Measuring Mutual Fund Performance with Characteristic-Based Benchmarks

KENT DANIEL, MARK GRINBLATT, SHERIDAN TITMAN, and RUSS WERMERS*

ABSTRACT

This article develops and applies new measures of portfolio performance which use benchmarks based on the characteristics of stocks held by the portfolios that are evaluated. Specifically, the benchmarks are constructed from the returns of 125 passive portfolios that are matched with stocks held in the evaluated portfolio on the basis of the market capitalization, book-to-market, and prior-year return characteristics of those stocks. Based on these benchmarks, "Characteristic Timing" and "Characteristic Selectivity" measures are developed that detect, respectively, whether portfolio managers successfully time their portfolio weightings on these characteristics and whether managers can select stocks that outperform the average stock having the same characteristics. We apply these measures to a new database of mutual fund holdings covering over 2500 equity funds from 1975 to 1994. Our results show that mutual funds, particularly aggressive-growth funds, exhibit some selectivity ability, but that funds exhibit no characteristic timing ability.

CURRENTLY, OVER ONE TRILLION dollars are invested in actively managed equity mutual funds. Assuming that the fees and expenses of these funds average about one percent of assets—a conservative estimate that ignores the expenses that funds generate from buying and selling stocks—the total costs generated by this industry exceed $10 billion per year. Although mutual funds provide a number of services, such as check-writing and bookkeeping services, more than half of the expenses of mutual funds arise because of their stock-selection efforts.[1]

This article examines whether mutual funds can systematically pick stocks that allow them to earn back a significant fraction of the fees and expenses that they generate. This question has been asked a number of times before, and has generated a great deal of controversy. Beginning with Jensen (1968),

* Daniel is from Northwestern University, Grinblatt is from the University of California Los Angeles, Titman is from the University of Texas at Austin, and Wermers is from the University of Colorado at Boulder. We thank Josef Lakonishok, Neil Pearson (the discussant), and participants of the "Performance and Behavior of Money Managers" session of the 1997 AFA meetings in New Orleans, Louisiana. We also thank Mark Carhart for providing data used in one section of this article, the Price Center for Entrepreneurial Studies, and The UCLA Academic Senate for providing funding, and Eduardo Schwartz for assistance in obtaining funding.

[1] A plausible estimate of the cost associated with active management can be derived by comparing the passive Vanguard Index 500 fund, which has an expense ratio of 20 basis points per year, with the funds in the Carhart (1997) database (mainly active funds), which have an average expense ratio of 130 basis points per year.

most academic studies have concluded that mutual funds do not have significant stock-picking abilities. More recent articles that come to the same conclusion include Malkiel (1995) and Gruber (1996). However, several studies have found evidence of persistence in mutual fund performance (Grinblatt and Titman (1992), Hendricks, Patel, and Zeckhauser (1993), and Goetzmann and Ibbotson (1994)), which suggests that superior portfolio managers who consistently outperform the market do exist. But this persistence has been attributed to either survival bias or benchmark errors (for example, see Brown, Goetzmann, Ibbotson, and Ross (1992), Malkiel (1995), Wermers (1997), and Carhart (1997)).

All of the aforementioned studies (except Wermers (1997)) assess the performance of mutual funds by examining the actual returns that investors realize from holding the funds. In contrast, studies by Grinblatt and Titman (1989a, 1993) take a somewhat different approach and examine the performance of the individual stocks held by funds. The Grinblatt and Titman studies are more generous than the other studies in their assessment of the abilities of mutual fund managers. Indeed, they find that in the 1975 to 1984 period, aggressive-growth and growth funds were able to outperform their benchmarks by an average of two to three percent, before expenses.

There are a number of advantages to directly evaluating the portfolio holdings of the mutual funds. First, by observing the portfolio holdings of funds, it is possible to design benchmarks that better capture the investment styles used by fund managers. Second, hypothetical returns generated from portfolio holdings do not include the fees, expenses, and trading costs that reduce the actual mutual fund returns used in other studies. Although these hypothetical returns overestimate the returns from holding the funds, they are appropriate in determining whether fund managers have any stock selection or timing abilities, since they are compared to a benchmark that also ignores transaction costs. Finally, it should be noted that the mutual fund samples considered in the Grinblatt and Titman studies are not contaminated by survival requirements.

The Grinblatt and Titman studies are, however, subject to some criticisms. First, the number of funds in these studies is relatively small (at most, 274) and the time period is relatively short (10 years). In addition, the Grinblatt and Titman benchmarks may not fully account for return anomalies, such as the size, book-to-market, and momentum effects. For example, Grinblatt, Titman, and Wermers (1995, GTW) suggest that the momentum effect, documented by Jegadeesh and Titman (1993), can significantly influence the performance measure used in Grinblatt and Titman (1993, GT). In particular, the GTW study indicates that the majority of mutual funds use momentum as a stock selection criterion, and that, on average, those funds that use such a criterion outperform the GT benchmark. By contrast, the average fund not using momentum as a selection criterion does not outperform the GT benchmark. Thus, the favorable returns documented in this earlier study may have been due to a simple momentum screen in the selection criteria of the funds, rather than

to special information or abilities that allow funds to ferret out underpriced stocks.

This article addresses these criticisms. The number of firms in our dataset is almost ten times the number in the dataset used by Grinblatt and Titman (1989a, 1993), and the time period for this study is twice as long.[2] This dataset is free from survivorship bias, and is the largest and most complete sample of mutual fund holdings examined to date.

We address the benchmark issue by introducing a new performance measurement method that forms benchmarks by directly matching the *characteristics* of the component stocks of the portfolio being evaluated. This method can be contrasted with an alternative characteristic/factor approach that forms factor portfolios based on characteristic-sorted stocks. Returns on these factor portfolios are then used as regressors in a traditional multifactor regression model.[3] Our more direct approach compares the returns of the equities held by a fund to the returns of portfolios of stocks with equivalent characteristics. Although our approach cannot be implemented without fund holdings, it is superior when holdings are available for a number of reasons:

1. The empirical evidence suggests that characteristics provide better ex-ante forecasts of the cross-sectional patterns of future returns (see Daniel and Titman (1997)).
2. Characteristic-matching also does a better job in matching future realized returns, meaning that the average fraction of the variance of the fund returns explained by the benchmark is higher and the standard error of the estimate of the fund's abnormal performance is lower. Thus, characteristic-matching should have more statistical power to detect abnormal performance than factor models.
3. The characteristics approach allows us to decompose fund returns into the components of Average Style (AS), Characteristic Selectivity (CS), and Characteristic Timing (CT). The sum of these measures is the overall hypothetical return of a fund. This decomposition provides a more accurate way to determine how funds are generating returns.
4. The calculation of the characteristic-based measures is straightforward.

Our application of the characteristic-based approach to this mutual fund sample suggests that, as a group, the funds showed some stock selection ability, but no discernable ability to time the different stock characteristics (e.g., buying high book-to-market stocks when those stocks have unusually high returns). Overall, the performance (CS + CT) added by fund managers is statistically significant, but not significantly greater than the difference between active and passive fund expenses.

[2] We examine the portfolio holdings of over 2500 equity mutual funds from 1975 to 1994.

[3] For example, Grinblatt and Titman (1989a) form eight extracted factors that use portfolios selected on the basis of size, dividend-yield, and past returns. Carhart (1997), following Fama and French (1993), forms extracted-factor portfolios based on the market, and on the characteristics of size, book-to-market, and momentum.

Finally, although the hypothetical returns of the mutual funds exhibit a great deal of persistence, the benchmark-adjusted returns do not. Our conclusions, in this regard, are consistent with the Carhart (1997) and Wermers (1997) evidence that suggests that the persistence in mutual fund performance is due to the use of simple momentum strategies by fund managers, rather than to certain managers having "hot hands" that allow them to pick winning stocks.

