Edwin J. Elton
Martin J. Gruber
New York University

Christopher R. Blake
Fordham University

The Persistence of Risk-Adjusted Mutual Fund Performance*

We examine predictability for stock mutual funds using risk-adjusted returns. We find that past performance is predictive of future risk-adjusted performance. Applying modern portfolio theory techniques to past data improves selection and allows us to construct a portfolio of funds that significantly outperforms a rule based on past rank alone. In addition, we can form a combination of actively managed portfolios with the same risk as a portfólio of index funds but with higher mean return. The portfolios selected have small but statistically significant positive risk-adjusted returns during a period where mutual funds in general had negative risk-adjusted returns.

There is overwhelming evidence that, post expenses, mutual fund managers on average underperform a combination of passive portfolios of similar risk.[1] The recent increase in the number and types of index funds that are available to individual investors makes this a matter of practical as well as theoretical significance. Numerous index funds, which track the Standard and Poor's (S&P) 500 Index or various small-stock, bond, value, growth, or international indexes, are now widely available to individual investors. These same choices have been available to institutional investors for some time. Given that there are sufficient index funds to span most investors' risk choices, that the index funds are available at low cost, and that the low cost of index funds means that a combination of index

* We thank the editor, Douglas Diamond, and the referee, Eugene Fama, for helpful comments. We are also grateful to Investment Company Data, Inc., and to Interactive Data Corporation for supplying data used in this study. This research was supported in part by a grant from Fordham University's Graduate School of Business Administration.

1. Of the few studies that find that managers or a subset of managers with a common objective (such as growth) outperform passive portfolios, most, if not all, would reach opposite conclusions when survivorship bias and/or correct adjustment for risk are taken into account.

(*Journal of Business*, 1996, vol. 69, no. 2)

0021-9398/96/6902-0001$01.50

funds is likely to outperform an active fund of similar risk, the question is, why select an actively managed fund?

An investor could rationally select an actively managed fund if he or she had the ability to select a fund that outperforms the average fund by a sufficient amount to beat a passive portfolio. The principal study to find results that might support this view is that of Hendricks, Patel, and Zeckhauser (1993), who find a "hot hands" effect in short-term predicted returns. Brown, Goetzmann, Ibbotson, and Ross (1992) have shown that survivorship bias can produce an appearance that performance is predictable even when there is no predictability and discuss the results of the Hendricks, Patel, and Zeckhauser study in this context.

This article examines mutual fund predictability for common stock funds, using a sample free of survivorship bias, and measures performance using risk-adjusted returns. We reconfirm the hot hands result that high return can predict high return in the short run. Like Hendricks, Patel, and Zeckhauser, we find that this is a short-run phenomenon. However, using *risk-adjusted* returns to rank funds, we find that past performance is predictive of future risk-adjusted performance in both the short run *and* longer run. Furthermore, when we utilize modern portfolio theory (MPT) techniques to allocate capital among funds, we can construct a portfolio of funds based on prior data that significantly outperforms a rule based on past rank alone and that produces a positive risk-adjusted excess return. In addition, we demonstrate the improvement in performance using MPT by selecting a combination of actively managed portfolios that has the same risk as a portfolio of index funds but has higher mean return. While consistent with past studies, our study finds that expenses account for only part of the differences in performance across funds. We find that there is still predictability even after the major impacts of expenses have been removed. Throughout our study we are able to construct portfolios of funds that have small but statistically significant positive risk-adjusted returns during a period where mutual funds in general had negative risk-adjusted returns.

The article is divided into six sections. In the first section we review the literature. In the second section we discuss our sample selection. This is followed by a section discussing how performance is evaluated and a section discussing our results. In the fifth section, we examine expenses and performance. The last section contains the conclusion.

I. Review of Literature

Many of the mutual fund studies have briefly looked at predictability of performance as part of a larger study of mutual fund performance.

These include Lehmann and Modest (1987), Grinblatt and Titman (1989), Blake, Elton, and Gruber (1993), and Elton, Gruber, Das, and Hlavka (1993). Recently, however, there have been several articles that have directly examined persistence in mutual fund performance. These include Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), Carhart (1994), Brown and Goetzmann (1995), Malkiel (1995), and Sharpe (1995).

Several issues are present in these studies. First, as Brown et al. (1992) show, survivorship bias causes an appearance of persistence even when none exists, and there is substantial potential for survivorship bias in common stock mutual funds (see Elton, Gruber, and Blake, in press). As the authors of the studies recognized, this affects the results reported in Lehmann and Modest (1987) and Grinblatt and Titman (1989) and the results for the larger of the two samples in Blake, Elton, and Gruber (1993). Second, results are not always risk-adjusted (see Hendricks, Patel and Zeckhauser 1993; and Malkiel 1995), or when they are risk-adjusted, the technique used is often inappropriate. For example, one technique that is used by several authors to adjust for risk is to regress returns on a portfolio composed of the top-performing funds, or the difference in return between portfolios of the top- and bottom-performing funds, on a set of indexes and to use the intercept from this regression as a measure of superior performance. The problem with this procedure is that the characteristics of the top-performing funds change significantly over time. In some periods, small-stock funds do best; in other periods, growth funds do best. This implies that the sensitivity of the portfolio of top-performing funds with a single- or multiple-index model is temporally unstable, and therefore a time-series estimate of the intercept on the portfolio is meaningless.

While the methodologies of Carhart (1994) and Sharpe (1995) differ significantly from ours, those studies are closest in spirit to our analysis. Sharpe employs a sample consisting of the 100 largest mutual funds. That sample includes bond funds and international funds, as well as stock funds. Sharpe uses quadratic programming to determine the sensitivities (betas) of a fund to 15 indexes. (This technique has become known as "return style analysis.") He then ranks on risk-adjusted excess return (alpha) and examines whether past alphas are related to future alphas. Carhart employs a sample of stock mutual funds. He uses a four-index model in a manner similar to ours to select high-performing funds. The four indexes he uses are different in definition and, in some cases, in spirit from the four indexes we employ. His methodology for evaluating performance is also different from ours. While both Sharpe and Carhart reach conclusions broadly similar to ours regarding the fact that past performance contains information about future performance, the samples and methodologies used differ among the three studies. Of major importance is that all three articles measure (evaluate) performance in very different ways.

In addition, our article differs from past articles on performance by including three important issues not generally discussed in prior studies. First, authors have not directly addressed the fundamental question of whether a portfolio of actively managed funds can be constructed that consistently beats index funds. Second, how much of the difference in stock fund performance is simply due to expenses? Blake, Elton, and Gruber (1993) find that, for bond mutual funds, a regression of risk-adjusted performance (alpha) on expenses has a slope of about minus one and an insignificant intercept, implying each percentage-point increase in expenses lowers performance by about 1 percentage point and that underperformance is primarily due to expenses. Given the great divergence in stock fund expenses, do those expense differences explain performance differences? Third, can any of the tools of modern portfolio theory be used to select a portfolio of mutual funds that will outperform index funds?

