Financial Analysts Journal
Volume 63 • Number 3
©2007, CFA Institute

Sector, Style, Region: Explaining Stock Allocation Performance

Raman Vardharaj and Frank J. Fabozzi, CFA

The importance of asset allocation policy in stock/bond portfolios is widely recognized. Drawing a parallel for equity-only portfolios, this study analyzed the importance of allocation by economic sector and by size and style in purely U.S. stock portfolios and the importance of regional allocation policy in international stock portfolios. The study found that allocation policy explains one-third to nearly three-quarters of among-fund variation in returns, nearly 90 percent of across-time variation, and more than 100 percent of the level of stock portfolio returns.

Several studies have investigated the importance of the asset allocation decision in balanced stock/bond portfolios. Among the first are two studies known as the "Brinson studies" (Brinson, Hood, and Beebower 1986; Brinson, Singer, and Beebower 1991). They reported that asset allocation explains more than 90 percent of the variation in a typical portfolio's performance.

Ibbotson and Kaplan (2000), noting that the results reported in the Brinson studies have been widely misunderstood, demonstrated that although more than 90 percent of return variability in a typical managed portfolio over time is explained by policy benchmarks, only about 40 percent of the variability *among* managed portfolios is explained by policy differences.[1]

Ibbotson and Kaplan (2000) also showed that the long-term "policy" return of a portfolio (that portion of the total return that comes from the asset allocation policy) slightly exceeds the "actual" return (a combination of the policy return and active return). Thus, the "policy portfolio" accounts for more than 100 percent of the return of a managed portfolio.

Being unaware of any similar studies that examined the importance of the allocation decision within a stock portfolio rather than among broad asset classes, we set out to explore such territory. One type of within-stock allocation would be to economic sectors—for example, consumer discretionary, health care, industrials. Active stock portfolios may attempt to generate alpha by differing (strategically and tactically) from their benchmarks in their sector allocations and by selecting securities (i.e., weighting them differently from the benchmark) within these sectors. One question we sought to answer, then, is how much of performance is explained by the long-term sector allocation decision.

Another type of within-stock allocation would be to market segments differentiated by size and style (in the sense of a value orientation versus a growth orientation). For example, based on Fama and French (1996) and other related findings, Kenneth French maintains a website that segments the U.S. stock market into two size categories (large and small) and three style categories (value, neutral, and growth), resulting in six size/style market segments.[2] Active fund managers may choose long-term policy allocations to these six market segments and make occasional tactical deviations from the allocations, and they may also actively choose securities within the segments. We empirically investigated the importance of the long-term allocation decision among these size/style segments.

Finally, for managers of international stock funds, one of the key allocation decisions is among regions of the world (e.g., Europe, Pacific Rim). We also investigated the importance of this allocation decision.

The market segment (or economic sector) allocation policy may be the single most important decision for equity portfolio managers. For example, during the technology bubble, the returns of the technology sector vastly outpaced all other sector returns. This pattern severely reversed during the subsequent bursting of the bubble. During both markets, decisions about allocation to the technology sector had enormous career ramifications for

Raman Vardharaj is a quantitative analyst at RS Investments, New York City. Frank J. Fabozzi, CFA, is professor in the practice of finance at Yale School of Management, New Haven, Connecticut.

many portfolio managers. Since the current Gulf War began in 2003, the energy sector has vastly outperformed all other sectors—badly hurting the skeptics of $60 per barrel oil. Because of the importance of sector allocation (and often the difficulty of the decision), some portfolio managers have entirely abandoned the effort and, instead, seek to stay neutral relative to benchmark sector allocations; their hope is to generate alpha through security selection.

Allocation based on economic sectors and allocation based on size/style segments are not entirely separate exercises. Over time, some sectors become overrepresented in certain size/style buckets relative to other sectors. Naturally, the question arises for the portfolio manager, is the portfolio following a sector allocation policy or a size/style allocation policy? We think, at the practitioner level, the answer depends on the organization of the manager's research efforts. On the one hand, most *fundamental* research analysts specialize in the economic sector because companies within a sector have much more in common with each other (in terms of a business model) than do companies within a size/style bucket. On the other hand, the academic literature and some quantitatively oriented analysts prefer to partition the stock market along the size/style dimension. In light of these differences, we decided to analyze the allocation along the sector dimension and along the size/style dimension. Interestingly, the results were qualitatively similar.

Kritzman and Page (2002, 2003) pointed out that the Brinson studies and the Ibbotson–Kaplan study are descriptive; that is, these studies reveal what investors *choose* to do with actual portfolios rather than what investors *should* do in a theoretical sense. Kritzman and Page, as well as Assoe, L'Her, and Plante (2006), used simulated portfolios to provide a normative evaluation of investment choice. Our study took a descriptive approach, like the Brinson and Ibbotson–Kaplan studies, because we believe that actual investment portfolios, unlike the simulated portfolios of a normative analysis, reflect important real-world constraints found in institutional investor settings, such as limits on tracking error, constraints on size or sector bets (to prevent style drift), prescribed maximum weights for a single stock or industry or country, and so on. These constraints reflect prudent risk management practices (as in limits on tracking error), marketing needs (e.g., style drift can change how the fund is categorized by such rating agencies as Morningstar or Lipper, which affects the fund's marketability), and legal requirements (as in limits on single stocks). Thus, actual investment portfolios differ from simulated portfolios in myriad ways that are hard to model but are important because the reasons underlying them are enduring. We shed light on the role of sector, size/style, and regional allocation policies in such real-world conditions.

