Performance and Persistence in Institutional Investment Management

JEFFREY A. BUSSE, AMIT GOYAL, and SUNIL WAHAL[*]

ABSTRACT

Using new, survivorship bias-free data, we examine the performance and persistence in performance of 4,617 active domestic equity institutional products managed by 1,448 investment management firms between 1991 and 2008. Controlling for the Fama–French (1993) three factors and momentum, aggregate and average estimates of alphas are statistically indistinguishable from zero. Even though there is considerable heterogeneity in performance, there is only modest evidence of persistence in three-factor models and little to none in four-factor models.

THE TWIN QUESTIONS OF WHETHER INVESTMENT MANAGERS generate superior risk-adjusted returns ("alpha") and whether superior performance persists are central to our understanding of efficient capital markets. Academic opinion on these issues revolves around the most recent evidence incorporating either new data or improved measurement technology. Although Jensen's (1968) original examination of mutual funds concludes that funds do not have abnormal performance, later studies provide evidence that relative performance persists over both short and long horizons.[1] Carhart (1997), however, reports that accounting for momentum in individual stock returns eliminates almost all evidence of persistence among mutual funds (with one exception, the continued underperformance of the worst performing funds (Berk and Xu (2004)). More recently, Bollen and Busse (2005), Cohen, Coval, and Pástor (2005), Avramov and Wermers (2006), and Kosowski et al. (2006) find predictability in

[1] See, for example, Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Brown and Goetzmann (1995), Elton, Gruber, and Blake (1996), and Wermers (1999).

performance even after controlling for momentum. But Barras, Scaillet, and Wermers (2009) and Fama and French (2008) find little to no evidence of persistence or skill, particularly in the latter part of their sample periods.

The attention given to the study of performance and persistence in retail mutual funds is entirely warranted. The data are good, and this form of delegated asset management provides millions of investors access to ready-built portfolios. At the end of 2007, 7,222 equity, bond, and hybrid mutual funds were responsible for investing almost $9 trillion in assets (Investment Company Institute (2008)). However, an equally large arm of delegated investment management receives much less attention, but is no less important. At the end of 2006, more than 51,000 plan sponsors (public and private retirement plans, endowments, foundations, and multi-employer unions) allocated more than $7 trillion in assets to about 1,200 institutional asset managers (Standard & Poor's (2007)). In this paper, we examine the performance and persistence in performance of portfolios managed by institutional investment management firms for these plan sponsors.

Institutional asset management firms draw fixed amounts of capital (referred to as "mandates") from plan sponsors. These mandates span a variety of asset classes, including domestic equity, fixed income, international equity, real estate securities, and alternative assets (including hedge funds and private equity). Our focus is entirely on domestic equity because, relative to other asset classes, it offers the most widely accepted benchmarks and risk-adjustment approaches. Within domestic equity, each mandate calls for investment in a product that fits a style identified by size and growth-value gradations. Multiple mandates from different plan sponsors can be managed together in one portfolio or separately to reflect sponsor preferences and restrictions. However, the essential elements of the portfolio strategy are identical and typically reflected in the name of the composite product (e.g., large-cap value). This "product" (rather than a derivative portfolio or fund) is our unit of observation.

Our data consist of composite returns and other information for 4,617 active domestic equity institutional investment products offered by 1,448 investment management firms between 1991 and 2008. The data are free of survivorship bias, and all size and value-growth gradations are represented. At the end of 2008, more than $2.5 trillion in assets were invested in the products represented by these data.

We assess performance by estimating factor models cross-sectionally for each product and by constructing equal- and value-weighted aggregate portfolios. Using the portfolio approach, the equal-weighted three-factor alpha based on gross returns is an impressive 0.35% per quarter with a *t*-statistic of 2.52. However, value-weighting turns this alpha into a statistically insignificant −0.01% per quarter. Correcting for momentum also makes a big difference: the equal-weighted (value-weighted) four-factor alpha drops to 0.20% (increases to 0.05%) and is not statistically significant. Fees further

decimate the returns to plan sponsors; the equal-weighted (value-weighted) net-of-fee four-factor quarterly alpha is 0.01% (−0.10%) and again not statistically significant.[2]

These aggregate results mask considerable cross-sectional variation in returns. The difference between equal-weighted and value-weighted results indicates that, to the extent that it exists, superior performance is concentrated in smaller products. And, the standard deviation of individual product alphas is high, at 0.78% per quarter for four-factor alphas. To disentangle the issue of whether high (or low) realized alphas are manifestations of skill or luck, we utilize the bootstrap approach of Kosowski et al. (2006), as modified by Fama and French (2008). We find very weak evidence of skill in gross returns, and net-of-fee excess returns are statistically indistinguishable from their simulated counterparts.

Despite these weak aggregate and average performance statistics, because there is large cross-sectional variation in performance, it may still be the case that institutional products that deliver superior performance in one period continue to do so in the future. Evidence of such persistence could represent a violation of efficient markets, and, for plan sponsors, represent an important justification for selecting investment managers based solely on performance. We judge persistence in two ways. First, we form deciles based on benchmark-adjusted returns and estimate alphas over subsequent intervals using factor models. We calculate alphas over short horizons (one quarter and 1 year) to compare them to the retail mutual fund literature, and over long horizons to address whether plan sponsors can benefit from chasing winners and/or avoiding losers. Second, we estimate Fama–MacBeth (1973) cross-sectional regressions of risk-adjusted returns on lagged returns over similar horizons. The latter approach allows us to introduce control variables (such as assets under management and flows).

For losers, there is evidence of reversals, but it is modest at best. This may come from look-ahead issues in some tests (see Carpenter and Lynch (1999) and Horst, Nijman, and Verbeek (2001)), and/or from economies of scale for smaller-sized portfolios. For winners, using the three-factor model, the alpha of the extreme winner decile 1 year (one quarter) after ranking is 0.96% (1.52%) with a *t*-statistic of 2.79 (3.55). However, after controlling for the mechanical effect of momentum (that winner products have winner stocks, which are likely to be in the portfolio during the post-ranking period), the 1-year (one-quarter) alpha shrinks to 0% (0.18%) per quarter and is statistically indistinguishable from zero. Persistence regressions show similar results over these horizons. Thus, at best (using three-factor models) there is modest evidence of persistence over 1 year; at worst (using four-factor models) there is no persistence in returns.

---

[2] By way of comparison, Gruber (1996) estimates a CAPM alpha of −13 basis points per month after expenses for mutual funds. Wermers (2000) estimates that mutual funds outperform the S&P 500 by an average of 2.3% per year before expenses and trading costs and underperform the S&P 500 by an average of 50 basis points per year net of expenses and trading costs.

Over evaluation horizons longer than 1 year, no measurement technique shows positive top-decile alphas.[3]

Earlier studies that examine performance and persistence in institutional investment management are hampered either by survivorship bias, a short time series (which limits the power of time series-based tests), or design. The first of these studies, Lakonishok, Shleifer, and Vishny (1992), examines the performance of 341 investment management firms between 1983 and 1989. They find that performance is poor on average, and acknowledge that although some evidence of persistence exists, data limitations prevent a robust conclusion. Coggin, Fabozzi, and Rahman (1993) also find that investment managers have limited skill in selecting stocks. Ferson and Khang (2002) use portfolio weights to infer persistence, and Tonks (2005) examines the performance of U.K. pension fund managers between 1983 and 1997. Both find some evidence of excess performance but with small samples. Christopherson, Ferson, and Glassman (1998) also find some evidence of persistence among 185 investment managers between 1979 and 1990, but their sample also suffers from survival bias. Goyal and Wahal (2008), while not directly interested in persistence, report that plan sponsors hire investment managers after large positive excess returns, but that post-hiring returns are zero; in contrast, pre-firing returns are not exclusively negative and post-firing returns display modest reversion. It is tempting to conclude that there is no persistence based on their results but such a conclusion does not necessarily follow. As they describe in their paper, the hiring and firing of investment managers is influenced by factors unrelated to performance. For example, investment managers may be fired because of personnel turnover at the investment management firm or reallocations of investment mandates from one asset class to another. In addition to agency considerations unrelated to performance, institutional frictions such as minority ownership of the investment manager, the use of an investment consultant, etc., can influence these decisions. This means that post-hiring and post-firing returns are also affected by selection mechanisms that are uncorrelated with performance, thereby making it difficult to make precise inferences about persistence in the universe of investment managers. In contrast, our paper tackles the subject head-on, with the largest sample to date that is uncontaminated by survivorship bias.

Our results are both of economic and practical significance. Clearly, economic interpretation in the context of efficient markets depends on the benchmark one chooses to consider. An investor happy with the CAPM or three-factor model might reasonably conclude that institutional investment managers deliver superior returns with some persistence. However, an investor intent on incorporating momentum into the analysis is unlikely to be so sanguine. Moreover, as we show in the robustness section of the paper, those partial to conditional methods in the spirit of Ferson and Schadt (1996) and more recent

---

[3] For mutual funds, Bollen and Busse (2005) report a four-factor alpha of 0.39% for the top decile in the post-ranking quarter. Kosowski et al. (2006) report a statistically significant monthly alpha of 0.14% in the extreme winner decile for the first year.

benchmarking methods that incorporate other passive portfolios (Cremers, Petajisto, and Zitzewitz (2008)) also face mixed evidence. To us, on balance, it is difficult to make the case for persistence.