II. Sample

To examine predictability, it is important to construct a sample that is free of survivorship bias. The necessity for this has been pointed out by Brown et al. (1992) and Elton, Gruber, and Blake (in press). Failing to account for survivorship bias can introduce the appearance of predictability when none exists. We designed a sample that is free of survivorship bias.

Our initial sample consisted of all funds that were categorized as "common stock" funds in the 1977 edition of Wiesenberger's *Investment Companies* and that had \$15 million or more in total net assets under management at the end of 1976. We excluded restricted funds from this sample. There were two types of restricted funds. The first type, called "variable annuities" (like the College Retirement Equities Fund), were listed in Wiesenberger but were primarily available through insurance plans. We eliminated these because their purchase was usually tied to an insurance product. The second type were restricted as to purchaser (e.g., a fund that could only be held by Lutherans). These were excluded because a general investor could not purchase them. After excluding funds with less than \$15 million in assets and the restricted funds, 188 funds remain in our sample. All 188 are followed from 1977 to the end of 1993, through name changes and mergers. Thus, our sample is free of survivorship bias.

Having determined our sample, we then collected returns for the funds. We calculate return for each fund on a monthly basis. In calculating return, dividends are assumed to be used to purchase additional shares in the fund at the reinvestment net asset value that is available to shareholders of the fund. This is the assumption made by Morningstar and Investment Company Data, Inc. (ICDI), in constructing their databases.

For funds that existed over the entire period, returns were supplied by ICDI. For funds that ceased to exist, returns were calculated from data supplied by Interactive Data Corporation, supplemented by information from the funds themselves. Merge terms (e.g., merge ratios) were obtained from the funds themselves. (For a more complete discussion of our sample, see Elton, Gruber, and Blake, in press.)

III. Measurement of Performance

In order to measure and compare performance, it is necessary to adjust for the risk of the fund. This is illustrated in Elton et al. (1993), where they showed that failure to include an index of firm size as a risk index leads to a substantial overestimate of the performance of funds that hold small stocks and an incorrect inference concerning average performance. They used a three-index model, including the S&P Index, a size index, and a bond index, to capture the relevant characteristics of performance. In this article, we continue to employ these influences, but we introduce one more index to account for the performance of growth versus value stocks. This new index has been added because of the establishment of a number of mutual funds that state either growth or value as an objective and because the growth and value distinction is highly correlated with book-to-market ratios, which have been shown by Fama and French (1993, 1994) to be empirically important in explaining common stock returns. A failure to account for this influence might result in our confounding the temporary performance of a type of fund (e.g., a "value" fund) with management skill. We have made one additional change to the performance model employed in Elton et al. (1993). In that article, the size index was orthogonalized with the S&P 500 Index. In this article, we measure size as the differential return between a portfolio of small stocks and large stocks, and we measure value or growth as the differential return between a portfolio of growth stocks and a portfolio of value stocks. Using differential returns has two benefits. First, this method produces indexes that are almost completely uncorrelated with each other. Second, the impact of these indexes on risk-adjusted performance is easy to understand, since they are zero-investment portfolios.

In this article, a fund's risk-adjusted performance is based on the intercept (a_i) from a four-index model. The model is

$$R_{it} = a_i + \beta_{iSP}R_{SPt} + \beta_{iSL}R_{SLt} + \beta_{iGV}R_{GVt} + \beta_{iB}R_{Bt} + \epsilon_{it}, \quad (1)$$

where

R_{it} = the excess return on fund i in month t (the return on the fund minus the 30-day Treasury-bill rate);

R_{SPt} = the excess return on the S&P 500 Index in month t;

R_{SLt} = the difference in return between a small-cap and large-cap stock portfolio,[2] based on Prudential Bache indexes in month t;

R_{GVt} = the difference in return between a growth and value stock portfolio based on Prudential Bache indexes in month t;

R_{Bt} = the excess return on a bond index in month t, measured by a par-weighted combination of the Lehman Brothers Aggregate Bond Index and the Blume/Keim High-Yield Bond Index;[3]

β_{ik} = the sensitivity of excess return on fund i to excess return on index $k(k = SP, SL, GV, B)$; and

ϵ_{it} = the random error in month t.

We distinguish between two time periods: the period where we rank and select funds (the "selection period"), and the period following the selection period, where we evaluate our selections of funds (the "performance period"). We use equation (1) to calculate both 1-year and 3-year risk-adjusted performance measures, which we will refer to as "alphas." The way alphas are calculated depends on which of the above 2 periods is being considered.

In order to calculate alpha in the selection period, we first calculate betas using 3 years of data. For example, if we were selecting funds on January 1, 1980, we would use the 3 years beginning January 1, 1977, through December 31, 1979, as our selection period to calculate betas. If we were ranking on 1-year alphas as of January 1980, we would estimate equation (1) over the prior 3 years and add the average monthly residual during 1979 to the estimated value of a_i to get a 1-year alpha. If we were ranking on 3-year alphas as of January 1980, we would simply use the value of a_i, estimated over the prior 3 years, for a 3-year alpha.

The alphas in the performance period were computed over each fund's full history if the fund existed over the full time frame or through the month of merger or policy change if the fund merged or changed investment policy. The alpha in the performance period is the value

2. The small-large index was formed by averaging the small-cap value index and the small-cap growth index and subtracting the average of the large-cap growth index and the large-cap value index. The growth-value index was formed by averaging the large-cap, mid-cap, and small-cap stock growth indexes and subtracting the average of the large-cap, mid-cap, and small-cap stock value indexes. Note that both of these indexes are in excess-return form. Since these indexes are based on differences, the riskless rate cancels out. Therefore, we do not subtract the riskless rate from these indexes, because doing so would in effect be double-counting the riskless asset.

3. See Elton, Gruber, and Blake (1995) for a detailed description of the bond index.

of a_i plus the average of the monthly residuals over the performance period. For example, if the performance period is 1 year, the alpha in the performance period is the overall a_i plus the average monthly residual during 1 year. If a fund that merges or changes policy is selected, the alpha in the performance period is a weighted average of the alpha and residuals on the selected fund through the month of merger or policy change and the average alpha plus average residuals on the surviving funds for the remaining months in the evaluation period. This is the return that would be earned by an investor who buys a fund at random from the population of funds that have existed since 1977 if the fund they own merges or changes policy.[4]

IV. Results

The first set of results involves ranking funds into 10 deciles each year according to a measure of past performance and then observing how well the deciles perform in subsequent periods. Does past performance (ranking) contain information about performance in future periods? While we start by examining whether ranking tends to be preserved over time, we quickly turn to questions about whether information in the extremes of the ranking is both economically and statistically significant.

A. *Ranking Results*

The funds ranked at each point in time are the funds available for purchase at that point in time. If a fund merges or changes investment policy (is no longer classified as a "common stock" fund) prior to or at that point in time, it is not a candidate for selection.

