Should Benchmark Indices Have Alpha?

Revisiting Performance Evaluation[*]

Martijn Cremers[†]

Antti Petajisto[‡]

Eric Zitzewitz[§]

November 6, 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 and propose 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, Wayne Ferson, Ken French, Juhani Linnainmaa, and Christopher Polk for comments, as well as conference and seminar participants at the Adam Smith Asset Pricing conference, University of Amsterdam, 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. This paper provides evidence that the practitioner and academic approaches can yield very different results, as the academic factor models assign large nonzero alphas even to the passive benchmark indices.

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 – after all, the S&P 500 and Russell 2000 together cover about 85% of the U.S. equity market value and are the two most common benchmark indices for fund managers. We argue that these nonzero index alphas are misleading, and that they are symptoms of biases that can significantly affect performance evaluation in general.

In this paper we investigate the Fama-French methodology to identify the sources of the 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 pricing and performance evaluation, and thus we propose them as good alternatives to the commonly used academic factor models. In order to avoid data mining, it is important not to blindly test a wide variety of candidate models; instead, our selection of alternatives is entirely guided by the issues we uncover in our analysis of the Fama-French methodology.

The main source of the nonzero index alphas is the 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, so the equal-weighted portfolios in the Fama-French factors give much more weight to a given unit of capitalization if it is in the Small size and High BM (i.e., value) portfolio. Such tilts in weights 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 the value-weighting of the market portfolio again severely distort the allocation within large-caps, generating a tilt in the benchmark portfolio 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 benchmark dominated by small-cap value stocks which has historically performed well, thereby explaining most of the negative index alpha. For both the S&P 500 and Russell 2000, this problem can be addressed simply by using a value-weighted SMB factor, which produces alphas much closer to zero.

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

---

[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) 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.

S&P 500 and other indices typically only include U.S. stocks, using the CRSP market proxy contributes to a positive alpha.

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. 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 exposures to the 100 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 the reconstitution effect also has an impact on index alphas.

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. 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. Our primary index-based alternative models are a four-factor model analogous to Carhart, except that it adds the S&P 500, Russell 2000, and an index-based value factor to the usual momentum factor, and a seven-factor model which adds the Russell Midcap and introduces separate index-based value factors for small, midcap, and large stocks. By construction, all of our alternative models eliminate or significantly reduce the alphas of common benchmark indices.

To better understand which model to recommend, we start by following the general approach of Fama and French (1993) and test how well various factor models explain common time-series variation in portfolio returns. Our test assets are U.S. all-equity mutual funds, representing a broad cross-section of actual investment portfolios. A model that closely tracks the returns for an individual fund will also produce low noise in the alpha estimate of the fund. We find that both the Carhart model and the index-based models capture significant common variation in fund returns, with the latter explaining an average of 91% of individual fund return variation. However, the index models do better: a four-factor index model decreases out-of-sample tracking error volatility by about 5% relative to Carhart, and a seven-factor index model decreases it by 10% on average, and more for larger or less active funds.

Next, we investigate how well the factor models explain the cross-section of average returns, starting with U.S. mutual funds. Sorting them into size and value groups, we find small funds outperforming large funds and value funds outperforming growth funds, consistent with the well-known size and value effects. A seven-factor index model largely eliminates these patterns across fund groups, producing alphas that are statistically indistinguishable from the average fund alpha.

In sharp contrast, the Carhart alphas indicate that small-cap funds underperformed large-cap funds by 2.13% per year, which arises from the fact that the Carhart alphas of the small-cap benchmark indices are on average an astounding 5.07% per year lower than those of the large-cap indices. If instead we control for the benchmark index of a fund, these results are completely reversed, with Carhart alphas suggesting that small-cap funds outperformed large-cap funds by 2.94%. Alternatively, adding the S&P 500 and Russell 2000 as factors to the Carhart model also reverses the pattern in alphas. This confirms that the sensitivity of the Carhart model can indeed have an economically very significant impact on performance evaluation of money managers.

As another set of test assets, we investigate the pricing of 100 Fama-French size-BM-sorted portfolios. The four-factor Carhart model has a cross-sectional R^2 of 29% over our time period. Far from being redundant assets, the S&P 500, Russell Midcap, and Russell 2000 increase the R^2 to 64% when added to the Carhart model. As an alternative to the Fama-French factors, a seven-factor index model 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.

