

Are mutual fund managers selecting the right benchmark index?

Bruce A. Costa^a, Keith Jakob^{b,*}

^a*Department of Finance, School of Business Administration, The University of Montana, Missoula, MT 59812-6808, USA*

^b*Department of Finance, Donald and Carol Jean Byrnes Professor, School of Business Administration, The University of Montana, Missoula, MT 59812-6808, USA*

Abstract

In this paper, we determine index suitability for mutual funds that specify the S&P 500 as their performance benchmark. Using a four-factor model, we calculate factor loadings for mutual funds and their benchmark index and measure deviations with respect to the risk factors in the model. We sort the funds into growth and value subsets. By comparing each fund with the appropriate growth or value index, we show that the inferences regarding abnormal fund performance change dramatically. Our results indicate that fine tuning of abnormal performance measurement can significantly alter inferences regarding the fund manager's contribution to mutual fund performance. Using this methodology, investors and academics can more accurately assess mutual fund manager performance relative to an appropriate benchmark, regardless of the target index selected by the management of the mutual fund. © 2011 Academy of Financial Services. All rights reserved.

JEL classification: G20

Keywords: Performance; Mutual funds; Benchmark index

1. Introduction

Practitioners and academics are keenly interested in understanding investment portfolio performance. One area that is particularly important is how to measure accurately the performance of mutual fund managers. There are three key characteristics that influence

* Corresponding author. Tel.: +1-406-243-6159; fax: +1-406-243-2086.
E-mail address: keith.jakob@business.umt.edu (K. Jakob).

mutual fund returns. Asset allocation and security selection are the first two important aspects of fund management, and are the main drivers of overall fund performance. In addition, fund performance is related to fees and the level of trading activity within the fund.

It is relatively simple to measure the raw performance of a mutual fund. All mutual funds must report their raw or unadjusted performance and their self-selected benchmark index for various periods of time, or holding periods. Currently, mutual funds do not generally report risk-adjusted performance metrics. Risk-adjusted performance metrics are not required partly because of the fact that both practitioners and academics have struggled with how to accurately measure them. To evaluate risk-adjusted mutual fund performance, it is necessary to have some type of nominal performance benchmark. Most large capitalization domestic equity mutual funds use the Standard & Poor's 500 index as their performance benchmark. Funds with investment characteristics that do not match the S&P 500 index should use different performance benchmarks. Fortin and Michelson (1999, 2002) examine raw returns and after-tax-returns for a variety of actively managed mutual fund investment categories. They compare category mean returns with the returns of an appropriate index associated with each category. They report that most equity fund categories underperform the relevant benchmark index and similar index-fund investments. This prior line of literature points out the underperformance of mutual funds when compared with their benchmark index. It is important to note that the comparison of a mutual fund's raw returns against its benchmark index returns does not control for the differences in the inherent risks between the mutual fund and its related benchmark index portfolio. To create a risk-adjusted metric, further analysis incorporating risk characteristics of fund returns is therefore necessary.

Using Carhart's (1997) four-factor model standard methodology for risk-adjustment, no direct comparison is performed between a fund and its reported benchmark. In this paper we expand on the risk-adjusted performance measures developed in Costa and Jakob (2010). By calculating factor loadings for both the fund and the benchmark index, we observe whether the fund deviates from the benchmark index with respect to the four risk factors in the model. Based on these observations we select an alternative benchmark that is more aligned with the risk factors for the fund. By comparing fund performance directly with this alternative benchmark, we show that when accounting for risk in this manner, inferences regarding abnormal fund performance change dramatically.

The pioneering work in risk-adjusted mutual fund performance measurement is the seminal paper of Jensen (1968). Jensen uses a Capital Asset Pricing Model (CAPM) approach to measure mutual fund manager performance. Following this method, the regression intercept, alpha, is designed to capture the risk-adjusted net return of a mutual fund. Using his model, Jensen finds, on average, actively managed mutual funds have negative alphas. Jensen interprets a negative alpha as evidence of fund manager underperformance.

After the critique of the CAPM by Roll (1977), many investigators begin to examine multiple risk factor models to more accurately explain financial asset returns. Fama and French (1993) identify five common risk factors associated with the overall returns of stocks and bonds. Three of the five common risk factors are equity related. The Fama and French model has a market return measure, a firm size measure, and a book-to-market factor. Gruber (1996) uses a four-factor model, composed of the three Fama and French stock-market related factors and a bond factor to examine mutual fund manager performance. Similar to

Jensen (1968), Gruber finds using his four-factor model that a majority of alphas are negative. His findings again imply that the average actively managed fund has negative risk-adjusted manager performance.¹

Carhart (1997) presents a different four-factor risk adjusted performance model that uses the three stock-market related factors, as well as an additional factor to capture Jegadeesh and Titman's (1993) one-year momentum anomaly. Jegadeesh and Titman document strategies that buy stocks that have performed well in the past and sell stocks that performed poorly in the past generate significantly positive returns over 3 to 12-month holding periods. According to the Carhart model, an alpha similar to the alpha in Jensen (1968) and Gruber (1996) captures the risk-adjusted net return of the mutual fund. The Carhart four-factor model has been cited in over 175 academic and practitioner peer-reviewed journal articles and is one of the most widely used risk-adjusted performance metrics for mutual fund returns. The Carhart model is often used to simply calculate risk adjusted returns, but it can also be used in a wide variety of techniques to classify equity funds based on the model's factor loadings. Detzel (2006) compares two methods of equity fund classification; factor loadings and the Morningstar's style box. He finds that factor loadings do a better job predicting mutual fund returns from 1994 to 2004. Hereafter, we will refer to the Carhart model as the four-factor model.