Each column of table 1 and table 2 shows the average risk-adjusted excess return (performance alpha) realized in subsequent periods when mutual funds are ranked and placed in deciles using a particular criterion. The column headings show the criteria used to rank funds. "Total Return" is ranking on annual total return calculated for 1-year selection periods; "3-Year Alpha" is ranking on 3-year selection alpha; "1-Year Alpha" is ranking on 1-year selection alpha; "*t* Alpha" is ranking on 3-year selection alpha divided by the standard error of 3-year selection alpha (the *t*-value of the intercept in eq. [1] using a

4. We considered and examined other rules for what happens when a fund merges or changes policy. Alternative choices had no significant effect on the results. This is in part because few of the selected funds merged or changed policy in the subsequent evaluation periods. For example, only 0.8% of 1,857 selected funds merged or changed policy in the first year after selection. In fact, among the top decile of selected funds, which we focus on in the next section, only five funds merged in the performance evaluation periods, none in the first year after selection and most near the ends of the 3-year performance evaluation periods, and none changed policy. Thus, our procedure for handling nonsurviving funds has virtually no effect on our results.

TABLE 1 **Average Realized 3-Year Alpha by Deciles for Different Ranking Criteria (Reported on a Monthly Basis, in %)**

	Deciles Formed on the Basis of:										
	Full Sample				High Expenses Eliminated*	Low R^2 Eliminated†			High Expenses and Low R^2 Eliminated‡		
Decile	Total Return	3-Year Alpha	1-Year Alpha	t Alpha	Total Return	3-Year Alpha	1-Year Alpha	t Alpha	3-Year Alpha	1-Year Alpha	t Alpha
Bottom 1	−.327	−.437	−.390	−.351	−.151	−.359	−.310	−.273	−.180	−.158	−.144
2	−.078	−.101	−.112	−.144	−.040	−.077	−.081	−.129	−.055	−.051	−.075
3	−.053	−.058	−.058	−.107	−.028	−.052	−.054	−.092	−.060	−.040	−.090
4	−.057	−.076	−.072	−.066	−.026	−.075	−.088	−.063	−.057	−.099	−.038
5	−.051	−.039	−.062	−.044	−.061	−.044	−.074	−.045	−.017	−.045	−.051
6	−.042	−.047	−.020	−.054	−.017	−.055	−.023	−.058	−.043	−.021	−.021
7	−.070	−.009	−.024	−.023	−.047	−.003	−.013	−.027	−.010	−.008	−.040
8	−.007	−.038	−.011	.012	−.016	−.043	−.018	.023	−.028	−.013	.034
9	−.016	.034	−.026	−.004	−.010	.042	−.012	.007	.036	.008	.001
Top 10	−.059	.009	.015	.023	−.049	.028	.038	.026	.019	.036	.029
Average	−.076	−.076	−.076	−.076	−.044	−.064	−.063	−.063	−.040	−.039	−.040
Rank correlation	.552	.952	.879	.976	.394	.915	.927	.976	.891	.927	.891
p-value	.0984	.0001	.0008	.0001	.2600	.0002	.0001	.0001	.0005	.0001	.0005

NOTE.—Shown is the average realized monthly risk-adjusted return (alpha computed for 3-year performance periods), where the deciles were formed using the ranking criteria shown at the top of each column and for the four samples shown above the columns.

* The top decile of funds with the highest past 3-year average expenses was eliminated each year.

† All funds for which the multi-index model explained less than 80% were eliminated.

‡ All funds for which the multi-index model explained less than 80% were eliminated, and then from the resulting sample the top decile of funds with the highest past 3-year average expenses was eliminated each year.

TABLE 2 Average Realized 1-Year Alpha by Forecast Deciles for Different Ranking Criteria (Reported on a Monthly Basis, in %)

	Deciles Formed on the Basis of:										
	Full Sample				High Expenses Eliminated*	Low R^2 Eliminated†			High Expenses and Low R^2 Eliminated‡		
Decile	Total Return	3-Year Alpha	1-Year Alpha	*t* Alpha	Total Return	3-Year Alpha	1-Year Alpha	*t* Alpha	3-Year Alpha	1-Year Alpha	*t* Alpha
Bottom 1	−.391	−.469	−.467	−.388	−.183	−.402	−.401	−.318	−.189	−.230	−.191
2	−.103	−.043	−.133	−.142	−.053	−.030	−.083	−.129	−.051	−.057	−.074
3	−.093	−.095	−.085	−.096	−.067	−.123	−.083	−.098	−.109	−.071	−.096
4	−.089	−.105	−.086	−.101	−.064	−.081	−.094	−.095	−.054	−.104	−.093
5	−.089	−.051	−.062	−.041	−.100	−.058	−.064	−.050	−.058	−.019	−.039
6	−.052	−.058	−.017	−.025	−.033	−.030	−.010	−.014	.001	−.026	.032
7	−.039	.056	.023	.000	−.024	.060	.030	.021	.061	.029	.031
8	.011	−.028	.049	.005	−.002	−.020	.024	.035	−.014	.034	.043
9	.014	.012	−.070	−.045	.000	.027	−.034	−.021	.032	−.015	−.018
Top 10	.051	−.004	.065	.050	.052	.030	.098	.051	.019	.107	.044
Average	−.078	−.078	−.078	−.078	−.047	−.063	−.062	−.062	−.036	−.035	−.036
Rank correlation	1.000	.770	.867	.867	.891	.806	.867	.927	.806	.903	.879
p-value	.0000	.0092	.0012	.0012	.0005	.0049	.0012	.0001	.0049	.0003	.0008

NOTE.—Shown is the average realized monthly risk-adjusted return (alpha computed for 1-year performance periods), where the deciles were formed using the ranking criteria shown at the top of each column and for the four samples shown above the columns.

* The top decile of funds with the highest past 3-year average expenses was eliminated each year.

† All funds for which the multi-index model explained less than 80% were eliminated.

‡ All funds for which the multi-index model explained less than 80% were eliminated, and then from the resulting sample the top decile of funds with the highest past 3-year average expenses was eliminated each year.

3-year selection period). Before examining performance predictability, it is worth noting that the average performance of funds is negative. Furthermore, when we examine the full sample, the average underperformance is statistically significant at the 1% level for the 3-year results and statistically significant at the 5% level for the 1-year results.

Let us examine whether information about past ranking tells us anything about future ranking. Table 1 presents the results where performance alphas are calculated over 3-year periods and reported as average monthly risk-adjusted return. The first four columns present results when we do not further restrict the group of funds being ranked. Note that, except for total return, any other ranking criterion studied (3-year selection alpha, 1-year selection alpha, or t-value on 3-year selection alpha) leads to a rank correlation coefficient that is significant at the 1% level. Ranking by total return is significant at the 10% level. We used two rules for eliminating certain funds from our sample to study the impact of elimination on the usefulness of rankings. First, we eliminated funds for which our risk model had low explanatory power in the selection period (adjusted R^2 below 0.8). We felt that an investor might recognize that funds for which the model fit poorly in the selection period might be less predictable in subsequent periods. A low R^2 could result from market timing or from very low diversification. Eliminating these funds had very little effect on the ranking results but did improve the performance of the upper and lower deciles. Second, after first eliminating funds with low adjusted R^2, from the resulting reduced sample we eliminated funds with high expenses (funds in the top decile of expense ratios in the reduced sample) to see if we were only picking up differences in expenses rather than differences in management performance. This had almost no effect on the rank correlations but improved average risk-adjusted returns and dramatically improved the risk-adjusted returns for the lowest deciles.