What are the practical consequences of this? If one takes the strong view that there is no persistence, then one logical conclusion might be that plan sponsors should engage in entirely passive asset management. Lakonishok et al. (1992) point out that if plan sponsors did not chase returns, they would have nothing to do. Given agency problems, exclusively passive asset management is an unlikely outcome. Moreover, French (2008) argues that price discovery, necessary to society, requires some degree of active management. These arguments imply that some degree of active management must exist and that plan sponsors, in equilibrium, should provide capital to such organizations.

Our paper proceeds as follows. Section I discusses our data and sample construction. We discuss the results on performance and persistence in Sections II and III, respectively. Section IV provides robustness checks, and Section V concludes.

I. Data and Sample Construction

A. Data

Our data come from Informa Investment Solutions (IIS), a firm that provides data, services, and consulting to plan sponsors, investment consultants, and investment managers. This database contains quarterly returns, benchmarks, and numerous firm- and product-level attributes for 6,040 domestic equity products managed by 1,661 institutional asset management firms from 1979 to 2008. Although the database goes back to 1979, it only contains "live" portfolios prior to 1991. In that year, data-gathering policies were revised such that investment management firms that exit due to closures, mergers, and bankruptcies were retained in the database. Thus, data after 1991 are free from survivorship bias. The average attrition rate over this period is 3.8% per year, which is higher than the 3% reported by Carhart (1997) for mutual funds. The coverage of the database is quite comprehensive. We cross-check the number of firms with two other similar data providers, Mercer Performance Analytics and eVestment Alliance. Both the time-series and cross-sectional coverages of the database that we use are better than the two alternatives.

Several features of the data are important for understanding the results. First, since investment management firms typically offer multiple investment approaches, the database contains returns for each of these approaches. For example, Aronson + Johnson + Ortiz, an investment management firm with $22 billion in assets, manages 10 portfolios in a variety of capitalizations and value strategies. The returns in the database correspond to each of these 10 strategies. Our unit of analysis is each strategy's return, which we refer to as a "product." Second, the database contains "composite" returns provided by the investment management firm. The individual returns earned by each plan-sponsor client (account) may deviate from these composite returns for a variety

of reasons. For example, a public-defined benefit plan may ask an investment management firm to eliminate "sin" stocks from its portfolio. Such restrictions may cause small deviations of earned returns from composite returns. Third, the returns are net of trading costs, but gross of investment management fees. Fourth, although the data are self-reported, countervailing forces ensure accuracy. The data provider does not allow investment management firms to amend historical returns (barring typographical errors) and requires the reporting of a contiguous return series. Further, the SEC vets these return data when it performs random audits of investment management firms. However, we cannot eliminate the possibility of backfill bias. We address this issue in Section IV.

In addition to returns, the database contains descriptive information at both the product level and the firm level. Roughly speaking, the descriptive information can be categorized into data about the trading environment, research, and personnel decisions. For each product, we obtain cross-sectional information on its investment style, a manager-designated benchmark, and whether it offers a performance fee. We also extract time-series information on assets under management, annual portfolio turnover, annual personnel turnover, and a fee schedule.

We impose simple filters on the data. First, we remove all products that are either missing style identification information or contain non-equity components such as convertible debt. This filter removes 874 products. Second, we remove all passive products (358 products) since our interest is in active portfolio management. Finally, we remove all products that are also offered as hedge funds (191 products). Our final sample consists of 4,617 products offered by 1,448 firms.

B. Descriptive Statistics

Table I provides basic descriptive statistics of the sample. Panel A shows statistics for each year, and Panel B presents similar information for each investment style. Style gradations are based on market capitalization (small, mid, large, and all cap) and investment orientation (growth, core, and value).[4]

In Panel A, the second column shows the number of active domestic equity institutional products from 1991 to 2008. The number of available products rises monotonically from 1991 to 2004, and then declines somewhat in the last 4 years. By the end of 2008, more than 2,500 products are available to plan sponsors. The third column shows average assets (in $ millions). Asset data are available for approximately 80% of the total sample. Average assets generally increase over time with the occasional decline in some years; most noticeable, and not surprising, is the decline in 2008. By the end of 2008, total assets exceed $2.5 trillion (2,572 products multiplied by average assets of $1 billion). The growth in the number of products and average assets mirrors that of the mutual fund industry, which also grew considerably during this time period

---

[4] Size break points in this database are as follows: Small caps are those less than $2 billion, mid-caps are between $2 and $7 billion, and large caps are larger than $7 billion.

Table I
Descriptive Statistics

The table presents descriptive statistics on the sample of institutional investment products. Asset size is in millions of dollars, turnover is in percent per year, and fees are in percent per year. The descriptives in Panel B are for the year 2008 only.

	Number Products	Average Asset Size	Average Turnover	Fees		
				$10M	$50M	$100M
Panel A: Descriptives Statistics by Year						
1991	1,201	621	60.7	0.79	0.65	0.61
1992	1,357	604	59.5	0.78	0.63	0.58
1993	1,572	628	62.0	0.79	0.65	0.60
1994	1,770	628	61.7	0.77	0.63	0.57
1995	1,953	802	65.5	0.78	0.63	0.58
1996	2,154	897	66.4	0.78	0.68	0.58
1997	2,309	1,068	68.9	0.79	0.64	0.58
1998	2,476	1,150	73.7	0.78	0.64	0.59
1999	2,655	1,254	76.0	0.78	0.65	0.59
2000	2,841	1,125	80.8	0.78	0.65	0.60
2001	3,001	990	78.2	0.79	0.66	0.60
2002	3,065	780	74.0	0.79	0.67	0.61
2003	3,137	1,050	73.6	0.80	0.68	0.62
2004	3,156	1,180	71.7	0.80	0.68	0.62
2005	3,080	1,319	70.4	0.81	0.69	0.63
2006	2,982	1,470	71.6	0.81	0.69	0.63
2007	2,877	1,395	73.3	0.81	0.69	0.63
2008	2,572	1,009	75.2	0.81	0.69	0.64
Panel B: Descriptives Statistics by Style						
Small Cap Growth	231	430	105.8	0.94	0.88	0.82
Small Cap Core	161	383	87.4	0.88	0.78	0.71
Small Cap Value	270	591	70.1	0.93	0.86	0.80
Mid Cap Growth	191	595	104.4	0.82	0.72	0.66
Mid Cap Core	89	269	87.3	0.78	0.67	0.60
Mid Cap Value	187	883	67.1	0.84	0.72	0.66
Large Cap Growth	369	1,281	74.7	0.76	0.62	0.57
Large Cap Core	327	1,091	61.1	0.69	0.56	0.50
Large Cap Value	417	2,115	59.6	0.71	0.58	0.51
All Cap Growth	66	763	94.1	0.85	0.75	0.71
All Cap Core	115	1,051	72.9	0.77	0.60	0.54
All Cap Value	149	694	49.5	0.80	0.68	0.63

(Investment Company Institute (2008)). Average portfolio turnover (shown in column 4) increases over time, from 60.7% in 1991 to 75.2% in 2008. The increase in turnover is gradual except for the occasional spike (e.g., during 2000). Wermers (2000) documents a similar increase in turnover for mutual funds during his 1975 to 2004 sample period.

The prototypical fee structure in institutional investment management is such that management fees decline as a step function of the size of the

mandate delegated by the plan sponsor. Although firms can have different breakpoints for their fee schedules, our data provider collects marginal fee schedules using standardized break points of $10 million, $50 million, and $100 million. The marginal fees for each break point are based on fee schedules; actual fees are individually negotiated between investment managers and plan sponsors. Larger plan sponsors typically are able to negotiate fee rebates. Some investment management firms offer most-favored-nation clauses, but our database does not contain this information. To our knowledge, no available database details actual fee arrangements, so we work with the pro forma fee schedules. The last three columns show average annual pro forma fees (in percent) assuming investment of $10 million, $50 million, and $100 million, respectively. Not surprisingly, average fees decline as investment levels increase. Fees are generally stable over time, varying no more than 7 basis points over the entire time period.

Panel B shows that all major investment styles are represented in our sample. As of the end of 2008, the largest number of products (417) reside in large-cap value, whereas the smallest are in all-cap growth (66). To allow for across-style comparisons without any time-series variation, we present values of assets, turnover, and fees as of the end of 2008. Generally, average portfolio sizes are biggest for large-cap products. Turnover is highest for small-cap products; the average turnover for small-cap growth is 105.8. Considerable variation also exists in fees across investment styles. Again, small-cap products have the highest fees, and large-cap products have the lowest fees. Although not shown in the table, intrastyle variation in fees is extremely small; almost all of the cross-sectional variation in fees is generated by investment styles.