Studying the predictive power of dividends, Jung and Shiller (2005) found that dividend yield ratios of individual companies have considerable ability to predict the future growth rate of real dividends (higher yields go together with lower future growth rates) but that the same is not true for an index of stocks.[3] The predictive ability of dividends at the individual-company level is suggestive of *micro-efficiency* (i.e., the market is efficient at the individual-stock level). The lack of predictability at the index level suggests a lack of *macro-efficiency* (i.e., efficiency at the aggregate stock market level). This seems to hint at a large role for asset allocation. This analysis, like all previous literature, is at the broad asset-class level; our study is from the standpoint of sectors within equity-only portfolios.

Our study differs from prior studies also in that we provide a test of statistical significance for the regression results. As discussed later, the standard *F*-test does not suffice for such a test because of the possibility that the regression may be detecting a spurious relationship as a result of the style analysis used to construct policy portfolios. So, we used a novel simulation approach to test for statistical significance.

Data and Methodology

We obtained large-capitalization, small-cap, and international equity mutual fund monthly returns from Lipper. Often, the same load fund was available in multiple share classes and they were not distinct portfolios. To screen out such duplication, we retained only front-end-load funds (class A shares, which are generally the most popular). Separately, we also obtained data on no-load funds.[4] We did not combine the front-end-load and the no-load funds into a single sample because doing so might have obscured the effect of 12b-1 fee differences between the two groups. The monthly net returns for the front-end-load funds were not adjusted for the up-front load (typically, 3–5 percent).

To perform style analysis on monthly returns for 10 years, we searched for all funds that had been in existence for at least 10 years as of the end of December 2004.[5] **Table 1** provides the number of large-cap, small-cap, and international funds that satisfied this criterion among front-end-load funds and no-load funds. For the 5-year style analysis, we searched for funds that had been in existence for at least 5 years through December 2004. Within each style category were a set of funds for the 10-year sample (1995–2004) and a set of funds for the 5-year

Table 1. Number of Funds in Various Categories for the 10-Year and 5-Year Samples

Segment and Sample Period	Front-End Load	No Load
Large cap		
10-Year sample	131	106
5-Year sample	303	187
Small cap		
10-Year sample	46	64
5-Year sample	150	156
International		
10-Year sample	41	39
5-Year sample	120	78
All categories combined		
10-Year sample	218	209
5-Year sample	573	421
Totals for load and no-load funds		
10-Year sample	427	
5-Year sample	994	

sample (2000–2004). Note that the 5-year samples are considerably larger than the 10-year samples because of the growth of the mutual fund industry during this period.

Overall, our samples are probably somewhat smaller than the true populations for several reasons. First, we did not retain back-end-load funds and other types of load funds because they are not portfolios distinct from the main class. Second, Lipper has some other equity categories that we did not include—for example, midcap, multicap, equity income, and emerging market funds. These categories tend to have fewer long-lived funds than the bigger categories (large, small, and international). Third, some funds may have changed in character since their classification by Lipper. Despite these issues, our samples are sufficiently large that a loss of generality is unlikely in our broad results.[6]

The 10-year sample has funds that survived the past 10 years, and the 5-year sample has funds that survived the past 5 years (and also includes the 10-year sample). The 5-year sample includes some funds that are not likely to survive the next 5 years and, therefore, are not as seasoned as the 10-year survivors. Yet, as we show later, our regression results for the most recent 5-year subperiod for the 10-year survivors are quite similar to the results for the 5-year sample. This finding suggests that survivorship bias had no significant impact on our regression results. Because nonsurviving funds underperform survivors, as noted by Malkiel (1995), inclusion of nonsurvivors would only strengthen our results in the analysis of return levels.

Exhibit 1 summarizes the funds in the sample and periods covered, and it provides lists of the Global Industry Classification Standard (GICS) sectors, the Fama–French size/style portfolios, and the breakdown of international regions in the study. We used Ibbotson Associates' Morningstar EnCorr software to obtain historical monthly returns to the 10 GICS economic sectors in the S&P 500 Index (large-cap funds) and S&P 600 Index (small-cap funds). Historical monthly value-weighted returns to the six Fama–French market segments (i.e., large or small × value, neutral, or growth) were obtained from French's website. In the style analysis of international funds, we used stock indices of the international regions given in Exhibit 1.

Using monthly returns, we performed style analysis for the 10 GICS economic sectors for each large-cap fund in order to identify the fund's policy benchmark for the 10-year period ended December 2004. The policy return, *PR*, for month *t* for the fund is the return to its estimated policy benchmark in that month and was calculated as

$$PR_{i,t} = w1_i R1_t + w2_i R2_t + ... + w10_i R10_t,$$

where $R1_t$, $R2_t$, . . ., $R10_t$ are the returns on the 10 economic sectors in month *t* and where $w1_i$, $w2_i$, . . ., $w10_i$ are the policy weights of fund *i* on the 10 economic sectors such that each weight is greater than 0 percent and less than 100 percent and they sum to 100 percent.[7]