Figure 1 plots expenses by decile when funds are ranked on 3-year alpha. Figure 1 shows that expenses are high in the low decile when we eliminate the 10% of the funds with the highest expenses. Expense differences are minimal across all deciles, although the lowest decile has slightly higher expense ratios than other deciles. Thus, while high expenses cause common stock funds to be in the lowest performance decile, they do not explain ranking across other deciles.

Table 2 presents the same analysis where average performance alpha (excess return) is measured for a period of 1 year after the rankings. The major difference between table 2 and table 1 is that, when using the 1-year performance evaluation period, ranking techniques involving 1 year of past data generally perform much better than ranking techniques involving 3 years of past data. Total return provides the best ranking technique (as opposed to risk-adjusted return), although, as

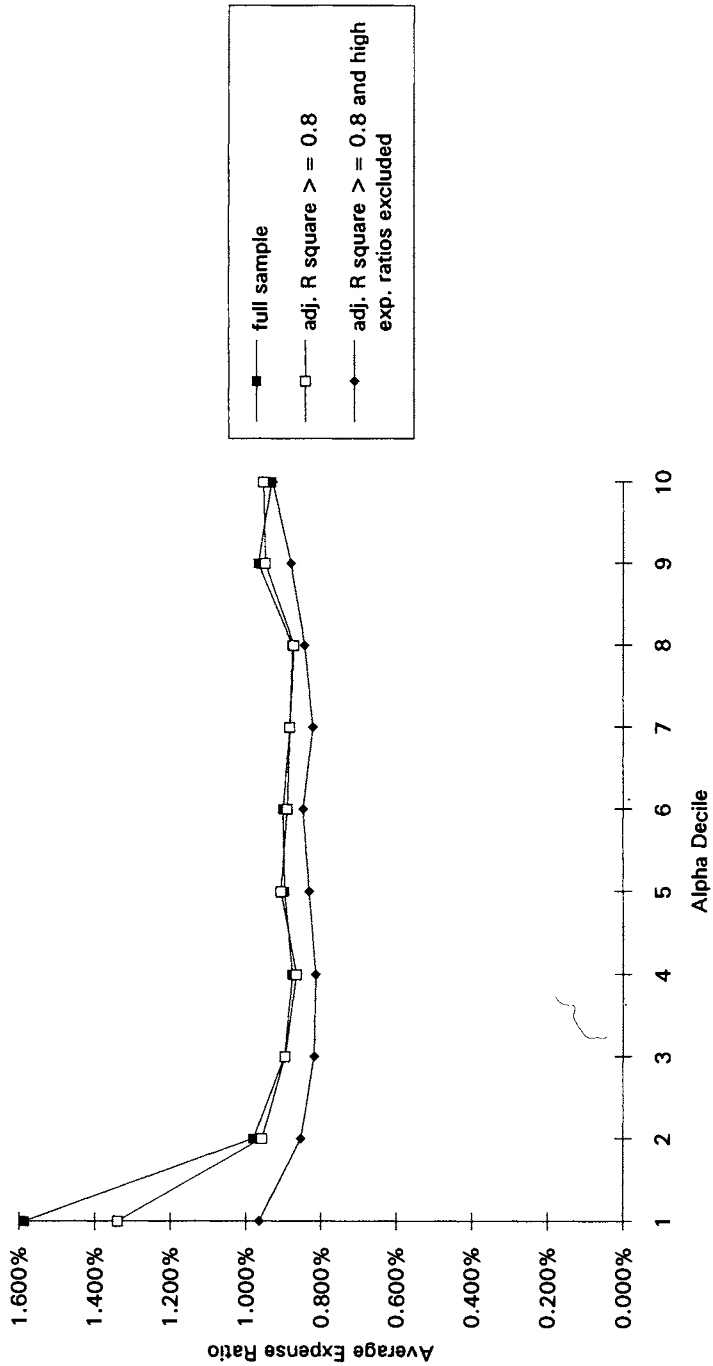

FIG. 1.—Performance and expenses

we will see shortly, it does not do as good a job at forecasting the tails of the distribution.

Comparing table 1 and table 2 shows that deciles formed on the basis of total return are highly correlated with future performance alpha when performance alpha is measured over a 1-year period, but the relationship deteriorates when future performance alpha is measured over 3 years. This is similar to the result in Hendricks, Patel, and Zeckhauser (1993). However, when ranking is done on a risk-adjusted basis, the predictability is increased when performance is measured over the longer (3-year) period.

B. Performance Results Using Ranks

While the preservation of rank across time is interesting, of perhaps more practical significance is an answer to the question, "Can information about past performance help earn a positive alpha in the future?" Tables 3 and 4 repeat the postselection performance alphas shown in tables 1 and 2 for the average fund and for deciles 1 and 10 and present the statistical significance of some differences. Table 3 presents the results for a 3-year holding period, while table 4 presents the 1-year holding period results.

We start by examining the section of panel A labled "full sample." Starting with the predictions based on selection alphas computed over the previous 3 years, we find that the top decile (decile 10) produces a positive excess return of 0.9 basis points per month if it is held for the following 3 years, while the bottom decile produces an excess return of −43.7 basis points per month. An investor selecting an equally weighted average of the funds in the top decile earns a positive (albeit tiny) excess return in an environment where most funds (and the average fund) earned a large negative excess return.

It is interesting to examine the performance of the top-ranked decile (portfolio) compared to both the bottom-ranked decile and the average fund. However, to see if these differences are statistically significant, differences in the mean return in each decile can be computed each year and a *t*-test performed for differences of the mean from zero. The results of this type of test are recorded in panel B. In addition, we can examine the number of time periods (out of a possible 12 times) that decile 10 (the top decile) outperforms decile 1 (the bottom decile) or the average fund. This is done in panel C. In interpreting this panel, it is useful to know that the chances of one decile having a higher estimated alpha 11 out of 12 times if the deciles being compared have the same true alpha is less than 1%; 10 of 12 times is statistically significantly different from zero at the 2% level, and nine out of 12 times is statistically significantly different from zero at the 6% level.

Returning to our analysis of ranking on 3-year selection alpha, we see that selecting the top decile is better (at the 1% significance level)

TABLE 3 Comparison of Realized 3-Year Alphas Using Deciles and Simple Rules for Different Ranking Criteria (Reported on a Monthly Basis)

	Deciles Formed on the Basis of:										
	Full Sample				High Expenses Eliminated[a]	Low R^2 Eliminated[b]			High Expenses and Low R^2 Eliminated[c]		
	Total Return	3-Year Alpha	1-Year Alpha	t Alpha	Total Return	3-Year Alpha	1-Year Alpha	t Alpha	3-Year Alpha	1-Year Alpha	t Alpha
A. Averages (in %):											
Average decile	−.076	−.076	−.076	−.076	−.044	−.064	−.063	−.063	−.040	−.039	−.040
Bottom decile	−.327	−.437	−.390	−.351	−.151	−.359	−.310	−.273	−.180	−.158	−.144
Top decile	−.059	.009	.015	.023	−.049	.028	.038	.026	.019	.036	.029
Simple rules	...	.066	.060	...	...	.065	.064	...	.058	.065	...
B. Differences (in %):											
Top vs. bottom	.267*	.446*	.405*	.374*	.102	.387*	.348*	.299*	.199*	.194*	.173*
Top vs. average	.017	.085*	.091*	.099*	−.004	.092*	.101*	.089*	.059	.075*	.068*
Simple rules vs. top	...	.057*	.044*	...	...	.037**	.026	...	.039*	.029	...
C. Number of 3-year periods (out of 12) that first listed group beat second listed group:											
Top vs. bottom	10	11	12	12	9	12	12	12	10	9	10
Top vs. average	5	9	9	12	5	10	11	11	9	11	9
Simple rules vs. top	...	10	10	...	...	9	7	...	...	...	...

