Mark Grinblatt
University of California, Los Angeles

Sheridan Titman
University of California, Los Angeles, and
Hong Kong University of Science and Technology

Performance Measurement without Benchmarks: An Examination of Mutual Fund Returns*

This article introduces a new measure of portfolio performance and applies it to study the performance of a large sample of mutual funds. In contrast to previous studies of mutual fund performance, the measure used in this study employs portfolio holdings and does not require the use of a benchmark portfolio. It finds that the portfolio choices of mutual fund managers, particularly those that managed aggressive growth funds, earned significantly positive risk-adjusted returns in the 1976–85 period.

I. Introduction

To measure portfolio performance, academic studies as well as professional performance evaluators have traditionally employed performance measures that compare the returns of managed portfolios to the returns of a benchmark portfolio. The most widely used measure in the academic literature, the Jensen Measure, is the intercept from a regression of the excess return (return minus the risk-free rate) of the managed portfolio on the excess return of a benchmark portfolio. Despite its wide use, this measure has been subject to considerable criticism. For example, Roll (1978) demonstrated that this measure can be sensitive to the choice of the benchmark portfolio. Moreover, as Jensen (1972) and others have shown, the Jensen Measure may provide a biased evaluation for market timers.

This traditional method of evaluating portfolio performance does not utilize information that is often available about the composition of the evaluated portfolio. When the composition of the evaluated portfolio is used, the performance

* The authors would like to thank Bret Pearlman and Russ Wermers for research assistance and seminar participants at UCLA, Boston College, the University of Utah, Dartmouth College, Stanford University, and University of California, Berkeley, for comments on earlier drafts.

(*Journal of Business*, 1993, vol. 66, no. 1)

0021-9398/93/6601-0005$01.50

evaluator can eliminate the need to compare returns to a benchmark portfolio, with all of its attendant problems. Cornell (1979) recognized this and suggested a performance measure that is an adaptation of the event study methodology. This measure was later refined and used by Copeland and Mayers (1982) in their analysis of Value Line rankings.

The Event Study Measure calculates the difference between the returns of assets when they are in the portfolio (the event period) with their returns at later dates (the comparison period). The basic idea is that the assets held by informed portfolio managers will have higher returns when they are included in the portfolio than when they are not included. The reason that later rather than earlier period returns are used for the comparison is that some portfolio managers pick their assets in part based on past returns, so that inference based on earlier returns is likely to be systematically biased. However, requiring the observation of asset returns in periods after they are held can create a different kind of bias if it forces the evaluator to ignore assets that lack returns in the comparison period. For portfolios that include the assets of firms that are either near-bankrupt or are takeover targets, the bias that is induced by this survival requirement can be severe.

The Event Study Measure provides an estimate of the sum of the time-series covariances between the portfolio weights and the subsequent returns of each asset included in the evaluated portfolio.[1] This sum represents an intuitive measure of performance since it equals the difference between the realized return of the managed portfolio and its expected return conditioned on the portfolio manager being uninformed (i.e., the return achievable if portfolio weights were uncorrelated with asset returns). In the next section, we introduce a new performance measure that can be viewed as an alternative way to compute this covariance. This new measure, unlike the Event Study Measure, is not subject to survivorship bias and has some computational advantages for statistical inference.

This new performance measure is used to evaluate the quarterly holdings of 155 mutual funds from December 31, 1974, to December 31, 1984. As such, this is the first study that we are aware of that provides estimates of the performance of managed portfolios that are not subject to the benchmark problems discussed by Roll and others. The mutual funds that we examine were previously evaluated by Grinblatt and Titman (1989*a*). In that paper, Jensen Measures were computed with four separate benchmarks: the Center for Research in Security Prices (CRSP) equally weighted index, the CRSP value-weighted

1. Grinblatt and Titman (1989*b*) showed this under the assumption that the event period weights are uncorrelated with the comparison period returns. This assumption is plausible if the information used by the manager to change his holdings is fully incorporated into market prices by the beginning of the comparison period.