To compare the total actual return with the policy return, we calculated the difference between the gross policy return and total return (both annualized over the 10 years) and compared that with reasonable estimates of replication costs. Ibbotson and Kaplan (2000) subtracted an approximate constant cost of replicating the policy mix through indexed mutual funds from the policy return. This approach facilitates a ratio comparison of the returns because the total return is itself net of active fees. We did not take this approach because when returns are negative (as they were during the recent bear market), the return *difference* approach is better than the ratio approach. Additionally, our method helps other researchers whose estimates of replication costs differ from ours to compare their assumed cost with the return difference we document and then conclude how much of the total return is explained by the policy return.[8]

Time-Series Variation

We regressed the 120 (for 10 years) monthly total returns to a large-cap fund against the corresponding monthly policy returns. The R^2 in this regression was averaged across all 131 front-end-load large-cap

Exhibit 1. Summary of Data

Funds

The fund sample consisted of the following:

- Lipper front-end-load funds
- Lipper no-load funds
- Large-cap funds (all styles)
- Small-cap funds (all styles)
- International funds (all styles)

No index funds were used. Monthly net returns from Lipper.

Sample period

The *10-year sample* consisted of funds that were in existence for the 10 years of January 1995–December 2004—with the first half being the 5 years of January 1995–December 1999 and the second half the 5 years of January 2000–December 2004. The *5-year sample* consisted of funds that were in existence for the 5 years January 2000–December 2004.

Economic sectors

The GICS sectors are as follows: consumer discretionary, consumer staples, energy, financials, health care, industrials, information technology, materials, telecommunications services, and utilities. Monthly returns for the sectors came from Ibbotson Associates.

Size/style segments

The six Fama–French market segments for the U.S. equity market are as follows: large-cap value, large-cap neutral, large-cap growth, small-cap value, small-cap neutral, and small-cap growth. Monthly returns by segment came from Kenneth French's website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

International regions

The international market was the MSCI World ex United States universe segmented as follows:

- United Kingdom
- Scandinavia (comprising Denmark, Finland, Norway, and Sweden)
- Europe ex United Kingdom and Scandinavia (comprising Austria, Belgium, France, Germany, Italy, the Netherlands, Spain, and Switzerland)
- Pacific Rim (comprising Australia, Hong Kong, Japan, New Zealand, and Singapore)
- Canada
- Emerging markets

Monthly returns for indices of all the regions except emerging markets came from Kenneth French's website; monthly returns for the emerging markets came from MSCI.

funds in the 10-year sample. The mean R^2 (89 percent) and median R^2 (91 percent) shown in Panel A of **Table 2** in the column "Return vs. Own Policy" for the 10-year sample of large-cap equity funds are the equivalent of the findings reported in the Brinson studies. Thus, analogous to their study, close to 90 percent of the average large-cap stock fund's return variability over time is explained by that fund's policy mix.

Has this explanatory power changed in recent years? To study this question, we divided the 10 years into two equal subperiods (January 1995 through December 1999 and January 2000 through December 2004) and performed separate style analyses for the two 60-month periods.[9] Then, for each subperiod, we regressed the 60 monthly total returns on the corresponding policy returns to obtain the R^2s and averaged them over the 131 funds, as we did for the full sample. The R^2s in Panel A for the subperiods show no significant change in explanatory power. When we repeated this process for the 303 front-end-load large-cap funds in the 5-year sample (funds in existence for at least 5 years through December 2004), the results remained the same.

We repeated this process for small-cap funds and the 10 economic sectors (see Panel B). For the 10-year sample, the R^2 has a mean that is slightly lower than that for the large-cap funds. Again, no significant degradation or improvement in the explanatory power has occurred over time.

Then, we combined the large-cap and small-cap samples and ran style analyses for each fund versus the six Fama–French market segments. As can be seen in Panel C of Table 2, the R^2s are again close to 90 percent. For the international funds (see Panel D), a similar analysis against the six regional indices yielded R^2s that averaged close to 90 percent.

Arguing along the lines of Ibbotson and Kaplan (2000), we note that the R^2s may be high simply because funds participate in stock markets in general, not because they follow a specific allocation policy as to sector, segment, or region. We studied this idea by regressing each fund's total returns against the returns to a common benchmark, rather than each against its own policy benchmark. For common benchmarks, we used the average of all policy benchmarks and the S&P 500. As shown in the middle and last columns of Panel A of Table 2, for large-cap funds, the mean R^2s for

Table 2. R^2s from Monthly Time-Series Regressions of Actual Returns on Policy Returns: All Relevant Front-End-Load Equity Funds, Data for 1995–2004

	Return vs. Own Policy		Return vs. Average Policy		Return vs. SPX			
Sample	Mean	Median	Mean	Median	Mean	Median		
A. Large-cap funds (vs. GICS sector benchmarks)								
10-Year sample								
10-Year policy weights	89%	91%	83%	84%	84%	85%		
First-half policy weights	91	92	85	87	85	87		
Second-half policy weights	90	92	82	85	83	85		
5-Year sample policy weights	90	92	83	85	84	86		
B. Small-cap funds (vs. GICS sector benchmarks)							Return vs. R2K	
							Mean	Median
10-Year sample								
10-Year policy weights	84%	86%	76%	78%	46%	47%	77%	79%
First-half policy weights	83	86	77	80	47	48	79	81
Second-half policy weights	87	89	77	79	48	49	77	82
5-Year sample policy weights	87	89	77	79	48	49	77	82
C. Combined large- and small-cap funds (vs. Fama–French benchmarks)								
10-Year sample								
10-Year policy weights	89%	90%	76%	78%	74%	80%		
First-half policy weights	90	92	79	81	75	83		
Second-half policy weights	89	90	75	79	74	79		
5-Year sample policy weights	88	90	75	77	72	78		
D. International funds (vs. regional benchmarks)							Return vs. EAFE	
							Mean	Median
10-Year sample								
10-Year policy weights	87%	89%	85%	88%	55%	57%	82%	85%
First-half policy weights	87	88	83	86	45	45	76	80
Second-half policy weights	90	92	88	92	62	64	87	89
5-Year sample policy weights	87	91	85	88	60	64	83	89