NOTE.—Simple rules solutions to optimization are applied to the top decile of either 3-year or 1-year alphas calculated in selection periods.

[a] The top decile of funds with the highest past 3-year average expenses was eliminated each year.

[b] All funds for which the multi-index model explained less than 80% were eliminated.

[c] All funds for which the multi-index model explained less than 80% were eliminated, and then from the resulting sample the top decile of funds with the highest past 3-year average expenses were eliminated each year.

* Significant at the 1% level.

** Significant at the 5% level.

TABLE 4 **Comparison of Realized 1-Year Alphas Using Deciles and Simple Rules for Different Ranking Criteria (Reported on a Monthly Basis)**

	Deciles Formed on the Basis of:										
	Full Sample				High Expenses Eliminated[a]	Low R^2 Eliminated[b]			High Expenses and Low R^2 Eliminated[c]		
	Total Return	3-Year Alpha	1-Year Alpha	*t* Alpha	Total Return	3-Year Alpha	1-Year Alpha	*t* Alpha	3-Year Alpha	1-Year Alpha	*t* Alpha
A. Averages (in %):											
Average decile	−.078	−.078	−.078	−.078	−.047	−.063	−.062	−.062	−.036	−.035	−.036
Bottom decile	−.391	−.469	−.467	−.388	−.183	−.402	−.401	−.318	−.189	−.230	−.191
Top decile	.051	−.004	.065	.050	.052	.030	.098	.051	.019	.107	.044
Simple rules	. . .	.073	.139	. . .	. . .	.083	.134	. . .	.069	.135	. . .
B. Differences (in %):											
Top vs. bottom	.442*	.465*	.532*	.439*	.235**	.431*	.498*	.369*	.208**	.337*	.235*
Top vs. average	.129**	.074	.143**	.128*	.099	.092**	.159**	.112*	.055	.142	.080**
Simple rules vs. top	. . .	.077*	.074*	. . .	. . .	.053*	.036	. . .	.050**	.028	. . .
C. Number of years (out of 12) that first listed group beat second listed group:											
Top vs. bottom	10	11	12	11	9	11	12	12	8	9	10
Top vs. average	9	9	8	11	7	10	9	10	9	9	8
Simple rules vs. top	. . .	10	10	. . .	. . .	10	8	. . .	9	8	. . .

NOTE.—Simple rules solutions to optimization are applied to the top decile of either 3-year or 1-year alphas calculated in selection periods.

[a] The top decile of funds with the highest past 3-year average expenses was eliminated each year.

[b] All funds for which the multi-index model explained less than 80% were eliminated.

[c] All funds for which the multi-index model explained less than 80% were eliminated, and then from the resulting sample the top decile of funds with the highest past 3-year average expenses was eliminated each year.

* Significant at the 1% level.

** Significant at the 5% level.

than selecting either the bottom decile or the average fund. It also outperforms the bottom decile 11 times and the average fund nine out of 12 times.

If we forecast 3-year performance alphas using the past 1-year selection alpha rather than the past 3-year selection alpha, we find results which are perhaps slightly improved. The top decile goes up slightly in return from 0.9 basis points to 1.5 basis points per month. This is an improvement, though the bottom decile also improves by about 4 basis points. The differences in risk-adjusted return between the top decile and either the bottom decile or average fund continue to be significant at the 1% level.

When we rank by t-value of the 3-year selection alpha, the return on the portfolio of the top decile funds is higher and is equal to 2.3 basis points per month. The differences in performance between the top decile and the average or bottom deciles are statistically significant at the 1% level. Here the top decile outperforms both the bottom decile and the average fund 12 out of 12 times.

The final selection metric we examined was total return in the year preceding the prediction. Unlike the cases with the other selection metrics, when we form deciles on total return we get a negative performance alpha over the following 3 years. While the top decile outperforms both the average fund and the bottom decile, differences from the average fund are no longer statistically significant.

The next section of tables 3 and 4 examines results when we have eliminated all funds in the selection period with an adjusted R^2 less than 0.8 on our four-index model. In all cases the risk-adjusted excess return on the top decile increases. It increases by a much smaller percentage for rankings on t-values, probably because the funds with low adjusted R^2's also had large standard errors and so were unlikely to be in the top decile for the t-value criterion when the full sample was used. When we discard funds with adjusted R^2 less than 0.8, the risk-adjusted return for the average fund also increases, indicating that funds with low adjusted R^2 tended to have negative risk-adjusted performance.

The last section of tables 3 and 4 shows performance when we delete the 10% of the funds with the highest expenses from the sample of funds having an adjusted R^2 not less than 0.8. The effect of removing these funds on the performance of the top decile is mixed, hurting performance when ranking with 3-year selection alphas and helping with the other ranking criteria. The major effect of removing the high-expense funds is to improve average performance and the performance of the lowest decile. Note that the performance of the bottom decile increases by more than 2.0% per year when the high-expense funds are deleted. This causes the differences between the top and bottom deciles to shrink; however, the differences are still statistically significant at the 1% level.

So far we have discussed ranking funds when the results of holding a decile (portfolio) is evaluated over a 3-year performance period. We now turn to an examination of results when the performance evaluation (holding) period for a portfolio is 1 year. While many of the results for the 1-year performance period are similar to those for the 3-year performance period, there are some interesting differences.

The first point to note from table 4 is that when evaluating performance over a 1-year period, the performance of the top decile based on 1-year ranking criteria improves markedly compared to forecasts prepared on the basis of 3 years of past data. Also note that whether one selects on the basis of total return, 1-year alpha, or t alpha, the top decile (portfolio) has fairly large monthly returns. For example, selecting on the basis of 1-year selection alphas gives a positive return of 78 basis points per year. Note that for all ranking techniques using the full sample, the top decile outperforms the bottom decile at the 1% level of statistical significance. In contrast to the 3-year performance evaluation period, total return seems to carry real information as a ranking criteria, though it does not perform as well as 1-year selection alpha.

Other results in table 4 are similar to those presented in table 3. Eliminating funds with a poorer fit (low adjusted R^2) during the selection period (when deciles are formed) results in a higher return for the top decile but decreases the difference between the top and bottom deciles. Eliminating high-expense funds sometimes lowers and sometimes raises the performance of the top decile but definitely raises the performance of the bottom decile.