II. Performance

A. Measurement Approach

Following convention in the mutual fund literature, our primary approach to measuring performance is to estimate factor models using time-series regressions. To generate aggregate measures of performance, we create equal- and value-weighted portfolio returns of all products available in that quarter. The weight used for value-weighting is based on the assets in that product at the end of December of the prior year. With these returns, we estimate:

$$r_{p,t} - r_{f,t} = \alpha_p + \sum_{k=1}^{K} \beta_{p,k} f_{k,t} + \epsilon_{p,t}, \tag{1}$$

where r_p is the portfolio return, r_f is the risk-free return, f_k is the kth factor return, and α_p is the abnormal performance measure of interest. To compute CAPM alphas, we use the excess market return as the only factor. For Fama–French (1993) alphas, we use market, size, and book-to-market factors. Since Fama and French (2004) maintain that momentum remains an embarrassment to the three-factor model, and since it appears to have become

the conventional way to measure performance, we also estimate a four-factor model. We obtain these four factors from Ken French's web site.[5]

We also calculate a variety of performance measures for each product. First, we estimate alphas using the factor models described above. This is only possible if the product has a long enough return history to reliably estimate the regression. We require 20 quarterly observations to estimate the alpha for each product. Since this requirement imposes a selection bias (potentially removing underperforming products), we do not interpret these results in assessing aggregate performance. Rather, our only purpose is to gauge cross-sectional variation in performance.

Second, we calculate benchmark-adjusted returns by simply subtracting a benchmark return from the quarterly raw return,

$$rx_{i,t} = r_{i,t} - r_{b,t}, \tag{2}$$

where r_i is the return on institutional product i, r_b is the benchmark return, and rx_i is the excess return. Such benchmark-adjusted returns are widely used by practitioners for evaluation purposes.

B. Aggregate Performance

Panel A of Table II shows estimates of aggregate measures of performance. In addition to equal- and value-weighted gross returns, we also present parallel results for net returns. To compute net returns, we first calculate the time-series average pro forma fee based on a $50m investment in that product. We then subtract one-quarter of this annual fee from the product's quarterly return.

The equal-weighted CAPM alpha is an impressive 0.57% per quarter with a *t*-statistic of 3.17. Since raw returns have significant exposure to size and value factors, the equal-weighted three-factor model alpha is reduced to 0.35% per quarter with a *t*-statistic of 2.52. Value-weighting the returns further reduces the alpha to −0.01 with a *t*-statistic of 0.05, suggesting that much of the superior performance comes from small products. Even using simple benchmark-adjusted returns, value-weighting makes a difference. Average equal-weighted benchmark-adjusted returns are 0.49% (with a *t*-statistic of 3.36), but value-weighted benchmark-adjusted returns are only 0.16% (with a *t*-statistic of 1.11). As with mutual funds, controlling for stock momentum makes a big difference— the equal-weighted four-factor alpha shrinks to 0.20% per quarter with a *t*-statistic of only 1.34, and the value-weighted four-factor alpha increases to

---

[5] Ken French's momentum factor is slightly different from the one employed by Carhart (1997). Carhart calculates his momentum factor as the equal-weighted average of firms with the highest 30% 11-month returns (lagged 1 month) minus the equal-weighted average of firms with the lowest 30% 11-month returns. French's momentum factor follows the construction of the book-to-market factor (HML). It uses six portfolios, splitting firms on the 50th percentile of NYSE market capitalization and on 30th and 70th percentiles of the 2- through 12-month prior returns for NYSE stocks. Portfolios are value-weighted, use NYSE break points, and are rebalanced monthly.

Table II
Distribution of Performance

The CAPM one-factor model uses the market factor. The three factors in the three-factor model are the Fama–French factors (market, size, and book-to-market). The four factors in the four-factor model are the Fama–French factors augmented with a momentum factor. We choose the benchmarks based on the investment style of the product to adjust raw returns. Panel A reports performance measures for portfolios along with their t-statistics in parentheses. We form portfolios from individual products. Portfolios are both equal- and value-weighted (we value weight based on asset size from December of the prior year). Returns are either gross or net of fees (for a mandate of $50 million). Panel B reports the percentiles for performance measures for gross returns of 3,842 individual products that have at least 20 quarters of available data. All numbers are in percent per quarter. The sample period is 1991 to 2008.

	Benchmark-Adjusted Returns	Factor Model Alphas		
		1-Factor	3-Factor	4-Factor
Panel A: Portfolio Performance				
EW Gross	0.49	0.57	0.35	0.20
	(3.36)	(3.17)	(2.52)	(1.34)
VW Gross	0.16	0.13	−0.01	0.05
	(1.11)	(1.06)	(−0.05)	(0.40)
EW Net	—	0.40	0.16	0.01
		(2.11)	(1.17)	(0.05)
VW Net	—	−0.02	−0.16	−0.10
		(−0.15)	(−1.39)	(−0.79)
Panel B: Individual Product Performance of Gross Returns				
		Alphas		
5th pcnt	−0.67	−0.71	−0.84	−0.96
10th pcnt	−0.39	−0.42	−0.60	−0.61
Mean	0.52	0.60	0.34	0.20
Median	0.43	0.51	0.21	0.18
90th pcnt	1.56	1.71	1.38	1.03
95th pcnt	1.99	2.20	1.94	1.45
		t-statistics		
5th pcnt	−1.13	−1.18	−1.63	−1.44
10th pcnt	−0.7	−0.68	−1.13	−1.04
Mean	0.77	0.81	0.44	0.33
Median	0.81	0.88	0.45	0.35
90th pcnt	2.16	2.22	2.02	1.69
95th pcnt	2.58	2.66	2.48	2.08

0.05% per quarter with a t-statistic of 0.40. Finally, as expected, incorporating fees shrinks both three- and four-factor alphas considerably and eliminates any statistical significance. The difference in alpha from equal-weighted gross and net returns is approximately 18 basis points per quarter, or 74 basis points per year. This roughly corresponds to the annual fees reported in Table I.

There is little evidence that, on aggregate, the products offered by institutional investment management firms deliver risk-adjusted excess returns. Of

course, it is entirely possible that *some* investment managers deliver superior returns. We turn to the distribution of product performance next.

C. Distribution of Performance

Panel B of Table II shows the cross-sectional distribution of performance measures using gross returns. We report the mean as well as the 5th, 10th, 50th (median), 90th, and 95th percentiles. Before proceeding, we urge caution in interpretation for two reasons. First, as indicated earlier, we require 20 quarterly observations to estimate a product's alpha. This naturally creates an upward bias in our estimates since short-lived products are more likely to be underperformers. Second, statistical inference is difficult. The individual alphas are cross-sectionally correlated. In principle, one could compute the standard error of the mean alpha (the cross-sectional average of the individual alphas). However, this would require an estimate of the $N \times N$ covariance matrix of the estimated alphas. Since our sample is large ($N = 4,617$), computational limitations preclude this approach. Therefore, we provide the percentiles of the cross-sectional distribution of individual t-statistics, rather than a single t-statistic for the mean alpha.

The average quarterly benchmark-adjusted return is 0.52% per quarter. If a plan sponsor evaluates the performance of institutional products using simple style benchmarks, then it might appear that, on average, institutional investment managers deliver superior performance. The cross-sectional distribution of alphas shows an interesting progression between the one-, three-, and four-factor models. For example, the mean alpha declines from 0.60% per quarter for the CAPM to 0.34% for the Fama–French (1993) three-factor model to 0.20% for the four-factor model. Similarly, the mean t-statistics decline from 0.81 for the CAPM to 0.44 for the three-factor model and eventually to 0.33 for the four-factor model. As with the aggregate results in Panel A, the sophistication of risk adjustment affects inference.

The tails of the distribution are interesting in their own right. Products that are in the 5th percentile have a four-factor alpha of −0.96%, and the 5th percentile of t-statistics is −1.44. However, the distribution is right-skewed. The four-factor alpha for the 95th percentile is 1.45% per quarter, and the corresponding t-statistic is 2.08. Thus, products in the top 5th (and perhaps even the top 10th) percentile deliver large returns.[6] Are these tails populated by truly skilled funds or by funds that just happened to get lucky? We examine this question next.[7]

---

[6] We remind the reader that the distribution is shifted to the right because of the requirement that at least 20 return observations be available to estimate alphas. This reduces the number of products for this exercise from 4,617 to 3,842. Not surprisingly, the average benchmark-adjusted return for the eliminated products is 0.17% lower than that for the remaining products, generating the right shift.