Recent papers question the accuracy and/or the relevance of these commonly used alpha residual risk-adjusted models. Costa and Jakob (2006) use the four-factor model and generate alphas and factor loadings for a large set of stock market indexes. Indexes are essentially unmanaged portfolios; therefore, according to the prior literature, their alphas should be insignificant. Because unmanaged indexes by design do not follow any particular momentum strategy, loadings for the momentum factor should also, theoretically, be insignificant. Contrary to these expectations, Costa and Jakob report many significant alphas and significant factor loadings on the momentum factor. Additionally, the index construction method (i.e., market, or equally weighted) and orientation (value or growth) impact the sign and magnitude of the alphas and risk factor loadings. In a similar article, Cremers, Petajisto, and Zitzewitz (in press) also find the four-factor model's alpha is significant when they look at the returns of unmanaged indexes. Cremers et al., hypothesize these alphas primarily arise from the disproportionate weight the Fama-French factors place on small value stocks that have performed well and from biases in the CRSP value-weighted market index.

In a subsequent paper, Costa and Jakob (2010) introduce and demonstrate a new methodology to improve the use and interpretation of the four-factor model. To achieve a true measure of the manager's contribution to the fund's performance, the methodology is extended to include a statistical comparison of the fund's alpha with the alpha of its representative benchmark index. This new methodology uses the four-factor model to compute alphas for both the benchmark index and the mutual fund. Costa and Jakob then perform a pair-wise *F* test between the alpha of the benchmark and the alpha of the mutual fund in question. This additional test reveals whether the mutual fund's alpha is statistically different from the alpha of the benchmark index. Using this method, the fund's alpha is compared directly to the alpha of the relevant benchmark index. In the prior literature, the alpha generated by the four-factor model is not related in any way to the performance of the benchmark index; however, the standard interpretation of a positive and significant alpha is

that the manager outperformed his benchmark. With Costa and Jakob's methodology, it is possible to have a negative alpha and still significantly outperform the relevant benchmark index.

In this paper, we use the Costa and Jakob (2010) methodology to focus on fund management benchmark index selection. All mutual funds must report returns for a benchmark index along with their fund returns. Our goal is to use the Costa and Jakob methodology to determine whether a fund is using the appropriate benchmark. With the original four-factor model, no direct comparison was performed between the fund and its reported benchmark. By calculating factor loadings for both the fund and the benchmark, it is now possible to measure how much the fund has deviated from the benchmark index with respect to the four risk factors in the model. By calculating factor loadings for other indexes over the same holding period, we can test whether the fund's performance is more closely aligned with its self-selected benchmark or another index that more appropriately matches a fund's risk characteristics.

Beyond measuring fund manager risk-adjusted performance, investors and regulators are concerned whether stated mutual fund investment style matches actual asset allocation of the fund. According to the Director of the SEC's Division of Enforcement, "Advisers simply cannot tell investors they are going to do one thing with their funds and then not follow through on those promises."² Our methodology examines a fund's investment style and asset allocation using a novel approach. By running a "horse race" directly with a large set of indexes including a fund's self-selected benchmark index, it is possible to measure how much the fund has deviated from its designated index and whether a different benchmark index is more appropriate with respect to the various risk factors in the four-factor model. Using the pair-wise F test, the alphas and the coefficients of the four risk factors can now be statistically compared to ascertain if there is a significant statistical difference between the performances of the underlying benchmark index and the mutual fund. Coefficients of each risk factor can be statistically compared with the corresponding factor loadings of the self-selected index and any other benchmark index. With this methodology, we find funds that deviate significantly from their benchmark not only in performance, but in asset allocation and risk characteristics.

The remainder of the paper is organized as follows. Section 2 presents our data, Section 3 describes our methodology, Section 4 discusses our results, and Section 5 summarizes our findings.

2. Data

Our goal is to utilize the described methodology on a subset of funds that has a well-defined benchmark, as well as obvious performance characteristics on a raw return basis. For this reason we look at top performing large-cap funds that also use the S&P 500 as their sole benchmark. We compare top performing large cap funds against the S&P 500 index, but the regression techniques in our study can be used with any fund against any benchmark index. Our paper presents additional regression results for a wide set of 18 benchmarks that can be used for comparisons of alpha and the four risk-factor loadings. We

start our data collection process by examining the *Lipper/The Wall Street Journal* top 300 large capitalization mutual funds with the best five-year performance as of December 31, 2009. The goal of the original Carhart model as well as our current methodology is to look for statistically significant under- or overperformance of individual mutual funds. To determine the abnormal performance of a fund with this type of regression analysis, one must use a minimum number of monthly return values (typically 36 or three years of data). We use 60 months of returns for each fund, which dramatically improves the overall significance of our results.