In order to be sure that our results were not due to unusual and persistent high performance of one or two funds, we eliminated the two funds that appeared most often in the top decile of the selection periods and recalcualted the performance period averages. The resulting average performance alpha for the remaining funds in the top decile sometimes increased and sometimes decreased, but all changes were negligible.

C. *Performance Results Employing Modern Portfolio Theory*

Up to now we have compared the performance of deciles (portfolios of funds) formed on the basis of alternative ranking criteria. In examining performance, the fraction of investor capital invested in each of the funds in a portfolio was assumed to be the same. But the literature of MPT gives us great insight into what optimal weights should be within a decile. In this section we examine two alternative evaluation schemes. Both use MPT to find the optimal weight on each fund in the top decile or top two deciles of funds. The first evaluation scheme compares performance on the basis of alpha. The second scheme evaluates on the basis of total return while holding risk constant.

Optimal portfolio weights. We start by assuming that the investor is going to hold a portfolio composed of the funds found in the top decile in the selection period. We then draw on the modern portfolio theory literature to find the optimal weighting of funds in that decile. It is well known that, if we assume returns are described by a model such as that shown in equation (1), then the optimal weight (fraction of money) to place in any fund is given by

$$X_i = \frac{(a_i/\sigma^2_{\epsilon_i})}{\sum_i (a_i/\sigma^2_{\epsilon_i})}, \tag{2}$$

where $\sigma^2_{\epsilon_i}$ is the variance of the random error term in equation (1). This is a generalization of the Treynor and Black (1973) and Elton, Gruber, and Padberg (1976) criteria and is derived for a multi-index model in Elton and Gruber (1992).

The results of employing this weighting across all securities in the top decile are shown in tables 3 and 4 under the row labeled "simple rules."[5] Note that employing a technique that optimizes weights leads to a large improvement in performance. Examining the results using a 3-year evaluation period, the optimal portfolio (selected on the basis of past performance) earns between 70 and 80 basis points per year in excess return. Note that eliminating funds with low adjusted R^2 over the selection period makes almost no difference in return. This comes about because funds with low adjusted R^2 tend to have high residual risk and thus receive lower weights in the optimal portfolio. Removing high-expense funds has, at most, a marginal effect on the performance of the optimal portfolio, since high-expense funds are normally not in the top decile.

If we turn to the results employing a 1-year performance evaluation period, we find that the results are even more dramatic. Employing the 1-year selection alpha to form deciles from which a portfolio is selected results in a portfolio that produces risk-adjusted excess return of more than 1.50% per year in the performance evaluation period.

In every case, both for the 1-year and 3-year performance evaluation periods, employing optimal weights results in a large increase in performance relative to using equal weights. In addition, in most cases the

5. None of the optimal portfolios selected by this rule involve short sales since all selection alphas in the top decile are positive. Note that a fund's 3-year selection alpha is the a_i appearing in eqq. (1) and (2). When 1-year selection alphas are used, we substitute the 1-year selection alpha for the a_i appearing in eq. (2). The residual variances in eq. (2) are those obtained from a 3-year selection period in either case.

results obtained with optimal weights are statistically significantly better than the results obtained with equal weights.

Before leaving this section we should comment on one more issue: the magnitude of the weights when a portfolio optimization technique is used. The percentage of time that each weight occurs is shown in figure 2. This figure makes it clear that while there are deviations of the weights from an equal-weighting scheme, the results do not come about by holding an extremely concentrated portfolio of funds. If we held equally weighted portfolios, the percent invested in any fund would average 6⅔%. The weights depicted in figure 2 show a high concentration in the 2%–11% range, with very few weights above 15%.

Active versus passive funds. We have one final way to test whether past mutual fund performance can be used to select funds that subsequently produce superior risk-adjusted returns. This involves a comparison of active and passive portfolios with the same risk. We select a set of risk levels (target betas) that seem relevant for investors in actively managed stock mutual funds. We then select both a portfolio of active funds and a portfolio of index funds that have the same target betas. Finally, we compare the return of the portfolio of active funds to that of the index portfolio to see which has a higher mean return.

Although an investor currently has a large number of index funds to select from, these funds are a recent phenomenon. Thus, we do not have actual fund histories for a large number of different types of index funds. For the S&P Index fund we selected the Vanguard S&P 500 Index fund, both because it has a long history and because it has the lowest expenses of the available S&P Index funds. To represent the return on other types of index funds, we had to use actual indexes as proxies. The indexes we chose were

1. the Wilshire Large Capitalization Value Portfolio for a value portfolio;
2. the S&P Barra Growth Portfolio for a growth portfolio;
3. the Wilshire Small Cap Growth Portfolio for a small-cap growth portfolio;
4. the Ibbotson Associates Small Capitalization Portfolio for a small-cap index;
5. the Ibbotson Associaties 30-day Treasury bill as a short-term money market portfolio; and
6. the Lehman Brothers Government Corporate Index as a bond index.

We first decided on a set of target betas that our hypothetical investor in index funds might choose. We used as our target betas the four average sensitivities to each of the indexes in our four-index model across all mutual funds in the sample and across all 3-year selection periods. These are the expected sensitivities for an investor who randomly selects from our population of funds. The average sensitivities