I. Data

Quarterly equity holdings for virtually all equity mutual funds that existed during any given quarter between December 31, 1974 and December 31, 1994 (inclusive) were purchased from CDA Investment Technologies, Inc., of Rockville, Maryland. The database lists the equity portion of each fund's holdings (i.e., the shareholdings by Committee on Uniform Security Identification Procedure (CUSIP) number, along with the self-declared investment objective (beginning June 30, 1980) at the end of each calendar quarter.[4] Over 5000 funds are included in this database. We exclude sector funds, bond and preferred funds, and any fund with an investment objective that is not oriented to general equity.[5] Over 2500 funds are represented in the remaining database. The collection procedure by CDA does not require that the funds survive.[6] Further details about the database may be found in Wermers (1997).

Table I examines how the different investment objectives, as self-reported by the funds, correspond with the stock characteristics chosen by the funds. In order to characterize the investment styles of the different funds, we identified the three characteristic quintile numbers corresponding to each stock held by a given mutual fund during a given year. We then computed the portfolio-weighted quintile numbers for that fund for that year. For example, if the Magellan Fund held half of the dollar value of its portfolio, in 1975, in stocks belonging to New York Stock Exchange (NYSE) size quintile 5 (large stocks), (NYSE) book-to-market quintile 5 (high book-to-market stocks), and (NYSE) momentum quintile 5 (high prior-year return stocks), and the other half in stocks belonging to size quintile 3, book-to-market quintile 3, and momentum quintile 3, then the average Magellan style for 1975 would be size quintile 4,

[4] We supplemented this data by adding the investment objectives of 274 funds on December 31, 1974. This represents data for all funds existing on that date for which we could find an objective.

[5] In all of our tests, we combined balanced funds and income funds into one category due to the relatively small number of funds in each of these categories.

[6] CDA attempts to identify every mutual fund in existence during each quarter. Several sources are consulted to maintain this "names" database, including publications such as *The Wall Street Journal*. CDA aggressively uses this names database to ensure that their holdings database is complete and up-to-date. However, holdings data for very small funds are often difficult to obtain, so some very small funds are missing from the database. Such omissions, we believe, do not materially affect the results of our study.

Table I

Characteristic (Style) Analysis

In the table below, the average characteristic (or style) is shown for funds in different investment-objective groups. To calculate this average characteristic, the quintile benchmark portfolio number (1 through 5) is identified, each year, for each stock held by each mutual fund for each of three different characteristics. These characteristics are market capitalization, book-to-market equity value, and prior-year return. Size portfolio 1 consists of small stocks, book-to-market portfolio 1 consists of low book-to-market stocks, and momentum portfolio 1 consists of low prior-year return stocks. Then, for each mutual fund, the portfolio-weighted average benchmark portfolio number is computed each year (for each characteristic). Finally, the fund average benchmark portfolio numbers are averaged across all funds existing in each category each year, and the resulting time-series is averaged over all years from 1975 to 1994. Before June 30, 1980, the investment objective of December 31, 1974 was used to classify funds.

	Average Size Quintile	Average Book-to-Market Qunitile	Average Momentum Quintile
All funds[a]	3.97	2.87	3.11
Aggressive growth	3.20	2.50	3.31
Growth	3.99	2.75	3.10
Growth and income	4.34	3.13	3.03
Balanced or income	4.34	3.29	3.03

[a] Limited to funds having an investment objective of aggressive-growth, growth, growth-in-come, income, or balanced.

book-to-market quintile 4, and momentum quintile 4.[7] The styles were then averaged across all funds during a given year. Finally, the time-series average was computed over the 1975 to 1994 period.

For all funds, the average size quintile number is 3.97, which means that on average, mutual funds hold stocks that fall just below the 4th quintile in size.[8] The stocks have approximately the median (NYSE) book-to-market ratio, falling on average in the 2.87 quintile, but show above-median momentum. Although there are differences in styles that are consistent with conventional wisdom (for example, aggressive-growth funds pick smaller stocks with lower book-to-market ratios and higher momentum), the differences are not as big as we had thought they would be. In unreported results, we recalculated the styles over five-year subperiods. The results showed a fairly consistent investment style by each investment objective group across subperiods. We interpret this as indicating that the funds, as a group, generally do not attempt to "time" the styles; that is, they do not try to invest more heavily in high momentum stocks just before a time period where the momentum effect is especially strong.

[7] Quintile breakpoints were based on NYSE stocks only, but NYSE, AMEX, and Nasdaq stocks were included in the style measurement. Further discussion on the formation of these portfolios will be presented shortly, and is also contained in the Appendix.

[8] This suggests that the average mutual fund picks stocks that are larger than the average stock. However, this is not surprising, since equilibrium adding-up constraints require funds, as a group, to hold more shares in large companies than in small companies.

II. Performance-Evaluation Methodology

For each fund, we first calculate the hypothetical monthly returns that would be generated by buying the number of shares of each Center for Research in Security Prices (CRSP)-listed stock held by the fund (as listed in the quarterly report of the fund) on the first day of each quarter and holding the portfolio until the first day of the following quarter (when the portfolio weights are updated based on the following quarterly report).[9] We report these hypothetical fund returns as an estimate of the gross returns of the fund. To arrive at a performance measure, we subtract the returns of a comparison or benchmark portfolio from these hypothetical fund returns. As discussed below, although the different performance measures are all based on this general principle, they generate different conclusions about performance because they use different benchmarks.

The performance measures used in this article include:

1. A characteristic-based approach that uses benchmark portfolios constructed to match the characteristics of the stocks held by a mutual fund. This approach yields new CS, CT, and AS measures.
2. The GT measure (Grinblatt and Titman 1993), which uses the past portfolio weights of a fund as a benchmark.
3. A Jensen Measure using the Carhart (1997) four factor portfolios as benchmarks.
4. A Jensen Measure using the CRSP value-weighted index as a benchmark.

The two Jensen Measures are based on traditional covariance-based pricing models of risk and return. In contrast, the GT model makes no assumptions about the relation between risk and return, but does assume that the expected return of each stock does not change systematically from period to period. The characteristic-based approach assumes that characteristics provide better estimates of expected returns than do factor sensitivities (see Daniel and Titman (1997)). Although this approach is not necessarily inconsistent with traditional covariance-based pricing models, it also provides an appropriate measure when these models are violated.

We also use the Carhart four-factor model in combination with the GT measure and the CS measure. All measures are described in detail in the following subsections.

A. The Characteristic-Based Performance Measures

We calculate characteristic-based measures that decompose the overall excess return of a fund into CS, CT, and AS measures. These three measures capture the three separate aspects of performance.

[9] In cases where a quarterly report is missing from the dataset, we generate returns using the number of shares listed in the most recently available quarterly report (adjusted for, e.g., stock splits), updating when the next report becomes available. Although many funds do not report every quarter, all funds are required to report shareholdings twice per year.