The general conclusion from our analysis is that benchmark indices matter 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 also 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 keep the number of factors smaller, a four-factor index model consisting of the S&P 500 and Russell 2000 together with value and momentum factors 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 it does not investigate model construction in any detail or evaluate alternative model specifications. Huij and Verbeek (2007) also question the use of common academic factors and instead advocate the use of factors based on mutual fund returns. Daniel, Grinblatt, Titman, and Wermers (1997) present a nonlinear benchmarking methodology based on characteristics-matched portfolios that avoids 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 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 document the long-term alphas of all common benchmark indices under the common academic factor models, and we identify the sources of the nonzero alphas. Furthermore, we propose improvements to the common academic factor models, as well as alternative models that are based on the common benchmark indices and therefore convenient for anyone to implement.

This paper proceeds as follows. Section 2 discusses the criteria for judging pricing and benchmarking models. Section 3 explains the data sources, including 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 presents the alternative factors we analyze. Section 6 examines the common variation in returns explained by various factor models. Section 7 explores how well each model explains the cross-section of average returns, using both mutual funds and Fama-French portfolios as test assets. We present our conclusions in Section 8. All tables and figures are in the appendix.

2 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 systematic 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. For example, even if value and size were not priced, they could 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; a more extreme example would be controlling for the average industry risk in a portfolio. Similarly, Fama and French (1993) propose two bond market factors in spite of the fact that their long-term risk premia are close to zero, in part because they explain so much time-series variation in returns but also because their risk premia may vary over time.

For the performance evaluation of a large pool of money managers, it is convenient to apply a generally applicable and relatively parsimonious model and not include ad hoc adjustments due to a manager's average exposure to non-priced factors. 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 and Russell 2000. While the mere subtraction of the benchmark index return may oversimplify performance evaluation, a set of multiple benchmark indices may also be used as convenient factors for pricing and benchmarking purposes.

To test how well a model can do as a benchmark for money managers, we test for the aforementioned properties. First, a new model should track the time series of returns better than the old models, producing lower tracking error volatility. This also means that the factors capture common variation in returns, which is a necessary condition for a nonzero factor premium in the Arbitrage Pricing Theory. Second, a model should explain the cross-section of average returns well, not generating significant alphas for large segments of the market such as large-caps or small-caps in general. This should hold for tests assets such as size and book-to-market-sorted portfolios but also for a cross-section of mutual funds, unless we consider it plausible that the average managerial skill varies from large positive to large negative values across market segments.

3 Data

3.1 Benchmark Indices

We include all non-specialized U.S. equity benchmark indices that are 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 these three index providers, covering monthly and daily index returns (including dividends) 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. 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 U.S. all-equity mutual funds in January 2007. The most common benchmark index is the S&P 500. Russell 2000 is the second-most popular benchmark, and its value and growth components are also relatively popular. The most common general mid-cap index is the S&P 400, although the Russell Midcap group of indices is collectively more popular.[3] Wilshire indices are less common in terms of the number of funds, but they each have a nontrivial amount of assets benchmarked to them.

Figure 1 shows the fraction of ordinary common stock 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 1996 to 2005. For S&P indices in 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 use strict market cap cutoffs.[4]

3.2 Other Data Sources

All stock data are from CRSP, supplemented with accounting data from Compustat. Mutual fund data items are primarily from CRSP, with the exception of quarterly holdings data from Thomson Financial, self-reported benchmark index data from Morningstar, and daily fund returns before 2001 from a survivorship-free database originally obtained from the Wall Street Web and used by Goetzmann, Ivkovic, and Rouwenhorst (2001). The CRSP and Thomson Financial mutual fund databases have been matched using MFLINKS. We pick a sample of U.S. all-equity mutual funds with at least $10M in assets, following the procedure in Cremers and Petajisto (2007). Fama-French factor and portfolio data are from Ken French's website.

[3] 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.

[4] The 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.

4 Alphas of Benchmark Indices

4.1 Baseline Results

Table 2 presents estimates of Carhart alphas for the major Russell, S&P, and Wilshire indices from 1980 to 2005.[5] As discussed in the introduction, alphas are positive and statistically 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 nonzero 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. In unreported results, index alphas estimated using betas from a trailing 60-month window are qualitatively similar.

4.2 Sources of Benchmark Alphas: Factor Construction

The standard Fama-French model makes a number of methodological choices. 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

[5] 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 (which is mechanically based on market capitalization).

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 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.

4.2.1 Definition of the Market Portfolio

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,[6] 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.[7] 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. This is explained by the underperformance of other securities such as foreign firms and closed-end funds, which have a surprisingly low Carhart alpha of -4.01% (t = 2.67) per year. The stocks included in the Fama-

---

[6] 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.