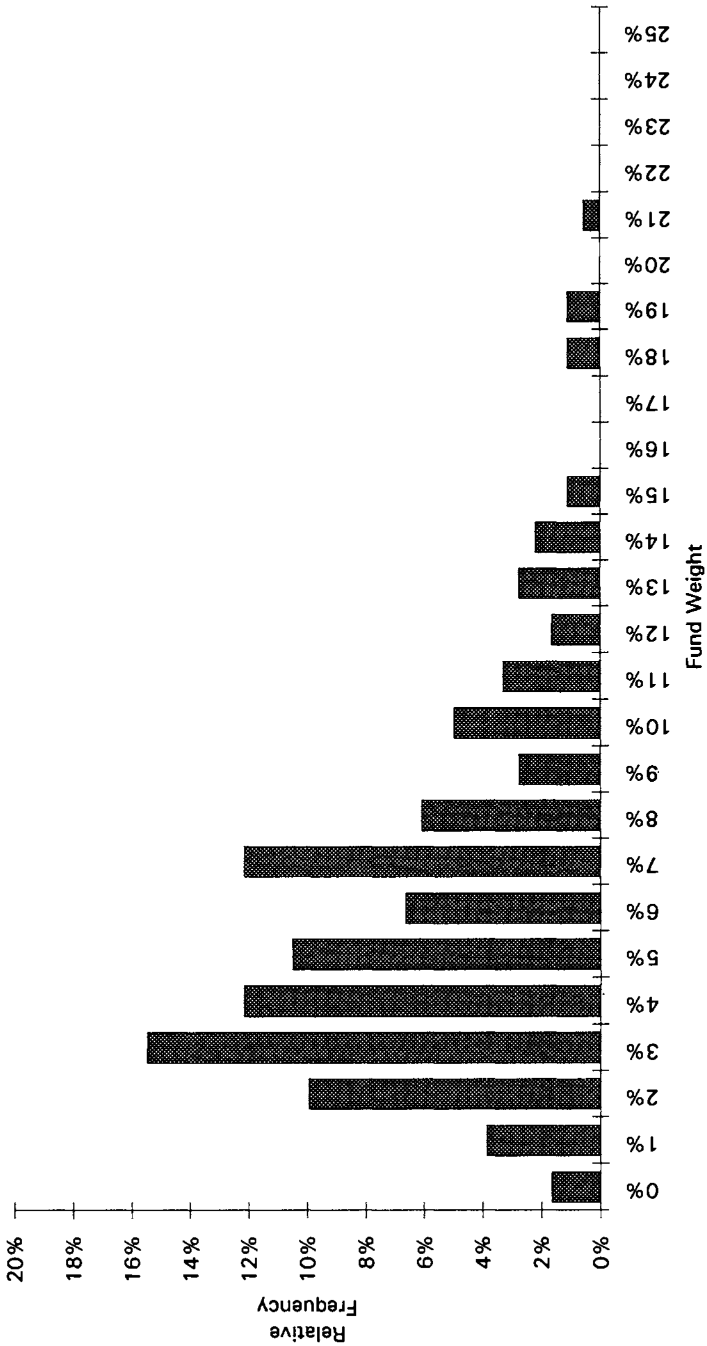

FIG. 2.—Frequency of composition weights in simple rules fund portfolios

are 0.86 on the S&P Index, 0.30 on the differential size index, 0.32 on the value-growth index, and 0.07 on the bond index. We constructed a portfolio of index funds that had these target betas as the passive portfolio for a hypothetical investor.[6]

We tried two alternative techniques for selecting portfolios from among active funds, one using maximization of 3-year selection alpha as an objective and one maximizing the average 3-year selection alpha over residual risk. The first criterion was chosen because it was a criterion investors might logically try and because it is consistent with much of the research in this area. We chose the second because Elton and Gruber (1992) show this criterion is optimal for selecting an active portfolio when security returns follow a multi-index model. To ensure that the results were realistic, short sales were not allowed. We initially selected funds using data from 1978 through 1980. The active portfolios were then reformed at the beginning of every year from 1982 through 1991 using the prior 3 years of data. Performance was measured over the period 1981–93.

We used an upper bound constraint of 10% investment in a fund. This forced some diversification. This upper bound constraint had to be relaxed on occasion in order to find a solution. The maximum upper limit was 20%, and even when this came into play there was a fair amount of diversification. Finally, we limited our choices to funds in the upper two deciles, since restricting choices to the top decile did not allow a solution in several of the selection periods.

The reason that we used the upper two deciles and occasionally had to allow the maximum fraction of capital invested in any fund to rise to 20% was because of the changing nature of stocks in the top two deciles. In certain years, the stocks in the top decile and in the top two deciles had sensitivities (betas) that differed considerably from the average sensitivities in our sample. For example, in 1 or more years high growth stocks may be good performers (have high predicted alphas), while in other years value stocks end up as the best performers.[7]

Our portfolio of active funds outperformed our passive-fund portfolio in each case. In terms of total return, the average outperformance

6. We solved a linear programming problem to minimize the difference between the betas on the portfolio of index funds and the target betas. The program was able to match the target betas exactly in 8 of 11 years and almost exactly in the other 3 years.

7. This intertemporal instability of betas over time in each decile casts some doubt over numerous other studies that measure alpha from a particular decile (or octile) over time by running a single- or multiple-index model on the returns of a particular decile over time. Since the betas are unstable, the results are not meaningful. This contrasts with the methodology used throughout this article, where alphas are estimated for individual funds and then aggregated for portfolios. As an example, for the four-index model we employ in this article, the sensitivity of the difference in return between the top and bottom performing decile on the small-stock index varies between +0.74 and −0.51.

was 22 basis points per year using 3-year selection alpha as an objective, and 71 basis points per year when funds were selected by maximizing the average 3-year selection alpha divided by the residual risk. The Vanguard S&P Index fund's return is reduced by fees and expenses; the returns on the other indexes are not. If we assume 30 basis points in fees and expenses for the index funds other than the Vanguard S&P fund, the return differences described above should be increased by 29 basis points per year.

While we were able to find a combination of active funds that outperformed a combination of index funds with average characteristics in each case we tried, this might not have been possible if the target betas had deviated considerably from the average sensitivities of the top-performing funds.

V. Size and Expenses

While we have found predictability in this study, an argument has been made that predictability should not exist. This argument proceeds from the assumption that successful managers subsequently raise fees in order to increase their income. The increase in fees could thus eliminate subsequent excess return, resulting in no continuity of performance even when managers have an ability to construct superior portfolios.

Managers' fees are almost universally stated as a fraction of total net assets. Thus, a manager's total compensation is equal to a stated percentage of total net asset value. Furthermore, growth in total net asset value is a function of prior performance (see Sirri and Tufano 1992). An increase in the fee percentage would hurt postexpense performance and therefore reduce subsequent growth in assets. If the relationship between asset growth and performance is strong enough, a successful manager might increase manager revenue more by not raising management fees and thereby having higher asset growth. Which of the above two behavioral models best describes actual performance is an empirical question.

To examine this question, we must control for the relationships between asset size and expenses and time and expenses. Table 5 shows the regression between expense ratios and the log of total net asset value for each year between 1977 and 1991. There is a significant negative relationship in each year, with the slope varying from -0.14 to -0.20 and an R^2 of between 0.24 and 0.44. Thus, expenses decline as size increases. The right-hand half of table 5 shows the predicted expense in percent for funds of different asset size. The numbers in the right part of the table are obtained by substituting the log of the size shown at the top of the column along with the corresponding parameter values shown in the left portion of the table into the equation described

Table 5 **Annual Expense Ratios and Fund Total Net Asset Sizes (Cross-Sectional Regression Results by Year; Annual Expense Ratio Regressed on Constant and Log of Size)**

						Predicted Expense Ratios Based on Regression Results (in %) (Total Net Assets in Millions $)				
Year	Adjusted R^2	Intercept (%)	t-Intercept	Slope	t-Slope	50	100	200	400	1,000
1977	.41	1.72	24.44	−.19	−11.93	.99	.87	.74	.61	.44
1978	.44	1.70	26.44	−.19	−12.71	.97	.85	.72	.59	.42
1979	.36	1.75	22.23	−.19	−10.62	1.01	.87	.74	.61	.44
1980	.29	1.75	18.77	−.18	−8.90	1.04	.91	.79	.66	.49
1981	.37	1.70	22.26	−.17	−10.53	1.04	.92	.81	.69	.53
1982	.35	1.79	19.55	−.19	−9.84	1.04	.91	.78	.65	.47
1983	.35	1.83	19.80	−.18	−9.79	1.11	.99	.86	.73	.56
1984	.30	1.71	18.25	−.16	−8.76	1.08	.97	.86	.75	.61
1985	.35	1.63	21.16	−.14	−9.52	1.07	.97	.87	.77	.64
1986	.31	1.98	16.24	−.20	−8.75	1.20	1.06	.93	.79	.60
1987	.37	1.74	20.06	−.16	−9.98	1.12	1.01	.90	.79	.64
1988	.23	1.70	15.19	−.14	−7.08	1.13	1.04	.94	.84	.71
1989	.24	1.78	16.03	−.15	−7.27	1.21	1.11	1.01	.91	.78
1990	.28	2.05	15.62	−.18	−8.04	1.33	1.21	1.08	.95	.78
1991	.31	2.00	16.94	−.17	−8.47	1.32	1.20	1.08	.96	.80

at the top of the table. There is a clear pattern of increasing expense ratios over time.