B. The CS Measure

The CS measure uses as a benchmark the return of a portfolio of stocks that is matched to the fund's holdings each quarter along the dimensions of size (market value of equity), book-to-market ratio, and momentum (the prior year return of the stock).[10] We match based on these three characteristics because past research has shown that these are the best ex ante predictors of cross-sectional patterns in common stock returns.[11] A CS measure of zero tells us that the performance of a fund could have been replicated, on average, by simply purchasing stocks with the same size, book-to-market, and momentum characteristics as the stocks that the fund held, and a positive and significant CS measure suggests that a manager had additional selectivity ability.[12]

The Appendix describes the construction of the 125 portfolios that are used as benchmark portfolios for the characteristic-based performance measures. To form these portfolios, we group the universe of common stocks listed from the NYSE, American Stock Exchange (AMEX), and Nasdaq into three quintile groupings based on the firm's size, book-to-market ratio, and prior-year return.[13] The $5 \times 5 \times 5$ sorting for the portfolios gives a total of 125 passive portfolios. The returns of each of these portfolios are calculated by value-weighting the stocks in the portfolio.[14]

Once we form these 125 passive portfolios, calculating the CS measure is straightforward. Each stock, in each quarter, is assigned to a passive portfolio according to its size, book-to-market, and momentum rank. The excess return of a particular stock is then calculated by subtracting the passive portfolio's return from the stock's return. These differenced returns are then multiplied by the portfolio weights of the different funds to obtain the abnormal or benchmark-adjusted returns for each of the funds for each month. The month t component of the CS measure is defined as

$$\mathrm{CS}_t = \sum_{j=1}^{N} \tilde{w}_{j,t-1}(\tilde{R}_{j,t} - \tilde{R}_t^{b_{j,t-1}}), \tag{1}$$

where $\tilde{w}_{j,t-1}$ is the portfolio weight on stock j at the end of month $t-1$, $\tilde{R}_{j,t}$ is the month t return of stock j, and $\tilde{R}_t^{b_{j,t-1}}$ is the month t return of the characteristic-based passive portfolio that is matched to stock j during month $t-1$.

[10] For a full description of the calculation of these measures, see the Appendix.

[11] See Fama and French (1992, 1996), Jegadeesh and Titman (1993), and Daniel and Titman (1997).

[12] We note that a similar procedure of comparing returns to the returns of stocks with similar size and book-to-market has been employed by Ikenberry, Lakonishok, and Vermaelen (1995) in an event-study context.

[13] The common stocks in our portfolios must also meet certain data requirements, which are discussed in the Appendix.

[14] We use NYSE-based breakpoints to determine the quintiles, and we value-weight the stocks within each of the 125 passive portfolios to avoid biases in computed returns from rebalancing and to avoid placing too much weight on extremely small stocks.

The time-series average, over all months that a fund exists, gives the CS measure for that fund.[15] In estimating the portfolio weight for a given month (for all measures used in this article), we use the most recent portfolio holdings available for a fund from the CDA database. These most recent holdings are usually the holdings at the end of the most recent past calendar quarter.[16]

For example, if Fidelity Magellan held IBM stock on March 31, 1993, we would subtract from IBM's April, 1993 return the April return of the portfolio that best matched IBM in terms of prior-year return, book-to-market ratio, and size. Subtracting the relevant portfolio return from IBM's stock return gives IBM's benchmark-adjusted return. We multiply IBM's benchmark-adjusted return by its weight in the Magellan portfolio. Repeating this procedure for each stock held by Magellan that month and summing gives Magellan's benchmark-adjusted performance for the month. Averaging over all months gives Magellan's overall selectivity ability.

C. The CT Measure

Note that this selectivity measure does not capture any performance driven by the ability to time the market. Fund managers can generate additional performance if size, book-to-market, or momentum strategies have time-varying expected returns that the manager can exploit by changing portfolio weights to exploit those styles when they are the most profitable.[17] To measure a fund manager's success at timing the different investment styles, we introduce a CT measure. The month t component of this measure is

$$\mathrm{CT}_t = \sum_{j=1}^{N} (\tilde{w}_{j,t-1}\tilde{R}_t^{b_{j,t-1}} - \tilde{w}_{j,t-13}\tilde{R}_t^{b_{j,t-13}}). \tag{2}$$

Note that the portfolio weight of stock j at month $t - 13$ is multiplied by $\tilde{R}_t^{b_{j,t-13}}$, the month t return of the characteristic-based benchmark portfolio that is matched to stock j during month $t - 13$. Thus, if the Magellan fund, for example, increased its weight in high book-to-market stocks at the beginning of a month in which the book-to-market effect was unusually strong, then the Magellan fund would have a positive CT component for that month. Again, the average over all months that Magellan existed is the CT measure for Magellan.

Our characteristic-based method has far higher power to separate out selectivity and timing ability than factor-based methods. For example, to determine whether timing ability is present using factor methods, one must determine whether changes in factor loadings correspond with the realizations of the

[15] Test statistics are also based on time-series methods.

[16] However, since some funds have a reporting calendar that does not coincide with normal calendar quarters, and since some funds only report twice per year, the most recent holdings data (in rare cases) may be up to six months old.

[17] However, Ferson and Schadt (1996) argue that if managers can time the market using publicly available information, it may not be appropriate to view this as performance.

associated factors. With the characteristic measure, one can look directly at whether shifts in the portfolio weights forecast future returns.

D. The AS Measure

Finally, to measure the returns earned by a fund due to that fund's tendency to hold stocks with certain characteristics, we create an AS return measure. The month t component of this measure is

$$\mathrm{AS}_t = \sum_{j=1}^{N} \tilde{w}_{j,t-13} \tilde{R}_t^{b_{j,t-13}}. \tag{3}$$

Each stock held by a fund at month $t - 13$ is matched with its characteristic-based benchmark portfolio of month $t - 13$. The month t return of this benchmark portfolio is then multiplied by the month $t - 13$ portfolio weight, and the resulting product is summed over all stocks held by the fund at month $t - 13$ to give the month t AS component. Finally, the average over all months gives the AS measure. Note that by lagging weights and benchmark portfolios by one year, we eliminate returns due to timing the characteristics. For example, a fund that successfully buys high book-to-market stocks when returns to such a strategy are unusually high will not exhibit an unusually high AS return, since this strategy will most likely involve moving into stocks within a year before the unusually high book-to-market return. However, a fund that systematically holds high book-to-market stocks to boost its portfolio return (without trying to time the effect) will exhibit a high AS Return. The sum of the CS, CT, and AS measures equals the total fund return.[18]

E. The GT Measure of Performance

The GT measure of performance for a given month is computed by multiplying the twelve-month change in each stock's portfolio weight by the following month return of that stock. The month t component for a given fund is

$$\mathrm{GT}_t = \sum_{j=1}^{N} (\tilde{w}_{j,t-1} - \tilde{w}_{j,t-13}) \tilde{R}_{j,t}. \tag{4}$$

The average over all months gives the GT measure for that fund. According to this measure, the benchmark used to adjust the return of a portfolio for its risk in a given month is the current return earned by the portfolio held 12 months prior to the current month's holdings. With sufficient stationarity, subtracting the return achieved by the prior year's holdings in the current month from the actual return in the current month yields a zero investment, zero systematic-

[18] In practice, this equivalence is only approximately true because of the additional requirement that a stock be listed in COMPUSTAT to be included in the calculation of the CS, CT, and AS measures for a fund.

risk portfolio. Hence, the expected difference between the portfolio return and its benchmark return should be zero.

F. The Carhart Measure

The Carhart benchmark is based on an extension of the Fama and French (1993) factor model, and is effectively a four-factor Jensen measure. It assumes that betas with respect to the returns of four zero-investment factor-mimicking portfolios,

1. High book-to-market minus low book-to-market (HML),
2. Small size minus big size (SMB),
3. High prior-year return less low prior-year return (PR1YR), and
4. CRSP value-weighted index less T-bills (RMRF),

are appropriate measures of multidimensional systematic risk (for details see Carhart (1997)). According to this model, in the absence of stock-selection or timing abilities the expected return for a fund is the sum of the risk-free return and the products of the betas with the factor risk-premia, which are simply the expected returns of each of these zero-investment portfolios. α_j, which is the Carhart measure for fund j, is calculated with the regression,

$$R_{j,t} - R_{F,t} = \alpha_j + b_j \cdot \mathrm{RMRF}_t + s_j \cdot \mathrm{SMB}_t + h_j \cdot \mathrm{HML}_t + p_j \cdot \mathrm{PR1YR}_t + e_{j,t}.$$

The Carhart approach, like the characteristic-based approach, in effect identifies a matching passive portfolio return for each fund return. This passive return, which is subtracted from the fund return to generate α_j, is a weighted average of the returns of the Carhart factor portfolios and the return of a one-month T-bill. However, in contrast to the characteristic-based approach, where the matching passive portfolio is based on the stock characteristics per se, the Carhart matching passive portfolio is based on the covariance of the fund returns with the returns of characteristic-based factor portfolios.