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

French size-BM-sorted portfolios have an alpha of 51 basis points per year, indicating underperformance by stocks with insufficient data or negative book value, which is also consistent with the general long-term underperformance of IPOs.[8]

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 equity mutual funds, although the funds represent a slightly smaller fraction of shareholders in REITs than in U.S. firms. 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.

Overall, these results indicate that the CRSP value-weighted market portfolio is a downward-biased benchmark for portfolios consisting of U.S. stocks. Instead, it would be more appropriate to benchmark actively managed U.S. equity portfolios with a market portfolio consisting of only U.S. equities.

4.2.2 Equal-Weighting in Fama-French factors

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). They then equal-weight across these six portfolios; 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, as illustrated in Panel B of Table 4. Fama and French exclude stocks with negative book equity or with no

[8] See Ritter (1991) for the long-term IPO performance, and 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."

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.

Panel A of Table 4 reports the average share of the CRSP market index represented by Fama and French's 2x3 portfolios as well as their average excess returns. Panel A also shows two portfolios of stocks which are excluded by the Fama-French factors but still included in the CRSP index. Just like in Fama and French (1993), two things are apparent: first, the growth portfolios have much more market capitalization than the value portfolios, and second, the best performance by far has been exhibited by the small value portfolio.

To see how this creates nonzero index alphas, let us consider two "target portfolios:" the Fama-French size decile 10 portfolio, which contains the typical large stocks in the S&P 500 index, and the size decile 4 portfolio, 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. If the benchmark portfolio has the same broad category exposures as the target portfolio, the alphas are likely to be zero; if the two differ significantly, this may be (but does not have to 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.

The left-hand side of Panel C shows the weights that 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 a large and negative weight of –19.1%; 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 (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 loading 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.

A simple way to address this problem is to value-weight the SMB component portfolios within the size groups. This affects the alphas in two ways: First, there is a direct effect coming from the high (and "exaggerated") average return on the equal-weighted SMB factor. Second, there is an even more significant indirect effect coming from the equal-weighted SMB factor, which distorts portfolio weights in large stocks in a way that induces an offsetting HML loading, which in turn contributes to a positive alpha for large stocks and negative alpha for small stocks. A value-weighted SMB factor avoids both problems. We quantify this effect in Section 4.3.

4.2.3 Single Value Factor across Size Groups

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% per year) than Big stocks (9.20–7.61 = 1.59% 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)).

4.2.4 Boundaries between Size and Value Groups

The fourth choice we 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 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 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 S&P 500 primarily includes stocks from NYSE size deciles 9 and 10, while midcap stocks are drawn mostly from deciles 6-8. The Russell 2000 includes stocks from size 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 BM 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, whereas the Fama-French portfolio cutoffs are based only on NYSE stocks which tend to have a value-bias relative to Nasdaq stocks.

Panels B and C in Table 5 report the SMB and HML betas of the Fama-French 10x10 Size-BM portfolios, along with two separate columns for stocks with inadequate book equity data or with other share codes. Three observations can be made. First, only the largest cap decile is clearly negatively correlated with SMB; the Midcaps (size 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 size 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 stocks in 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.

4.3 Impact of Alternative Models on Benchmark Alphas

In this subsection, we examine how alternative choices in constructing factors affect the index alphas as well as their implied loadings on Size-BM portfolios. Panel A of Table 6 contains the results for the S&P 500 and Panel B for the Russell 2000.[9] Each panel estimates several alternative models and calculates the weights implied by the resulting betas on a 3x4 set of Size-BM portfolios (Large, Mid, and Small size; Low, Medium, High, and No BM).[10] These implied weights are then compared with both the weights estimated from flexible models (which include each of the 12 portfolios as a factor; the "NNLS" model further restricts all weights to be nonnegative) 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. While such mismatches need not necessarily contribute to index alphas, in practice it turns out that models producing close portfolio weight matches also produce smaller index alphas.

The first column in Panel A of Table 6 estimates the standard Carhart four-factor model for the S&P 500. 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

[9] 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.

[10] 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).

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.

The alpha of the S&P 500, which is 82 basis points per year in the Carhart model, 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 it also brings the model-implied 3x4 portfolio weights closer to the actual index weights.