index, the Lehmann and Modest (1987) 10-factor benchmark, and an eight-portfolio benchmark developed from stocks grouped on the basis of size, dividend yield, and past returns. The abnormal returns of the mutual funds, calculated with our new measure, are very similar to the abnormal returns calculated with the Grinblatt and Titman eight-portfolio benchmark. They are quite different, however, from those calculated with the other three benchmarks. Like the earlier paper, we find positive performance on average, with the aggressive growth and growth funds performing the best. (The venture capital/special situations category also performed well, but this category contained only three funds.)

II. Measuring Performance with Observable Portfolio Weights

A. *Motivating the Measure*

Our measure is based on the assumption that, from the perspective of uninformed investors, the vector of expected asset return is constant over time. This implies that the portfolio holdings of an uninformed investor cannot be correlated with future asset returns. However, since an informed investor can predict when certain assets will have either higher or lower than average returns, from his perspective, the vector of expected returns changes over time. He can therefore profit from these changing expected returns by tilting his portfolio weights over time in favor of assets with expected returns that have increased and away from assets with expected returns that have decreased.

B. *Relation to Prior Work*

An informed manager whose percentage holdings of an asset is monotonically increasing in its conditional expected return will exhibit a positive unconditional covariance between the proportional holdings of a particular asset and the subsequent returns of that asset. A particular aggregation of these covariances that has some intuitive appeal as well as theoretical support is the sum, over all investments, of the covariances between portfolio weights and returns,

$$\text{cov} = \sum_{j=1}^{N} (E[w_j R_j] - E[w_j]E[R_j]). \tag{1}$$

This sum can be thought of as the actual expected return of the investor's portfolio less what the portfolio's expected return would be if his portfolio weights and asset returns were uncorrelated. Alternatively, the latter term inside the sum can be thought of as the appropriate adjustment for risk, since it represents the expected return of a constant weight portfolio with the same average risk as the evaluated portfolio. Under plausible conditions, the expression in (1) will be positive for informed investors and zero for uninformed investors.[2]

Rewriting the expression in (1) in two other ways leads to other interesting interpretations that prove to be useful for performance evaluation:

$$\text{cov} = \sum_{j=1}^{N} E[w_j(R_j - E[R_j])], \tag{1a}$$

and (1b)

$$\text{cov} = \sum_{j=1}^{N} E[(w_j - E[w_j])R_j].$$

Expression (1a) is the foundation for the Event Study Measure. Expression (1b) is the foundation for our new measure. In contrast to the Event Study Measure, which requires an estimate of the unconditional return $E(R_j)$, our new measure requires an estimate of the expected weight.

The identities between (1), (1a), and (1b) hold at the sample covariance level as well, since

$$\text{scov}(w_j, R_j) = \sum_{t=1}^{T} w_{jt}(R_{jt} - \overline{R}_j)/T = \sum_{t=1}^{T} (w_{jt} - \overline{w}_j)R_{jt}/T, \tag{2}$$

where

scov = sample covariance between the weights and returns of asset j,

w_{jt} = the portfolio weight at the beginning of period t (with sample mean w_j), and

R_{jt} = the portfolio return from date t to date $t + 1$ (with sample mean $\overline{R}_j$).

C. Estimating the Measure

The last subsection presented two equivalent ways to compute a covariance between portfolio weights and the return of a security—the first approach uses an estimate of the expected return for each asset, the second uses an estimate of the asset's expected portfolio weight. This section focuses on the best way to estimate these expectations.

2. Note that each of the N terms in eq. (1) need not be positive for an informed investor. This is because, if two assets are correlated, and both are forecasted to have higher than normal returns, an increase in the proportion held of one asset might require a decrease in the proportion held of another asset for hedging purposes. However, Grinblatt and Titman (1989*b*) show that nonincreasing Rubinstein absolute risk aversion is a sufficient condition for the sum of the N terms in eq. (1) to be positive.