Depending on the structure of returns and the behavior of the portfolio manager, characteristic-timing ability, as described above, could bias this performance measure downward. The reason is that, with positive timing ability, the estimated factor betas will be biased upwards.[19]

G. The CAPM-Based (One-Factor) Jensen Measure

The capital asset pricing model (CAPM) risk adjustment simply subtracts, from the fund return, the sum of (1) the risk-free return and (2) the product of the CRSP value-weighted risk premium and the fund beta. Again, this can be thought of as matching each fund in a given month with a passive portfolio, comprised of a weighted-average of T-bills and the value-weighted index, where the weight is given by the fund's CAPM beta. As with the four-factor

[19] See Jensen (1972), Dybvig and Ross (1985), and Grinblatt and Titman (1989b).

measure, timing the market, e.g., increasing beta when the CRSP value-weighted index has a large risk-premium, could bias this measure downward.

H. Carhart-Adjusted GT and CS Measures

As a consistency check of the GT and CS measures, we will consider additional measures that apply the four-factor Carhart regression model to the benchmark-adjusted returns generated using the GT and CS measures. Recall that the GT and CS measures calculate the returns of zero-cost portfolios that investors can, at least in principle, create. In theory, these zero-cost portfolios should have zero factor risk, so that their mean returns should provide a measure of abnormal return. However, in reality, the portfolios may not have zero factor risk. Therefore, in order to further eliminate the effect of factor risk on these benchmark-adjusted returns, we regress the returns on the returns of the Carhart factors and measure performance as the intercept of the regression.

III. Results

A. Mutual Fund Performance From 1975 to 1994

We first present an overview of the performance of our sample of mutual funds over the 20-year period from 1975 to 1994. Table II provides a year-by-year comparison of the gross returns of the average fund with the CRSP NYSE/AMEX value-weighted and equal-weighted indices. The column labeled "Gross Return" shows the average pre-expense hypothetical portfolio return, each year, for all funds having an investment objective of aggressive-growth, growth, growth-income, income, or balanced. In classifying funds each month of a given year, we used the self-declared investment objective from the most recent quarter-end, except for the period before June 30, 1980, where we included all funds.[20] The gross return shown for each year is the average monthly return for that year, annualized to percent per year.

Inspection of the yearly returns indicates that the average fund return generally falls between the value-weighted and equally-weighted index returns, which is due to mutual funds holding more small stocks than is represented by the value-weighted portfolio. The average fund outperformed the value-weighted index nine out of ten years from 1975 to 1984, with a yearly average of 2.8 percent above the index. By comparison, the average fund outperformed the index six out of ten times from 1985 to 1994, with a yearly average of 1.3 percent.

Also presented in Table II are results for two measures of performance for the average mutual fund during this 20-year period. The first measure presented is the Grinblatt and Titman (1993) measure (GT). Yearly results for the GT

[20] We included all funds before June 30, 1980 because CDA began collecting investment objective information on that date. However, sector funds represented a small proportion of all funds before 1980. In the remainder of this article, we include, prior to June 30, 1980, only funds with an investment objective of AG, G, GI, I, or B at the beginning of 1975. We collected this investment objective information from another source to supplement the CDA data.

Table II

Mutual Fund Returns, 1975–1994

For each year (row) below, the average buy-and-hold monthly return (annualized to percent per year) is presented for the Center for Research in Security Prices value weighted New York Stock Exchange/American Stock Exchange (CRSP VW) portfolio, CRSP equally weighted NYSE/AMEX portfolio (CRSP EW), and the equally weighted portfolio of all funds existing during the year with a self-declared investment objective of aggressive-growth, growth, growth-income, income, or balanced. In addition, the total number of funds in these five investment-objective categories at the end of each year is shown. For the period before 1980, all funds (regardless of investment objective) are included in the averages. Finally, the GT and CS (Characteristic Selectivity) performance measures are presented for the equally weighted portfolio of the same funds. The GT measure is calculated by subtracting the time t return of the portfolio held at month $t - 13$ from the time t return of the portfolio held at $t - 1$. The CS measure is the difference between the time t return of the portfolio held at time $t - 1$ and the time t return of the time $t - 1$ matching control portfolio. All portfolios are rebalanced monthly. Time-series t-statistics are shown in parentheses.

			Equally Weighted Portfolio of Funds[a]			
Year	CRSP VW w/Dividends	CRSP EW w/Dividends	Number	Gross Return	GT Performance	CS Performance
1975	34.00	57.23	383	36.28	NA	0.78
						(0.49)
1976	24.64	43.53	383	26.97	1.47	0.38
					(1.33)	(0.27)
1977	−3.81	14.90	588	0.82	2.01*	0.91
					(1.93)	(0.89)
1978	9.00	21.90	547	14.19	2.21	3.00*
					(1.17)	(2.02)
1979	22.33	35.84	545	29.50	2.12	2.50*
					(1.75)	(1.96)
1980	30.50	32.73	485	34.60	3.67	1.06
					(1.65)	(0.55)
1981	−3.48	5.73	479	−1.39	0.25	0.72
					(0.13)	(0.60)
1982	20.21	27.40	481	24.64	5.39***	2.59
					(3.47)	(1.37)
1983	21.37	32.54	481	22.40	0.75	0.99
					(0.46)	(0.93)
1984	5.86	−2.07	508	0.96	0.25	−1.40
					(0.29)	(−0.78)
1985	28.11	24.85	520	32.91	1.99	−0.23
					(1.63)	(−0.16)
1986	17.07	12.46	582	15.67	1.55	0.06
					(0.98)	(0.04)

measure show that the average fund, before fees and expenses, has consistently provided a positive performance measure. GT performance is positive in 17 years, and is significantly positive in three years. By contrast, GT performance is negative (and insignificant) in only two years. The 20-year average GT performance is 1.91 percent per year, with very similar results across the two ten-year subperiods. These results are very close to those for a smaller set of funds analyzed in

Table II—*Continued*

Year	CRSP VW w/Dividends	CRSP EW w/Dividends	Equally Weighted Portfolio of Funds[a] Number	Gross Return	GT Performance	CS Performance
1987	7.29	0.05	653	8.01	2.20	2.04
					(1.00)	(0.74)
1988	16.66	19.59	735	17.43	−0.85	−0.51
					(−0.47)	(−0.30)
1989	26.61	15.53	786	26.34	4.04***	0.28
					(4.23)	(0.30)
1990	−3.49	−21.35	829	−5.95	2.87	0.11
					(1.75)	(0.05)
1991	27.95	35.27	984	34.84	3.02	1.15
					(1.36)	(0.72)
1992	7.94	20.10	1,049	9.73	0.18	0.05
					(0.17)	(0.04)
1993	10.61	21.16	1,421	13.12	3.65	0.75
					(1.60)	(0.60)
1994	0.14	−2.87	1,973	−0.44	−0.40	0.19
					(−0.52)	(0.17)
1975–1994[b]	14.98	19.73		17.03	1.91***	0.77**
					(5.19)	(2.23)
1975–1984[b]	16.06	26.97		18.90	2.01***	1.15**
					(3.88)	(2.45)
1985–1994	13.89	12.48		15.17	1.82***	0.39
					(3.47)	(0.77)

* Significant at the 10 percent significance level.

