large-cap group consists of funds with the S&P500, Russell 1000, Russell 3000, or Wilshire 5000 as their benchmarks; the mid-cap indices are the S&P400, Russell Midcap, and Wilshire 4500; the

[22] Results available upon request.

small-cap indices are S&P600 and Russell 2000; and the value and growth groups are determined from the corresponding style indices.

We again examine both excess returns and benchmark-adjusted returns for a few reasons. First, benchmark-adjusted return is the performance measure that most investors focus on, because their natural investment alternative is a low-cost index fund which replicates the index return, and it is also the measure that fund managers focus on, because beating the index is their explicit self-declared investment objective. Second, if a benchmark model gives very different results for excess returns and benchmark-adjusted returns, it can only come from nonzero alphas assigned to the benchmark indices themselves. Because we want to avoid attributing any skill to the passive benchmark index, a good benchmark model should produce similar alphas for both excess returns and benchmark-adjusted returns.

Table 14 shows the fund alphas across the Fama-French and Carhart models. The time period is from 1996 to 2005 so that all indices are available to us over the entire sample. Each fund group represents an equal-weighted portfolio of funds. We estimate betas and alphas from monthly returns on these portfolios of funds and the benchmark factors. Fund returns are net returns, i.e., after all fees and expenses.

Panel A shows the excess returns and benchmark-adjusted returns on funds. Over this ten-year sample, small-cap funds beat large-cap funds by 2.79% per year, and value funds beat growth funds by 1.90% per year. Controlling for the benchmark index returns, we see that the average fund lost to its benchmark by 0.80% per year. Furthermore, the benchmark-adjustment eliminates the return spread between growth and value funds, and it reduces the return spread between small-cap and large-cap funds from 2.79% to 2.02%.

The most interesting patterns occur for the Carhart model (Panel B). With excess returns, the model shows the small-cap funds with alphas that are 2.13% below the large-cap fund alphas, but with benchmark-adjusted returns, the small-cap fund alphas are 2.94% *above* the large-cap fund alphas. The simple benchmark-adjustment therefore changes the small and large-cap alphas by 5.07% for the Carhart model. This is a truly dramatic effect, especially in the context of mutual fund alphas which are very close to zero on average, and it is certainly large enough to potentially reverse the conclusions of performance analysis. These numbers are also very similar with the Fama-French model, and they can only come from nonzero alphas that the two models assign to the benchmark indices. We argue that this finding casts severe doubt on the validity of the standard Carhart alpha estimates across the size dimension. Across the value dimension, there is no such unambiguous effect.

Panel C shows further evidence of model misspecification. It reports the alphas from a six-factor model including the Carhart factors as well as the S&P500 and Russell 2000, which are the most common benchmark indices for mutual funds. Adding these two factors changes the spread in four-factor excess-return alphas between small and large-cap funds by 2.60%.[23] In other words, when we let the data speak in this type of a horse race, funds tends to load on the two indices instead of getting their small and large-cap exposure from the market portfolio and the SMB factor.

Panels D and E in Table 14 report the alphas from pure index models. In contrast to Carhart, now the fund alphas are very similar across excess returns and benchmark-adjusted returns, especially with the seven-factor model. This arises from the fact that the index models produce exactly zero alphas for the constituent indices and only small alphas for the other indices. Like in the tracking error analysis, this has the important implication that the seven-factor index model can be applied to the excess returns on all fund returns regardless of a fund's style or benchmark index.

In terms of the magnitude of alphas, the seven-factor index model produces relatively plausible values. The average fund has underperformed by -0.88%, with large-cap funds underperforming by -1.29% and small-cap funds actually slightly outperforming by 0.37%. There is no pattern across value groups. The slight outperformance by small-cap funds is consistent with conventional wisdom; furthermore, in equilibrium with costly information acquisition we would also expect small stocks to be less efficiently priced than large and liquid stocks, although that source of alpha could in turn be fully offset by higher fees and trading costs. Perhaps the most reassuring thing about the alphas is that there are no fund groups with large positive or negative values – such outliers in either direction would represent clear inefficiencies in the mutual fund market. This stands in contrast to the Carhart model which produces a -3.99% alpha for small-cap growth and -3.09% for small-cap core funds. Furthermore, the seven-factor index model produces alphas that are surprisingly similar to the benchmark-adjusted returns, suggesting that even the simple subtraction of the benchmark index return may be a better benchmark model than the standard academic three- or four-factor models.

[23] The difference in Panel B is –3.20 – (–1.07) = –2.13%. In Panel C, it is –0.13 – (–0.60) = 0.47%.

6.3 Robustness of Prior Studies: Performance Persistence, Active Management, and Smart Money

This section examines whether the major prior findings of the mutual fund literature are robust to controlling for benchmark alphas. In general the answer is yes, with an important caveat we discuss below. Most of the mutual fund results that we examined are cross-sectional comparisons of returns that are largely within, rather than across, investment objectives (e.g., Small Growth). Since correcting for benchmark alphas has a roughly uniform effect on funds in a given investment objective, within-objective comparisons are largely unaffected.

Table 14 replicates a selected set of results from the literature using the models examined in this paper. A natural place to start is Carhart (1997). Carhart's starting point was an earlier finding (e.g., Hendricks, Patel, and Zeckhauser (1993)) that funds with high past returns tend to have higher future returns. He finds that this was partly due to high-past-return funds' exposure to the momentum factor. Controlling for momentum explains most of the performance persistence among the top 90 percent of funds, although the bottom decile continues to exhibit a puzzling underperformance.

Panel A replicates Carhart's main result for a latter time period, and then examines its robustness to: 1) modifying the Fama-French factors as discussed above, 2) switching to an index model, or 3) adding widely followed benchmark indices (specifically, the difference between S&P 500 and Russell 2000 returns and the risk-free rate) to the standard or modified Carhart model. While switching from the original Carhart model to our modified models or index models does not materially affect conclusions about performance persistence, adding controls for Russell 2000 and S&P 500 exposure to the original or modified Carhart models eliminates the performance difference between the top and second-worst decile, and reduces by one-third the gap between the two worst performing deciles. Almost all of this difference is due to adding the Russell 2000, which captures the fact that funds that persistently perform poorly tend to include the small-cap stocks in the Russell 2000 more than the small-cap stocks in the Fama-French portfolios.[24] Since the Russell 2000 has roughly the same impact when added to the original Carhart model and our modified model, this suggests that a component of the performance

---

[24] Adding controls for other indices, such as the Russell Midcap, does not affect results in any of the panels in Table 14. Results are also similar for MOD4 and MOD7 and for the various index models, and so only one modified and one index model are reported.

persistence remaining in Carhart's analysis may be explained by persistence in how funds are affected by exposure to the Russell 2000 (e.g., by reconstitution effects).

Panel B repeats the often asked question of whether it is worth paying for active management, comparing the returns of low-expense ratio index funds with the lowest, median, and highest expense ratio actively managed funds.[25] A comparison of Carhart alphas suggests that the most expensive actively managed funds underperform the index funds by 1.74 percent per year. The differences in (net of expense) alphas roughly corresponds to the differences in average expense ratios of the four groups of funds, which are 19, 65, 121, and 209 basis points per year for index and low, median, and high-expense active funds, respectively.

About half of the difference in Carhart alpha between inexpensive and expensive actively managed funds is eliminated by switching to a modified Carhart or index model. Expensive funds are more likely to be small-cap funds which, as discussed, have their performance understated by the unmodified Carhart model. Furthermore, controlling for exposure to the Russell 2000 and S&P 500 eliminates all or most of the remaining difference between index and inexpensive actively managed funds and expensive actively managed funds. Inexpensive funds load more on the S&P 500 and, perhaps surprisingly, less on the Russell 2000, and this explains much or all of the better after-expense performance that has been found in the past.

Panel C repeats the tests for "smart money" of Gruber (1996), Zheng (1999), and Sapp and Tiwari (2004). Following Sapp and Tiwari, funds are divided into two portfolios based on whether they experienced positive or negative dollar inflows in the prior calendar quarter. Here, we find results consistent with prior work across all of our models. Portfolios that attracted inflows outperformed those with outflows relative to the Fama-French model. Both inflow and outflow-receiving portfolios load positively and to roughly equal extents on the benchmark indices. In unreported results, we also found that varying the model did not alter past results on the negative correlation between performance and portfolio turnover, the positive correlation between performance and portfolio size among large-cap funds, and the negative correlation between performance and portfolio size among small-cap funds.

In general, the conclusions of prior studies are most susceptible to the benchmarking issues documented in this paper whenever the investigated fund characteristic is correlated with

---

[25] In this analysis, "low-expense" is defined as being below the 30th percentile expense ratio among index funds. The average of this cutoff in our 1981-2005 time period is 25 basis points. About 59 percent of low-expense index funds track the S&P 500, compared with 55 percent of all index funds in our sample. Enhanced index funds are excluded from all portfolios analyzed in Table 15.

the location of stocks or portfolios in the size-value grid. Past conclusions about mutual fund performance differences are largely unaffected if they are conducted within the same investment objective. That said, including both indices and academic factors in a model can yield insights about the extent to which a result is explained by index exposure. Individual portfolios are more likely to be correlated with the size-value grid, which can make the choice of benchmarking model important. For example, if an active trading strategy ends up being long large growth stocks and short large value stocks in the Fama-French 2x3 grid, the Fama-French and Carhart models turn a return premium of 1.66% per year in favor of value stocks to an alpha premium of 4.33% or 3.90% per year in favor of growth stocks.[26] A researcher or investor who believes that using the common factor models properly accounts for value and size exposures can therefore draw an incorrect conclusion about the profitability of the strategy. Naturally, the same mechanism can lead to wrong conclusions in fund selection as well.