Note: SPX = S&P 500; R2K = Russell 2000; EAFE = MSCI Europe/Australasia/Far East Index.

the 10-year sample versus the average of all policy benchmarks and versus the S&P 500 are lower than the 89 percent found for the regression against the funds' own policy benchmarks but still quite high.

The R^2 for small-cap funds (Panel B) has a mean of just 46 percent versus the S&P 500, of course, but with the Russell 2000 used as the benchmark, the R^2 rises to an average of 77 percent. Similarly, for international funds, the R^2 has a mean of just 55 percent versus the S&P 500 but with the MSCI Europe/Australasia/Far East (EAFE) Index used as the benchmark, it rises to 82 percent.

Next, we linked the monthly returns (total returns as well as policy returns) to create quarterly returns. Then, following the same procedure as for monthly returns, we ran time-series regressions of total returns on policy returns. As shown in **Table 3**, the R^2 in the 40-quarter regression of large-cap funds on economic sectors has a mean not far from the mean of 89 percent when monthly returns were used. Generally, all the R^2s in Table 3 are quite similar to (and often marginally higher than) the ones in Table 2. The R^2s are quite high even when benchmark index returns were used instead of "own policy" returns, which is similar to our findings for the monthly regressions. Hence, we conclude that the high R^2s in these time-series regressions occur primarily from the funds' participation in the stock market in general, not from the specific allocation policies of each fund. Results for no-load funds (not reported here) were quite similar to those for front-end-load funds.

Table 3. R^2s from Quarterly Time-Series Regressions of Actual Returns on Policy Returns: Relevant Front-End-Load Equity Funds, Data for 1995–2004

	Return vs. Own Policy		Return vs. Benchmark[a]	
Sample	Mean	Median	Mean	Median
A. Large-cap funds (vs. GICS sector benchmarks)				
10-Year sample				
10-Year policy weights	91%	93%	88%	87%
First-half policy weights	91	93	85	88
Second-half policy weights	93	94	87	89
5-Year sample policy weights	93	94	88	90
B. Small-cap funds (vs. GICS sector benchmarks)				
10-Year sample				
10-Year policy weights	84%	88%	79%	82%
First-half policy weights	83	87	77	79
Second-half policy weights	89	92	84	85
5-Year sample policy weights	90	92	84	85
C. Combined large- and small-cap funds (vs. Fama–French benchmarks)				
10-Year sample				
10-Year policy weights	91%	92%	80%	84%
First-half policy weights	91	93	77	84
Second-half policy weights	91	93	85	86
5-Year sample policy weights	91	93	85	86
D. International funds (vs. regional benchmarks)				
10-Year sample				
10-Year policy weights	90%	93%	85%	88%
First-half policy weights	90	91	79	83
Second-half policy weights	94	96	93	94
5-Year sample policy weights	93	94	90	94

[a]Benchmarks as follows: Large-cap stocks, S&P 500; small-cap stocks, Russell 2000; combined large- and small-cap stocks, S&P 500; international stocks, EAFE.

Cross-Sectional Variation

Ibbotson and Kaplan (2000) advised that to understand how much of the variation in returns *among* funds is explained by allocation policy differences, one must compare funds with each other through cross-sectional regression analyses. If funds had a wide range of long-term sector (or market segment or world region) allocation policies but were passively managed within these allocations, then all of the variation in returns among the funds would be explained by their allocation policies.

To investigate cross-sectional variation among funds, we performed the following type of regression: In the 10-year sample with 10 economic sectors, we regressed the total return during January 1995 for all front-end-load large-cap funds on their policy returns during January 1995. The R^2 in this cross-sectional regression indicates how much of the variation in large-cap fund return during this month was explained by the long-term differences in sector allocation among funds. The unexplained portion of the variation is the result of other reasons that shape portfolio performance—such as tactical allocation shifts (i.e., temporary deviations from long-term allocations) that may have occurred that month in some funds, differences in security selection within the sectors, and fees. We repeated this regression for all the other 119 months (through December 2004).