** Significant at the 5 percent significance level.

*** Significant at the 1 percent significance level.

[a] Includes funds with a self-declared investment objective of aggressive-growth, growth, growth-income, income, and balanced. Before June 30, 1980, all funds, regardless of investment objective, were included in the averages.

[b] GT measure begins in 1976.

the GT article, which found an average GT performance of 1.93 percent per year for 261 funds during the first ten-year period.

As noted earlier, GTW suggest that the positive GT measures may result from funds employing a simple momentum screen. To assess this, in Panel A of Table III we regress the monthly GT measures on the four Carhart factor portfolios. The first row presents results for the equally weighted portfolio of all equity funds (we will discuss the results for aggressive-growth funds and for growth funds in the next section of this article). The results clearly show that the GT measure does not account for cross-sectional return differences due either to momentum or to market capitalization. The resulting Carhart-adjusted GT measure is 56 basis points per year, as compared to an unadjusted GT measure of 191 basis points (see Table II).[21]

[21] We also regressed the monthly GT performance on only the momentum portfolio (PR1YR). The Carhart-adjusted GT measure from that regression was 77 basis points per year, which

The second performance measure presented in Tables II and III is the CS measure, which should not be subject to the same biases as the GT measure. As discussed in Section II.B, this measure should eliminate performance due to cross-sectional differences in stock returns attributable to the size, book-to-market, and momentum anomalies. Any remaining performance should therefore be evidence that mutual funds have selectivity ability.

The results presented in Table II indicate that the average fund has a significantly positive selectivity measure, with a 20-year average of 77 basis points per year. However, the evidence indicates that this performance is mainly concentrated in the first ten-year period, which exhibits a performance level nearly three times that of the second ten years.

To verify that the positive CS measure is not a result of loading on priced factor risk, we regress the monthly CS measures on Carhart portfolio monthly returns. As shown in Panel B of Table 3, the Carhart factors have only a modest effect on measured performance. The Carhart-adjusted CS measure for all funds, shown in the first row, is 56 basis points per year, as compared with the unadjusted CS measure of 77 basis points per year (Table II), and is statistically significant at the 10 percent level.[22] This result suggests that mutual fund portfolio managers are in fact beating the mechanical strategies represented by our benchmarks. However, the average benchmark-adjusted returns are fairly small—about the same size as the typical management fee.

B. A Closer Look at the Measures of Performance

Table IV provides a more comprehensive analysis of performance. For comparison, we add results for two additional measures that were not included in Table II: a one-factor Jensen measure and a Carhart four-factor Jensen measure. Also, we break down the performance by fund type.

To calculate the Carhart measures, we regressed the time-series of monthly buy-and-hold excess returns for an equally-weighted portfolio of funds existing at the beginning of each month on the four Carhart factor portfolio monthly returns.[23] To calculate the one-factor Jensen measure, we regressed the same dependent variable on the first Carhart factor, which is the monthly buy-and-hold excess return associated with a value-weighted market proxy portfolio. Separate regressions are shown for each five-year subperiod, along with the entire 20-year period. Note that 20-year averages are slightly different than those shown in Table II, as we now require funds to have investment-objective data at the beginning of 1975 to be included in the pre-1980 series.

suggests that the momentum effect is chiefly responsible for the magnitude of the GT performance of our mutual funds. This finding is consistent with the correlation between GT performance and the tendency to invest on momentum documented by GTW.

[22] Since the CS measure is the fund return net of the return of a portfolio of stocks with similar size, book-to-market, and momentum characteristics, it would seem that the loadings of the benchmark-adjusted returns on the four factors should be zero. However, this will not be the case if managers are able to pick stocks that, for example, have high book-to-market ratios but whose returns do not behave like the returns of other high book-to-market stocks.

[23] Panel A reflects results for all funds, while Panels B through E reflect results for investment-objective subgroups.

Table III

Carhart-Adjusted Measures

The dependent variables in these regressions are the time-series of monthly GT performance components that would accrue to buying an equally weighted portfolio, each month, of all funds having the appropriate investment objective during that month. Before June 30, 1980, the investment objective of December 31, 1974 was used to classify funds. The four independent variables are the time-series of monthly returns associated (1) with a value-weighted market proxy portfolio minus T-bills (RMRF), (2) with the difference in returns between small and large capitalization stocks (SMB), (3) with the difference in returns between high and low book-to-market stocks (HML), and (4) with the difference in returns between stocks having high and low prior-year returns (PR1YR). See Carhart (1997) for further description of these variables. The intercept has been annualized to percent per year. The *t*-statistics are shown in parentheses below coefficient estimates.

	Independent Variables				
Dependent Variable	Intercept	RMRF	SMB	HML	PR1YR
Panel A. Regressions of GT Measures on Carhart Factors (1976–1994)					
All funds[a]	0.56*	0.01	0.03***	0.01	0.11***
	(1.91)	(1.40)	(2.71)	(1.04)	(13.11)
Aggressive-growth funds	1.76***	0.03***	0.03*	0.01	0.18***
	(3.24)	(2.89)	(1.72)	(0.64)	(11.70)
Growth funds	0.47	0.01	0.04***	0.01	0.12***
	(1.43)	(1.12)	(3.72)	(0.76)	(13.18)
Panel B. Regressions of CS Measures on Carhart Factors (1975–1994)					
All funds[a]	0.56*	0.04***	0.03***	−0.06***	0.01
	(1.87)	(6.24)	(2.91)	(−5.66)	(0.93)
Aggressive-growth funds	1.41**	0.10***	0.08***	−0.20***	0.005
	(2.07)	(7.26)	(3.55)	(−8.65)	(0.27)
Growth funds	0.54	0.05***	0.04***	−0.06***	0.01
	(1.60)	(6.97)	(3.92)	(−5.33)	(1.19)

* Statistically significant at the 10-percent level.

** Statistically significant at the 5-percent level.

*** Statistically significant at the 1-percent level.

[a] Limited to funds having an investment objective of aggressive-growth, growth, growth-income, income, or balanced.

Table IV

Comparison of Performance Measures

In each panel below, five measures of performance are presented. To calculate the gross return, the GT performance, and the CS performance, an equally weighted portfolio of all funds existing during each month and having a self-declared investment objective as specified in each panel during the most-recent quarter-end is formed. Before June 30, 1980, the investment objective of December 31, 1974 is used. The buy-and-hold return (or performance) of that portfolio during that month is then computed, and, finally, the time-series average across all months in a given time period is computed. For the one-factor Jensen measures, the monthly buy-and-hold excess return on the equally weighted portfolio of funds was regressed on the time-series of monthly returns associated with a value-weighted market proxy portfolio minus T-bills. For the four-factor Jensen measure, the same time-series was regressed on three additional monthly time-series, associated with (1) the difference in returns between small- and large-capitalization stocks, (2) the difference in returns between high and low book-to-market stocks, and (3) the difference in returns between stocks having high and low prior-year returns. See Fama and French (1993) and Carhart (1997) for further description of these variables. All returns have been annualized to percent per year. Also, the number of funds existing at the end of each time-period is shown. Time-series t-statistics are shown in parentheses.