7 Conclusions

The standard Fama-French and Carhart models, which have been widely adopted in academic research for asset pricing and performance evaluation purposes, suffer from serious biases. Because of their construction methodology, both SMB and HML portfolios assign disproportionate weight to extreme value stocks, especially among small stocks. Since that small corner of the market, with only 2% of total market capitalization, has also produced the highest returns, these benchmarks are tough to beat for any manager with a tilt toward small stocks, and conversely, they are relatively easy to beat with large-stock tilt. As HML represents an average of the large-cap and small-cap value effects, it is an easy benchmark for small-cap value funds and a tough benchmark for large-cap value funds (and vice versa for growth funds). Furthermore, the CRSP value-weighted market index, which includes other securities besides U.S. stocks, contributes to a positive bias to all alpha estimates for U.S. stocks.

One of the most striking pieces of evidence for this bias comes from the four-factor Carhart alphas of passive benchmark indices. The most common large-cap indices, S&P 500 and Russell 1000, exhibit economically and statistically significant positive alphas of 0.82% and 0.47% per year, respectively, from 1980 to 2005. The corresponding small-cap indices, Russell 2000 and S&P 600, have earned significant negative alphas of -2.41% and -2.59% per year.

[26] From 1980 to 2005, large value in the Fama-French 2x3 grid has outperformed large growth by 1.66% per year. Yet the three and four-factor models produce alphas of –2.21% and –1.70%, respectively, for large value, and alphas of +2.12% and +2.20% for large growth.

Naturally, one would expect passive benchmark indices to have zero alphas; in fact, one could even define alpha relative to a set of passive indices which are the low-cost alternatives to active management. This is exactly what we test as an alternative to the common three and four-factor models.

We start our model comparison with standard asset pricing tests for 10x10 size-and-book-to-market sorted portfolios. Replacing SMB and HML with index-based factors considerably increases the R^2 of a cross-sectional regression of portfolio returns on factor betas.

Our analysis of mutual fund returns reveals the dramatic impact that the benchmark alphas can have on inferences about performance. When comparing small and large-cap funds, adjusting for the benchmark index has about 5% per year impact on their Fama-French and Carhart alphas, fully reversing the conclusions about skill between small and large-cap funds. Index-based models do not exhibit similar biases and generally produce much less extreme alphas across all fund groups.

We also analyze the tracking error volatility across funds to see which models best track the time-series of fund returns. Again the index-based models outperform the traditional Fama-French and Carhart models in terms of their out-of-sample tracking error.

Overall, the results strongly suggest using alternative models for pricing and performance evaluation. Mutual fund returns are best explained by a seven-factor model consisting of the S&P500, Russell Midcap, and Russell 2000, separate value-minus-growth factors for each index, and a momentum factor. If we need to economize on the number of factors, an index-based four-factor model with the S&P500, Russell 2000, R3V-R3G, and UMD factor dominates the Carhart model. The pricing tests also suggest the index-based seven-factor model as the best one, and the pure index-based four-factor model as an improvement over the Carhart model.

We have not searched over an exhaustive set of alternative index-based models, so even better model specifications may well exist. Our main objective has been to expose the weaknesses in existing models and propose a new direction that represents an improvement for performance evaluation.

References

Ahn, S., and C. Gadarowski, 2003, "Small Sample Properties of the Model Specification Test Based on the Hansen-Jagannathan Distance," *Journal of Empirical Finance*, forthcoming.

Ammann, M., and H. Zimmermann, 2001, "Tracking Error and Tactical Asset Allocation," *Financial Analysts Journal* 57- 2, 32-43.

Barber, B.M., and J.D. Lyon, 1997, "Detecting long-run abnormal stock returns: the empirical power and specification of test statistics," *Journal of Financial Economics* 43, 341-372.

Beneish, M.D., and R.E. Whaley, 1996, "An Anatomy of the 'S&P Game': The Effects of Changing the Rules," *Journal of Finance* 51-5, 1909-1930.

Berk, J.B. and R.C. Green, 2004, "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy* 112-6, 1269-1295.

Biktimirov, E., A. Cowan, and B. Jordan, 2004, "Do Demand Curves for Small Stocks Slope Down?" *Journal of Financial Research* 27-2, 161-17.

Bollen, N.P.B. and J.A. Busse, 2004, "Short-Term Persistence in Mutual Fund Performance," *Review of Financial Studies* 18-2, 569-597.

Boyer, B.H., 2006, "Comovement Among Stocks with Similar Book-to-Market Ratios," working paper.

Brinson, G.P., L.R. Hood, and G.L. Beebower, 1986, "Determinants of Portfolio Performance," *Financial Analysts Journal* 42-4, 39-44.

Brown, S.J. and W.N. Goetzmann, 1995, "Performance Persistence," *Journal of Finance* 50-2, 679-698.

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

Carhart, M., 1997, "On Persistence in Mutual Fund Returns," *Journal of Finance* 5211, 57-82.

Chen, J., H. Hong, M. Huang, and J.D. Kubik, 2004, "Does Fund Size Erode Performance? Organizational Diseconomies and Active Money Management," *American Economic Review* 94-5, 1276-1302.

Chen, H., G. Noronha, and V. Singal, 2004, "The Price Response to S&P 500 Index Additions and Deletions: Evidence of Asymmetry and a New Explanation," *Journal of Finance* 59-4, 1901-1929.

Chen, H., G. Noronha and V. Singal, 2006, “Index Changes and Losses to Index Fund Investors,” *Financial Analysts Journal* 62-4, 31-47.

Cremers, K.J.M. and A. Petajisto, 2007, “How active is your fund manager? A new measure that predicts performance,” *Review of Financial Studies*, forthcoming.

Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, “Measuring Mutual Fund Performance with Characteristic-Based Benchmarks,” *Journal of Finance* 52-3, 1035-1058.

Greenwood, R., 2004, "Short- and Long-Term Demand for Stocks: Theory and Evidence on the Dynamics of Arbitrage," *Journal of Financial Economics* 75-3, 607-649.

Hansen, L., and R. Jagannathan, 1997, “Assessing specification errors in stochastic discount factor models,” *Journal of Finance* 52, 557-590.

Hodrick, R.J., and X. Zhang, 2001, “Evaluating the Specification Errors of Asset Pricing Models,” *Journal of Financial Economics* 62-2, 327-276.

Huij, J., and M. Verbeek, 2007, “On the Use of Multifactor Models to Evaluate Mutual Fund Performance,” Working paper, Erasmus University Rotterdam.

Jagannathan, R., and Z. Wang, 1998, “An asymptotic theory for estimating beta-pricing models using cross-sectional regression,” *Journal of Finance* 57, 1285-1309.

Lehmann, B.N., and D.M. Modest, 1987 “Mutual fund performance evaluation: a comparison of benchmarks and benchmark comparisons,” *Journal of Finance* 42, 233-265.

Moor, L. De, and P. Sercu, 2006, “The Small Firm Anomaly: US and International Evidence,” Working paper, Katholieke Universiteit Leuven.

Petajisto, A., 2006, “The Index Premium and Its Hidden Cost for Index Funds,” Yale ICF working paper.

Ritter, J.R., 1991, “The long-run performance of initial public offerings,” *Journal of Finance* 46, 3–28.

Ross, S..,1976, “The Arbitrage Theory of Capital Asset Pricing,” *Journal of Economic Theory* 13, 341-360

Sensoy, B., 2007, “Performance Evaluation and Self-Designated Benchmarks in the Mutual Fund Industry,” USC Working Paper

Sensoy, B. and S. Kaplan, 2007, "Do mutual funds time their benchmarks?" NBER Working Paper

Shanken, J., 1992, "On the Estimation of Beta-Pricing Models," *Review of Financial Studies* 5-1, 1-33.

Sharpe, W.F., 1992, “Asset allocation: Management style and performance measurement,” *Journal of Portfolio Management* 18-2, 7-19.

Wermers, R., 2000, “Mutual Fund Performance: An Empirical Decomposition into Stock-Picking Talent, Style, Transactions Costs, and Expenses,” *Journal of Finance* 55-4, 1655-1695.

Appendix A: Tables

Table 1. The most common benchmark indices.

For each index, the second column is the number of US all-equity mutual funds reporting the index as their primary benchmark in January 2007. The last column is the sum of total net assets across all such funds. The data source is Morningstar. Some funds have a missing primary benchmark in the database.

Index	Number of mutual funds	Mutual fund assets ($M)
S&P 500	1,318	2,130,000
Russell 2000	251	214,712
Russell 1000 Growth	180	162,710
Russell 1000 Value	177	249,537
Russell 2000 Growth	132	48,579
Russell Midcap Growth	107	73,563
Russell 2000 Value	106	65,066
S&P 400	74	102,241
Russell Midcap Value	62	85,629
Russell 1000	53	56,660
Russell 3000	48	43,344
Russell Midcap	35	23,260
Russell 3000 Growth	31	67,130
S&P 600	27	14,326
Russell 3000 Value	26	63,722
Wilshire 5000	20	114,092
S&P 500 Value	8	6,307
Wilshire 4500	5	16,254
S&P 500 Growth	5	345
S&P 400 Value	4	10,869
S&P 400 Growth	3	192
S&P 600 Value	3	181
S&P 600 Growth	2	57

Table 2. Alphas of benchmark indices.

This table shows the Carhart four-factor alphas for benchmark indices. Alphas are computed from monthly data. The numbers shown are expressed in percent per year, with t-statistics in parentheses. The sample period is January 1980 to December 2005, except for the following indices whose return data begin later: S&P 400 (2/1981), Wilshire 4500 (1/1984), S&P 600 (3/1984) and the Growth and Value components of the Russell Midcap (2/1986), S&P 400 (6/1991), and S&P 600 (1/1994).