As shown in Panel A of **Table 4**, the 120-month average of the R^2s in these cross-sectional regressions was only 32 percent for front-end-load large-cap funds. That is, long-term sector allocation

Table 4. R^2s from Cross-Sectional Regressions of Actual Returns on Policy Returns: Front-End-Load and No-Load Funds, Data for 1995–2004

	Front-End-Load Funds			No-Load Funds		
Sample	Monthly Returns	Quarterly Returns	Full Period Compounded	Monthly Returns	Quarterly Returns	Full Period Compounded
A. Large-cap funds (vs. GICS sector benchmarks)						
10-Year sample						
10-Year policy weights	32%	33%	*3%*	34%	35%	*3%*
First-half policy weights	36	39	78	35	37	75
Second-half policy weights	39	39	84	42	44	78
5-Year sample policy weights	39	41	79	39	42	73
B. Small-cap funds (vs. GICS sector benchmarks)						
10-Year sample						
10-Year policy weights	28%	25%	40%	32%	29%	*16%*
First-half policy weights	27	29	*25*	34	31	33
Second-half policy weights	38	34	80	40	39	70
5-Year sample policy weights	36	34	72	34	34	69
C. Combined large- and small-cap funds (vs. Fama–French benchmarks)						
10-Year sample						
10-Year policy weights	40%	41%	*15%*	42%	41%	*18%*
First-half policy weights	44	44	55	45	43	56
Second-half policy weights	40	41	72	43	44	76
5-Year sample policy weights	43	44	76	42	42	73
D. International stock funds (vs. regional benchmarks)						
10-Year sample						
10-Year policy weights	19%	23%	*1%*	25%	26%	*18%*
First-half policy weights	29	34	56	32	35	75
Second-half policy weights	21	23	*26*	28	28	30
5-Year sample policy weights	23	28	*5*	23	25	*17*

Note: Values in italic are not significantly different from zero; values not in italic are significant at the 5 percent level.

differences accounted for only about one-third of the variation in the monthly returns. (In Table 4, the numbers not in italic are significant at the 5 percent level of significance.)

Starting with these 120 monthly returns, we linked the returns to produce 40 quarterly returns. We performed similar cross-sectional regressions for the 40 quarterly returns and averaged them over all the periods. As shown in Table 4, the R^2s for the cross-sectional regressions for front-end-load large-cap funds changed little from the R^2s for the monthly regressions.

When we split the 10 years into two subperiods we found that the 120-month average R^2 for large-cap funds, as Panel A shows, was slightly different from the two 60-month average R^2s. This result indicates changes in allocation policy over time, probably as a result of changes in managers or fund investment processes. Notice that the monthly and quarterly R^2s are generally similar.

We repeated the same analysis for small-cap funds. As shown in Panel B of Table 4, sector allocation policy seems to have slightly lower cross-sectional explanatory power than in the case of the large-cap funds in the 10-year sample. Some difference is noticeable between the first and second subperiods; during the most recent subperiod, the explanatory power of sector allocation in small-cap equity has been nearly the same as in large-cap equity. When large- and small-cap funds were combined (Panel C) and the analysis was run against the Fama–French market segments, the explanatory power turned out to be slightly higher than in either large-cap or small-cap stocks alone.

In general, we found that allocation policy in U.S. equity funds has explained about a third of the monthly and quarterly return variation among large-cap and small-cap funds in recent years. In Table 4, the results for no-load funds are quite similar to those for front-end-load funds for the monthly and quarterly data.

Allocation policy among global regions (Panel D of Table 4) appears to explain about a fourth (ranging from 19 percent to 35 percent) of the return variation among international funds based on the monthly and quarterly data. A question that arises is whether one should study allocation to individual countries rather than to world regions because the aggregation that occurs for regions obscures country-specific issues. A problem with a country approach, however, is that 15–20 countries would need to be included and, given the limited number of international funds in the sample, the risk of overfitting the data would arise.

Senechal (2004) found that since 2000, the cross-sectional dispersion of asset returns has declined and the percentage of portfolio risk coming from common factors (as opposed to idiosyncratic risk) has increased. In Table 4, Panels A and B indicate that sector allocation policy explained slightly more return variation during the most recent 5 years relative to the prior 5 years. This finding reverses in Panels C and D, however, which report R^2s for size/style and regional allocation policy. So, whether a uniform change in the explanatory power of allocation policy in the two regimes has occurred is not clear.

Finally, for the 10-year sample of front-end-load large-cap funds, Panel A shows that when we regressed the full period's compound actual returns on policy returns for the 131 funds in the sample, we found the R^2 to be only 3 percent. When we repeated this analysis for the subsamples and the 5-year sample, however, we found much higher R^2s—78 percent during the first half and 84 percent during the second half of the 10-year sample. When we repeated the analysis with the 5-year sample of front-end-load large-cap funds (303 funds), the R^2 was 79 percent.

Table 4 shows, in large-cap and small-cap funds, that sector allocation policy explained nearly three-quarters of the variation in returns across funds over the past 5 years. The same is true with the size/style allocation policy (Panel C). In the earlier subperiod also, allocation policy explained between one-third to three-quarters of cross-sectional variation among these funds (except for front-end-load small-cap funds). The allocation policy generally explained little of the cross-sectional variation in the 10-year samples for all fund categories, however, as is evident from the fact that none of the R^2s (except for front-end-load small-cap funds) were statistically significant. In the international fund category, regional allocation policy generally explained little of the cross-sectional variation over the past 5 years but explained a significant amount during the prior 5 years.[10]

One reason allocation policy generally explained more 5-year cross-sectional return variation than 10-year return variation may be that the average tenure of a fund manager is closer to 5 years than to 10 years. So, allocation policy inferred from style analysis for 5-year data may be closer to the true (unobserved) allocation policy of a fund (which is set by its manager) than style analysis for 10-year data. To the extent that only small and infrequent tactical deviations occur from the policy, the result will be higher explanatory power with 5-year data. Alternatively, fund allocation policy to sectors, styles, and regions may shift over time, so the 10-year policy on a fund is not up-to-date and, therefore, not very informative.