	Number	Gross Return	GT Performance	CS Performance	One-Factor Jensen Measure	Four-Factor Jensen Measure
Panel A. All Funds (Returns are Monthly, Annualized to Percent Per Year)[a]						
1975–1979[b]	214	21.33	2.06***	1.58**	2.78**	1.44
			(2.89)	(2.21)	(2.40)	(1.50)
1980–1984	508	16.31	2.10**	0.79	0.62	0.98
			(2.41)	(1.10)	(0.55)	(0.97)
1985–1989	786	20.07	1.79**	0.33	−0.80	0.83
			(2.46)	(0.43)	(−0.69)	(0.96)
1990–1994	1,973	10.26	1.86**	0.45	−0.16	−0.36
			(2.43)	(0.66)	(−0.16)	(−0.68)
1975–1994[c]		16.99	1.94***	0.79**	0.60	0.39
			(5.02)	(2.20)	(1.07)	(0.87)
Panel B. Aggressive-Growth Funds (Returns are Monthly, Annualized to Percent Per Year)						
1975–1979[b]	59	25.43	3.44***	2.71*	5.39**	1.67
			(2.85)	(1.99)	(2.25)	(1.01)
1980–1984	97	15.50	4.16***	1.07	−1.37	1.04
			(2.92)	(0.63)	(−0.46)	(0.56)
1985–1989	150	18.09	5.20***	0.12	−4.95	0.96
			(3.59)	(0.06)	(−1.58)	(0.57)
1990–1994	211	12.49	3.15**	2.07	0.62	0.66
			(2.63)	(1.05)	(0.21)	(0.48)
1975–1994[c]		17.88	4.02***	1.49*	−0.21	0.40
			(6.01)	(1.66)	(−0.14)	(0.43)
Panel C. Growth Funds (Returns are Monthly, Annualized to Percent Per Year)						
1975–1979[b]	69	20.30	1.99***	2.21**	1.37	1.20
			(2.74)	(2.34)	(0.96)	(1.02)
1980–1984	219	16.07	2.35**	0.85	0.25	1.20
			(2.38)	(0.97)	(0.19)	(1.08)
1985–1989	366	20.41	1.58**	0.65	−0.75	1.19
			(2.17)	(0.75)	(−0.57)	(1.31)
1990–1994	1,207	10.11	1.83**	0.42	−0.33	−0.42
			(2.22)	(0.61)	(−0.35)	(−0.80)
1975–1994[c]		16.72	1.93***	1.03**	0.12	0.32
			(4.66)	(2.43)	(0.19)	(0.63)

Table IV—*Continued*

	Number	Gross Return	GT Performance	CS Performance	One-Factor Jensen Measure	Four-Factor Jensen Measure
Panel D. Growth-Income Funds (Returns are Monthly, Annualized to Percent Per Year)						
1975–1979[b]	49	18.98	1.21**	0.62	1.36*	1.30
			(2.40)	(1.11)	(1.75)	(1.62)
1980–1984	122	16.47	1.15	0.72	1.23	1.28
			(1.61)	(1.22)	(1.46)	(1.28)
1985–1989	172	20.51	0.09	−0.12	0.94	0.17
			(0.17)	(−0.35)	(1.01)	(0.18)
1990–1994	354	9.26	1.32**	−0.48	−0.34	−0.91
			(2.46)	(−1.19)	(−0.47)	(−1.48)
1975–1994[c]		16.30	0.93***	0.19	0.80*	0.24
			(3.18)	(0.76)	(1.96)	(0.56)
Panel E. Balanced and Income Funds (Returns are Monthly, Annualized to Percent Per Year)						
1975–1979[b]	37	19.67	1.03**	−0.27	3.15**	1.84
			(2.32)	(−0.45)	(2.52)	(1.63)
1980–1984	70	17.57	0.53	0.39	3.02**	0.52
			(0.89)	(0.39)	(2.08)	(0.43)
1985–1989	98	21.10	0.46	0.19	2.37*	0.67
			(1.03)	(0.30)	(1.90)	(0.63)
1990–1994	201	9.39	1.03*	−0.23	−0.35	−0.67
			(1.68)	(−0.61)	(−0.52)	(−1.02)
1975–1994[c]		16.93	0.74***	0.02	2.09***	0.82
			(2.74)	(0.05)	(3.40)	(1.49)

* Significant at the 10 percent significance level.

** Significant at the 5 percent significance level.

*** Significant at the 1 percent significance level.

[a] Limited to funds having an investment objective of aggressive-growth, growth, growth-income, income, or balanced.

[b] 1976–1979 for the GT measure.

[c] 1976–1994 for the GT measure.

The results in Panel A show that, over the entire sample period, both the one-factor and the four-factor Jensen measures of performance are positive, but insignificant. The one-factor Jensen measure itself is only 24 percent smaller than the CS measure, but the *t*-statistic is considerably smaller because of the higher standard error on this coefficient. The Carhart four-factor measure, which is 39 basis points per year, is similar in magnitude to the Carhart-adjusted GT and CS measures presented in the last section, which were both 56 basis points per year. However, the simple Carhart measure, with its higher standard error, is insignificant, while the Carhart-adjusted GT and CS measures were both significant.

Panels B, C, D, and E present results for aggressive-growth, growth, growth-income, and balanced and income funds, respectively. The GT measure for the average aggressive-growth fund, 4.02 percent per year, is more than twice that of the average fund, 1.94 percent per year. This difference could result from either superior selectivity ability or to the greater use of momentum strategies by aggressive-growth funds.

The CS measure for aggressive-growth funds (Panel B), which controls for the momentum effect, is almost double that of the average fund and is just significant at the 10 percent level.[24] However, probably because these funds are employing a momentum screen, the CS measure is considerably lower than the GT measure.

Some evidence in support of this hypothesis is found in Panel A of Table III. The regression of the GT measure for the portfolio of aggressive-growth funds on the four Carhart factors exhibits a very high momentum loading. However, the Carhart-adjusted GT measure, 1.76 percent/year, is highly significant and is more than three times the measure for all funds, suggesting that, at least after using the Carhart factors as a benchmark, the funds still exhibit considerable selectivity ability.[25]

Results for growth funds, shown in Panel C of Table IV and in Panels A and B of Table III, indicate that growth funds are more like the average fund than the average aggressive-growth fund. The GT performance, CS performance, and the Carhart-adjusted GT and CS performance for growth funds are very similar to the measures for all funds. Finally, Panels D and E of Table IV indicate that growth-income and balanced and income funds are the worst performing groups in our sample, whether measured with the GT or the CS performance measures.

C. Persistence of Mutual Fund Performance

Up to this point, we have examined fairly broad averages of the different performance measures. The results are consistent with a Grossman and Stiglitz (1980) view of efficient markets that allow investors to earn just enough to compensate them for their time and effort. In this section we examine whether certain funds consistently outperform other funds. Specifically, we examine the "hot hands" phenomena described in Hendricks, Patel, and Zeckhauser (1993).

Table V presents the average returns, along with the benchmark-adjusted returns, of a strategy that buys the stocks held by the mutual funds with the highest return over the previous calendar year. For example, at the end of 1975, we sorted all mutual funds having at least a one-year track record by their average monthly hypothetical returns during 1975. Then, quintile portfolios were formed, with P1 containing 1975's best funds and P5 containing

[24] The CS performance of aggressive-growth funds also seems to be concentrated in the 1975 to 1979 period, as shown in Panel B of Table IV.

[25] The simple Carhart measure for the aggressive-growth funds, in Panel B of Table IV, is considerably smaller than the Carhart-adjusted GT and CS measures shown in Panels A and B of Table III. This is interesting because the differences between the Carhart-adjusted measures and the simple Carhart measures equals the Carhart measures of the GT and CS benchmarks. Hence, the CS- and GT-benchmarks for the AG funds exhibit negative performance with the Carhart measure, which is indicative of a problem with the Carhart measure. The CS benchmark is selected only on the basis of size, book-to-market, and past-return, and therefore, if the Carhart measure adequately controls for these effects, the measure should be zero. Similarly, since the GT benchmark is selected on the basis of the fund's holdings of the previous-year, the Carhart measure should also be zero for this benchmark.