From the initial dataset, we eliminate any passively managed index funds. Only Class A shares are included when there are duplicate funds differing by fee structure within fund families. We also eliminate any “funds of funds” and any funds that are missing data during the 60-month study period. We examine the prospectus of each of the 114 remaining funds to determine what index or indexes are specified as benchmarks for the fund. Seventy mutual funds list the S&P 500 as their sole benchmark index. This group of 70 mutual funds is our final data sample set. These 70 funds actually represent many more funds with different ticker symbols that are essentially the same fund with slightly different fee structures. For example ING Core Equity Research A (AAGIX), one of the funds in our sample, has five other ticker symbols for its other fund classes. The 70 funds in the final sample had total net assets as of December 31, 2009 of \$75.2 billion. Our next step is to collect monthly returns for these 70 mutual funds, the S&P 500, and a set of alternative benchmark indexes.

Our mutual fund monthly return data comes from the CRSP Survivor-Bias Free U.S. Mutual Fund Database. Carhart developed this unique database for his dissertation entitled “Survivor Bias and Persistence in Mutual Fund Performance.”³ We compile monthly return data from January, 2005, through December, 2009. For the same period we gather monthly return data for the S&P 500 directly from Standard and Poor’s. Russell Investments and Standard and Poor’s provide a large variety of popular benchmark indexes. From these two sources, we collect 60 months of return data for various alternative benchmarks to the S&P 500. Finally, we use the U.S. Research Returns data obtained directly from Kenneth French’s Website for the values of the monthly risk factors for the four-factor model.⁴

3. Methodology

To quantitatively measure manager performance, we begin with the Carhart (1997) four-factor regression method for estimating mutual fund and benchmark index performance. The four-factor model uses Fama and French’s (1993) three-factor model with an additional factor to capture Jegadeesh and Titman’s (1993) one-year momentum anomaly. With this regression method, an alpha, similar to the alpha in Jensen (1968), is designed to capture the risk-adjusted net return of a mutual fund.

For our initial risk-adjustment analysis we use the following four-factor model:

$$r_i = \alpha_i + \beta_{1i}RMRF_t + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}PR1YR_t + \varepsilon_i \quad (1)$$

where r_i is the monthly mutual fund return or benchmark index return minus the 30-day treasury bill return; RMRF is the excess return on the value-weighted CRSP index including

all NYSE, AMEX, and Nasdaq stocks; and SMB, HML, and PR1YR are returns on value-weighted zero-investment, factor-mimicking portfolios for size, book-to-market equity, and one-year momentum in stock returns, respectively. We run the four-factor model regression for the 70 mutual funds in our sample, as well as the S&P 500 and the other alternative indexes.

Running the four-factor model generates alphas and coefficients for the four risk factors. The first key additional step that we use is to compare, statistically, the alpha and coefficients from each fund with the alpha and coefficients from the S&P 500 using a pair-wise F test. Using the S&P 500 data, the F test comparing alphas indicates whether the manager statistically outperformed or underperformed his self-selected benchmark index over the 60-month holding period. By examining the regression results for the fund and the S&P 500 side-by-side, we can determine whether the factor loadings for the various risk factors statistically differ between the fund and their benchmark. By performing the alpha comparison analysis with other alternative indexes, the F test indicates whether the manager statistically outperformed or underperformed an alternative benchmark index over the 60-month holding period. As an initial screening to better understand how benchmark index choice impacts risk adjusted return, we compare each fund's returns with two straightforward benchmark alternatives. Because our sample includes funds that focus solely on the S&P 500 as a benchmark, we use the returns from the S&P 500 growth and the S&P 500 value indexes as basic alternative benchmarks. This analysis gives us an indication of whether the full S&P 500 or a related value or growth benchmark serves as a more appropriate measure of style and expected performance.

With the standard four-factor model methodology, the loadings on the four risk factors, RMRF, SMB, HML, and PR1YR, indicate how much of the fund's returns are derived from each measure of risk. The alpha is the portion of the fund's returns that cannot be explained by the four risk factors in the model. In the standard four-factor model the alpha is the measure of risk-adjusted return for a fund. This dissection of fund returns into four unique risk characteristics is an important feature of the four-factor model, but it does not address the most common concern of investors and regulators regarding portfolio asset allocation and stated fund objectives.