Main index	Style component		
	Value	All	Growth
Russell 3000	-0.55	0.18	1.02
	(-1.01)	(0.96)	(2.05)
Russell 1000	-0.45	0.47	1.50
	(-0.83)	(2.58)	(2.73)
Russell Midcap	-0.52	0.17	1.61
	(-0.54)	(0.24)	(1.34)
Russell 2000	-1.25	-2.41	-3.41
	(-1.31)	(-3.35)	(-3.87)
S&P 500	-0.35	0.82	1.82
	(-0.69)	(2.95)	(2.76)
S&P Midcap 400	0.84	1.44	0.64
	(0.51)	(1.33)	(0.32)
S&P Smallcap 600	-1.49	-2.59	-3.05
	(-0.89)	(-2.20)	(-1.39)
Wilshire 5000		0.05	
		(0.43)	
Wilshire 4500		-0.56	
		(-0.79)	

Table 3. Four-factor alphas by CRSP share code, 1980-2005.

This table aggregates the share codes reported in CRSP into groups. The CRSP value-weighted index consists of all share codes except ADRs. The table reports the average share of the CRSP VW index accounted for by each group from 1980-2005, along with their four-factor alphas. The four-factor alpha of the CRSP value weighted index is of course zero by construction. The table also reports, based on December 2004 data, the case of each group's capitalization that is a member of three indices (the S&P 500, Russell 3000, and Wilshire 5000) and the share that is reported as holdings by equity mutual funds (those holding more than 50% of TNA in equities) on SEC form 12D. T-stats from robust standard errors are in parentheses.

			Four factor alphas		Percent of capitalization held by:			
Group	Share codes (descending order of market cap)	Average share of CRSPVW	Percent per year	t-stat	S&P 500	Russell 3000	Wilshire 5000	Equity funds
U.S. common stocks	11,10	92.68%	0.23	(2.00)	77.4	97.0	98.9	10.12
Subset included in FF portfolios	11,10	87.87%	0.51	(2.68)				
Subset not included in FF portfolios	11,10	4.81%	-2.74	(1.66)				
All other securities in CRSP index	See below	7.32%	-4.01	(2.67)	12.4	14.8	24.0	4.88
Non-US stocks, units, and SBIs	12, 72, 42	4.76%	-3.74	(2.00)	14.6	0.9	12.3	5.57
Closed-end funds	14, 44, 15, 74, 24	1.06%	-1.65	(1.02)	0.0	0.1	0.1	0.09
REITs	18, 48	0.74%	-0.75	(0.37)	21.3	97.2	99.8	8.35
Other (certificates, SBIs, units)	71, 23, 73, 70, 41, 21, 40, 20	0.76%	-3.39	(1.85)	0.0	0.5	12.4	0.79
CRSP value-weighted index	All except ADRs	100%	0.00	(0.00)	69.6	87.0	89.8	9.49
ADRs (excluded from CRSPVW)	31, 30	3.31%	4.25	(1.55)	0.0	0.0	0.0	

Table 4. Fama-French factor component portfolios.

This table explores the impact of alternative construction methodologies on the average returns of the SMB factor. The original Fama-French SMB is an equal-weighting of Small-Low, Small-Med, and Small-High less an equal-weighting of Big-Low, Big-Med, and Big-High. Panel A displays the average annualized returns of these six portfolios, as well as Big-None and Small-None portfolios which capture U.S. stocks that are included in the CRSP value-weighted index but are excluded from the other portfolios, for example due to missing or negative book equity. Panel B reports the average capitalization weights of these portfolios in the CRSP VW index. Panel C calculates versions of SMB using Small and Big portfolios that equal weight their Low, Medium, and High components, that value-weight the 3 components, and that value-weight these 3 components as well as the None portfolios.

Panel A. Average excess return (% per year)

	Book-to-Market			
Size	Low (Growth)	Medium	High (Value)	None or <0
Big (decile 6-10)	7.61	8.62	9.20	8.93
Small (decile 1-5)	4.85	11.77	13.21	8.32

Panel B. Average capitalization weights in CRSPVW index (%)

Big (decile 6-10)	44.72	28.21	12.69	2.26
Small (decile 1-5)	4.04	3.66	2.21	2.20

Panel C. Average returns of alternative SMB factors (% per year)

	Method for calculating returns for size portfolios		
	EW (Low, Med, High)	VW (Low, Med, High)	VW (Low, Med, High, None)
Big (decile 6-10)	8.47	7.87	7.86
Small (decile 1-5)	9.94	9.24	8.99
Small minus Big	1.47	1.37	1.13

Table 5, Comparing actual portfolios with their Fama-French benchmarks.

This table shows the benchmark portfolio holdings implied by the three-factor Fama-French model. These holdings are contrasted with the true holdings of the target portfolios we are trying to explain. As target portfolios, we pick the FF size deciles 10 (large-cap stocks) and 4 (small-cap stocks) within the 100 FF portfolios, since they represent the typical S&P 500 and Russell 2000 constituent stocks, respectively. Panels A and B show the portfolio weights of the three FF factors, together with the excess return on the 2x3 portfolio components. Since the MktRf factor includes CRSP securities that are not part of the 2x3 FF grid, we include these stocks in a separate “None” column. Panel C shows the true weights that each of the two target portfolios (size deciles) have on the extended 2x4 grid, alongside the weights implied by the three-factor model. The implied weights can be derived from the three-factor betas multiplied by the factor portfolio weights; the regression betas are shown above the implied portfolio weights.

Panel A: Market portfolio weights and component returns (%)

MktRf weights						Average excess return per year					
	None	Gro	Med	Val	All		None	Gro	Med	Val	All
Big	7.8	42.6	25.5	11.1	86.9	Big	5.92	7.61	8.62	9.20	7.72
Small	4.2	3.5	3.4	2.0	13.1	Small	6.47	4.85	11.77	13.21	8.29
All	12.0	46.1	28.9	13.0	100.0	All	5.87	7.20	8.95	10.02	7.64

Panel B: Fama-French factor portfolio weights (%)

SMB						HML					
	None	Gro	Med	Val	All		None	Gro	Med	Val	All
Big	0.0	-33.3	-33.3	-33.3	-100.0	Big	0.0	-50.0	0.0	50.0	0.0
Small	0.0	33.3	33.3	33.3	100.0	Small	0.0	-50.0	0.0	50.0	0.0
All	0.0	0.0	0.0	0.0	0.0	All	0.0	-100.0	0.0	100.0	0.0

Panel C: Target portfolio weights vs. their three-factor benchmark weights (%)

Target portfolio: Size decile 10						Benchmark portfolio: 0.967 x MktRf - 0.318 x SMB - 0.086 x HML					
	None	Gro	Med	Val	All		None	Gro	Med	Val	All
Big	0.0	60.0	29.2	10.8	100.0	Big	7.5	56.1	35.2	17.0	115.8
Small	0.0	0.0	0.0	0.0	0.0	Small	4.1	-2.9	-7.3	-13.0	-19.1
All	0.0	60.0	29.2	10.8	100.0	All	11.6	53.2	27.9	4.0	96.7

Target portfolio: Size decile 4						Benchmark portfolio: 1.055 x MktRf + 0.799 x SMB + 0.226 x HML					
	None	Gro	Med	Val	All		None	Gro	Med	Val	All
Big	0.0	0.0	0.0	0.0	0.0	Big	8.2	7.0	0.3	-3.7	11.8
Small	0.0	40.7	40.5	18.7	100.0	Small	4.5	19.1	30.2	40.0	93.8
All	0.0	40.7	40.5	18.7	100.0	All	12.7	26.1	30.5	36.3	105.5

Table 6. Alphas and betas of 10x12 size-BM portfolios.

This table reports the four-factor Carhart alphas as well as SMB and HML betas for 10x12 Size-BM portfolios. The 10x10 portfolio returns are as computed following the methodology on Kenneth French's website. The "None" book-to-market column includes U.S. common stocks (share codes 10 and 11) from the CRSP dataset that are excluded from the Fama-French portfolios because they have negative book value or insufficient historical data. The "Other" column includes all other securities (excluding U.S. common stocks) that are included in the CRSP market index. The sample extends from 1980 to 2005. The numbers in Panel A are in basis points per year.

Panel A: Four-factor alpha

	Book-to-market deciles														
	Other	None	Growth	2	3	4	5	6	7	8	9	Value	1-10	N-10	All
Large	-320	106	371	296	-66	104	-147	-58	-100	-427	-245	-286	101	102	83
9	-175	504	216	158	-130	-138	131	-203	64	-287	-70	-18	9	6	-2
8	-609	-326	281	25	-59	-187	35	-158	-68	63	124	-56	4	-12	-46
7	-505	300	361	161	-151	-114	-187	54	-33	-205	-66	49	28	15	-30
6	-112	-185	-108	324	-93	32	-140	-89	-194	14	85	609	-39	-70	-66
5	-376	498	-156	-31	157	103	-26	28	112	-134	199	-124	-52	-10	-62
4	-388	111	-437	-55	-231	-87	-39	273	82	327	74	-380	-121	-111	-154
3	-476	33	-562	12	110	-179	158	215	97	137	20	-4	-47	-45	-127
2	-224	-65	-945	-51	-167	210	185	17	501	117	203	-9	-77	-74	-117
Small	-324	-378	-928	-332	184	237	325	321	267	360	550	409	17	-74	-125
All	-401	-52	160	199	-103	-44	-91	-37	34	-242	-23	-43	36	23	0