Panel B of Table 6 conducts the same exercise for the Russell 2000. Switching from an equal to a value-weighted SMB in column 3 increases the estimated alpha by full percentage point per year, from -2.66% to -1.62%. However, even in the more flexible models, the negative alpha of the Russell 2000 remains significant. As we show later, the remaining alpha is concentrated in June and July, suggesting that it is related to the annual reconstitution of the index at the end of June.

Table 7 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. Again the largest improvements in alphas come from the first two steps between (1) and (3), which includes switching to a market portfolio with only U.S. stocks as well as to a value-weighted SMB factor. Alphas decline further between (4) and (7), but this requires adding more factors which may potentially offset the benefit of reduced index alphas.

4.4 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 decompose index alphas into two sources: exposure to passive size- and value-sorted portfolios and stock selection within these 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 size and value factor model, but the rest of the alpha in principle could be explained as it comes from passive and broad-based portfolios of stocks.

Suppose that the assets held by a portfolio *p* are each members of exactly one of *J* benchmark portfolios. The alpha of benchmark portfolio *j* is α_{jt}^{b}. We denote the share of portfolio *p* accounted for by members of benchmark portfolio *j* in month *t* as w_{jt}, and their weighted average alpha as α_{jt}^{a}. Given this setup, the alpha of portfolio *p* can be decomposed into alphas due to the *J* benchmark portfolios, and alphas due to the relative performance $\alpha_{jt}^{a} - \alpha_{jt}^{b}$ of the members of each benchmark portfolio *j* held by *p*. We can call the latter "selection alphas," as they arise from stock selection within a benchmark portfolio, and the former "style alphas" due to asset allocation across benchmark portfolios:[11]

$$\alpha_{pt} = \sum_{j=1}^{J} \underbrace{w_{jt}\alpha_{jt}^{a}}_{\text{active alpha}} = \sum_{j=1}^{J} \underbrace{w_{jt}\left(\alpha_{jt}^{a} - \alpha_{jt}^{b}\right)}_{\text{selection alpha}} + \sum_{j=1}^{J} \underbrace{w_{jt}\alpha_{jt}^{b}}_{\text{style alpha}}. \tag{1}$$

In Panel A of Table 8, we show for the S&P 500 the total alpha contribution ("active alpha" in equation (1)) coming from S&P 500 stocks within each benchmark portfolio, as well as the selection alpha of the index stocks relative to other stocks in the corresponding benchmark portfolios. As benchmarks 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. stocks as well as the securities with other share codes, as discussed earlier. All numbers in the table are time-series averages across the sample period.

The four-factor alpha of the S&P 500 comes overwhelmingly from the top market cap decile, which accounts for 73 bp out of the 81 bp alpha of the index. This is mostly due to the two extreme growth portfolios within the top size decile which have large positive four-factor Carhart

---

[11] We could also further decompose style alpha to timing alpha and average style alpha, similarly in spirit to Daniel, Grinblatt, Titman, and Wermers (1997). In unreported results, we found that timing has a slightly negative contribution to the returns on both the S&P 500 and Russell 2000.

alphas of 371 bp and 296 bp per year (Panel A in Table 5) and which contain about 35% of the value of the S&P 500 index.[12] The second part of Panel A indicates that stock selection by the S&P 500 within the 10x12 benchmark portfolios accounts for only 11 bp of its alpha. Hence, almost 90% of the S&P 500 alpha comes from its exposure to passive benchmark portfolios and not from any well-informed stock selection by the S&P index selection committee.[13]

The Russell 2000 (Panel B) exhibits some negative "stock selection" which amounts to 69 bp per year. 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 and could potentially be explained by a factor model. The remaining stock selection alpha comes almost entirely from the upper and lower boundaries of the index (size deciles 2 and 5-6, while deciles 3-4 show very little selection alpha.

4.5 Index Reconstitution

Index reconsitution effects present another possible explanation for the negative alpha of the small-cap indices (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.[14] Stocks being added to the Russell 2000 outperform those being deleted in June due to the anticipation of large index fund trading at the end of the month, and some of the excess performance 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 rebalancing 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

[12] 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. 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.

[13] 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 alpha of the marginal portfolio (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.

[14] 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.

Table 9, 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 two 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% 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 nonzero alphas are December (positive) and January (negative), consistent with the well-known January effect.

5 Selection of Alternative Factors

What would we then propose as better factors models that are not subject to the aforementioned issues? We try two different approaches. First, we modify the original Fama-French factors as discussed before: we restrict the market portfolio to U.S. stocks, value-weight the SMB factor, introduce separate value factors for different size groups, and modify the Small/Big and High/Low-BM cutoffs. Second, we introduce size and value factors based on common benchmark indices such as the S&P 500 and Russell 2000, as these indices are already value-weighted and not affected by the issues we discussed.