Fund managers are required to specify a relevant benchmark and clearly describe their overall investment strategies in their prospectus. To determine whether the manager is following through on his stated objectives, the alpha and the factor loadings based on the fund's returns must be compared directly to the alpha and the factor loadings for the self-selected benchmark index as well as alternative indexes. Comparing the fund with the self-selected benchmark, the S&P 500, the pair-wise F test on alpha indicates whether the manager significantly under- or out-performed their specified benchmark index on a risk-adjusted basis. The pair-wise F -tests for the risk factor loadings on RMRF, SMB, HML, or PR1YR indicate whether the manager has significantly deviated from the fund's stated objectives and specified benchmark with regard to each risk factor. If we find large significant differences between the factor loadings for the S&P 500 and a particular fund, this suggests that the fund management has potentially not chosen the most appropriate benchmark index for comparison purposes in the prospectus. For such funds, additional compar-

Table 1 Descriptive statistics for raw returns of the S&P 500 Index and 70 large cap growth and income mutual funds from January 2005 through December 2009

Number of funds	70
Mean 60-month cumulative return	10.62%
Median 60-month cumulative return	8.00%
Minimum 60-month cumulative return	1.07%
Maximum 60-month cumulative return	46.91%
Standard deviation 60-month cumulative return	8.81%
S&P 500 60-month cumulative return	2.11%
Average Fund excess performance	8.51%
Number of funds that outperformed S&P 500	65
Number of funds that underperformed S&P 500	5

Cumulative returns for the five-year holding period for 70 funds and the S&P 500 from January 2005 through December 2009. All 70 funds specify the S&P 500 as their sole benchmark index in their prospectus. Excess return is the difference between the 60 month cumulative return for the funds and the S&P 500.

isons of factor loadings with the factor loadings for alternative indexes could help determine a more appropriate benchmark index.

3. Results

As a first step to understand fund manager performance, it is customary to look at raw returns relative to an appropriate benchmark index. For this study, we examine 70 Large Cap Domestic mutual funds against their self-selected benchmark index, the S&P 500. Table 1 presents the raw returns for our mutual fund sample and the S&P 500 for the 60-month period from January, 2005 through December, 2009. The mean and median 60-month total returns for the fund sample are 10.62% and 8.00%, respectively. On an annualized basis, the mean fund return is equivalent to a 2.04% annual return. The S&P 500 for the same period had a 60-month total return of 2.11% or an annual return of 0.42%. It is important to note that the returns used for the mutual funds are after fund expenses. For the entire 60-month holding period, the funds on average outperformed the S&P 500 by 8.51% or 1.65% annually. Of the 70 funds in the sample, 65 outperformed the S&P 500 on a raw return basis over the 60-month holding period. Based on these measures 93% of the funds would report that they beat their benchmark over the 60-month holding period. The fact that the vast majority of our sample mutual funds beat the S&P 500 on a raw returns basis for the holding period is not a surprise, because our sample is a subset of the top performing 300 large cap mutual funds according to *Lipper* and *The Wall Street Journal*.

The initial analysis examines raw returns for the 70 mutual funds in our sample. These fund returns are not adjusted for the inherent risk within each mutual fund portfolio. The assumption behind raw return comparisons is that the risk of the fund and the risk of the relevant benchmark are identical. Mutual fund prospectuses must report raw fund performance against their specified benchmark index. To test whether the fund manager is reporting returns of an appropriate benchmark along with their fund returns, some type of risk characteristic assessment or adjustment is necessary. To quantitatively measure manager

Table 2 Standard four-factor model summary results

	Alpha (intercept)	RMRF	SMB	HML	PR1YR
Mean	.000149	.903779	-.09852	.034195	.004165
Median	-.00023	.91554	-.09804	.05135	.01251
Min	-.00186	.62654	-.39894	-.39487	-.33298
Max	.00421	1.13744	.11299	.4164	.25678
Observations positive	30	70	13	46	44
Observations positive and significant	4	70	0	21	27
Observations negative	40	0	57	24	26
Observations negative and significant	2	0	41	8	14

Monthly excess, risk-adjusted return coefficients, based on the four-factor model for 70 large cap funds from January 2005 through December 2009. Where r_i is the monthly return for the mutual fund minus the one-month treasury bill return, $RMRF$ is the excess return on the value-weighted CRSP index including all NYSE, AMEX, and Nasdaq stocks; and SMB , HML , and $PR1YR$ are returns on value-weighted zero-investment, factor-mimicking portfolios for size, book-to-market equity, and one-year momentum in stock returns, respectively. Significance is defined as 10% or better.

$$r_i = \alpha_i + \beta_{1i}RMRF_t + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}PR1YR_t + \varepsilon_i$$

risk-adjusted performance, we begin with the four-factor model. Using this regression approach, the alpha captures the risk-adjusted net return of a mutual fund.

Table 2 presents the regression results for the standard four-factor model. The mean alpha of 0.000149 indicates, after adjusting for risk, that the average manager's contribution to fund performance is 0.0149% per month, or about 0.2% per year. Based on these results, the previously accepted interpretation of the four-factor model suggests that the 70 Large Cap fund managers did not, on average, add an economically significant contribution to the performance of their funds. Of the 70 funds, in our sample, only 30 funds have positive alphas, and only four of the positive alphas are significantly different from zero. There are 40 funds with negative alphas, and two are significantly different from zero. The typical interpretation of the four-factor risk-adjusted performance methodology for the 70 funds in our sample suggests that only four fund managers significantly added value to their managed fund and two fund managers significantly underperformed.