Panel B: SMB beta

	Book-to-market deciles														
	Other	None	Growth	2	3	4	5	6	7	8	9	Value	1-10	N-10	All
Large	-0.01	0.01	-0.40	-0.26	-0.27	-0.36	-0.28	-0.29	-0.23	-0.23	-0.19	-0.27	-0.31	-0.31	-0.29
9	0.20	0.39	0.14	-0.01	-0.03	-0.09	-0.08	-0.05	0.04	-0.11	-0.11	0.09	0.05	0.09	0.10
8	0.27	0.47	0.38	0.20	0.35	0.21	0.17	0.13	0.15	0.13	0.08	0.22	0.28	0.36	0.35
7	0.38	0.52	0.45	0.25	0.16	0.23	0.24	0.21	0.45	0.25	0.14	0.20	0.34	0.39	0.39
6	0.53	0.74	0.66	0.52	0.45	0.22	0.36	0.26	0.32	0.29	0.44	0.50	0.46	0.50	0.50
5	0.33	0.65	0.81	0.90	0.71	0.60	0.47	0.49	0.56	0.53	0.35	0.67	0.69	0.69	0.63
4	0.42	0.97	1.01	0.77	0.83	0.96	0.63	0.74	0.57	0.94	0.75	0.85	0.81	0.85	0.78
3	0.41	0.94	1.14	0.98	0.91	0.75	0.75	0.83	0.73	0.69	1.10	0.79	0.89	0.90	0.81
2	0.54	1.01	1.23	1.29	1.14	1.45	0.85	0.90	0.96	0.94	0.89	1.09	1.11	1.09	0.97
Small	0.83	1.04	1.41	1.41	1.42	1.13	1.08	1.24	1.08	0.99	1.07	1.07	1.20	1.16	1.10
All	0.22	0.63	-0.19	-0.07	-0.05	-0.08	-0.05	0.03	0.13	0.07	0.11	0.32	-0.06	-0.02	0.00

Panel C: HML beta

	Book-to-market deciles														
	Other	None	Growth	2	3	4	5	6	7	8	9	Value	1-10	N-10	All
Large	-0.04	-0.23	-0.54	-0.04	0.06	0.22	0.22	0.36	0.50	0.70	0.54	0.63	-0.09	-0.09	-0.09
9	0.30	-0.41	-0.62	-0.02	0.27	0.43	0.53	0.53	0.56	0.73	0.71	0.74	0.15	0.11	0.12
8	0.24	-0.44	-0.72	-0.03	0.14	0.42	0.51	0.58	0.68	0.63	0.67	0.91	0.19	0.10	0.12
7	0.30	-0.54	-0.69	0.01	0.39	0.34	0.53	0.59	0.52	0.59	0.74	0.72	0.18	0.10	0.11
6	0.25	-0.33	-0.75	-0.13	0.22	0.29	0.53	0.60	0.77	0.65	0.77	0.75	0.18	0.10	0.12
5	0.26	-0.46	-0.77	-0.13	0.23	0.30	0.46	0.72	0.66	0.85	0.80	0.82	0.20	0.10	0.13
4	0.37	-0.33	-0.62	0.00	0.22	0.30	0.51	0.44	0.62	0.57	0.79	0.85	0.24	0.13	0.16
3	0.32	-0.24	-0.51	-0.23	-0.02	0.31	0.32	0.42	0.56	0.64	0.66	0.83	0.19	0.11	0.15
2	0.38	-0.03	-0.51	-0.30	-0.06	0.09	0.23	0.44	0.43	0.49	0.64	0.75	0.15	0.12	0.15
Small	0.30	0.17	-0.39	-0.31	-0.15	0.13	0.15	0.08	0.33	0.35	0.46	0.64	0.17	0.17	0.18
All	0.13	-0.29	-0.55	-0.03	0.13	0.28	0.35	0.39	0.53	0.69	0.65	0.80	0.01	-0.01	0.00

Table 7A. Weights on 3x4 Size-BM portfolios implied by models – S&P 500.

The Carhart model is estimated for various versions of the SMB and HML factors and the average implied weights the model places on each of the 3x4 Size-Book-to-Market (BM) portfolios are calculated. This is compared with a Flexible model in which the excess returns of the index are regressed on those of the 3x4 portfolios. Model 1 is the standard Carhart model. Model 2 excludes share codes other than 10 and 11 (U.S. common stocks) from the CRSP-VW index. Model 3 replaces the equal-weighted SMB factor with one where the Small and Big portfolios are value-weighting of their Low, Medium, and High BM components. Model 4 includes the "No or Negative" BM components (called "None" in the table below) in Small and Big. Model 5 calculates separate HML factors for Big and Small (e.g., BHML = Big_High - Big_Low). Model 6 splits SMB into "Mid minus Large" (deciles 6-8 minus deciles 9 and 10) and "Small minus Mid". Model 7 splits BHML into one for Large stocks (NYSE deciles 9 and 10) and Midcap stocks (deciles 6-8). Model 8 includes both Medium BM stocks with High BM stocks when constructing the HML factors. T-stats based on robust standard errors are in parentheses.

Model	Carhart	(2)	(3)	(4)	(5)	(6)	Flexible	NNLS	Actual weights	Avg weights
Share codes in market factor	CRSPVW	10/11	10/11	10/11	10/11	10/11				
SMB weighting	EW	EW	VW	VW	VW	VW				
SMB stocks included	As in FF	As in FF	As in FF	All	All	All				
Small-Big cutoff	50th pct	50th pct	50th pct	50th pct	50th pct	N/A				
Size deciles included in BHML	N/A	N/A	N/A	N/A	Top 5	Top 2				
Size deciles included in SHML	N/A	N/A	N/A	N/A	Btm 5	Btm 5				
BM deciles included in H	Top 3	Top 3	Top 3	Top 3	Top 3	Top 7				
Obs	312	312	312	312	312	312	312	312	312	312
Adjusted R-sq	0.9924	0.9934	0.9934	0.9939	0.9941	0.9957	0.9882	0.9882	N.M	N.M
Constant (% per year)	0.82	0.59	0.33	0.32	0.11	0.21	0.07	0.16	0.35	0.04
	(2.78)	(2.12)	(1.23)	(1.24)	(0.43)	(0.91)	(0.20)	(0.47)	(1.79)	(0.10)
UMD	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01	-0.01
	(3.28)	(3.43)	(3.57)	(3.49)	(3.77)	(3.67)	(2.83)	(3.03)	(4.00)	(1.26)
MktRF	1.00	0.99	1.00	1.00	1.01	1.01				
	(0.14)	(0.77)	(0.01)	(0.27)	(0.91)	(2.16)				
SMB	-0.21	-0.19	-0.18	-0.19	-0.18					
	(23.88)	(23.05)	(23.80)	(24.95)	(21.38)					
Mid minus Large (MML)						-0.21				
						(22.45)				
Small minus Mid (SMM)						-0.09				
						(6.68)				
HML	0.01	0.02	0.06	0.05						
	(0.69)	(1.87)	(5.07)	(4.54)						
BHML					0.00	-0.01				
					(0.20)	(0.72)				
SHML					0.05	0.04				
					(4.85)	(2.33)				
MidHML						0.04				
						(2.01)				
Average weights on 3x4 portfolios implied by models							Flex	NNLS	Actual	Market
Large_Low - RF	0.451	0.440	0.453	0.457	0.477	0.524	0.541	0.544	0.507	0.396
Large_Med - RF	0.285	0.279	0.278	0.278	0.276	0.295	0.247	0.245	0.278	0.230
Large_High - RF	0.153	0.152	0.139	0.136	0.116	0.125	0.122	0.130	0.112	0.098
Large_None - RF	0.012	0.012	0.012	0.014	0.014	0.015	0.016	0.016	0.009	0.012
Mid_Low - RF	0.083	0.081	0.083	0.084	0.087	-0.014	0.020	0.000	0.031	0.073
Mid_Med - RF	0.081	0.079	0.079	0.079	0.078	0.045	0.015	0.000	0.035	0.065
Mid_High - RF	0.055	0.054	0.050	0.049	0.041	0.024	0.008	0.014	0.020	0.035
Mid_None - RF	0.012	0.012	0.012	0.014	0.014	0.004	0.032	0.030	0.002	0.012
Small_Low - RF	-0.033	-0.033	-0.062	-0.045	-0.069	-0.025	-0.017	0.000	0.001	0.042
Small_Med - RF	-0.033	-0.027	-0.030	-0.019	-0.016	0.037	-0.086	0.000	0.002	0.038
Small_High - RF	-0.044	-0.032	0.011	0.013	0.042	0.022	0.066	0.000	0.002	0.023
Small_None - RF	0.023	0.023	0.023	-0.011	-0.009	0.008	0.027	0.014	0.000	0.023

Table 7B. Weights on 3x4 Size-BM portfolios implied by models – Russell 2000.

The Carhart model is estimated for various versions of the SMB and HML factors and the average implied weights the model places on each of the 3x4 Size-Book-to-Market (BM) portfolios are calculated. This is compared with a Flexible model in which the excess returns of the index are regressed on those of the 3x4 portfolios. Model 1 is the standard Carhart model. Model 2 excludes share codes other than 10 and 11 (U.S. common stocks) from the CRSP-VW index. Model 3 replaces the equal-weighted SMB factor with one where the Small and Big portfolios are value-weighting of their Low, Medium, and High BM components. Model 4 includes the "No or Negative" BM components (called "None" in the table below) in Small and Big. Model 5 calculates separate HML factors for Big and Small (e.g., BHML = Big_High - Big_Low). Model 6 splits SMB into "Mid minus Large" (deciles 6-8 minus deciles 9 and 10) and "Small minus Mid". Model 7 splits BHML into one for Large stocks (NYSE deciles 9 and 10) and Midcap stocks (deciles 6-8). Model 8 includes both Medium BM stocks with High BM stocks when constructing the HML factors. T-stats based on robust standard errors are in parentheses.