[7] We also estimate regressions of four-factor alphas on a variety of variables that proxy for research activities, trading, and the composition of human capital. These regressions are interesting but noisy so we report them in the Internet Appendix rather than in the body of the paper. The

D. Skill or Luck

It is possible that some of the estimated alphas are high because of luck. To disentangle luck from true skill, we utilize the approach of Kosowski et al. (2006). Kosowski et al. bootstrap the returns of products under the null of zero alpha and then base their inference on the entire cross-section of simulated alphas and their t-statistics. We implement their procedure with the modification proposed by Fama and French (2008).[8] The reader is referred to these papers for further details on the simulation technique.

We use the four-factor model to conduct this experiment. The simulation can be conducted using alphas or t-statistics. We use both in the interest of completeness but advocate caution in interpretation of results based on alphas. Alphas are estimated imprecisely and simulation results based on t-statistics are inherently superior because they control for the precision of each estimate of alpha. This weighting is all the more important because our (relatively short) time series relies on quarterly (not monthly) returns. Therefore, we conduct our inference largely from t-statistic-based simulations. Table III presents the results. Panel A presents the results for alphas from gross returns, while Panel B shows the same results for alphas from net returns. In each panel, we show the percentiles of actual and simulated (averaged across 1,000 simulations) alphas and their t-statistics. We also show the percentage of simulation draws that produce an alpha/t-statistic greater than the corresponding actual value. This column can be interpreted as a p-value of the null that the actual value is equal to zero.

For the vast majority of the percentiles of alphas that we report, the actual alphas are less than the simulated alphas in more than 5% of the cases. Using t-statistics, which as described above are preferable, there is only one case (the 60th percentile) in which the simulated t-statistics are less than the actual t-statistic in less than 5% of the draws. Using alphas, this is also the case in only one situation (the 99th percentile). Examining net returns to plan sponsors (Panel B), the distribution of actual alphas or t-statistics is indistinguishable from their simulated counterparts. Overall, there seems to be very little evidence of skill.

dependent variable (four-factor alpha) is measured over the entire return history of the product and the independent variables are measured over the entire time series or at the end of the time series. The two most interesting results from these regressions is that (a) firms that use sell-side Wall Street research have lower returns, and (b) firms that employ Ph.D.s have higher returns. We stress that these are simply correlations; these results imply no causality. For instance, we can only infer that better performing firms have more Ph.D.s. We cannot infer that having Ph.D.s improved these firms' returns, or that higher returns allowed them to hire Ph.D.s.

[8] Kosowski et al. (2006) present their main results when they bootstrap the residuals for each product *independently*. Fama and French (2008) sample the product and factor returns *jointly* to better account for common variation in product returns not accounted for by factors, and correlated movement in the volatilities of factor returns and residuals.

Table III
Luck versus Skill in Performance

Performance is measured using four-factor alphas, similar to that in Table II. The table shows percentiles of actual and (average) simulated alphas and their t-statistics. The details of the simulation are described in the text. We also show the percentage of simulation draws that produce an alpha/t-statistic greater than the corresponding actual value. Alphas are in percent per quarter. The sample period is 1991 to 2008.

	Alphas			t-Statistics		
Pct	Actual	Sim	% (Sim > Actual)	Actual	Sim	% (Sim > Actual)
Panel A: Gross Returns (Number of Products = 3,842)						
1	−1.65	−2.01	18.20	−2.36	−2.74	22.20
2	−1.34	−1.58	23.40	−1.96	−2.34	19.70
3	−1.19	−1.37	27.80	−1.73	−2.11	16.80
4	−1.06	−1.22	27.70	−1.55	−1.95	14.60
5	−0.96	−1.11	26.50	−1.44	−1.82	14.60
10	−0.61	−0.80	15.70	−1.04	−1.40	13.00
20	−0.32	−0.49	14.10	−0.58	−0.92	11.30
30	−0.13	−0.30	9.20	−0.25	−0.59	8.80
40	0.02	−0.15	7.40	0.05	−0.31	7.20
50	0.18	−0.02	5.40	0.35	−0.05	5.50
60	0.32	0.10	5.80	0.63	0.20	4.80
70	0.46	0.25	7.60	0.90	0.48	6.30
80	0.66	0.43	9.10	1.23	0.80	7.10
90	1.03	0.74	8.50	1.69	1.27	8.90
95	1.45	1.06	7.50	2.08	1.67	10.80
96	1.61	1.16	6.70	2.21	1.79	11.10
97	1.79	1.31	7.00	2.39	1.95	11.90
98	2.06	1.52	6.80	2.58	2.17	15.50
99	2.74	1.94	4.10	2.91	2.55	20.80
Panel B: Net Returns (Number of Products = 2,987)						
1	−1.96	−1.97	53.60	−2.68	−2.67	52.00
2	−1.58	−1.57	55.30	−2.33	−2.30	55.00
3	−1.44	−1.35	64.70	−2.08	−2.08	51.80
4	−1.32	−1.21	68.90	−1.95	−1.92	54.80
5	−1.19	−1.10	65.80	−1.80	−1.80	52.00
10	−0.79	−0.79	53.00	−1.40	−1.38	53.20
20	−0.51	−0.48	56.40	−0.92	−0.91	52.00
30	−0.32	−0.30	56.20	−0.58	−0.58	49.70
40	−0.15	−0.15	48.40	−0.28	−0.31	46.40
50	0.00	−0.02	40.60	0.00	−0.05	40.60
60	0.15	0.10	34.10	0.29	0.20	35.90
70	0.29	0.24	34.60	0.57	0.47	33.00
80	0.49	0.43	29.70	0.92	0.79	29.50
90	0.86	0.73	22.30	1.36	1.25	33.10
95	1.25	1.04	19.30	1.77	1.65	33.70
96	1.41	1.15	17.30	1.92	1.77	32.20
97	1.54	1.29	19.00	2.05	1.92	34.10
98	1.80	1.50	17.70	2.23	2.13	38.80
99	2.47	1.89	9.70	2.61	2.49	38.10

III. Persistence

Persistence in performance is important from an economic and practical perspective. From an economic view, if prior-period performance can be used to predict future returns, this represents a significant challenge to market efficiency. From a plan sponsor's perspective, performance represents an important (but not the only) screening mechanism. If little or no persistence exists in institutional product returns, then any attempt to select superior performers is likely futile.

We use two empirical approaches to measure persistence. Our first approach follows the mutual fund literature, with minor adjustments to accommodate certain facets of institutional investment management. The second approach uses Fama–MacBeth (1973) style cross-sectional regressions to get at persistence while controlling for other variables.

A. Persistence across Deciles

We follow Carhart (1997) and form deciles during a ranking period and then examine returns over a subsequent post-ranking period. However, unlike Carhart (1997), we form deciles based on benchmark-adjusted returns rather than raw returns for two reasons. First, plan sponsors frequently focus on benchmark-adjusted returns, at least in part because expected returns from benchmarks are useful for thinking about broader asset allocation decisions in the context of contributions and retirement withdrawals. Second, sorting on raw returns could cause portfolios that follow certain types of investment styles to systematically fall into winner and loser deciles. For example, small cap value portfolios may fall into winner deciles in some periods, not because these portfolios delivered abnormal returns, but because this asset class generated large returns over that period (see Elton et al. (1993)). Using benchmark-adjusted returns to form deciles circumvents this problem.

Beginning at the end of 1991, we sort portfolios into deciles based on the prior annual benchmark-adjusted return. We then compute the equal-weighted return for each decile over the following quarter. As we expand our analysis to examine persistence over longer horizons, we compute this return over appropriate future intervals (e.g., for the first year, our holding period is from quarter one through quarter four; for the second year, the holding period is from quarter five through quarter eight, etc.). We then roll forward, producing a non-overlapping set of post-ranking quarterly returns. Concatenating the post-ranking period quarterly returns results in a time series of post-ranking returns for each portfolio; we generate estimates of abnormal performance from these time series.

Similar to our assessment of average performance in the previous section, we assess post-ranking abnormal performance by regressing the post-ranking gross returns on K factors as follows:

$$r_{d,t} - r_{f,t} = \alpha_d + \sum_{k=1}^{K} \beta_{d,k} f_{k,t} + \epsilon_{d,t}, \tag{3}$$

<div align="center">

Table IV

Performance Persistence across Deciles

</div>

We sort products in deciles according to the benchmark-adjusted return during the ranking period of 1 year. We hold the decile portfolios for post-ranking periods ranging from one quarter to 3 years. We rebalance the portfolios at the end of every quarter when the holding period is one quarter and at the end of every year otherwise. Factor models are the same as those in Table II. All alphas are in percent per quarter, and *t*-statistics are reported in parentheses next to alphas. Decile 1 contains the worst-performing products, and decile 10 contains the best-performing products. The sample period is 1991 to 2008.