Table 3 presents the alpha and four-factor coefficients for the S&P 500 and other popular benchmark indexes for our 60-month holding period. This table presents the customary indexes (referred to as blends) used in the mutual fund marketplace, as well as their growth and value orientation subsets. For the S&P 500 blend, also referred to simply as the S&P 500, the alpha and three of the four coefficients for the risk factors are significantly different from zero. The negative (-0.00113) and significant ($t = -1.78$) alpha for the S&P 500 in Table 3 is precisely why further analysis is warranted. According to prior literature, an alpha that is significantly different from zero indicates that the manager's contribution to the fund's overall performance was significant. In this case, as in Costa and Jakob (2006, 2010) and Cremers, Petajisto and Zitzewitz (in press), we find a significant alpha for an unmanaged index. Because the four-factor model finds significant alpha and factor loadings for the S&P 500 index, we must reinterpret the outputs for our sample funds to deal with this benchmark performance interpretation issue.

To determine whether the mutual funds in our sample are using the proper benchmark

Table 3 Standard four-factor model results for indexes

		alpha (intercept)	RMRF	SMB	HML	PR1YR
Russell 200 blend	Parameter estimate	-.00084	.91162	-.22658	.04764	.00417
	<i>t</i> -value	-1.56	42.68	-5.92	1.38	.25
Russell 200 growth	Parameter estimate	.00015	.93826	-.11282	-.29226	-.00480
	<i>t</i> -value	.13	33.01	-2.22	-6.36	-.22
Russell 200 value	Parameter estimate	-.00268	.88582	-.34427	.39059	.01147
	<i>t</i> -value	-1.93	25.20	-5.47	6.87	.43
S&P 500 blend	Parameter estimate	-.00113	.94024	-.13328	.05290	-.00032
	<i>t</i> -value	-1.78	58.23	-4.61	2.03	-.03
S&P 500 growth	Parameter estimate	.00082	.97757	.05451	-.10517	-.08387
	<i>t</i> -value	.43	20.01	.62	-1.33	-2.24
S&P 500 value	Parameter estimate	-.00243	.89753	-.18333	.39525	-.04094
	<i>t</i> -value	-1.64	29.91	-2.73	6.51	-1.42
Russell 1000 blend	Parameter estimate	-.00079	.96013	-.08566	.02624	-.00193
	<i>t</i> -value	-1.61	77.09	-3.84	1.30	-.20
Russell 1000 growth	Parameter estimate	.00364	.99806	.00163	-.30311	-.00394
	<i>t</i> -value	.44	47.93	.04	-9.00	-.25
Russell 1000 value	Parameter estimate	-.00201	.91984	-.17525	.36013	-.00492
	<i>t</i> -value	-2.04	36.68	-3.90	8.88	-.26
Russell mid-cap blend	Parameter estimate	.00072	1.07965	.27898	-.02584	-.02468
	<i>t</i> -value	.058	34.50	4.98	-.051	-1.03
Russell mid-cap growth	Parameter estimate	.00124	1.14602	.30918	-.31991	-.00505
	<i>t</i> -value	.82	29.89	4.50	-5.16	-.17
Russell mid-cap value	Parameter estimate	-.00015	1.00298	.26158	.27913	-.05397
	<i>t</i> -value	-.11	30.33	4.42	5.22	-2.13
Russell 2500 blend	Parameter estimate	-.00034	1.02032	.63337	.05233	-.00271
	<i>t</i> -value	-.046	54.40	18.86	1.73	-.19
Russell 2500 growth	Parameter estimate	.00064	1.1197	.68188	-.25852	.02269
	<i>t</i> -value	.64	43.99	14.96	-6.28	1.16
Russell 2500 value	Parameter estimate	-.00149	.91372	.59373	.35468	-.03667
	<i>t</i> -value	-1.36	32.73	11.88	7.86	-1.71
Russell 2000 blend	Parameter estimate	-.00142	.98732	.88439	.15308	.02802
	<i>t</i> -value	-1.77	48.48	24.25	4.65	1.80
Russell 2000 growth	Parameter estimate	-.00047	1.09043	.94306	-.18292	.05545
	<i>t</i> -value	-.54	48.89	23.61	-5.07	3.25
Russell 2000 value	Parameter estimate	-.00237	.87948	.82956	.48020	-.00660
	<i>t</i> -value	-1.67	24.37	12.84	8.23	-.24

Monthly excess, risk-adjusted return coefficients and *t*-values based on the four-factor model for the S&P 500 index and other popular benchmark indexes from January 2005 through December 2009. Where r_i is the monthly return for the index minus the one-month treasury bill return, *RMRF* is the excess return on the value-weighted CRSP index including all NYSE, AMEX, and Nasdaq stocks; and *SMB*, *HML*, and *PR1YR* are returns on value-weighted zero-investment, factor-mimicking portfolios for size, book-to-market equity, and one-year momentum in stock returns.

$$r_i = \alpha_i + \beta_{1i}RMRF_t + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}PR1YR_t + \varepsilon_i$$

index to gauge their performance, we examine the four-factor model coefficients for the 18 different indexes in Table 3. The indexes are listed in descending order based on the market capitalization or size of the component firms within the index. The Russell 200 is a benchmark index for the two hundred largest market capitalization United States based firms. The Russell 200 is a more concentrated index than the S&P 500, and