Model	Carhart	(2)	(3)	(4)	(5)	(6)	Flexible	NNLS	Actual weights	Avg weights
Share codes in market factor	CRSPVW	10/11	10/11	10/11	10/11	10/11				
SMB weighting	EW	EW	VW	VW	VW	VW				
SMB stocks included	As in FF	As in FF	As in FF	All	All	All				
Small-Big cutoff	50th pct	50th pct	50th pct	50th pct	50th pct	N/A				
Size deciles included in BHML	N/A	N/A	N/A	N/A	Top 5	Top 2				
Size deciles included in SHML	N/A	N/A	N/A	N/A	Btm 5	Btm 5				
BM deciles included in H	Top 3	Top 3	Top 3	Top 3	Top 3	Top 7				
Obs	312	312	312	312	312	312	312	312	312	312
Adjusted R-sq	0.9686	0.9695	0.9838	0.9796	0.9795	0.9819	0.9862	0.9859	N.M	N.M
Constant (% per year)	-2.41	-2.66	-1.62	-1.53	-1.50	-1.61	-2.13	-2.17	-1.07	-1.23
	(3.21)	(3.64)	(2.92)	(2.44)	(2.36)	(2.83)	(4.12)	(4.16)	(2.50)	(2.40)
UMD	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.02	0.02	0.00	-0.01
	(0.28)	(0.33)	(0.46)	(0.50)	(0.49)	(0.43)	(2.17)	(1.84)	(0.29)	(0.88)
MktRF	1.06	1.06	1.03	1.02	1.02	1.02				
	(4.34)	(4.18)	(2.97)	(2.02)	(1.88)	(1.31)				
SMB	0.80	0.82	0.81	0.81	0.81					
	(30.89)	(32.13)	(46.67)	(44.26)	(35.78)					
Mid minus Large (MML)						0.78				
						(26.10)				
Small minus Mid (SMM)						0.70				
						(19.75)				
HML	0.20	0.21	0.06	0.09						
	(6.03)	(6.53)	(2.59)	(3.78)						
BHML					0.05	0.03				
					(1.84)	(1.02)				
SHML					0.04	0.06				
					(2.00)	(1.28)				
MidHML						0.02				
						(0.49)				
Average weights on 3x4 portfolios implied by models							Flex	NNLS	Actual	Market
Large_Low - RF	0.110	0.097	0.027	0.018	0.015	-0.043	-0.030	0.000	0.000	0.396
Large_Med - RF	0.036	0.030	0.030	0.033	0.033	0.011	-0.045	0.000	0.000	0.230
Large_High - RF	-0.020	-0.021	0.034	0.048	0.051	0.005	0.039	0.008	0.000	0.098
Large_None - RF	0.012	0.012	0.012	0.002	0.002	0.000	-0.004	0.000	0.000	0.012
Mid_Low - RF	0.020	0.018	0.005	0.003	0.003	0.082	0.120	0.062	0.040	0.073
Mid_Med - RF	0.010	0.008	0.009	0.009	0.009	0.104	0.143	0.105	0.032	0.065
Mid_High - RF	-0.007	-0.007	0.012	0.017	0.018	0.056	-0.008	0.000	0.010	0.035
Mid_None - RF	0.013	0.013	0.012	0.002	0.002	0.017	-0.027	0.000	0.007	0.012
Small_Low - RF	0.213	0.213	0.343	0.268	0.271	0.221	0.302	0.322	0.323	0.042
Small_Med - RF	0.308	0.314	0.338	0.285	0.284	0.288	0.413	0.418	0.321	0.038
Small_High - RF	0.391	0.403	0.233	0.218	0.213	0.173	0.092	0.116	0.173	0.023
Small_None - RF	0.024	0.024	0.024	0.171	0.171	0.151	0.030	0.002	0.093	0.023

Table 8. Alphas and sum of squared differences between weights on 3x4 portfolios produced by the models and those from the flexible model.

This table summarizes results from for multiple indices. For each model and index reported in Table 7, this table reports the alphas and the sum of the squared differences between the actual average index holdings of the 3x4 portfolios and those implied by the model. For subsets of indices, the table also reports the sum of squared average alphas and the sum of sum-of-squared differences in portfolio weights.

Model	Carhart	(2)	(3)	(4)	(5)	(6)	Flexible	NNLS	Actual	Avg
Share codes in market factor	CRSPVW	10/11	10/11	10/11	10/11	10/11				
SMB weighting	EW	EW	VW	VW	VW	VW				
SMB stocks included	As in FF	As in FF	As in FF	All	All	All				
Small-Big cutoff	50th pct	50th pct	50th pct	50th pct	50th pct	N/A				
Size deciles included in BHML	N/A	N/A	N/A	N/A	Top 5	Top 2				
Size deciles included in SHML	N/A	N/A	N/A	N/A	Btm 5	Btm 5				
BM deciles included in H	Top 3	Top 3	Top 3	Top 3	Top 3	Top 7				
Panel A: Alphas							Flex	NNLS	Actual	Avg
S&P 500	0.82	0.59	0.33	0.32	0.11	0.21	0.07	0.16	0.35	0.04
S&P 500 Growth	1.82	1.58	1.25	1.23	-0.01	-0.11	-0.13	-0.64	-0.53	-0.60
S&P 500 Value	-0.35	-0.58	-0.76	-0.76	0.07	0.37	0.12	0.49	1.05	0.42
Russell 2000	-2.41	-2.66	-1.62	-1.53	-1.50	-1.61	-2.13	-2.17	-1.07	-1.23
Russell 2000 Growth	-3.41	-3.66	-2.51	-2.43	-1.09	-1.13	-1.77	-1.91	-1.34	-1.58
Russell 2000 Value	-1.25	-1.50	-0.63	-0.54	-1.89	-1.80	-2.18	-1.61	-0.71	-0.62
Russell Midcap	0.17	-0.08	0.21	0.24	0.30	0.45	-0.17	-0.09	0.86	0.52
Russell Midcap Growth	1.61	1.56	1.95	1.97	2.79	1.34	0.43	-0.50	0.69	0.27
Russell Midcap Value	-0.52	-0.62	-0.50	-0.48	-0.59	0.02	-0.64	0.09	1.11	0.59
Panel B: Sums of squared average alphas										
All 9 indices	26.00	28.71	15.68	14.91	15.29	9.29	13.11	11.90	7.41	5.64
Panel C: Sum of squared differences in 3x4 portfolio weights										
S&P 500	0.016	0.016	0.015	0.012	0.014	0.005	0.017	0.006		
S&P 500 Growth	0.250	0.247	0.203	0.201	0.122	0.012	0.056	0.002		
S&P 500 Value	0.115	0.121	0.136	0.132	0.115	0.058	0.052	0.053		
Russell 2000	0.079	0.082	0.014	0.018	0.018	0.026	0.044	0.027		
Russell 2000 Growth	0.198	0.198	0.092	0.065	0.059	0.050	0.081	0.043		
Russell 2000 Value	0.130	0.134	0.098	0.129	0.173	0.103	0.095	0.039		
Russell Midcap	0.120	0.124	0.119	0.116	0.114	0.050	0.043	0.035		
Russell Midcap Growth	0.306	0.303	0.350	0.328	0.428	0.162	0.274	0.167		
Russell Midcap Value	0.307	0.319	0.293	0.297	0.311	0.178	0.088	0.064		
All 9 indices avg	0.169	0.172	0.147	0.144	0.150	0.071	0.083	0.048		

Table 9. Attribution analysis of benchmark indices.

Panel A shows how the Carhart alpha of the S&P500 index arises from the contributions of index stocks in 100 Fama-French portfolios selected by market capitalization and book-to-market ratio, as well as size portfolios for U.S. stocks with insufficient BM data ("None") and for other CRSP securities ("Other"). For each cell, the Carhart betas and monthly alphas of index stocks are computed, then monthly alphas are multiplied by the monthly weight of the index in that cell, and finally the monthly alpha contributions are added up across all months. The alpha contribution of index stocks is also shown relative to all stocks in each cell, using the same weights on the 120 component portfolios as the S&P 500. Panel B repeats the analysis for the Russell 2000. All numbers are in basis points per year.

Panel A: S&P 500

Contribution to alpha

	Other	None	Growth	2	3	4	5	6	7	8	9	Value	All
Large	-5	6	74	45	-11	-12	-13	-13	-6	-11	-2	-3	73
9	0	1	1	4	-2	-2	2	-2	-1	-3	-1	1	-2
8	0	2	4	1	-1	-1	2	-3	0	1	1	-1	7
7	0	0	0	0	-1	0	0	1	-1	-1	1	-1	-3
6	0	0	-1	1	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	-1
4	0	0	0	0	0	0	0	0	0	0	0	0	1
3	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0
Small	0	0	0	0	0	0	0	0	0	0	0	0	0
All	-5	7	79	49	-14	-14	-12	-14	-5	-21	-5	-2	81

Alpha relative to Fama-French benchmark

	Other	None	Growth	2	3	4	5	6	7	8	9	Value	All
Large	3	2	-3	2	0	1	1	-2	-8	2	0	0	3
9	0	1	-3	1	0	0	0	0	-1	1	0	1	0
8	0	2	2	1	0	0	2	-1	1	2	0	0	10
7	0	0	0	-1	0	1	0	1	0	0	0	0	-1
6	0	0	-1	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0
Small	0	0	0	0	0	0	0	0	0	0	0	0	0
All	3	3	-4	4	0	1	4	-2	-4	1	1	0	11

Panel B: Russell 2000

Contribution to alpha

	Other	None	Growth	2	3	4	5	6	7	8	9	Value	All
Large	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	-1	0	0	0	0	0	0	0	0	0	-3
7	0	-3	6	-1	-1	0	-1	-1	1	0	0	0	2
6	-2	-5	-12	-7	0	-3	-5	-1	-6	-2	-1	-1	-60
5	0	-1	-14	-2	-2	-1	-2	-2	0	-3	4	1	-29
4	-4	-1	-19	0	-7	-1	1	5	1	6	1	-4	-33
3	-3	-8	-16	1	2	-5	4	3	0	2	-2	-1	-29
2	-1	-7	-15	-6	-7	-2	1	-2	3	-2	1	-1	-50
Small	-1	-1	-13	1	-2	3	2	0	0	0	1	-1	-29
All	-14	-35	-93	-15	-20	-14	-5	-3	-3	-1	5	-9	-238

Alpha relative to Fama-French benchmark

	Other	None	Growth	2	3	4	5	6	7	8	9	Value	All
Large	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	-2	0	0	0	0	0	0	0	0	0	-4
7	0	-2	0	-1	-1	1	-1	-1	0	0	0	0	-2
6	0	-2	-5	-9	3	-1	-1	-1	-2	-1	-2	-2	-41
5	2	-6	-6	-1	-2	-3	0	-3	0	-1	4	2	-12
4	2	-3	-1	1	0	2	1	-1	0	1	-2	-1	-2
3	2	-8	1	0	0	-1	1	-1	0	-1	-1	-1	-8
2	0	-6	5	-2	-4	-2	0	-2	0	-2	-1	1	-15
Small	0	6	4	3	-1	1	1	0	0	-1	-1	-1	-4
All	4	-32	-7	-6	-3	-3	-1	-10	-4	-3	-2	-3	-69

Table 10. Russell 2000 alphas in June and July.