Decile	1st Quarter	1st Year	2nd Year	3rd Year
		Panel A: 1-Factor Alphas		
1	0.30 (0.91)	0.44 (1.70)	0.70 (3.07)	0.52 (2.07)
2	0.30 (1.14)	0.40 (1.56)	0.59 (2.73)	0.52 (2.55)
3	0.27 (1.21)	0.34 (1.67)	0.41 (2.32)	0.52 (2.81)
5	0.35 (2.02)	0.38 (2.24)	0.42 (2.58)	0.32 (1.84)
8	0.52 (2.58)	0.49 (2.45)	0.46 (2.22)	0.33 (1.49)
9	0.63 (2.49)	0.63 (2.76)	0.34 (1.41)	0.31 (1.19)
10	1.17 (2.26)	0.78 (1.79)	0.24 (0.63)	0.28 (0.76)
10–1	0.87 (1.30)	0.34 (0.71)	−0.45 (−1.22)	−0.24 (−0.72)
		Panel B: 3-Factor Alphas		
1	−0.18 (−0.66)	0.08 (0.37)	0.52 (2.82)	0.35 (1.65)
2	−0.10 (−0.46)	−0.01 (−0.06)	0.29 (1.65)	0.28 (1.75)
3	−0.05 (−0.28)	0.05 (0.29)	0.15 (1.11)	0.29 (2.06)
5	0.07 (0.49)	0.10 (0.76)	0.19 (1.43)	0.07 (0.63)
8	0.26 (1.65)	0.23 (1.48)	0.22 (1.37)	0.09 (0.56)
9	0.52 (2.62)	0.49 (2.80)	0.11 (0.62)	0.09 (0.41)
10	1.52 (3.55)	0.96 (2.79)	0.17 (0.57)	0.20 (0.61)
10–1	1.70 (2.88)	0.88 (2.09)	−0.35 (−0.97)	−0.16 (−0.47)
		Panel C: 4-Factor Alphas		
1	0.32 (1.20)	0.30 (1.36)	0.37 (1.85)	0.07 (0.30)
2	0.27 (1.19)	0.16 (0.72)	0.20 (1.04)	0.10 (0.58)
3	0.16 (0.77)	0.19 (0.96)	0.12 (0.74)	0.13 (0.88)
5	0.07 (0.47)	0.14 (0.90)	0.15 (1.02)	0.06 (0.43)
8	0.06 (0.35)	0.07 (0.40)	0.16 (0.91)	0.06 (0.34)
9	−0.06 (−0.42)	0.04 (0.30)	0.09 (0.45)	0.01 (0.05)
10	0.18 (0.66)	0.00 (0.00)	−0.04 (−0.13)	−0.09 (−0.25)
10–1	−0.15 (−0.39)	−0.30 (−0.95)	−0.41 (−1.03)	−0.15 (−0.41)

where r_d is the return for decile d, and f_k is the kth factor return. We use factors identical to those described above in equation (1).

Table IV shows alphas corresponding to CAPM, three-factor, and four-factor models in Panels A, B, and C, respectively. We report four different post-ranking horizons: one quarter to draw inferences about short-term persistence and the first, second, and third year after ranking. Note that the second- and third-year alphas are not multiyear alphas; for example, the second-year alpha pertains

to the performance of the decile in the second year after ranking, not the performance over the first *and* second year. This follows convention and avoids overlapping observations in the statistical tests.[9]

We estimate alphas for each decile and horizon, but this generates a large number of statistics to report. To avoid overwhelming the reader with a barrage of numbers, we only report alphas for some of the deciles. We report select loser deciles (1, 2, and 3) for comparison with the mutual fund literature. We also report alphas for three winner deciles (8, 9, and 10) and one intermediate decile (5). The variation in benchmark-adjusted returns used to create deciles is large (ranging from −13.6% in decile 1 to 22.2% for decile 10).

We start with a discussion of losers. There is a modest reversal in fortunes for products in the extreme loser deciles, relative to the CAPM, and to some degree, the three-factor model. For instance, the extreme loser decile has an alpha of 0.52% (0.35%) per quarter in the second (third) year after ranking, with a *t*-statistic of 2.82 (1.65). Controlling for momentum, these estimates and their statistical significance drop sharply: The corresponding four-factor alpha is 0.37% (0.07%) with a *t*-statistic of 1.85 (0.30). Modest reversion could arise from two sources: economies of scale and look-ahead issues. With regard to the first, if a product performs poorly and loses assets, its future performance may improve simply because it is better able to manage a smaller asset base. For example, this can arise from proportionately lower trading costs. However, measuring this requires detailed information on the cost function of each product. Such data are not available to us, but broad evidence for mutual funds reported in Chen et al. (2004) is consistent with such an effect; in the Internet Appendix,[10] we report regressions that are similar to those of Chen et al. (2004) for our sample. We can, however, get a handle on look-ahead issues. As Carpenter and Lynch (1999) and Horst et al. (2001) point out, look-ahead issues in nonsurvivorship-biased samples could generate spurious reversals. Since poorly performing products are more likely to disappear, the expected attrition rate in decile 1 is higher than decile 10. In our data, by the third year after decile formation, decile 1 has lost 19% of its constituents, whereas decile 10 has lost only 8%. The average benchmark-adjusted return in the last year before disappearing for portfolios in decile 1 is −5.4% and 3.5% for decile 10.[11] We gauge the impact of differential attrition on loser deciles with a simple exercise. Specifically, we assume the population alpha is normally distributed with mean zero and standard deviation σ_α. We then calculate the mean of a truncated normal distribution for various values of σ_α and degrees of truncation. The Appendix at the end of this article reports the results of this exercise.

---

[9] For readers interested in multiyear (2- and 3-year) alphas, we provide these in the Internet Appendix.

[10] An Internet Appendix for this article is available online in the "Supplements and Datasets" section at http://www.afajof.org/supplements.asp.

[11] We also compare these attrition rates and last-year excess returns to those in mutual fund portfolios and find that both are higher for mutual funds than for institutional funds. For instance, the 3-year attrition rate in domestic equity mutual funds is 25% (versus 18% for our sample), and the last-year excess return is −22% (versus −6% for our sample).

Under a true null of zero alpha, and assuming a σ_α of 0.8% (inferred from the cross-sectional distribution of alphas in Table II), 20% truncation in population imparts an upward bias in alpha of 0.28%. This adjustment would reduce the one-, three-, and four-factor alphas of the loser decile in the third year after portfolio formation to 0.24, 0.07, and −0.21%, respectively (versus 0.52, 0.35, and 0.07 reported in Table IV), all of which would be statistically insignificant.

With respect to winners, over a one-quarter horizon, some evidence of persistence exists in winner deciles, at least based on the one- and three-factor models. For example, the alpha for decile 10 under the CAPM (three-factor) model is 1.17% (1.52%) per quarter with a *t*-statistic of 2.26 (3.55). The spread between the extreme winner and loser decile (10-1) is also high, and in the case of the three-factor model, statistically significant. However, as Carhart (1997) shows and Fama and French (2008) confirm, momentum plays an important role in persistence. Winner deciles are more likely to have winner stocks, and given individual stock momentum, this mechanically generates persistence among winners. Consistent with this, persistence in the extreme winner decile falls dramatically to a statistically insignificant 0.18% after controlling for momentum. By comparison, Bollen and Busse (2005) report a one-quarter post-ranking four-factor alpha of 0.39% for the winner decile of mutual funds.

Although short-term (one-quarter) persistence is interesting from an economic perspective, plan sponsors do not deploy capital for such short horizons. The transaction costs (known as transition costs) from exiting a product after one quarter and entering a new one are large, in addition to adverse reputation effects from rapidly trading in and out of institutional products. Persistence over long horizons is also more important from a practical and economic perspective. Long-horizon persistence represents a violation of market efficiency and a potentially value-increasing opportunity for plan sponsors.

One year after decile formation, the three-factor alpha for the extreme winner decile is high (0.96%) and statistically significant (*t*-statistic of 2.79). But once again, controlling for momentum reduces the alpha to 0.00% and eliminates the statistical significance. In the second and third year after decile formation, no evidence of persistence exists using either the three- or four-factor model in any of the winner deciles. Thus, over the long horizons over which plan sponsors typically conduct their investments, minimal evidence of persistence exists.

B. Regression-Based Evidence

Our second approach to measuring persistence involves estimating Fama–MacBeth (1973) regressions of future returns on lagged returns at various horizons and aggregating coefficients over time:

$$
\begin{aligned}
r_{p,t+k_1:t+k_2} &- r_{b,t+k_1:t+k_2} \\
&= \gamma_{0,t} + \gamma_{1,t}(r_{p,t-k:t} - r_{b,t-k:t}) + \gamma_{2,t}Z_{p,t} + e_{p,t}, \quad \text{or}
\end{aligned}
\tag{4}
$$

$$(r_{p,t+k_1:t+k_2} - r_{f,t+k_1:t+k_2}) - \sum_{k=1}^{K} \widehat{\beta_{p,k}} f_{k,t+k_1:t+k_2} \tag{5}$$

$$= \gamma_{0,t} + \gamma_{1,t}(r_{p,t-k:t} - r_{b,t-k:t}) + \gamma_{2,t}Z_{p,t} + e_{p,t},$$

where $t + k_1$ to $t + k_2$ is the horizon over which we measure future returns, $t - k$ to t is the horizon over which we measure the lagged returns, and Z_p represents control variables. The first specification considers persistence in benchmark-adjusted returns, while the second specification considers the same in risk-adjusted returns. Following Brennan, Chordia, and Subrahmanyam (1998), we estimate the betas in the second specification with a first-pass time-series regression using the whole sample. In addition, because the betas in the second specification are estimated, we adjust the γ coefficients using the Shanken (1992) errors-in-variables correction. Finally, we adjust the Fama–MacBeth (1973) standard errors for autocorrelation because of the overlap in the dependent and/or independent variables.