Table 4 SMB factor loadings based on index size and style

Style	Large cap			Mid cap	Small cap	
	Russell 200	S&P 500	Russell 1000	Russell mid cap	Russell 2500	Russell 2000
Growth	-0.11282	0.05451	0.00163	0.30918	0.68188	0.94306
Blend	-0.22658	-0.13328	-0.08566	0.27898	0.63337	0.88439
Value	-0.34427	-0.18333	-0.17525	0.26158	0.59373	0.82956

The SMB coefficient based on the four-factor model for 18 separate indexes from January 2005 through December 2009. The indexes are positioned in the table based on size and style. Where r_i is the monthly return for the index minus the one-month treasury bill return, $RMRF$ is the excess return on the value-weighted CRSP index including all NYSE, AMEX, and Nasdaq stocks; and SMB , HML , and $PR1YR$ are returns on value-weighted zero-investment, factor-mimicking portfolios for size, book-to-market equity, and one-year momentum in stock returns.

$$r_i = \alpha_i + \beta_{1i}RMRF_t + \beta_{2i}SMB_t + \beta_{3i}HML_t + \beta_{4i}PR1YR_t + \varepsilon_i$$

represents approximately two-thirds of the total market capitalization of all United States listed stocks. At the other end of the spectrum, the Russell 2000 is a small-cap stock market index that is used as the most common benchmark for “small-cap” funds. To proceed we must determine whether any of the four factor coefficients has a clear relationship with index size or index growth/value orientation. After looking at all four factors, the SMB coefficient has the clearest relationship to these metrics. This clear relationship exists, because the indexes themselves are constructed based on market capitalization of the included stocks. To help visualize these relationships, Table 4 presents the SMB coefficients separately across the 18 different indexes. Within Table 4, there are two evident trends. First, there is an inverse relationship between firm market capitalization within the index and the SMB coefficient. And, second, within each index growth-blend-value group, the SMB coefficient decreases as you move from the growth subset to the blend and finally to the value subset. This second relationship indicates that the growth subset for each major index contains firms with smaller market capitalizations than the value subset.

Table 5, Panel A summarizes the findings regarding alpha of the 70 funds in the sample versus the alpha of the S&P 500. When the S&P 500 and the funds are both run through the four-factor model, there are 15 fund alphas that are significantly greater than the alpha of the S&P 500 and one fund alpha that is significantly less than the alpha of the S&P 500. By statistically comparing fund alphas directly with the S&P 500 benchmark alpha, the number of funds and level of significant abnormal performance changes dramatically from the standard four-factor model results (see Table 2).

Based on the SMB relations shown in Table 4, we divide our sample into two subsets: growth and value. Relative to the S&P 500, the funds in the growth (value) oriented subset have an SMB coefficient greater (less) than the SMB coefficient for the S&P 500. In Table 5 Panel B, 52 of the 70 funds have SMB coefficients greater than the SMB coefficient for the S&P 500; 19 of these fund SMB coefficients are significantly greater than the SMB coefficient for the S&P 500. These findings indicate that a large majority of the funds were more growth oriented in style than their stated benchmark, the S&P 500. The remaining

Table 5 Summary comparison of alpha and SMB coefficients for 70 large cap funds and the S&P 500 Index

Panel A	Alpha (intercept)
Observations > S&P 500	65
Observations > S&P 500 and significant	15
Observations < S&P 500	5
Observations < S&P 500 and significant	1
Panel B	SMB
Observations > S&P 500	52
Observations > S&P 500 and significant	19
Observations < S&P 500	18
Observations < S&P 500 and significant	8

Alpha and SMB monthly excess, risk-adjusted return coefficients, based on the four-factor model for 70 large cap funds from January 2005 through December 2009. Panel A presents the direction and significance of fund alphas versus the alpha of the S&P 500. Panel B presents the direction and significance of fund SMB coefficients versus the SMB coefficient of the S&P 500. Significance is defined as 10% or better.

eighteen funds have SMB coefficients that are less than the SMB coefficient for the S&P 500. Eight of these funds have an SMB coefficient that is significantly less than the SMB coefficient for the S&P 500. Our results with respect to the SMB coefficient suggest that many of the funds within the sample are specifying a benchmark that is not accurately aligned with their actual orientation or style.

To examine fund performance relative to alternative benchmarks, we compare each fund against the S&P 500 growth or value index based on the magnitude of the fund's SMB coefficient. Table 6 presents the raw returns for the growth and value subsets. The 52 funds with the largest SMB coefficients (greater than the S&P 500 SMB coefficient) are directly compared with the S&P 500 growth index. The 18 funds with the smallest SMB coefficients are directly compared with the S&P 500 value index. On a raw return basis, 49 of the 52 growth oriented funds have higher returns than the S&P 500 growth index. On a raw return basis, all 18 of the value oriented funds have higher returns than the S&P 500 value index. Fund managers often use raw return as their sole metric of performance to compare against their specified benchmark index. The raw return measure fails to adjust for the risk of the fund.