In this table, the regression models (1), (4), and (8) from Table 7 are run including an indicator variable for June and July. Only the constant and June-July coefficients are reported; the other coefficients are very similar to those reported earlier (and a similar table for Russell 2000 Growth). T-stats from robust standard errors are in parentheses.

	Russell 2000			Russell 2000 Growth		
Model	(1)	(4)	(8)	(1)	(4)	(8)
Constant	-0.106	-0.058	-0.064	-0.133	-0.080	-0.025
	(1.65)	(1.07)	(1.24)	(1.84)	(1.32)	(0.45)
June-July dummy	-0.582	-0.422	-0.395	-0.923	-0.748	-0.515
	(3.86)	(3.52)	(3.46)	(4.84)	(4.75)	(4.05)
Total alpha per year	-2.432	-1.542	-1.559	-3.440	-2.450	-1.331

Table 11. Correlations across factors.

Panel A reports the time series correlations of the Fama-French factors with our modified versions of those factors. Panel B reports the correlations of the Fama-French factors with factors based on common benchmark indices: the S&P 500 (S5), Russell 2000 (R2), Russell Midcap (RM), and Russell 3000 (R3). The value and growth components of the indices are represented by V and G. For example, “R2-S5” is long Russell 2000 and short S&P 500, while “R2V-R2G” is long Russell 2000 Value and short Russell 2000 Growth. The time period is 2/1986–12/2005.

Panel A: Original FF factors with modified FF factors

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	MktRF as in FF											
(2)	SMB as in FF	18.9										
(3)	HML as in FF	-49.0	-43.2									
(4)	MktRF share codes 10/11	99.9	17.7	-49.2								
(5)	SMB with 50% cutoff	15.6	97.5	-29.6	14.3							
(6)	BHML, Size top 5, H top 3	-37.3	-25.3	90.7	-37.8	-8.2						
(7)	SHML, Size btm 5, H top 3	-51.9	-52.3	93.3	-51.8	-43.5	69.6					
(8)	MML	19.9	85.9	-22.5	18.4	88.8	-0.2	-38.3				
(9)	SMM	7.7	86.4	-29.7	7.0	88.0	-14.5	-38.4	56.4			
(10)	BHML, Size top 2, H top 7	-35.7	-23.8	86.3	-36.4	-7.7	91.8	69.2	3.1	-17.2		
(11)	SHML, Size btm 5, H top 7	-54.8	-55.4	91.8	-54.6	-46.3	68.5	98.3	-39.7	-41.9	69.8	
(12)	MidHML	-39.9	-57.0	90.3	-39.7	-44.5	77.1	88.5	-39.5	-38.6	76.4	89.4

Panel B: Original FF factors with index factors

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	MktRF as in FF										
(2)	SMB as in FF	18.9									
(3)	HML as in FF	-49.0	-43.2								
(4)	S5	98.1	2.0	-41.5							
(5)	R2-S5	14.6	93.3	-24.2	-2.4						
(6)	RM-S5	10.7	70.8	-3.9	-4.6	85.7					
(7)	R2-RM	14.7	91.8	-36.0	0.0	90.2	55.1				
(8)	R3V-R3G	-44.2	-36.0	90.5	-37.9	-15.1	5.0	-28.6			
(9)	S5V-S5G	-21.9	-16.1	72.5	-20.0	5.6	23.8	-10.9	84.4		
(10)	RMV-RMG	-48.2	-53.0	89.2	-37.5	-36.4	-20.4	-41.9	89.8	63.3	
(11)	R2V-R2G	-55.6	-53.8	89.7	-44.6	-40.0	-23.7	-45.0	83.2	55.1	92.9

Table 12. Cross-sectional pricing results.

Panel A presents the results for various cross-sectional OLS regressions where mean excess returns of 100 Fama-French size-BM-sorted test portfolios (10x10 sort) are regressed on their factor betas. The multivariate factor betas of each test portfolio are estimated in a time-series regression. For each model, we report the coefficients in the first row and their t-statistics (in parentheses) below, where standard errors are adjusted for the estimation risk in betas (Shanken (1992)). We also report the Hansen-Jagannathan statistic and its asymptotic p-value of pricing errors being as large or larger under the null of the model holding exactly. Panel B repeats the same tests for purely index-based models. The time period for both panels is 2/1986–12/2005.

Panel A: Modified Fama-French models						
	(1)	(2)	(3)	(4)	(5)	(6)
H-J statistic	0.69	0.68	0.68	0.66	0.65	0.63
p-value	9.4%	11.6%	12.4%	21.2%	22.1%	43.5%
R^2	28.6%	34.4%	59.2%	63.5%	63.5%	47.5%
Constant	0.17	0.08	0.30	0.23	0.22	0.32
	(3.38)	(1.58)	(5.65)	(5.40)	(5.53)	(4.99)
UMD	0.49	0.41	0.05	-0.01	-0.01	0.20
	(4.67)	(4.48)	(0.62)	(0.12)	(0.11)	(2.92)
MktRF	-0.14	-0.04	-0.28	-0.21	-0.20	-0.30
	(2.42)	(0.75)	(3.82)	(3.17)	(3.50)	(3.91)
SMB	0.09	0.07	0.07	0.06	0.06	
	(2.45)	(2.00)	(1.99)	(1.78)	(1.84)	
MML (Mid minus Large)						0.14
						(4.05)
SMM (Small minus Mid)						-0.03
						(1.80)
HML	0.10	0.08	0.02	0.03	0.03	
	(3.14)	(2.40)	(0.72)	(1.16)	(1.20)	
BHML (Big HML)						0.15
						(3.31)
SHML (Small HML)						0.06
						(2.08)
MidHML						-0.01
						(0.39)
S5		-0.02	-0.28	-0.20	-0.20	
		(0.44)	(3.68)	(3.01)	(3.41)	
RM-S5			0.06	0.08	0.08	
			(2.02)	(2.56)	(2.77)	
R2-RM				-0.03	-0.03	
				(1.32)	(1.40)	
R3V-R3G					0.06	
					(1.26)	

Table 12. (continued)

Panel B: Index-based models								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
H-J statistic	0.71	0.70	0.68	0.66	0.75	0.74	0.70	0.69
p-value	4.5%	5.0%	12.6%	20.7%	0.6%	0.7%	4.3%	9.1%
R^2	32.6%	48.4%	49.1%	58.2%	24.1%	48.3%	47.8%	57.8%
Constant	0.07	0.28	0.24	0.21	0.34	0.31	0.33	0.26
	(1.18)	(4.69)	(3.99)	(3.60)	(5.08)	(4.71)	(5.57)	(4.79)
S5	-0.09	-0.32	-0.27	-0.20	-0.36	-0.35	-0.37	-0.25
	(-1.40)	(-4.24)	(-3.62)	(-2.95)	(-4.70)	(-4.64)	(-4.63)	(-3.78)
R2-S5	0.14				0.12			
	(3.13)				(2.76)			
RM-S5		0.17	0.16	0.12		0.18	0.17	0.13
		(4.29)	(4.46)	(4.06)		(4.38)	(4.57)	(4.13)
R2-RM		-0.02	-0.01	-0.03		-0.02	-0.02	-0.04
		(-1.04)	(-0.72)	(-1.57)		(-1.15)	(-1.21)	(-1.83)
R3V-R3G	0.09	0.06			0.10	0.06		
	(2.64)	(1.84)			(2.71)	(1.78)		
S5V-S5G			0.10	0.09			0.10	0.09
			(1.89)	(1.75)			(1.91)	(1.75)
RMV-RMG				-0.11				-0.11
				(-2.34)				(-2.32)
R2V-R2G			-0.01	0.05			0.01	0.06
			(-0.37)	(1.56)			(0.23)	(2.00)
UMD	0.57	0.18	0.30	0.14				
	(4.82)	(2.72)	(4.25)	(2.37)				

Table 13. Mutual fund tracking error across benchmark models.

This table shows the out-of-sample tracking error volatility for US all-equity mutual funds 1996-2005. Whenever a fund reports its positions (semiannually or quarterly), its prior twelve-month daily returns are regressed on each of the factor models to determine its betas. Using those betas, the fund's monthly out-of-sample predicted return and the difference between the predicted and actual fund return are computed. Each fund's tracking error is computed as the time-series volatility of that difference over the sample period. Each number in the table represents an equal-weighted average of those tracking errors across funds. Panel B uses only funds with low Active Share. Panel C shows the results for different lengths and sampling intervals of the estimation period.