This approach has three advantages. First, it offers more flexibility in exploring horizons over which persistence may exist. Second, it allows us to examine persistence in benchmark-adjusted returns. To the extent that practitioners use benchmark adjustments to help them decide where to allocate capital, examining persistence in benchmark-adjusted returns may be of interest. Third, it allows us to directly control for product-level attributes (such as total assets) that may affect future returns and persistence. This is particularly useful in the presence of diseconomies of scale.

Table V presents the results of these regressions. We use four dependent variables corresponding to benchmark-adjusted returns, and one-, three-, and four-factor model risk-adjusted returns described in equations (4) and (5). As in the decile-based tests, the horizons of the dependent variables are the first quarter and the first, second, and third year ahead.

In Panel A, the explanatory variables are the previous quarter's return, the prior year's return, the prior 2-year holding period return, or the prior 3-year holding period return. Panel B augments these regressions with 1-year lagged assets under management for the product as well as the firm, and with prior-year flows.

Consistent with the results reported in Table IV, modest evidence of persistence exists at the one-quarter and 1-year horizon with one- and three-factor model adjusted returns; the coefficients on lagged returns over these horizons are positive and generally statistically significant. For instance, when we regress the 1-year-ahead three-factor model return against lagged 1-year returns, the average coefficient is 0.090 with a t-statistic of 2.07. The addition of control variables in Panel B, however, reduces this coefficient to 0.072 and the t-statistic to 1.53. Even without these control variables, the addition of momentum drops the corresponding coefficient to 0.071 and the t-statistic to 1.41. Over horizons longer than a year for either the dependent variable or the independent variables, there is no statistical evidence of persistence.

<div align="center">

Table V

Persistence Regressions

</div>

We estimate the following Fama-MacBeth (1973) cross-sectional regression:

$$r_{p,t+k_1:t+k_2} - r_{b,t+k_1:t+k_2}$$

$$= \gamma_{0,t} + \gamma_{1,t}(r_{p,t-k:t} - r_{b,t-k:t}) + \gamma_{2,t}Z_{p,t} + e_{p,t}, \text{ or}$$

$$(r_{p,t+k_1:t+k_2} - r_{f,t+k_1:t+k_2}) - \sum_{k=1}^{K} \widehat{\beta_{p,k}} f_{k,t+k_1:t+k_2}$$

$$= \gamma_{0,t} + \gamma_{1,t}(r_{p,t-k:t} - r_{b,t-k:t}) + \gamma_{2,t}Z_{p,t} + e_{p,t},$$

where $t + k_1$ to $t + k_2$ is the horizon over which we measure future returns, $t - k$ to t is the horizon over which we measure the lagged returns, and Z_p represents control variables. The first specification uses benchmark-adjusted returns, while the second uses risk-adjusted returns (we estimate the betas in the second specification using a first-pass time-series regression using the whole sample). The controls included in Panel B are lagged asset size at the product- and firm-level and lagged cash flows. We estimate the regressions each quarter when the dependent variable is a quarterly return and each year otherwise. The table reports the time-series average of the γ_1 coefficient along with t-statistics (corrected for autocorrelation) in parentheses. In addition, because the betas in the second specification are estimated, we adjust the γ coefficients using the Shanken (1992) errors-in-variables correction. The sample period is 1991 to 2008.

Horizon of Returns			Future Return Adjustment		
				Factor Model	
Future $t + k_1 : t + k_2$	Lagged $t - k : t$	Benchmark	1-Factor	3-Factor	4-Factor
Panel A: No Additional Controls					
1st quarter	1-quarter	0.077 (1.90)	0.031 (0.99)	0.047 (2.06)	0.040 (2.49)
	1-year	0.046 (3.51)	0.015 (1.18)	0.029 (3.38)	0.025 (3.99)
	2-years	0.025 (3.24)	0.010 (1.09)	0.016 (2.76)	0.010 (2.36)
	3-years	0.015 (2.78)	0.007 (1.13)	0.009 (2.22)	0.003 (0.93)
1st year	1-quarter	0.441 (3.51)	0.149 (1.64)	0.231 (2.24)	0.183 (2.01)
	1-year	0.147 (2.40)	0.064 (1.49)	0.090 (2.07)	0.071 (1.41)
	2-years	0.082 (3.25)	0.027 (0.64)	0.044 (1.66)	0.015 (0.53)
	3-years	0.030 (1.50)	0.011 (0.36)	0.018 (1.02)	−0.004 (−0.19)
2nd year	1-quarter	−0.170 (−0.94)	−0.182 (−1.72)	−0.133 (−1.71)	−0.078 (−0.80)
	1-year	0.071 (1.13)	−0.016 (−0.17)	0.021 (0.38)	−0.019 (−0.28)
	2-years	0.012 (0.43)	−0.011 (−0.22)	0.001 (0.05)	−0.028 (−0.73)
	3-years	0.024 (1.20)	−0.014 (−0.37)	0.001 (0.06)	−0.017 (−0.56)
3rd year	1-quarter	−0.119 (−1.10)	−0.184 (−1.06)	−0.166 (−1.28)	−0.138 (−0.88)
	1-year	0.002 (0.04)	−0.003 (−0.06)	−0.019 (−0.65)	−0.045 (−1.27)
	2-years	0.046 (1.20)	−0.011 (−0.19)	0.002 (0.09)	−0.018 (−0.41)
	3-years	0.009 (0.39)	−0.002 (−0.07)	0.009 (0.47)	−0.010 (−0.39)
Panel B: With Additional Controls					
1st year	1-quarter	0.430 (3.41)	0.126 (1.25)	0.229 (2.05)	0.185 (1.99)
	1-year	0.136 (2.15)	0.049 (0.93)	0.072 (1.53)	0.064 (1.21)
	2-years	0.082 (3.12)	0.018 (0.38)	0.037 (1.23)	0.009 (0.28)
	3-years	0.028 (1.36)	0.004 (0.10)	0.011 (0.54)	−0.011 (−0.41)

Table V—*Continued*

Horizon of Returns			Future Return Adjustment		
			Factor Model		
Future $t + k_1 : t + k_2$	Lagged $t - k : t$	Benchmark	1-Factor	3-Factor	4-Factor
Panel B: With Additional Controls					
2nd year	1-quarter	−0.203 (−1.15)	−0.209 (−1.72)	−0.160 (−1.99)	−0.090 (−0.90)
	1-year	0.041 (0.53)	−0.074 (−0.60)	−0.022 (−0.28)	−0.057 (−0.57)
	2-years	0.003 (0.08)	−0.031 (−0.49)	−0.016 (−0.44)	−0.045 (−0.82)
	3-years	0.020 (0.95)	−0.024 (−0.51)	−0.010 (−0.38)	−0.027 (−0.65)
3rd year	1-quarter	−0.147 (−1.08)	−0.236 (−1.08)	−0.220 (−1.24)	−0.185 (−0.87)
	1-year	−0.014 (−0.32)	−0.026 (−0.46)	−0.058 (−1.38)	−0.077 (−1.42)
	2-years	0.040 (0.97)	−0.024 (−0.36)	−0.015 (−0.41)	−0.036 (−0.62)
	3-years	0.000 (−0.01)	−0.003 (−0.08)	0.001 (0.05)	−0.019 (−0.61)

It is common practice among plan sponsors and investment consultants to focus on performance and persistence using benchmark-adjusted returns. The results using benchmark-adjusted returns are, in fact, different from those using factor models. Panels A and B show considerable persistence in 1-year forward benchmark-adjusted returns. This predictability is evident with 1- and 2-year holding period returns (and marginally even 3-year holding period returns). Thus, using benchmark-adjusted returns, a plan sponsor could reasonably argue that using prior performance to pick institutional products and accordingly allocate capital is appropriate.

Overall, judging a variety of results across different factor models and estimation techniques, we believe that the results show little to no evidence of persistence in institutional portfolios.