Return Level

Ibbotson and Kaplan (2000) pointed out that many market commentators mistakenly assumed that the Brinson studies also showed that nearly 90 percent of the return *level* was explained by asset allocation policy, although the Brinson studies did not address this question. To study this issue, our approach was to calculate for each fund the return difference between the compound annual policy return and compound annual total return.[11] If the difference exceeded reasonable costs of replicating the policy weights through passive portfolios, then we inferred that the fund had underperformed its policy and that *more than* 100 percent of the return level was explained by the allocation policy.

As shown in Panel A of **Table 5**, for large-cap front-end funds versus economic sector benchmarks in the 10-year sample, the compound policy return exceeded the compound total return by an average of 3.73 percent. This amount is clearly far larger than any reasonable cost (say, 0.25–0.75 percent) of replicating the sector policy mix passively. When policy weights were calculated for the second subperiod and for the 5-year sample, the underperformance was smaller but still far in excess of any reasonable replication cost. Results for the large-cap no-load funds are similar but slightly less extreme.

Interestingly, in small-cap funds versus their economic sector benchmarks (Panel B of Table 5), for the early subperiod, the total return exceeded the policy return. (The outperformance would be even greater when adjusted for the replication cost.) This period was followed by heavy underperformance during the most recent 5 years of the study period (moderate underperformance for no-load funds). Small-cap managers outperformed and the sector allocation policy explained less of the return variation among funds during the first 5

Table 5. Excess of Policy Benchmark Returns over Actual Returns: Compound Annualized Basis, Data for 1995–2004

	Front-End-Load Funds		No-Load Funds	
Sample	Mean	Median	Mean	Median
A. Large-cap stock funds (vs. GICS sector benchmarks)				
10-Year sample				
10-Year policy weights	3.73%	3.69%	3.44%	3.38%
First-half policy weights	3.58	3.41	3.48	3.13
Second-half policy weights	2.75	3.03	2.26	2.16
5-Year sample policy weights	2.75	2.96	1.84	1.73
B. Small-cap stock funds (vs. GICS sector benchmarks)				
10-Year sample				
10-Year policy weights	2.87%	2.41%	0.47%	0.17%
First-half policy weights	–2.20	–1.92	–2.65	–1.04
Second-half policy weights	6.13	5.01	1.10	2.48
5-Year sample policy weights	4.72	4.53	2.11	1.56
C. Combined large- and small-cap stock funds (vs. Fama–French benchmarks)				
10-Year sample				
10-Year policy weights	2.45%	2.07%	1.41%	1.34%
First-half policy weights	1.99	2.05	1.40	2.29
Second-half policy weights	2.31	2.12	0.86	0.57
5-Year sample policy weights	1.61	1.70	0.16	0.27
D. International stock funds (vs. regional benchmarks)				
10-Year sample				
10-Year policy weights	2.77%	2.89%	0.73%	0.76%
First-half policy weights	1.64	1.67	0.69	0.99
Second-half policy weights	4.04	4.56	1.85	2.54
5-Year sample policy weights	3.30	3.63	1.58	2.51

years. This pattern reversed during the next 5 years. If replication costs were at the high end, the no-load small-cap funds would have slightly outperformed their policy returns in the 10-year sample.

When we analyzed combined large- and small-cap funds versus Fama–French sector benchmarks (Panel C of Table 5), we continued to find underperformance relative to the policy but, now, the underperformance was lower. If replication costs were at the high end, then among no-load funds, actual returns would have been nearly the same as policy returns during the second half of the 10-year sample and slightly higher in the 5-year sample.[12] International stock funds (Panel D) generally trailed their policy benchmarks by more than any reasonable replication cost. Their underperformance has increased in recent years.

Thus, on average, equity funds have not added value above their policy benchmarks. This result is stronger in front-end-load funds than in no-load funds. The return data do not take load into account. Thus, the underperformance of front-end-load funds would probably have been even larger if the load had been taken into account. The underperformance may stem from the drag arising from management fees and 12-b1 fees, high trading expenses, poor security selection, and/or mistimed tactical allocation shifts. Consequently, we conclude that, generally, more than 100 percent of the level of fund return is explained by the policy return. This result may reflect a basic fact of active investing that Sharpe (1991) pointed out: The logic of the marketplace dictates that the return on the actively managed dollar will equal the return on the passively managed dollar before costs but will be less after costs.

Statistical Significance

Conventional statistical tests cannot be used to determine the significance of the regression results shown in our various tables. Policy returns are style weights times the sector returns. Style weights were obtained in constrained regressions of total returns on sector returns. So, we cannot rule out the possibility that at least a partially spurious relationship has been introduced between the policy returns and total returns through the style analysis. That is, even random data may have generated some relationship between total returns and policy returns. Hence, we chose a simulation approach to tackle the issue of statistical significance.

We mimicked all of the analyses underlying Tables 2, 3, and 4 with the use of random data that matched the original true data in its distribution.[13] A true data R^2 was considered statistically significant at the 5 percent level of significance if it exceeded random data R^2s in the simulation 95 percent of the time. By this method, all the true data R^2s except those shown in italics in Table 4 were found to be statistically significant.[14]

Conclusion

We examined what part of a U.S. large-cap or small-cap equity fund's performance is explained by its allocation policy for economic sectors and market segments (characterized by size and style). Similarly, we examined what part of an international equity fund's performance is explained by its allocation policy among world regions. We found that the answer has three parts.