Tracking error volatility (% per year)						
Panel A: All funds						
Model	None	CAPM	FF	Carhart	+S5+R2	MOD7
Excess return	17.35	8.28	6.50	6.44	6.10	6.15
Benchmark-adjusted	6.91	6.58	6.18	6.14	5.95	5.99
Model	MOD4	IDX4	IDX5	IDX6a	IDX7	IDX6b
Excess return	6.40	6.15	6.12	5.96	5.80	5.82
Benchmark-adjusted	6.12	6.03	5.86	5.79	5.71	5.76
Panel B: Active Share < median						
Model	None	CAPM	FF	Carhart	+S5+R2	MOD7
Excess return	16.02	6.53	5.19	5.20	4.90	4.95
Benchmark-adjusted	5.33	5.13	4.78	4.75	4.67	4.61
Model	MOD4	IDX4	IDX5	IDX6a	IDX7	IDX6b
Excess return	5.20	4.73	4.82	4.68	4.49	4.47
Benchmark-adjusted	4.73	4.62	4.49	4.44	4.36	4.39
Panel C: All funds, alternative estimation periods						
Daily data, 6 months						
Model	None	CAPM	FF	Carhart	+S5+R2	MOD7
Excess return	17.35	8.23	6.49	6.48	6.19	6.21
Benchmark-adjusted	6.91	6.55	6.18	6.21	6.03	6.08
Model	MOD4	IDX4	IDX5	IDX6a	IDX7	IDX6b
Excess return	6.49	6.19	6.19	6.00	5.85	5.84
Benchmark-adjusted	6.18	6.02	5.90	5.83	5.75	5.79
Monthly data, 3 years						
Model	None	CAPM	FF	Carhart	+S5+R2	MOD7
Excess return	17.35	8.77	6.82	6.82	6.79	6.94
Benchmark-adjusted	6.91	6.87	6.74	6.78	6.96	7.18
Model	MOD4	IDX4	IDX5	IDX6a	IDX7	IDX6b
Excess return	6.76	6.48	6.57	6.56	6.57	6.54
Benchmark-adjusted	6.76	6.80	6.78	6.91	7.05	7.02
Monthly data, 5 years						
Model	None	CAPM	FF	Carhart	+S5+R2	MOD7
Excess return	17.35	8.64	6.90	6.86	6.64	6.77
Benchmark-adjusted	6.91	6.83	6.74	6.75	6.82	6.96
Model	MOD4	IDX4	IDX5	IDX6a	IDX7	IDX6b
Excess return	6.75	6.42	6.46	6.43	6.45	6.45
Benchmark-adjusted	6.71	6.71	6.66	6.73	6.85	6.83

None	-	MOD4	MKT2, SMB2, HML, UMD
CAPM	MKT	IDX4	S5, R2-S5, R3V-R3G, UMD
FF	MKT, SMB, HML	IDX5	S5, R2-S5, S5V-S5G, R2V-R2G, UMD
Carhart	MKT, SMB, HML, UMD	IDX6a	S5, RM-S5, R2-RM, S5V-S5G, R2V-R2G, UMD
+S5+R2	MKT, SMB, HML, UMD, S5, R2	IDX7	S5, RM-S5, R2-RM, S5V-S5G, RMV-RMG, R2V-R2G, UMD
MOD7	MKT2, MMB, SMM, BHML, MHML, SHML, UMD	IDX6b	S5, RM-S5, R2-RM, S5V-S5G, RMV-RMG, R2V-R2G

Table 14. Mutual fund alphas.

This table shows the alphas of net return for US all-equity mutual funds 1996-2005. Funds are sorted into groups based on their estimated benchmark indices: the size groups represent small, mid, and large-cap stocks, and the value groups represent growth, core, and value stocks. Alphas are computed with excess return (i.e., fund return minus risk-free rate) or benchmark-adjusted return (i.e., fund return minus benchmark index return) as left-hand-side variables and various benchmark models on the right-hand side. The numbers show the annualized alpha, with t-statistics in parentheses below.

Excess return					Benchmark-adjusted return				
Size	Value group				Size	Value group			
group	1	2	3	All	group	1	2	3	All
Panel A: No model									
3	4.43	5.22	6.59	5.36	3	-1.12	-1.15	-0.73	-1.13
	(0.79)	(1.11)	(1.48)	(1.11)		(-0.83)	(-1.94)	(-0.93)	(-1.38)
2	6.70	8.86	9.00	7.88	2	-1.61	-1.79	-1.33	-1.65
	(0.90)	(1.59)	(1.94)	(1.24)		(-1.21)	(-1.58)	(-1.14)	(-1.64)
1	7.39	8.41	10.23	8.15	1	2.90	-1.04	0.22	0.89
	(0.91)	(1.49)	(2.06)	(1.29)		(1.97)	(-1.02)	(0.19)	(0.95)
All	5.63	6.38	7.53	6.39	All	-0.46	-1.21	-0.60	-0.80
	(0.88)	(1.32)	(1.70)	(1.22)		(-0.46)	(-2.56)	(-0.92)	(-1.30)
Panel B: Carhart (MKT, SMB, HML, UMD)									
3	-1.24	-1.01	-1.30	-1.07	3	-3.28	-1.81	-0.92	-2.26
	(-1.51)	(-2.33)	(-1.26)	(-2.06)		(-3.31)	(-4.42)	(-1.31)	(-3.99)
2	-2.37	-1.69	-0.53	-1.69	2	-3.35	-2.22	-0.31	-2.58
	(-1.24)	(-1.23)	(-0.37)	(-1.14)		(-3.03)	(-2.46)	(-0.34)	(-3.18)
1	-3.99	-3.09	-1.20	-3.20	1	1.69	-0.73	1.06	0.68
	(-2.06)	(-2.15)	(-0.91)	(-2.26)		(1.39)	(-0.79)	(0.91)	(0.82)
All	-2.08	-1.75	-1.27	-1.69	All	-2.34	-1.60	-0.43	-1.70
	(-1.83)	(-2.56)	(-1.20)	(-2.21)		(-3.09)	(-4.47)	(-0.70)	(-3.68)
Panel C: MKT, SMB, HML, UMD, S5, R2									
3	-0.93	-1.02	0.01	-0.60	3	-1.03	-1.05	-0.70	-1.03
	(-1.00)	(-3.22)	(0.01)	(-1.39)		(-1.33)	(-3.33)	(-0.97)	(-2.50)
2	1.23	1.20	1.58	1.56	2	-2.76	-2.77	-0.78	-2.50
	(0.70)	(1.09)	(1.23)	(1.28)		(-2.53)	(-2.67)	(-0.82)	(-2.93)
1	-0.17	-0.20	0.73	-0.13	1	0.40	-2.34	-0.31	-0.76
	(-0.12)	(-0.17)	(0.61)	(-0.12)		(0.34)	(-2.90)	(-0.30)	(-1.04)
All	-0.38	-0.69	0.21	-0.23	All	-1.22	-1.58	-0.64	-1.25
	(-0.34)	(-1.27)	(0.22)	(-0.36)		(-1.82)	(-4.79)	(-0.99)	(-2.96)
Panel D: S5, R2-S5, R3V-R3G, UMD									
3	-1.52	-1.22	-0.51	-1.10	3	-2.02	-1.23	-0.22	-1.38
	(-2.05)	(-3.57)	(-0.87)	(-2.32)		(-2.63)	(-3.69)	(-0.35)	(-3.22)
2	-0.64	0.92	1.90	0.36	2	-2.82	-1.67	-0.06	-2.09
	(-0.39)	(0.81)	(1.67)	(0.28)		(-2.64)	(-1.97)	(-0.07)	(-2.63)
1	-0.74	0.75	3.24	0.46	1	1.55	-0.95	0.74	0.44
	(-0.49)	(0.71)	(2.68)	(0.48)		(1.31)	(-1.11)	(0.68)	(0.58)
All	-1.15	-0.58	0.45	-0.54	All	-1.54	-1.21	0.01	-1.13
	(-1.18)	(-1.13)	(0.68)	(-0.88)		(-2.41)	(-3.77)	(0.01)	(-2.85)
Panel E: S5, RM-S5, R2-RM, S5V-S5G, RMV-RMG, R2V-R2G, UMD									
3	-1.93	-1.45	-0.66	-1.29	3	-1.71	-1.44	-0.84	-1.44
	(-3.32)	(-5.44)	(-1.16)	(-3.60)		(-3.40)	(-5.46)	(-1.64)	(-4.40)
2	-1.40	-0.14	0.70	-0.41	2	-1.44	-0.67	0.66	-0.90
	(-1.61)	(-0.16)	(0.78)	(-0.55)		(-1.73)	(-0.71)	(0.76)	(-1.45)
1	0.29	0.12	1.64	0.37	1	0.29	-0.38	1.64	0.32
	(0.23)	(0.11)	(1.53)	(0.39)		(0.24)	(-0.45)	(1.59)	(0.41)
All	-1.51	-1.07	-0.11	-0.88	All	-1.30	-1.07	-0.16	-0.99
	(-2.46)	(-2.26)	(-0.20)	(-1.94)		(-2.17)	(-3.41)	(-0.30)	(-2.66)

Table 15. Performance Persistence, Expenses and Returns, and Smart Money.

This table highlights the impact of alternative models on conclusions about performance persistence, expenses and returns, and smart money. We sort equity funds from 1980 to 2005 into equal-weighted portfolios according to their prior-year returns, their most recently disclosed expense ratio, and whether they received inflows in the prior calendar quarter. For each portfolio of funds, we estimate excess returns using a variety of models. Panel A includes actively managed funds only; Panels B and C include both actively managed and index funds; all panels exclude enhanced index funds. A "low-expense" index fund is defined as one that is in the lowest 30 percentile for all index funds in that calendar year (the cutoff averages 40 basis points). T-statistics for return differences across portfolios are heteroskedasticity robust and adjust for clustering within month.