IV. Robustness

A. Data Veracity

A natural concern in the measurement of performance and persistence is the degree to which the data accurately represent the population. A common concern is whether the data adequately represent poorly performing products. This issue could arise from backfill bias (Liang (2000)), the use of incubation techniques by investment management firms (Evans (2007)), or similar selective reporting. These problems likely can never be perfectly detected or eliminated. Consequently, we follow two approaches to determine their potential impact on our results. First, we follow Jagannathan, Malakhov, and Novikov (2006) and eliminate the first 3 years of returns for each product and re-run our main tests. Our regressions are inevitably noisier, but our basic results are unchanged and

reported in the Internet Appendix.[12] Second, we return to the Appendix at the end of this article to determine the impact of truncation. Recall that the three-factor (four-factor) alphas in our sample for equal-weighted gross returns are 0.35% (0.20%) per quarter. If we assume a σ_α of 0.80% based on analysis reported in Panel B of Table II, then the Appendix shows that over 25% (13%) of the left tail of the distribution would have to be truncated to produce an alpha of 0.35% (0.20%). To us, such a high degree of truncation seems implausible.

B. Performance Measurement

In addition to data issues, performance measurement is subject to a variety of concerns about choice of asset pricing models, model specification, and estimation issues. Despite numerous rejections of the standard CAPM, we continue to report CAPM alphas because it at least represents an equilibrium model of expected returns. We also report Fama–French (1993) alphas because of their ability to capture cross-sectional variation in returns and Carhart's (1997) four-factor alphas because of their widespread acceptance in this literature. However, at least two other approaches are important to consider: using conditioning information and expanded factor models.

Christopherson et al. (1998) argue that unconditional performance measures are inappropriate for two reasons (see also Ferson and Schadt (1996)). First, they note that sophisticated plan sponsors presumably condition their expectations based on the state of the economy. Second, to the extent that plan sponsors employ dynamic trading strategies that react to changes in market conditions, unconditional performance indicators may be biased. They advocate conditional performance measures and show that such measures can improve inference. We follow their prescription and estimate conditional models in addition to the unconditional models described earlier. We estimate the conditional models as

$$r_{p,t} = \alpha_p^C + \sum_{k=1}^{K} \left(\beta_{p,k}^0 + \sum_{l=1}^{L} \beta_{p,k}^l Z_{l,t-1} \right) f_{k,t} + \epsilon_{p,t}, \tag{6}$$

where the Z's are L conditioning variables.

We use four conditioning variables in our analysis. We obtain the 3-month T-bill rate from the economic research database at the Federal Reserve Bank in St. Louis. We compute the default yield spread as the difference between BAA- and AAA-rated corporate bonds using the same database. We obtain the dividend-price ratio, computed as the logarithm of the 12-month sum of dividends on the S&P 500 index divided by the logarithm of the index level from Standard & Poor's. Finally, we compute the term yield spread as the difference between the long-term yield on government bonds and the T-bill yield using data from Ibbotson Associates.

---

[12] We cannot use the approach recommended by Evans (2007) because institutional products do not have tickers supplied by NASD. However, the Jagannathan et al. (2006) procedure of removing data close to inception dates is similar in spirit and the 3-year horizon is quantitatively close to Evans (2007).

Table VI
Alternative Factor Models

Conditional alphas are calculated from

$$r_{d,t} = \alpha_d^C + \sum_{k=1}^{K} \left(\beta_{d,k}^0 + \sum_{l=1}^{L} \beta_{d,k}^l Z_{l,t-1} \right) f_{k,t} + \epsilon_{d,t},$$

where f's are K factors and Z's are L instruments. Instruments include the 3-month T-bill rate, the dividend price ratio for the S&P 500, the term spread, and the default spread. Seven-factor alphas are calculated from the factor model proposed by Cremers et al. (2008). All alphas are in percent per quarter, and t-statistics are reported in parentheses next to alphas. The sample period is 1991 to 2008.

Decile	1st Quarter	1st Year	2nd Year	3rd Year
	Panel A: Conditional 3-Factor Alphas			
1	0.09 (0.37)	0.29 (1.42)	0.67 (3.17)	0.51 (1.82)
2	0.02 (0.08)	0.06 (0.35)	0.37 (2.26)	0.35 (1.91)
3	0.06 (0.35)	0.06 (0.38)	0.29 (2.36)	0.27 (1.64)
5	0.16 (1.38)	0.13 (1.20)	0.27 (2.22)	0.13 (1.18)
8	0.21 (1.46)	0.29 (2.06)	0.32 (2.06)	0.15 (1.18)
9	0.53 (2.40)	0.57 (3.19)	0.14 (0.79)	0.11 (0.54)
10	1.58 (3.08)	1.24 (3.45)	0.14 (0.41)	0.32 (0.98)
10–1	1.49 (2.27)	0.95 (2.20)	−0.53 (−1.33)	−0.20 (−0.54)
	Panel B: Conditional 4-Factor Alphas			
1	0.37 (1.68)	0.20 (0.88)	0.40 (2.07)	0.08 (0.34)
2	0.25 (1.37)	0.08 (0.42)	0.19 (1.13)	0.12 (0.69)
3	0.20 (1.13)	0.11 (0.59)	0.13 (1.10)	0.10 (0.63)
5	0.12 (0.91)	0.10 (0.87)	0.13 (1.05)	0.06 (0.48)
8	−0.03 (−0.24)	0.10 (0.68)	0.16 (0.96)	0.10 (0.76)
9	0.00 (0.01)	0.15 (1.23)	0.04 (0.20)	0.01 (0.04)
10	0.30 (1.07)	0.43 (1.77)	−0.13 (−0.40)	−0.11 (−0.34)
10–1	−0.07 (−0.19)	0.23 (0.65)	−0.54 (−1.42)	−0.19 (−0.59)
	Panel C: Unconditional 7-Factor Alphas			
1	0.46 (2.12)	0.50 (2.55)	0.50 (2.59)	0.23 (1.07)
2	0.32 (1.90)	0.28 (1.54)	0.26 (1.67)	0.19 (1.33)
3	0.16 (0.99)	0.19 (1.16)	0.13 (1.13)	0.20 (1.76)
5	0.09 (0.76)	0.11 (1.06)	0.17 (1.60)	0.10 (1.10)
8	0.10 (0.75)	0.12 (0.99)	0.26 (1.95)	0.18 (1.34)
9	0.12 (0.88)	0.18 (1.51)	0.23 (1.50)	0.16 (0.88)
10	0.59 (2.14)	0.37 (1.48)	0.27 (0.96)	0.18 (0.62)
10–1	0.13 (0.34)	−0.13 (−0.39)	−0.23 (−0.59)	−0.05 (−0.12)

Table VI reports the results of these regressions. Panel A (B) shows conditional alphas for the persistence tests based on three- (four-) factor models. We urge some caution in interpretation because of potential overparameterization. Four factors and four conditioning variables result in 21 independent variables, which obviously reduces the degrees of freedom in our regressions. At the same time, we find that the average $\bar{R}^2$ improvement from conditional

models over their unconditional counterparts is very modest. Despite these issues, the results largely mirror the unconditional estimates reported earlier; there is evidence of persistence in three-factor models 1 year after ranking but it shrinks appreciably in both magnitude and statistical significance once we control for momentum. For instance, the 1-year three-factor post-ranking alpha of the extreme winner decile is 1.24% per quarter with a t-statistic of 3.45, and the extreme winner-minus-loser spread is 0.95 (t-statistic = 2.20). But, controlling for the mechanical effects of momentum, the four-factor alpha for the extreme winner decile drops to 0.43% with a t-statistic of 1.77 (and the winner-minus-loser spread declines to 0.23 with a t-statistic of 0.65). Similar to the conditional models, there is no persistence beyond the first year.

We also consider factor models augmented by index portfolios. Cremers et al. (2008) report that, when evaluated using three- and four-factor models, passive portfolios such as the S&P 500 and Russell 2000 produce economically meaningful alphas. They argue that this is because these factor models place disproportionate weight on small and/or value stocks; one of their proposed "solutions" is to estimate a seven-factor model that is based on index returns. This model, labeled "IDX7" in their paper, includes the following factors: the S&P 500, the difference between the Russell Midcap and the S&P 500, the difference between the Russell 2000 and the Russell Midcap, the difference between the S&P 500 Value and S&P 500 Growth, the difference between the Russell Midcap Value and Russell Midcap Growth, the difference between the Russell 2000 Value and Russell 2000 Growth, as well as the momentum factor. We repeat our analysis with this seven-factor model and report the results in Panel C. Here the results are weaker than the conditional models. The 1-year post-ranking alpha of the extreme winner decile is 0.37% with a t-statistic of 1.48. This is higher than the four-factor unconditional model results but lower than the conditional model results reported in Panels A and B.

Thus, while there are inevitably differences in parameter estimates between the unconditional and conditional models, and between the four- and seven-factor models, the basic flavor of the results is similar.

C. Statistical Significance

Finally, we gauge statistical significance in the paper with regular t-statistics. Kosowski et al. (2006) recommend a bootstrap procedure to alleviate problems of small sample and skewness in the returns data. We follow their approach to calculate bootstrapped standard errors. Specifically, we assume that equations (1) and (6) with zero alpha describe the data-generating process. We generate each draw by choosing at random (with replacement) the saved residuals. We then compute alphas and their t-statistics from the bootstrapped sample. We repeat this procedure 1,000 times to obtain the empirical distribution of t-statistics. The statistical significance using critical values from this bootstrapped distribution is very similar to that reported in the rest of the paper.