Panel A of Table 10 reports the time series correlations of the traditional Fama-French factors together with their alternative versions. The market factor with only U.S. stocks has an almost perfect correlation of 99.9% with the CRSP market index, in spite of the 23 bp difference in average return. The modified SMB factor with value weights and all U.S. stocks also has a very high correlation of 97.5% with the original SMB factor in spite of the considerable differences in portfolio weights; however, the modifications reduce its correlation with HML from -43.2% to -29.6%, which is desirable when the factors are used together in a model.

When SMB is split into a Mid-minus-Big (MMB) and Small-minus-Mid (SMM), the two new factors have a correlation of 56.4%. Splitting HML into BHML (deciles 9-10), MidHML (deciles 6-8), and SHML (deciles 1-5) also produces correlated factors, with correlations ranging from 69.8% between SHML and BHML to as high as 89.4% between SHML and MidHML.

Panel B of Table 10 reports the correlations between our index-based factors. To keep them as comparable as possible to the Fama-French factors, we maintain the long-short structure for the factor portfolios. Hence, our index-based version of the Carhart model includes the S&P500 as the market, Russell 2000 minus S&P 500 as the small-minus-big factor, Russell 3000 Value minus Russell 3000 Growth as the value factor, and the same momentum factor. A more comprehensive seven-factor model would split R2-S5 into R2-RM and RM-S5, thus distinguishing between smallcaps and midcaps, as well as dividing the value factor into S5V-S5G, RMV-RMG, and R2V-R2G (large, mid, and small-cap value factors), while keeping the same momentum factor to facilitate comparison with Carhart.

The index-based models have generally similar correlations to the modified Fama-French factors. The main exception is our value factor based on Russell 3000 (R3V-R3G), which uses value weights rather than the 50-50 weights between small and large stocks in the original HML factor. The original HML factor is more of a small-cap value factor, as its correlation is 89.7% with R2V-R2G and only 72.5% with S5V-S5G. In contrast, R3V-R3G has a similar correlation with all three value factors; in particular, its improved correlation with large-caps is helpful if we care about explaining portfolios with a large amount of market capitalization.

6 Explaining Common Variation in Returns

6.1 Methodology

A factor model should capture significant time-series variation in portfolio return. This is not only a necessary condition in Arbitrage Pricing Theory for a factor to be priced, but also useful for benchmarking purposes. Specifically, a benchmark that closely tracks a portfolio return over time also produces tight standard errors on alpha. Hence, regressing portfolio returns on a factor model in a time series, we should observe a high R^2 and small residual error volatility. We focus on the latter, commonly called tracking error volatility (or just “tracking error” for simplicity), because this quantity conveniently indicates the standard deviation of a money manager’s realized alpha.

We test this using a sample of U.S. all-equity mutual funds. This sample represents not only a large cross-section of portfolios, varying from small-cap to large-cap and from value to growth, but it also includes the kind of actual investment portfolios encountered in practical applications.

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, adding 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 6); and a seven-factor model with separate value factors for large-caps, mid-caps, and small-caps (MOD7, or column 6 in Table 6).

As for the index-based factors, 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. 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 (see Appendix A for more discussion of beta estimation). 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. Using out-of-sample returns brings the added benefit of penalizing a model for overfitting the data, so having too many factors will actually weaken the out-of-sample performance of a model.

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, which 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.

6.2 Results

Panel A of Table 11 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 120-140 bp 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. Panel C is discussed in Appendix A.

7 Explaining the Cross-Section of Average Returns

7.1 Cross-section of Mutual Fund Returns

7.1.1 Methodology

Having established that the proposed factors indeed capture significant common variation in returns, we proceed to test how well these factors explain the cross-section of average returns, again using all-equity mutual funds as test assets. 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 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, the 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.

7.1.2 Results

Table 12 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%.[15] 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.

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

Panels D and E in Table 12 report the alphas from pure index models. In contrast to the Carhart model, 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. 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.

7.2 Cross-Section of Stock Returns

7.2.1 Methodology

In this subsection, we investigate the cross-sectional pricing of stocks using three other sets of test assets: (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.[16] For cross-sectional pricing, 25 portfolios may be a very small sample, so our main analysis focuses on the set of 100 size-BM portfolios. To save space, only results for this set are reported (the other results are available upon request).