Panel A: Prior-year return persistence

	Alpha by model (% per year)					
Prior-year return decile	FF	4F (Carhart)	MOD7	IDX7	4F+R2+S5	MOD7+R2+S5
10 (highest)	2.11	-0.77	0.07	0.58	0.15	-0.26
2	-2.85	-1.48	-1.69	-1.31	0.20	-0.57
1 (lowest)	-4.65	-2.90	-2.78	-2.38	-0.73	-1.37
Decile 10 - Decile 2	4.96	0.72	1.76	1.89	-0.05	0.31
	(2.35)	(0.35)	(0.79)	(0.83)	(0.03)	(0.16)
Decile 2 - Decile 1	1.80	1.41	1.09	1.06	0.93	0.80
	(3.32)	(2.57)	(2.01)	(2.09)	(1.78)	(1.51)
Decile 10 - Decile 1	6.76	2.13	2.85	2.95	0.88	1.11
	(2.96)	(0.98)	(1.23)	(1.22)	(0.44)	(0.56)

Panel B: Expenses and returns

	Alpha by model (% per year)					
Group	FF	4F (Carhart)	MOD7	IDX7	4F+R2+S5	MOD7+R2+S5
Low-cost index fund	-0.02	0.27	-0.02	0.43	-0.16	-0.05
Active, decile 1 (low)	-0.46	-0.40	-0.67	-0.24	0.13	-0.41
Active, decile 5-6	-0.83	-0.90	-0.95	-0.74	0.20	-0.47
Active, decile 10 (high)	-1.25	-1.47	-1.09	-0.79	0.05	-0.59
Index - Decile 1	0.44	0.67	0.66	0.68	-0.29	0.37
	(1.02)	(1.35)	(1.38)	(1.38)	(0.64)	(0.77)
Decile 1 - Decile 10	0.78	1.07	0.42	0.55	0.07	0.18
	(1.57)	(2.04)	(0.88)	(1.26)	(0.16)	(1.39)

Panel C: Prior-quarter inflows and returns

	Alpha by model (% per year)					
Group	FF	4F (Carhart)	MOD7	IDX7	4F+R2+S5	MOD7+R2+S5
Positive inflows	0.02	-0.36	-0.38	-0.31	0.60	0.09
Negative inflows	-2.29	-2.08	-2.35	-1.99	-1.43	-1.87
Difference	2.31	1.72	1.96	1.67	2.03	1.96
	(2.76)	(1.73)	(1.82)	(1.95)	(2.06)	(1.96)

Appendix B: Figures

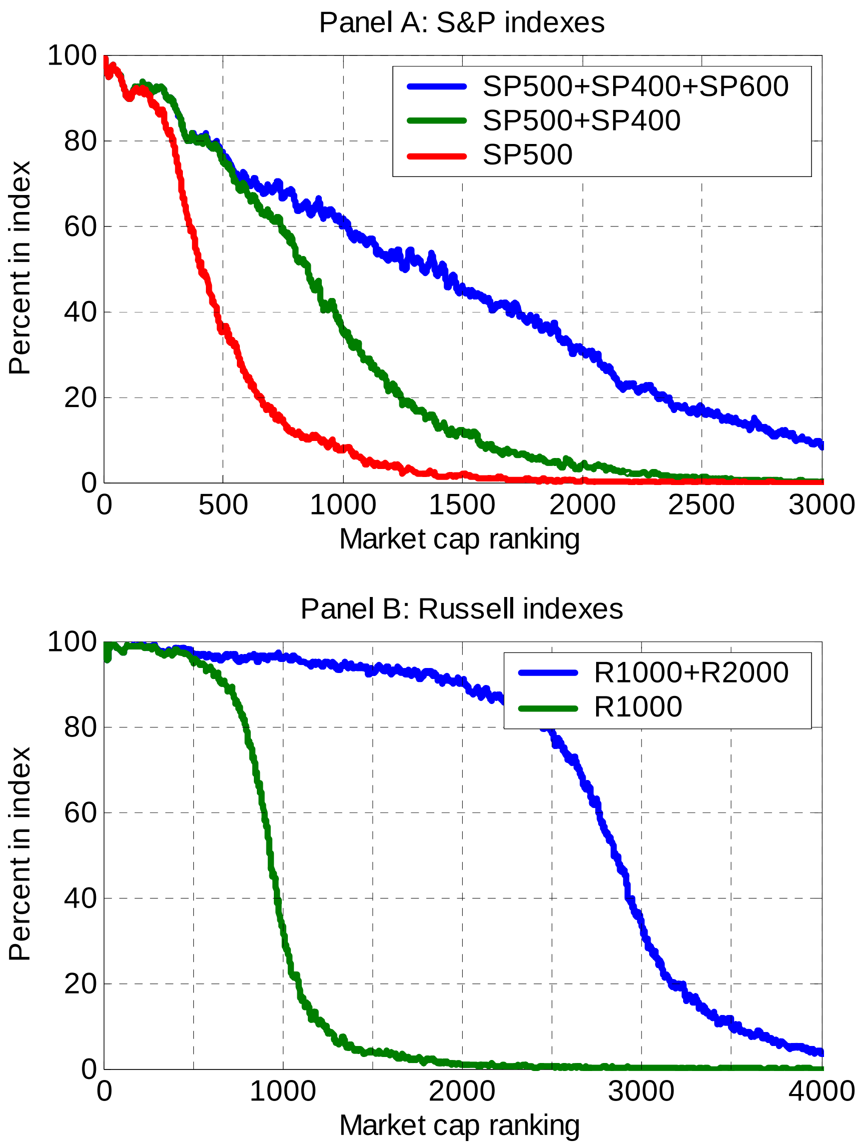

Figure 1. Index membership as a function of market capitalization.
All US stocks in CRSP are sorted each month based on their market cap. For each market cap rank, we include 10 stocks above and below and then compute the percentage of those 20 stocks that are index constituents that month. The figures show the averages across 120 months from 1996 to 2005.

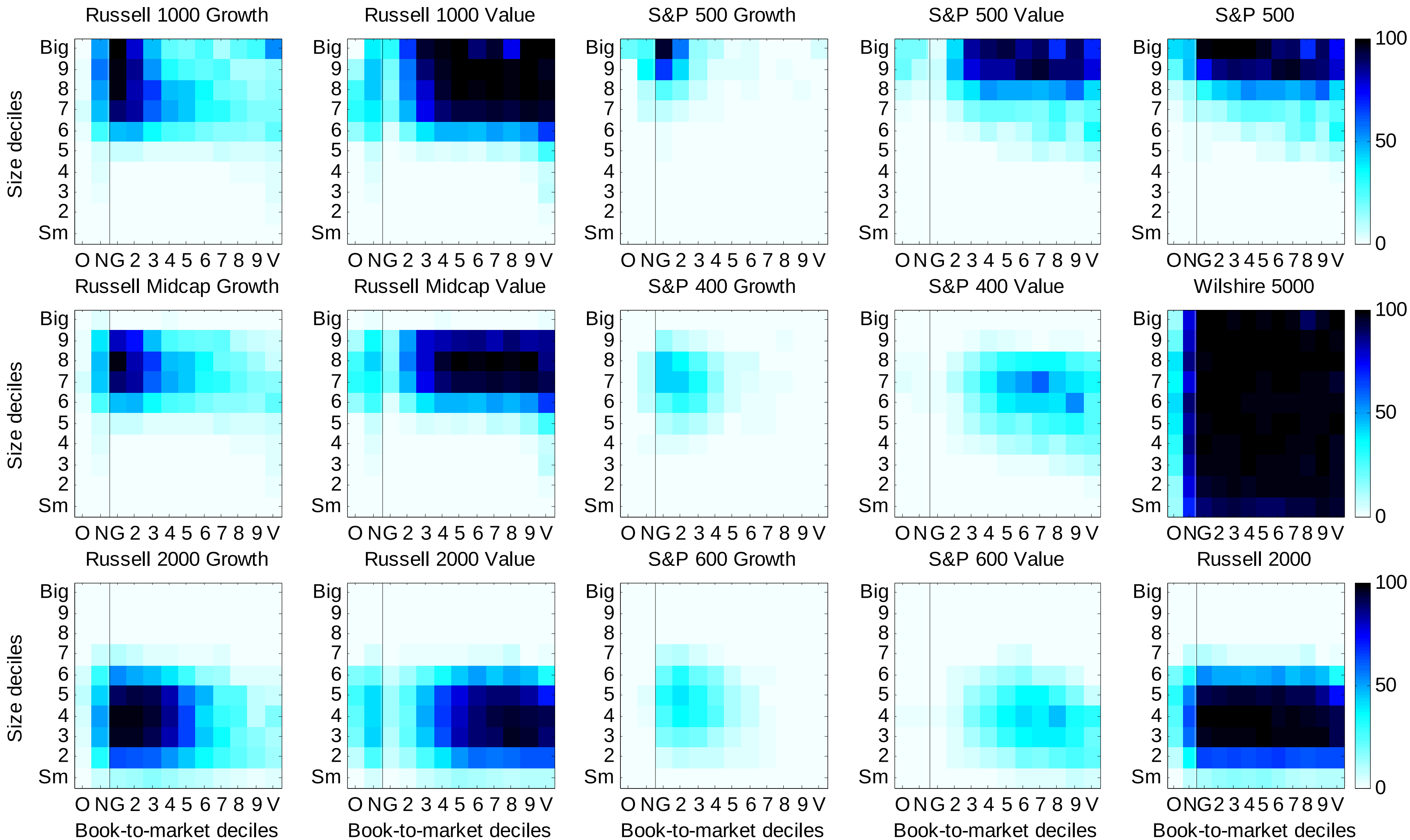

Figure 2. Index membership across size and value groups.

All securities on CRSP are divided into 10 size groups and one of 12 value groups. For each 10x12 component portfolio, the figures shows the fraction of market capitalization that is included in the benchmark index. The component portfolios are determined once a year based on market equity and book-to-market, following the methodology of Fama and French (1993). We also add two new value groups: "N" for those US stocks where the Fama-French inclusion criteria are not satisfied (typically relatively new listings), and "O" for all other stocks. The figures show the mean value from 1997 to 2005, computed across all months. Only ADRs are excluded to mimic the inclusion criteria of the CRSP market index.