V. Conclusion

In this paper, we examine the performance and persistence in performance of 4,617 active domestic equity institutional products managed by 1,448 investment management firms between 1991 and 2008. Plan sponsors use these products as investment vehicles to delegate portfolio management in investment styles across all size and value-growth gradations.

Before fees, we find little evidence of superior performance, either in aggregate or on average. However, even if no evidence of aggregate superior performance exists, it could be the case that some investment managers deliver superior returns over long periods. In addition to its own economic interest, such persistence is of enormous practical value, since plan sponsors routinely use performance in allocating capital to these firms. Our estimates of persistence are sensitive to the choice of models. Three-factor models show modest evidence of persistence and would cause an investor to use performance as a screening device. In contrast, four-factor models that incorporate the mechanical effects of stock momentum show little to no persistence. Conditional four-factor models and seven-factor models paint a similar picture.

Appendix: Mean of a Truncated Normal Distribution of Alphas

We assume the population α is distributed normally with mean zero and standard deviation σ_α. The table reports the mean of a truncated distribution where the left tail of the distribution is truncated (unobserved).

	Fraction of Population That Is Left-Truncated										
σ_α	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
0.1%	0	0.01	0.02	0.03	0.03	0.04	0.05	0.06	0.06	0.07	0.08
0.2%	0	0.02	0.04	0.05	0.07	0.08	0.10	0.11	0.13	0.14	0.16
0.3%	0	0.03	0.06	0.08	0.10	0.13	0.15	0.17	0.19	0.22	0.24
0.4%	0	0.04	0.08	0.11	0.14	0.17	0.20	0.23	0.26	0.29	0.32
0.5%	0	0.05	0.10	0.14	0.17	0.21	0.25	0.28	0.32	0.36	0.40
0.6%	0	0.07	0.12	0.16	0.21	0.25	0.30	0.34	0.39	0.43	0.48
0.7%	0	0.08	0.14	0.19	0.24	0.30	0.35	0.40	0.45	0.50	0.56
0.8%	0	0.09	0.16	0.22	0.28	0.34	0.40	0.46	0.52	0.58	0.64
0.9%	0	0.10	0.18	0.25	0.31	0.38	0.45	0.51	0.58	0.65	0.72
1.0%	0	0.11	0.19	0.27	0.35	0.42	0.50	0.57	0.64	0.72	0.80
1.1%	0	0.12	0.21	0.30	0.38	0.47	0.55	0.63	0.71	0.79	0.88
1.2%	0	0.13	0.23	0.33	0.42	0.51	0.60	0.68	0.77	0.86	0.96
1.3%	0	0.14	0.25	0.36	0.45	0.55	0.65	0.74	0.84	0.94	1.04
1.4%	0	0.15	0.27	0.38	0.49	0.59	0.70	0.80	0.90	1.01	1.12
1.5%	0	0.16	0.29	0.41	0.52	0.64	0.75	0.85	0.97	1.08	1.20
1.6%	0	0.17	0.31	0.44	0.56	0.68	0.79	0.91	1.03	1.15	1.28
1.7%	0	0.18	0.33	0.47	0.59	0.72	0.84	0.97	1.09	1.22	1.36
1.8%	0	0.20	0.35	0.49	0.63	0.76	0.89	1.03	1.16	1.30	1.44
1.9%	0	0.21	0.37	0.52	0.66	0.81	0.94	1.08	1.22	1.37	1.52
2.0%	0	0.22	0.39	0.55	0.70	0.85	0.99	1.14	1.29	1.44	1.60

REFERENCES

Avramov, Doron, and Russ Wermers, 2006, Investing in mutual funds when returns are predictable, *Journal of Financial Economics* 81, 339–377.

Barras, Laurent, Olivier Scaillet, and Russ Wermers, 2009, False discoveries in mutual fund performance: Measuring luck in estimated alphas, *Journal of Finance* 65, 179–216.

Berk, Jonathan, and Jing Xu, 2004, Persistence and fund flows of the worst performing mutual funds, Working paper, University of California at Berkeley.

Bollen, Nicolas, and Jeffrey Busse, 2005, Short-term persistence in mutual fund performance, *Review of Financial Studies* 18, 569–597.

Brennan, Michael, Tarun Chordia, and Avanidhar Subrahmanyam, 1998, Alternative factor specifications, security characteristics and the cross-section of expected returns, *Journal of Financial Economics* 49, 345–374.

Brown, Stephen, and William Goetzmann, 1995, Performance persistence, *Journal of Finance* 50, 679–698.

Carhart, Mark, 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.

Carpenter, Jennifer, and Anthony Lynch, 1999, Survivorship bias and attrition effects in performance persistence, *Journal of Financial Economics* 54, 337–374.

Chen, Joseph, Harrison Hong, Ming Huang, and Jeffrey Kubik, 2004, Does fund size erode mutual fund performance? The role of liquidity and organization, *American Economic Review* 94, 1276–1302.

Christopherson, Jon, Wayne Ferson, and Debra Glassman, 1998, Conditioning manager alphas on economic information: Another look at the persistence of performance, *Review of Financial Studies* 11, 111–142.

Coggin, T. Daniel, Frank Fabozzi, and Shafiqur Rahman, 1993, The investment performance of U.S. equity pension fund managers: An empirical investigation, *Journal of Finance* 48, 1039–1055.

Cohen, Randolph, Joshua Coval, and Ľuboš Pástor, 2005, Judging fund managers by the company they keep, *Journal of Finance* 60, 1057–1096.

Cremers, Martijn, Antti Petajisto, and Eric Zitzewitz, 2008, Should benchmark indices have alpha? Revisiting performance evaluation, Working paper, Yale University.

Elton, Edwin, Martin Gruber, and Christopher Blake, 1996, The persistence of risk-adjusted mutual fund performance, *Journal of Business* 69, 133–157.

Elton, Edwin, Martin Gruber, Sanjiv Das, and Matthew Hlavka, 1993, Efficiency with costly information: A reinterpretation of the evidence for managed portfolios, *Review of Financial Studies* 6, 1–22.

Evans, Richard, 2007, The incubation bias, Working paper, University of Virginia.

Fama, Eugene, and Kenneth French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, Eugene, and Kenneth French, 2004, The CAPM: Theory and evidence, *Journal of Economic Perspectives* 76, 25–46.

Fama, Eugene, and Kenneth French, 2008, Mutual fund performance, Working paper, University of Chicago.

Fama, Eugene, and James MacBeth, 1973, Risk, return, and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.

Ferson, Wayne, and Kenneth Khang, 2002, Conditional performance measurement using portfolio weights: Evidence for pension funds, *Journal of Financial Economics* 65, 249–282.

Ferson, Wayne, and Rudi Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425–462.

French, Kenneth, 2008, The cost of active investing, Working paper, Dartmouth College.

Goyal, Amit, and Sunil Wahal, 2008, The selection and termination of investment management firms by plan sponsors, *Journal of Finance* 63, 1805–1847.

Grinblatt, Mark, and Sheridan Titman, 1992, The persistence of mutual fund performance, *Journal of Finance* 47, 1977–1984.

Gruber, Martin, 1996, Another puzzle: The growth in actively managed mutual funds, *Journal of Finance* 51, 783–810.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, 1993, Hot hands in mutual funds: Short-run persistence of performance, 1974–1988, *Journal of Finance* 48, 93–130.

Horst, Jenke, Theo Nijman, and Marno Verbeek, 2001, Eliminating look-ahead bias in evaluating persistence in mutual fund performane, *Journal of Empirical Finance* 8, 345–373.

Investment Company Institute, 2008, *2008 Investment Company Fact Book*, Investment Company Institute.

Jagannathan, Ravi, Alexey Malakhov, and Dmitry Novikov, 2006, Do hot hands persist among hedge fund managers? An empirical evaluation, Working paper, Northwestern University.

Jensen, Michael, 1968, The performance of mutual funds in the period 1945–1964, *Journal of Finance* 48, 389–416.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White, 2006, Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis, *Journal of Finance* 61, 2551–2595.

Lakonishok, Josef, Andrei Shleifer, and Robert Vishny, 1992, The structure and performance of the money management industry, *Brookings Papers: Microeconomics*, 339–391.

Liang, Bing, 2000, Hedge funds: The living and the dead, *Journal of Financial and Quantitative Analysis* 35, 309–326.

Shanken, Jay, 1992, On the estimation of beta-pricing models, *Review of Financial Studies* 5, 309–326.

Standard & Poor's, 2007, *The Money Market Directory of Pension Funds and their Investment Managers 2007*, Standard & Poor's.

Tonks, Ian, 2005, Performance persistence of pension fund managers, *Journal of Business* 78, 1917–1942.

Wermers, Russ, 1999, Mutual fund herding and the impact on stock prices, *Journal of Finance* 54, 581–622.

Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses, *Journal of Finance* 55, 1655–1695.