Table V

Tests of Performance Persistence

This table presents the average monthly buy-and-hold gross return over 1976–1994 (annualized to percent per year) for equally weighted portfolios of mutual funds which were ranked on their gross return during the prior year. To select these funds each year, all funds existing during the entire prior year and having an investment objective of AG, G, GI, I, or B were ranked on their average monthly return of the prior year. Quintile portfolios were formed, and the gross return for the equally weighted portfolio of funds in each quintile was measured over the following year. All mutual funds existing during a given month were included in the following-year returns calculation, even if the fund did not survive the entire year. Then, the entire sort process was repeated for the following year. Finally, the time-series average return for each portfolio was calculated. The portfolio labeled "Best" is the quintile with the highest prior-year return, while "Worst" is the portfolio with the lowest prior-year return. For example, the gross return for a strategy of buying last year's best quintile of funds, each year, was 18.40 percent per year. Also presented are four measures of performance for the gross-returns sort strategy.

	Best *P1*	*P2*	*P3*	*P4*	Worst *P5*	*P1* Minus *P5*
Gross return	18.40	16.50	15.89	14.89	13.80	4.60***
						(2.63)
GT performance	1.79***	1.34***	1.62***	1.86***	2.87***	−1.08**
	(3.96)	(3.34)	(4.40)	(4.70)	(5.41)	(−2.00)
CS performance	1.08**	0.99**	0.87**	0.55	0.18	0.90
	(2.01)	(2.50)	(2.38)	(1.33)	(0.31)	(1.19)
One-factor Jensen measure	2.70**	1.09	0.70	−0.21	−1.51	4.21**
	(2.27)	(1.62)	(1.40)	(−0.36)	(−1.44)	(2.39)
Four-factor Jensen measure	0.16	0.12	0.94**	0.72	0.77	−0.61
	(0.17)	(0.22)	(2.00)	(1.25)	(0.79)	(−0.37)

* Significant at the 10 percent significance level.

** Significant at the 5 percent significance level.

*** Significant at the 1 percent significance level.

Note: Limited to funds having an investment objective of aggressive-growth, growth, growth-income, income, or balanced.

1975's worst funds. Finally, the hypothetical return, GT performance, CS performance, one-factor Jensen measure, and Carhart four-factor Jensen measure were computed for the equally weighted portfolio of funds within each quintile during 1976. All funds existing during any given month were included in the monthly calculations for 1976, even if the fund did not survive until the end of 1976. All quintile portfolios were rebalanced monthly.

The table reveals that the hot hands phenomena found in other data sets also holds here. Stocks held by last-year's best funds outperform other stocks in the following year. Specifically, stocks in the best quintile (P1) outperform stocks in the worst quintile (P5) by 4.6 percent per year.

Among the measures of performance, however, only the one-factor Jensen measure is significantly higher for P1 than for P5. The CS performance is not significantly different between P1 and P5, which suggests that persistence can be explained by the fact that the stocks held by the funds with the best

performance have characteristics that are associated with high returns. This evidence is consistent with earlier evidence in Carhart (1997) and Wermers (1997), who suggest that persistence can be explained by the momentum anomaly. While the GT measure is larger for P5 than for P1, this is due to the GT measure picking up performance associated with momentum.[26]

D. Performance Attribution Analysis

In order to determine the different sources of performance of the funds, we separated fund performance into three components: the CS measure already presented, the CT, and the AS returns. As mentioned in section II, the CT measure is designed to detect performance that is due to a mutual fund successfully timing stock characteristics (e.g., holding high book-to-market stocks when these stocks have their highest returns). The AS measure is designed to detect returns that are due to a mutual fund systematically holding stocks with certain style characteristics (e.g., holding high book-to-market stocks all the time). Table VI presents results for these measures.

The AS components of the fund returns are quite similar across the different investment objectives. However, the CS measure is, on average, about one percent per year higher for aggressive-growth and growth funds than for growth-income and balanced and income funds. Again, this finding appears to be mainly driven by the first ten-year subperiod.[27]

Finally, we find that the CT performance is insignificant for all categories of funds in our analysis of the entire sample period and is never significantly positive in any sample period for any category of funds. This indicates that the average mutual fund is not able to effectively time the different stock characteristics. In unreported results, we calculated the CT performance for individual funds and found no convincing evidence of individual funds successfully timing the characteristics.

IV. Summary and Conclusions

When a professional portfolio manager proposes an investment strategy based on the fundamental analysis of equities, the presumption is that he or

[26] Note that the GT measure calculates returns of a zero-cost portfolio that has a long position in stocks that the fund bought over the preceding year, and a short position in stocks that the fund sold over the preceding year. If last year's best (worst) funds sell some stocks in their portfolios, they are likely to be selling stocks with positive (negative) momentum, since most stocks in their portfolio will have done well (poorly). If these funds randomly choose the stocks they buy, then the zero-cost portfolio corresponding to last year's best (worst) funds contains stocks with negative (positive) momentum on average, resulting in negative (positive) returns, as described by Jegadeesh and Titman (1993).

[27] One can only speculate why the first ten years generates larger risk-adjusted returns than the second ten years. Two hypotheses are likely. One is that financial markets became more efficient over time, perhaps driven by the expansion in professionally managed money. Another is that this expansion diluted the performance of good mutual fund managers because the proportion of such managers decreased appreciably over time relative to the total number of managers.

Table VI

Performance Attribution Analysis

In each panel below, three performance attribution components are presented for funds in different investment objective categories. These three components are calculated as follows. The "average style attribute" (AS) is calculated, for month t, by matching each stock held by a mutual fund, on month $t - 13$, with the proper control portfolio at month $t - 13$. Then, the measure for a fund is computed by applying each stock weight at $t - 13$ to the matching control portfolio return at t. The "characteristic timing attribute" (CT) is computed, for each fund, by matching stocks held at time $t - 13$ and at time $t - 1$ with the proper control portfolios at time $t - 13$ and time $t - 1$, respectively. Next, the portfolio-weighted return of the time $t - 13$ matching portfolio, at time t, is subtracted from the portfolio-weighted return of the time $t - 1$ control portfolio, also at time t. The "characteristic-selectivity attribute" (CS) is calculated as described in the notes to Table II. All numbers presented below are computed by averaging, each month, the given measure over all funds having the appropriate self-reported investment objective during that month. Before June 30, 1980, the investment objective of December 31, 1974 was used to classify the funds. The time-series average and t-statistic are presented below (t-statistic in parentheses).

	AS Attribute	CT Attribute	CS Attribute[a]
Panel A. 1976–1994 (Monthly, Annualized to Percent Per Year)			
All funds[b]	14.85	0.02	0.79**
		(0.11)	(2.20)
Aggressive-growth	14.89	0.23	1.49*
		(0.90)	(1.66)
Growth	14.56	−0.06	1.03**
		(−0.29)	(2.43)
Growth-income	14.75	−0.02	0.19
		(−0.11)	(0.76)
Balanced or income	15.23	0.04	0.02
		(0.14)	(0.05)
Panel B. 1976–1984 (Monthly, Annualized to Percent Per Year)			
All funds[b]	15.01	0.39	1.18**
		(1.39)	(2.34)
Aggressive-growth	15.87	0.40	1.89*
		(0.90)	(1.75)
Growth	14.41	0.26	1.53**
		(0.72)	(2.37)
Growth-income	14.35	0.51	0.67*
		(1.59)	(1.66)
Balanced or income	15.42	0.56	0.06
		(1.06)	(0.10)
Panel C. 1985–1994 (Monthly, Annualized to Percent Per Year)			
All funds[b]	14.71	−0.31	0.39
		(−1.65)	(0.77)
Aggressive-growth	14.02	0.08	1.09
		(0.28)	(0.76)
Growth	14.69	−0.35	0.54
		(−1.58)	(0.97)
Growth-income	15.11	−0.50**	−0.30
		(−2.00)	(−1.14)
Balanced or income	15.06	−0.43*	−0.02
		(−1.87)	(−0.06)

* Significant at the 10 percent significance level.