Our analysis consists of three parts. First, we add the common benchmark indices (plus the value and growth component of the Russell 3000) directly to the four-factor Carhart model.

[16] 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.

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.

Panel A of Table 13 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.

As our econometric approach, we use 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).[17]

7.2.2 Results

In Panel A of Table 13, 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

[17] 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.

significant jump in the cross-sectional R^2. The addition of RM-S5 in column 3 also 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 6, 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.[18] Further, the alternative construction of the Fama-French market, SMB and HML factors (not reported) makes no meaningful difference for cross-sectional pricing of these test portfolios.

In Panel B of Table 13, 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).

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 included in none of the

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

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 13).

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.

8 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 historically, 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 a relatively easy benchmark for small-cap value funds but a tougher 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. 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.

As alternatives to the well-known three and four-factor models, we test models with modified versions of the Fama-French factors as well as models based on the common benchmark indices. We analyze tracking error volatility across a broad cross-section of mutual funds to see which models best explain the common variation in returns and thus most closely track the time-series of fund returns. The index-based models produce the lowest out-of-sample tracking error, thus outperforming the traditional Fama-French and Carhart models.

When applied to the cross-section of average mutual fund returns, the index-based models explain average returns well, producing alphas close to zero for all fund groups. The Carhart model produces slightly larger alphas in general, but its biggest weakness is its sensitivity to a seemingly innocuous adjustment: when comparing small-cap and large-cap funds, adjusting for the benchmark index has a drastic 5% per year impact on their Carhart and Fama-French alphas, fully reversing the conclusions about skill between small and large-cap funds. The index-based models do not exhibit similar biases, as they do not produce significant nonzero alphas for large-cap stocks and small-cap stocks in general.

We also compare models in standard asset pricing tests for 10x10 size-and-book-to-market-sorted portfolios. Replacing SMB and HML with index-based factors increases the R^2 of a cross-sectional regression of portfolio returns on factor betas, indicating that the index models explain average returns better.

Overall, the results support the use of 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. Economizing 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 cross-sectional pricing tests with 90 or 100 size-BM Fama-French portfolios also indicate that the index-based seven-factor model performs best, and that the pure index-based four-factor model is an improvement over the Carhart model.

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* 52-1, 57-82.

Chan, L.K.C., S.G. Dimmock, and J. Lakonishok, 2006, "Benchmarking money manager performance: Issues and evidence," working paper.

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.

Fama, E.F. and K.R. French, 1993, “Common risk factors in the returns on stocks and bonds,” *Journal of Financial Economics* 33, 3-56.

Goetzmann, W.N., Z. Ivkovic, and K.G. Rouwenhorst, 2001, “Day Trading International Mutual Funds: Evidence and Policy Solutions,” *Journal of Financial and Quantitative Analysis* 36-3, 287-309.

Gruber, M.J., 1996, “Another Puzzle: The Growth in Actively Managed Mutual Funds,” *Journal of Finance* 51-3, 783-810.

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,” *Financial Management*, forthcoming.

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.

Hendricks, D., J. Patel, and R. Zeckhauser, 1993, “Hot hands in mutual funds: Short-run persistence of relative performance, 1974-1988,” *Journal of Finance* 48-1, 93-130.

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.

Sapp, T. and A. Tiwari, 2004, “Does stock return momentum explain the ‘smart money’ effect?” *Journal of Finance* 59-6, 2605-2622.

Sensoy, B., 2007, “Performance Evaluation and Self-Designated Benchmarks in the Mutual Fund Industry,” *Journal of Financial Economics*, forthcoming.

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.

Zheng, L., “Is money smart? A study of mutual fund investors’ fund selection ability,” *Journal of Finance* 54-3, 901-933.

Appendix A: Robustness of Beta Estimation

When estimating betas to compute the tracking error of a fund in Section 6, 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 11 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.[19]

---

[19] Results available upon request.