To properly measure performance against an appropriate benchmark, we examine the risk adjusted performance for the growth and value subsets against the S&P 500 growth and value indexes. By running the funds and these alternative benchmarks through the four-factor model, we statistically compare the 52 growth (18 value) oriented funds directly with the S&P 500 growth (value) index. Table 7 summarizes the results for these new tests, as well as the two prior benchmark adjusted performance tests using the four-factor model. The stand-alone four-factor model interpretation assumes that the manager is making a significantly positive (negative) contribution to fund performance when alpha is positive (negative) and significantly different from zero. According to this model, and as shown in Table 7, for the 52 growth oriented funds, three managers add value and one manager has a negative abnormal performance. For the 18 value oriented

Table 6 Raw returns for growth and value subsets

Number of growth oriented funds	52
Mean 60-month cumulative return	10.48%
Median fund return	7.95%
Minimum fund return	1.67%
Maximum fund return	46.91%
Standard deviation 60-month return	8.96%
S&P 500 growth 60-month return	2.11%
Average Fund excess performance	8.37%
Number of funds that outperformed S&P 500 growth	49
Number of value oriented funds	18
Mean 60-month cumulative return	10.61%
Median fund return	8.03%
Minimum fund return	1.07%
Maximum fund return	33.38%
Standard deviation 60-month return	8.70%
S&P 500 value 60-month return	−3.33%
Average fund excess performance	13.94%
Number of funds that outperformed S&P 500 value	18

The cumulative returns for the five-year holding period for the 52 growth oriented funds and the 18 value oriented funds as well as the S&P 500 value and growth indexes from January 2005 through December 2009. Funds are divided into growth and value subsets based on their SMB coefficient. Growth (value) funds have an SMB coefficient greater (less) than the SMB coefficient for the S&P 500. All 70 funds specified the S&P 500 as their sole benchmark index in their prospectus.

funds, one manager has positive and one manager has negative abnormal performance. It is important to remember that the original four-factor model does not include any comparison with the fund specified benchmark.

When the funds are run against their specified benchmark, the S&P 500, within the growth oriented subset eight of the managers had a positive contribution and one manager had a negative contribution. Within the value subset, seven of the fund managers had a positive contribution. The changes noted in manager contribution are likely because of the fact that the S&P 500 for the 60-month holding period had a significant negative alpha relative to zero. This revised benchmark test suggests that more managers are making a significant contribution to fund performance once their specified benchmark is included in the analysis. However, when we now compare the growth and value subsets with the alphas of the alternative S&P 500 growth or value indexes, the results change again. None of the 52 managers from the growth oriented subset make a significant contribution to performance when the S&P 500 growth oriented index is used as their benchmark. For the 18 value oriented funds, the number of managers with a significant positive contribution to return drops from seven to six when the S&P 500 value index is used as the benchmark.

Academics and practitioners are interested in properly gauging the manager contribution for mutual fund performance. We believe that to ideally measure manager contribution, the correct benchmark must be used in the analysis. By examining the SMB coefficients from the four-factor model, it is possible to find an appropriate benchmark for a particular fund. Table 7 clearly shows that the choice of benchmark can dramatically impact the measurement of, and conclusions made with regards to fund manager contribution to returns.

Table 7 Pair-wise *F*-test results for funds vs. the S&P 500 growth or S&P 500 value index

Orientation based on SMB	Observations	4-Factor Model significant alphas	Significant alpha vs. S&P 500 _{Blend}	Significant alpha vs. S&P 500 _{Growth}	Significant alpha vs. S&P 500 _{Value}
$SMB_{fund} > SMB_{S\&P\ 500}$ (growth)	52 Funds	3 Positive	8 Greater than	0 Greater than	
$SMB_{fund} > SMB_{S\&P\ 500}$ (value)	18 Funds	1 Negative 1 Positive	1 Less than 7 Greater than	0 Less than	n/a 6 Greater than
		1 Negative	0 Less than	n/a	0 Less than

Funds are divided into growth and value subsets based on their SMB coefficient. Growth (value) funds have an SMB coefficient greater (less) than the SMB coefficient for the S&P 500. All 70 funds specified the S&P 500 as their sole benchmark index in their prospectus. Significant alphas based on the standard four-factor model are reported for all funds. Using the *F* test methodology all fund alphas are compared with the alphas for the S&P 500 index. In the last two columns growth (value) fund alphas are also compared with the S&P 500 growth (value) index. Significance is defined as 10% or better.

4. Conclusion

Of the 70 funds in our sample, 65 outperformed the S&P 500 on a raw return basis over the 60-month holding period. Based on these measures 93% of the funds would report that they beat their benchmark over the 60-month holding period. As an initial metric of fund manager risk-adjusted performance, we use the Carhart (1997) four-factor model. According to prior literature, the alpha in this regression model captures the risk-adjusted net return of a mutual fund. Costa and Jakob (2010) add an additional step that statistically compares the alphas and coefficients of a mutual fund with the alpha and coefficients from the self-selected benchmark index using a pair-wise *F* test. When comparing alphas, the *F* test indicates whether the manager statistically outperformed or underperformed the relevant benchmark index over the specified holding period.