To analyze whether successful managers increase their expense ratios over time, we used the equation described at the top of table 5 along with each fund's asset size to determine predicted expense ratios. These predicted expense ratios are shown in table 6 in the first column. The table analyzes expenses for all funds in the top decile using 3-year alphas in the selection period. Actual expense ratios are in the second column, and the difference is in the third column. The first row is expense ratios at the time of selection. For example, if we were selecting on January 1, 1980, the expense ratios are for the end of 1979. Rows 2–6 are the expense ratios at the end of the 5 years subsequent to selection (1980–84 in the above example). The expenses at the time funds are selected are exactly what would have been predicted given the fund size and year of selection. In the first year of the evaluation period, their expense ratios are again exactly what would be predicted given the fund size and year of selection. Expense ratios increase slightly in subsequent years, although the magnitude of the increase is small. Furthermore, examining the second column of table 6, and examining table 5 to note how expenses increase with time, shows that, ignoring the increase in size, selected funds increase fees no more than the average fund. However, the increase in fees is faster than their increase in size would suggest. Fees could also be increased by imposing loads. Although 19 funds changed from no-load to load funds, none of these funds were in the top selection decile in the year before the change. In fact, the greatest number of these funds were in the *bottom* selection decile in the year before the change! In conclusion, the fees of top-performing funds exhibit at most a slight increase in years subsequent to their top ranking, and clearly not enough to affect performance. On average then, managers of successful funds

TABLE 6 **Comparison of Average Size and Average Actual and Predicted Annual Expense Ratios (for Top Decile Ranked by Alphas Calculated during 3-Year Selection Periods)**

Year t	Predicted Expense Ratio (%)	Actual Expense Ratio (%)	Difference (%)	Asset Size (Millions $)
Prior	.89	.89	.00	189.87
1	.87	.87	.00	239.67
2	.86	.90	−.05	294.40
3	.86	.91	−.05	346.87
4	.86	.92	−.05	405.20
5	.88	.92	−.04	457.87

NOTE.—Prior year is the end of the last year prior of the selection period, year 1 is the end of the first year after selection period, year 2 is the end of the second year after the selection period, etc.

increase their total revenues by having the sizes of their funds increase, not by increasing expenses.

VI. Conclusions

There are a number of lessons that can be learned from this study.

1) The past carries information about the future. Funds that did well in the past tend to do well in the future on a risk-adjusted basis.

2) Both 1- and 3-year alphas convey information about future performance.

3) When future performance is evaluated over 3-year periods, selection on prior 3-year alpha conveys no less, and perhaps more, information about future performance than selection using other time horizons.

4) When future performance is evaluated over a 1-year period, selection of funds based on the prior year's data conveys much more information about performance than selection based on data from the prior 3 years.

5) There is definite information about future performance conveyed by past performance, and this information works for periods 3 years into the future as well as 1 year into the future. "Hot hands" may be an important phenomenon, but there is a longer persistence in performance than noted in the hot hands literature.

6) Employing modern portfolio theory to form optimal portfolios based on past information leads to the selection of portfolios of mutual funds that have a positive and both economically and statistically significant return compared to a portfolio that places an equal amount in each fund that is considered.

7) The very bad performance of the lowest decile is largely accounted for by the fact that it contains the majority of the funds with very high expenses. Expense ratios are about the same for all other deciles. When high-expense funds are removed from the sample, we see that past performance still tells us a lot about future performance. Differences in risk-adjusted return between the top and bottom decile are partially differences in selection skill and partially differences in expenses.

8) Successful funds do not increase their fees compared to less successful funds.

References

Blake, C. R.; Elton, E. J.; and Gruber, M. J. 1993. The performance of bond mutual funds. *Journal of Business* 66 (July): 371–403.

Brown, S. J., and Goetzmann, W. N. 1995. Performance persistence. *Journal of Finance* 50 (June): 679–98.

Brown, S. J.; Goetzmann, W. N.; Ibbotson, R. G.; and Ross, S. A. 1992. Survivorship bias in performance studies. *Review of Financial Studies* 5 (Winter): 553–80.

Carhart, M. M. 1994. On persistence in mutual fund performance. Working paper. Chicago: University of Chicago.
Elton, E. J., and Gruber, M. J. 1992. Portfolio analysis with non-normal multi-index return-generating process. *Review of Quantitative Finance and Accounting* 2 (March): 5–17.
Elton, E. J.; Gruber, M. J.; and Blake, C. R. 1995. Fundamental economic variables, expected returns, and bond fund performance. *Journal of Finance* 50 (September): 1229–56.
Elton, E. J.; Gruber, M. J.; and Blake, C. R. In press. Survivorship bias and mutual fund performance. *Review of Financial Studies.*
Elton, E. J.; Gruber, M. J.; Das, S.; and Hlavka, M. 1993. Efficiency with costly information: A reinterpretation of evidence for managed portfolios. *Review of Financial Studies* 6 (Spring): 1–22.
Elton, E. J.; Gruber, M. J.; and Padberg, M. 1976. Simple criteria for optimal portfolio selection. *Journal of Finance* 31 (December): 1341–57.
Fama, E. F., and French, K. R. 1993. Common risk factors in the returns on bonds and stocks. *Journal of Financial Economics* 33 (February): 3–53.
Fama, E. F., and French, K. R. 1994. Size and book-to-market factors in earnings and returns. Working paper. Chicago: University of Chicago.
Grinblatt, M., and Titman, S. 1989. Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Business* 62 (July): 393–416.
Grinblatt, M., and Titman, S. 1992. Performance persistence in mutual funds. *Journal of Finance* 47 (December): 1977–84.
Hendricks, D.; Patel, J.; and Zeckhauser, R. 1993. Hot hands in mutual funds: Short-run persistence of performance, 1974–88. *Journal of Finance* 48 (March): 93–130.
Lehmann, B. N., and Modest, D. 1987. Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *Journal of Finance* 42 (June): 233–65.
Malkiel, B. G. 1995. Returns from investing in equity funds. *Journal of Finance* 50 (June): 549–72.
Sharpe, W. F. 1995. The styles and performance of large seasoned U.S. mutual funds. Working paper. Palo Alto, Calif.: Stanford University.
Sirri, E. R., and Tufano, P. 1992. The demand for mutual fund services by individual investors. Working paper. Cambridge, Mass.: Harvard Business School.
Treynor, J., and Black, F. 1973. How to use security analysis to improve portfolio selection. *Journal of Business* 46 (January): 66–86.
Wiesenberger Financial Service. 1977–92. *Investment Companies.* Boston: Warren, Gorham & Lamont.