Appendix B: 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 U.S. equity mutual funds 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. 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	None	Gro	Med	Val	All
Big	7.8	42.6	25.5	11.1	86.9
Small	4.2	3.5	3.4	2.0	13.1
All	12.0	46.1	28.9	13.0	100.0

Average excess return per year	None	Gro	Med	Val	All
Big	5.92	7.61	8.62	9.20	7.72
Small	6.47	4.85	11.77	13.21	8.29
All	5.87	7.20	8.95	10.02	7.64

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

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

HML	None	Gro	Med	Val	All
Big	0.0	-50.0	0.0	50.0	0.0
Small	0.0	-50.0	0.0	50.0	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	None	Gro	Med	Val	All
Big	0.0	60.0	29.2	10.8	100.0
Small	0.0	0.0	0.0	0.0	0.0
All	0.0	60.0	29.2	10.8	100.0

Benchmark portfolio: 0.967 x MktRf - 0.318 x SMB - 0.086 x HML	None	Gro	Med	Val	All
Big	7.5	56.1	35.2	17.0	115.8
Small	4.1	-2.9	-7.3	-13.0	-19.1
All	11.6	53.2	27.9	4.0	96.7

Target portfolio: Size decile 4	None	Gro	Med	Val	All
Big	0.0	0.0	0.0	0.0	0.0
Small	0.0	40.7	40.5	18.7	100.0
All	0.0	40.7	40.5	18.7	100.0

Benchmark portfolio: 1.055 x MktRf + 0.799 x SMB + 0.226 x HML	None	Gro	Med	Val	All
Big	8.2	7.0	0.3	-3.7	11.8
Small	4.5	19.1	30.2	40.0	93.8
All	12.7	26.1	30.5	36.3	105.5

Table 5. 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 6, Panel A. 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 6, Panel B. 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 7. 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 6, 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 8. 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 9. Russell 2000 alphas in June and July.

In this table, the regression models (1), (4), and (8) from Table 6 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 10. 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 11. 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 12. 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 group	Value group 1	Value group 2	Value group 3	All	Size group	Value group 1	Value group 2	Value group 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 13. 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 13. (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)				

Appendix C: Figures

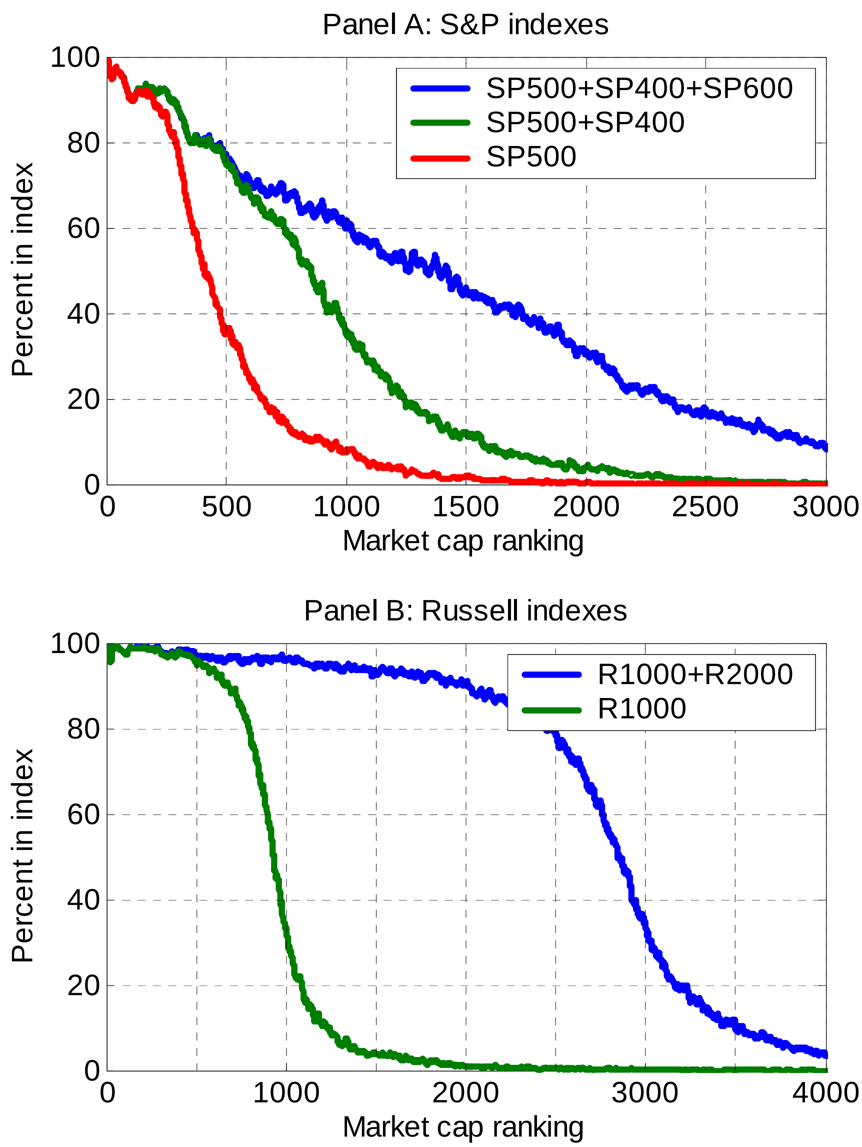

Figure 1. Index membership as a function of market capitalization.
All US stocks in CRSP are sorted each month based on their market cap. For each market cap rank, we include 10 stocks above and below and then compute the percentage of those 20 stocks that are index constituents that month. The figures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