In this paper, we expand on the Costa and Jakob (2010) methodology by determining whether a fund is using the appropriate benchmark. With the stand-alone four-factor model, no direct comparison is made between the fund and its reported benchmark. We calculate factor loadings for both the fund and the benchmark and measure whether the fund has deviated from the benchmark index with respect to the four risk factors in the model. We also calculate factor loadings on a set of 17 alternative benchmark indexes. We find an inverse relationship between firm market capitalization within the index and the SMB coefficient. Also, within each index growth-blend-value group, the SMB coefficient decreases as we move from the growth subset to the blend and to the value subset.

With this in mind, we sort our 70 top performing large cap funds into growth and value subsets based on the magnitude of their SMB coefficient. By comparing each fund with the relevant growth or value alternative index, we show that the inferences regarding abnormal fund performance change dramatically. For our sample of 70 mutual funds, 52 funds have a

more growth oriented style than their specified benchmark based on the SMB coefficient. The remaining 18 funds appear more value oriented than the S&P 500. Using the standard four-factor model, four funds exhibit positive abnormal performance and two funds have negative abnormal performance. For the holding period, the benchmark index, the S&P 500, has a significantly negative alpha. When the alphas for each fund are statistically compared with the alpha of their self-selected benchmark index, we find 15 funds that significantly outperform their stated benchmark. One fund significantly underperforms its benchmark. When funds with large SMB coefficients are compared with an alternative benchmark, the S&P 500 growth index, all significant abnormal performance disappears. Six of the funds with small SMB coefficients outperform the S&P 500 value alternative benchmark index. When using appropriate benchmarks and adjusting for risk, less than 9% of these top *Lipper/The Wall Street Journal* funds have positive and significant abnormal performance.

Our results indicate that fine tuning of abnormal performance measurement can significantly alter inferences regarding fund manager contribution. With our methodology, investors and academics can more accurately assess manager performance relative to an appropriate benchmark regardless of the benchmark index selected by fund management. Almost all mutual funds only report raw returns, and with this basic information individual investors and financial planners cannot directly infer if a fund's performance is because of a manager's stock selection skill or the risk characteristics of the portfolio. With the methodology described in this paper, individual investors can now accurately measure the risk characteristics and statistically significant performance of their mutual funds and their specified benchmark index as well as alternative indexes. Individual investors and financial planners can verify that mutual fund managers are truly investing in portfolios with risk characteristics similar to their stated benchmarks. If the coefficients on the four risk factors are significantly different from those of the benchmark, individual investors can assess performance against more appropriate benchmarks based on the magnitude of the SMB factor. In a best case scenario regulating bodies would require funds to report risk adjusted performance and the risk characteristics of the fund returns and benchmark indexes. With this information investors and financial planners could make straightforward and informed decisions about how to invest their money.

Notes

1. Our tests do not include a bond factor because we focus exclusively on equity mutual funds. Fama and French (1993) note that bond-related factors are only important in capturing the returns for bond funds and add no explanatory power to the model when measuring the performance of equity.
2. "SEC Charges Mutual Fund Manager for Violating Socially Responsible Investing Restrictions" U.S. SEC Press Release 2008–157. July 30, 2008.
3. The data is obtained from the Center for Research in Security Prices (CRSP) from the University of Chicago's Graduate School of Business. The data base is a unique because it does not eliminate mutual funds that disappear via merger, liquidation, or

other means. The records for these disappearing funds remain in the electronic database as dead funds.

4. For more detailed information about the Fama/French factors see http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Acknowledgment

The authors appreciate the suggestions and help from Patrick Barkey, Ron Premuroso, and Barbara Reider. This work was supported, in part, by a School of Business Administration Research Grant and a Professorship from Donald and Carol Jean Byrnes.

References

- Carhart, M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52, 57–86.
- Costa, B., & Jakob, K. (2006). Do Stock indexes have abnormal performance? *The Journal of Performance Measurement*, 11, 8–22.
- Costa, B., & Jakob, K. (2010). Enhanced performance measurement of mutual funds: Running the benchmark through the hurdles. *Journal of Applied Finance*, 20, 95–102.
- Cremers, K.J., Petajisto, A., & Zitzewitz, E. Should benchmark indices have alpha? *Revisiting Performance Evaluation*, SSRN Working Paper.
- Detzel, F. L. (2006). Determining a mutual fund's equity class. *Financial Services Review*, 15, 199–212.
- Fama, E., & French, K. (1993). Common risk factors in the returns on stocks and bonds. *The Journal of Financial Economics*, 33, 3–55.
- Fortin, R., & Michelson, S. (1999). Mutual fund indexing versus active management. *Journal of Financial Planning*, 12, 74–81.
- Fortin, R., & Michelson, S. (2002). Indexing versus active mutual fund management. *Journal of Financial Planning*, 15, 82–94.
- Gruber, M. (1996). Another puzzle: The growth in actively managed mutual funds. *The Journal of Finance*, 51, 783–810.
- Jegadeesh, N., & Titman S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48, 65–91.
- Jensen, M. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, 23, 389–416.