** Significant at the 5 percent significance level.

*** Significant at the 1 percent significance level.

[a] CS attribute starts in 1975.

[b] Limited to funds having an investment objective of aggressive-growth, growth, growth-income, income, or balanced.

she expects the strategy to outperform simpler, purely mechanical, strategies based on stock characteristics like book-to-market, size, and momentum. As companies like Vanguard have demonstrated, simple mechanical portfolio strategies can be implemented at substantially lower cost than the more subjective strategies used by most mutual funds. Therefore, if the active mutual funds fail to beat the mechanical strategies, they may be wasting resources.

This article introduces a characteristic-based benchmark that is designed to measure whether mutual funds pick stocks that outperform simple mechanical rules. The evidence presented in this article suggests that the average mutual fund does, in fact, succeed along this dimension. However, we find that the amount by which the average mutual fund beats a mechanical strategy is fairly small (under 100 basis points) and is approximately equal to the average management fee. Aggressive-growth and growth funds, which exhibit the highest performance, probably also generate the largest costs. This evidence is consistent with an equilibrium, like that of Grossman and Stiglitz (1980), where informed traders are able to outperform the market just enough to earn back their fees.

The much stronger performance numbers described in Grinblatt and Titman (1993), and replicated with the larger data set in this study, can be attributed to the fact that the benchmark used by Grinblatt and Titman does not control for performance due to momentum investing. However, momentum investing does not entirely explain why aggressive-growth and growth funds tend to outperform growth-income and income funds. Although the aggressive-growth and growth funds do pick stocks with higher momentum, on average, the residual performance after controlling for momentum is still somewhat higher for these funds than for growth-income and balanced and income funds.

We have discussed our characteristic-based benchmark with portfolio managers who have performed very well by implementing momentum and high book-to-market strategies. As one might expect, these individuals are not particularly enthusiastic about a benchmark that gives them no credit for having been insightful enough to implement such strategies. The characteristic-based selectivity measure assigns no significant abnormal performance to those investors who simply follow the same mechanical characteristic-based strategy over the entire time period. This is true even if that strategy did extremely well.

Our characteristic-based approach does, however, attribute abnormal performance to those portfolio managers who change their investment styles over time, implementing the styles when they have the highest expected returns. These investors will show abnormal performance with our style timing measure. We find no evidence that funds are successful style timers.

Although average abnormal performance is relatively small, it might still be the case that some funds perform extremely well. Following a number of authors, we address this issue by examining whether past performance is indicative of future performance. Our initial evidence suggests that the "hot hands" phenomena documented by Hendricks, Patel, and Zeckhauser (1993)

can be explained by our different benchmarks. In future work, we plan on looking at individual funds in more detail to determine which funds are exhibiting abnormal performance, and the extent to which the better-performing funds can be identified, ex ante.

Appendix

Construction of the Characteristic-Sorted Benchmark Portfolios

This appendix discusses the formation of the 125 size, book-to-market, and momentum sorted benchmark characteristic portfolios.

Beginning in July 1972, and in each following July, we place every common stock listed on NYSE, AMEX, and Nasdaq into portfolios, provided these firms meet our data requirements. Our criteria for inclusion are similar to those spelled out in Fama and French (1993). We require that COMPUSTAT data be available for at least two years prior to the inclusion of the firm in the sample, and that the firm have market value data available on CRSP at the end of December and the end of June preceding the formation date. In addition, we require that the firm have at least six monthly returns available on CRSP in the 12 months preceding the formation date (for the momentum calculation). The portfolios are all value-weighted, buy-and-hold portfolios.

The composition of each of the 125 portfolios is based on a triple-sort on each firm's market equity value (or size), book-to-market ratio, and momentum. Each formation date, the universe of common stocks is first sorted into quintiles based on each firm's market equity just prior to the formation date (i.e., on the last day of June). The breakpoints for this sort are based on NYSE firms only, although NYSE, AMEX, and Nasdaq stocks are included in the analysis.[28] Then, the firms within each size quintile are further sorted into quintiles based on their book-to-market ratio. The book-to-market ratio is the ratio of the book-value at the end of the firm's fiscal year during the calendar year preceding the formation date to the market value at the end of the preceding December. Here, we "industry adjust" the book-to-market ratios by subtracting the long-term industry average book-to-market ratio from each individual firm's ratio, following Cohen and Polk (1995). We define 50 industries depending on the underlying firm's principal Standard Industrial Classification (SIC) code as reported by CRSP.[29]

Finally, the firms in each of the 25 size/BM portfolios are then sorted into quintiles based on their preceding twelve-month return, giving us a total of 125 portfolios. However, here the preceding twelve-month return is calculated through the end of May (that is, the return up through one month prior to the formation date). This method avoids problems associated with the bid-ask bounce and monthly return reversals (see Jegadeesh (1990)).

[28] That is, the size breakpoints are designed so that there will be an equal number of NYSE firms in each of the five portfolios.

[29] Our 50 industries are the same as those defined in Fama and French (1994).

Summary statistics on the mean returns, sizes, and average characteristics of these portfolios are available upon request from the authors.

REFERENCES

Brown, Steven J., William N. Goetzmann, Roger G. Ibbotson, and Stephen A. Ross, 1992, Survivorship bias in performance studies, *Review of Financial Studies* 5, 553–580.

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

Cohen, Randolph B., and Christopher K. Polk, 1995, An investigation of the impact of industry factors in asset-pricing tests, Working paper, University of Chicago.

Daniel, Kent D., and Sheridan Titman, 1997, Evidence on the characteristics of cross-sectional variation in common stock returns, *Journal of Finance*, 52, 1–33.

Dybvig, Philip H., and Stephen A. Ross, 1985, Differential information and performance measurement using a security market line, *Journal of Finance* 40, 383–399.

Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *Journal of Finance* 47, 427–465.

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

Fama, Eugene F., and Kenneth R. French, 1994, Industry costs of equity, Working paper, University of Chicago.

Fama, Eugene F., and Kenneth R. French, 1996, Multifactor explanations of asset pricing anomalies, *Journal of Finance* 51, 55–84.

Ferson, Wayne E., and Rudi W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *Journal of Finance* 51, 425–461.

Goetzmann, William N., and Roger G. Ibbotson, 1994, Do winners repeat: Predicting mutual fund performance, *The Journal of Portfolio Management* 20, 9–18.

Grinblatt, Mark, and Sheridan Titman, 1989a, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of Business* 62, 393–416.

Grinblatt, Mark, and Sheridan Titman, 1989b, Portfolio performance evaluation: Old issues and new insights, *Review of Financial Studies* 2, 393–422.

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

Grinblatt, Mark, and Sheridan Titman, 1993, Performance measurement without benchmarks: An examination of mutual fund returns, *Journal of Business* 66, 47–68.

Grinblatt, Mark, Sheridan Titman, and Russ Wermers, 1995, Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior, *American Economic Review* 85, 1088–1105.

Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *American Economic Review* 70, 393–408.

Gruber, Martin J., 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 relative performance, 1974–1988, *Journal of Finance* 48, 93–130.

Ikenberry, David, Josef Lakonishok, and Theo Vermaelen, 1995, Market underreaction to open market share repurchases, *Journal of Financial Economics* 39, 181–208.

Jegadeesh, Narasimhan, 1990, Evidence of predictable behavior in security prices, *Journal of Finance* 45, 881–898.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Jensen, Michael C, 1968, The performance of mutual funds in the period 1945–1964, *Journal of Finance* 23, 389–416.

Malkiel, Burton G., 1995, Returns from investing in equity mutual funds 1971 to 1991, *Journal of Finance* 50, 549–572.

Wermers, Russ, 1997, Momentum investment strategies of mutual funds, performance persistence, and survivorship bias, Working paper, University of Colorado.