First, allocation policy explains nearly 90 percent of the monthly or quarterly return variability over time. Second, as for return variation among funds, which is to us the more interesting question, we found for large- and small-cap U.S. funds that allocation policy explains about one-third of monthly or quarterly return variation. And for international funds, allocation policy among regions explains about one-fourth of monthly or quarterly return variation. In the most recent 5-year period, 2000–2004, close to three-quarters of the variation among funds in compound annual returns was explained by the allocation policy for large- and small-cap U.S. funds. A simulation showed that all these R^2s are statistically significant. Finally, we found that, generally, policy returns exceed total returns, even after accounting for reasonable estimates of cost for passive replication of the policy mix. Thus, generally more than 100 percent of fund return level is explained by the policy allocation.

Some active managers focus exclusively on bottom-up stock picking and thus do not purposefully pick the resulting sector or style allocation for their portfolios. Such managers would be surprised by our finding that in recent years, nearly three-quarters of the cross-sectional variation of 5-year returns of U.S. funds can be explained by sector or style allocation policy. Thus, a factor that they probably ignored played a large role in their performance relative to peers.

In the past, many studies of mutual fund performance have found that actively managed portfolios, on average, have negative alphas after their returns are adjusted for exposure to certain risk factors that have been identified in the academic literature—beta, size, and value. But many active managers do not pay sufficient attention to their exposures to these risk factors or their performance adjusted for these factors. Often, top-down active managers construct and manage their portfolios partly or even fully based on proprietary strategies for long-term policy and short-term tactical allocations to economic sectors, styles, and regions. Such managers might be taken aback to find that active managers as a group underperform portfolios that passively replicate their self-chosen average allocation strategies. In a sense, this finding mirrors the past findings based on risk factors.

In recent years, mutual fund rating agencies, such as Lipper and Morningstar, to control for marketwide factors, have moved toward ranking or rating funds in relation to their size/style counterparts rather than in relation to all equity funds. Given our finding that nearly three-quarters of the cross-sectional variation of recent 5-year returns among U.S. funds can be explained by allocation policy to Fama–French size/style market segments, the approach of the fund rating agencies is validated.

We thank Gregory Bomash and Aleksandra Royzen of Guardian Life Insurance (the parent of RS Investments) for programming assistance.

This article qualifies for 1 PD credit.

Notes

1. The Brinson studies did not address variation among funds, but many analysts wrongly assume that the 90 percent figure is applicable to that issue.
2. French's website is http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Morningstar uses a segmentation method broadly similar to the one used by French. It segments the U.S. stock market into three size groups (large, mid, and small) and three style groups (value, blend, and growth): www.morningstar.com.
3. Jung and Shiller's (2005) interesting analysis was based on a suggestion by Samuelson (1998) that the stock market is micro-efficient but not macro-efficient.
4. We removed index funds from our study because the allocation decisions for such funds are not voluntary.
5. Sharpe (1992) proposed returns-based style analysis.
6. Using the same criteria as described in the text, we screened for funds in the Morningstar database. The resulting number of distinct qualifying funds among front-end-load as well as no-load funds was comparable to our fund sample sizes. The overall returns were also comparable. The detailed Lipper versus Morningstar comparison is available from the authors upon request. In the screening process, we also verified Lipper returns by examining returns from another source, such as Bloomberg or FactSet Research Systems.
7. The policy weighting assumes no short sales or leverage and assumes that any allocation to cash is small, all of which are reasonable assumptions for equity mutual funds.
8. The regression R^2 results were not qualitatively affected by whether or not the constant replication costs were subtracted up front.
9. Separate analyses by subperiod were necessary because the policy benchmark weights for a fund obtained for the 120-month period might not have been the same as the fund's policy benchmark weights during the two subperiods. Managerial changes and other changes in investment processes can account for differences.
10. An interesting question in regard to the international funds is why all the monthly and quarterly cross-sectional R^2s are statistically significant whereas some of the 5-year compound and generally all of the 10-year compound R^2s are not. The reason is that we are reporting the average of 120 in the 10-year sample (60 in the 5-year sample) monthly, 40 (20) quarterly, and the sole full-period compound cross-sectional regressions that can be performed in the 10-year (5-year) samples. There is less variability around an average statistic (average of many months/quarters) than around a point statistic (sole full period). Similarly, we noticed in our simulations that the variability declined as the number of funds in the sample increased, which led to stronger statistical inferences for the larger samples.
11. Ibbotson and Kaplan (2000) calculated the percentage of the managed portfolio return level explained by the policy return for that portfolio as the ratio of compound annual policy return divided by the compound annual total return. Thus, a managed portfolio that outperformed its policy would have a ratio less than unity.
12. The excess returns (policy less total) are lower in no-load funds than in front-end-load funds. The primary reason is that in the Lipper data, the average total returns of no-load funds generally exceed those of front-end-load funds (even without including an adjustment for their load). We confirmed this finding by using Morningstar data. (Details are available upon request.) Nanda, Wang, and Zheng (2005) showed that front-end-load funds that introduced additional share classes in the 1990s experienced a significant drop in performance about two years after introducing the new classes. During the 1990s, the vast majority of load funds switched from one class (often the front-end-load) to multiple classes. This phenomenon is probably related to our finding about the underperformance of front-end-load funds.
13. For example, in Table 4, Panel A, we showed an R^2 (averaged over 120 months) of 32 percent in the cross-sectional regressions of monthly actual returns on policy returns for the sample of front-end-load large-cap funds with 10 years of returns. To determine the statistical significance of this number (i.e., 32 percent), we proceeded as follows. We assumed the presence of a similar number of funds (131 in this case) and 10 sectors and generated 120 random monthly returns for funds and sectors under the assumption that they were all independent and lognormally distributed. Random fund monthly log returns were drawn from a population calibrated to have the same mean and standard deviation as that of all large-cap U.S. funds over the 120 months. Random sector monthly log returns were similarly drawn from a population calibrated to have the same mean and standard deviation as that of large-cap GICS sectors over the 120 months. Thus, the random sample matched the actual sample in all key aspects. We then followed the same steps that we had followed with actual funds to produce cross-sectional regression R^2s. That is, we performed constrained regressions of randomly generated fund total returns on randomly generated sector returns to obtain style weights. Next, we calculated fund policy returns by using the style weights and sector returns. Then, for each month, we performed cross-sectional regressions of total returns on policy returns for all funds in the simulation sample and averaged the resulting R^2 over all months. We repeated this process all over again with a fresh set of randomly generated fund total returns and sector returns to obtain a new average R^2. We ran the simulation 1,000 times. The 95th percentile for the average monthly cross-sectional regression R^2 in this simulation was 17.4 percent. The implication is that there was only a 5 percent chance that an R^2 greater than or equal to 17.4 percent could have been obtained purely by chance in this analysis. Because 32 percent exceeds 17.4 percent, we concluded that the actual R^2 is statistically significant at the 5 percent level of significance. The detailed simulation results are available from the authors upon request.
14. Our regression R^2 results can be divided into three groups: (1) High R^2: statistically significant but not economically meaningful and, therefore, misleading. The time-series results fell into this group. (2) R^2 ranging from moderate to high: statistically significant and economically meaningful. The cross-sectional results, other than those for 10-year full-period compound returns, fell into this group. (3) Low R^2: statistically insignificant. The cross-sectional results for 10-year full-period compound returns fall into this group. Thus, statistical significance was necessary but not sufficient in determining the usefulness of a result. Economic reasoning was required to interpret the results. To help the reader make these distinctions in our results, we had to test all the R^2s for significance.