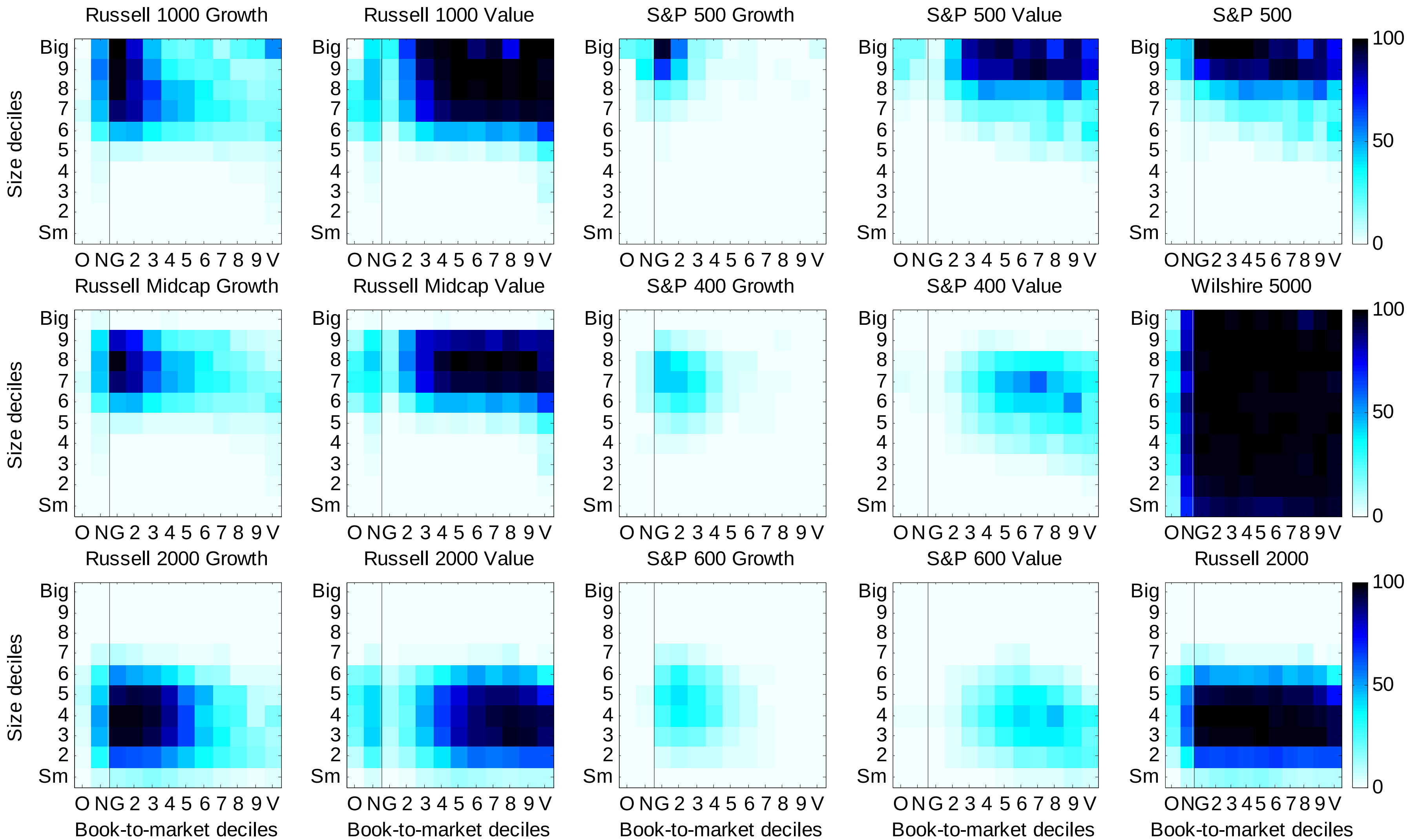

Figure 2. Index membership across size and value groups.

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.