References

Assoe, Kodjovi, Jean-François L'Her, and Jean-François Plante. 2006. "The Relative Importance of Asset Allocation and Security Selection." *Journal of Portfolio Management*, vol. 33, no. 4 (Fall):46–55.

Brinson, Gary P., Randolph Hood, and Gilbert L. Beebower. 1986. "Determinants of Portfolio Performance." *Financial Analysts Journal*, vol. 42, no. 4 (July/August):39–48.

Brinson, Gary P., Brian D. Singer, and Gilbert L. Beebower. 1991. "Determinants of Portfolio Performance II." *Financial Analysts Journal*, vol. 47, no. 3 (May/June):40–48.

Fama, Eugene, and Kenneth French. 1996. "Multifactor Explanations of Asset Pricing Anomalies." *Journal of Finance*, vol. 51, no. 1 (March):55–84.

Ibbotson, Roger D., and Paul Kaplan. 2000. "Does Asset Allocation Policy Explain 40, 90, or 100 Percent of Performance?" *Financial Analysts Journal*, vol. 56, no. 1 (January/February):26–33.

Jung, Jeeman, and Robert J. Shiller. 2005. "Samuelson's Dictum and the Stock Market." *Economic Inquiry*, vol. 43, no. 2 (April):221–228.

Kritzman, Mark, and Sébastien Page. 2002. "Asset Allocation versus Security Selection: Evidence from Global Markets." *Journal of Asset Management*, vol. 3, no. 3 (December):202–212.

———. 2003. "The Hierarchy of Investment Choice." *Journal of Portfolio Management*, vol. 29, no. 4 (Summer):11–23.

Malkiel, Burton G. 1995. "Returns from Investing in Equity Mutual Funds 1971 to 1991." *Journal of Finance*, vol. 50, no. 2 (June):549–572.

Nanda, Vikram, Z. Jay Wang, and Lu Zheng. 2005. "The ABCs of Mutual Funds: On the Introduction of Multiple Share Classes." Working paper (January).

Samuelson, Paul A. 1998. "Summing Up on Business Cycles: Opening Address." In *Beyond Shocks: What Causes Business Cycles.* Edited by Jeffrey C. Fuhrer and Scott Schuh. Boston: Federal Reserve Bank of Boston.

Senechal, Edouard. 2004. "A Model Commentary: The Challenges of Declining Cross-Sectional Volatility." *Barra Newsletter* (Autumn):2–10.

Sharpe, William F. 1991. "The Arithmetic of Active Management." *Financial Analysts Journal*, vol. 47, no. 1 (January/February):7–9.

———. 1992. "Asset Allocation: Management Style and Performance Measurement." *Journal of Portfolio Management*, vol. 18, no. 2 (Winter):7–19.