Using the first approach, the expression in (2) summed over all assets aggregates to the performance measure

$$\Sigma\,\Sigma\,(R_{jt} - \overline{R}_j)\,w_{jt}/T. \tag{3}$$

If the entire time series is used to estimate the means of the asset returns, the covariance estimate in (3) may provide erroneous inferences for certain portfolio strategies in small samples. For example, a contrarian strategy of picking stocks that have previously experienced a price decline induces a positive sample covariance between portfolio weights and returns since it picks stocks that tend to have downward-biased sample means. Thus, past returns cannot be used to estimate expected returns in (1a) if we desire an unbiased estimate of the covariance. For similar reasons future portfolio holdings should not be used to estimate the expected holdings in (1b) to compute the covariance.

To simplify our comparison of the two measures we will assume that the period $t + k$ return for each asset is used as a proxy for its period t expected return in the Event Study Measure and that its period $t - k$ portfolio holding is used as a proxy for its expected holding for our new measure. The Event Study Measure and our new measure, which we henceforth term, "the Portfolio Change Measure," can thus be expressed as follows:

$$\text{Event Study Measure} = \Sigma\,\Sigma\,[w_{jt}(R_{jt} - R_{j,t+k})]/T, \tag{4a}$$

and

$$\text{Portfolio Change Measure} = \Sigma\,\Sigma\,[R_{jt}(w_{jt} - w_{j,t-k})]/T. \tag{4b}$$

Under the null hypothesis of no superior information, both current and past weights are uncorrelated with current returns. In this case, the measures presented in (4a) and (4b) converge to zero in large samples. In addition, under the alternative hypothesis that an investor is informed, the measures in (4a) and (4b) converge to the average covariance under the assumption that the $w_{j,t-k}$'s are uncorrelated with the R_{jt}'s. However, if past weights and current returns are positively correlated, the covariance estimate is biased toward zero.

It can easily be seen that for a constant universe of risky assets the large sample values of the two measures are virtually identical. (In finite samples they differ at the k first and k last time-series entries.) However, the Portfolio Change Measure has a statistical advantage that derives from its being the average dollar return (i.e., end-of-period value per unit of investment) of a zero-cost portfolio. Under the null hypothesis that the portfolio manager is uninformed (as defined above), the dollar returns of this zero-cost portfolio are serially uncorrelated, which makes the computation of test statistics for the significance of the average dollar return a trivial exercise. In contrast, the Event Study Measure uses future returns as a performance benchmark, which induces serial correlation in the time series of return differences in (4a). Tests of statistical significance are thus more difficult with the Event Study Measure.

Another weakness of the Event Study Measure is its previously mentioned sensitivity to the future survival of assets currently in the evaluated portfolio. If a particular asset fails to exist shortly after it is included in an evaluated portfolio, the investor's holding of that asset cannot be used to assess the portfolio's performance. This creates a bias in large samples as well as small samples. The Portfolio Change Measure, which for each time period applies current and past portfolio weights to returns in the coming period, cannot have survivorship bias by construction.

D. Caveats

As pointed out earlier, the key assumption in this analysis is that the mean returns of assets are constant over the sample period. If this assumption is violated, the measures in equations (4a) and (4b) can be gamed by investors. Specifically, portfolios that include assets when their expected returns are higher than usual (perhaps because they are temporarily riskier) will realize positive "performance" with the measures. This problem is likely to be particularly severe for portfolios that specialize in takeover targets or bankrupt stocks as well as for managers who exploit serial correlations in stock returns.[3]

In addition, there is a potential small sample problem with our measure that can arise for portfolios with unconditional expected returns that trend upward over time. Consider, for example, a fund manager who continually increases the beta (and hence the expected return) of his portfolio over the observation period. The zero-cost portfolio used to evaluate this manager's performance will thus have an average beta (and thus an average risk premium) that is positive, implying a positive expected performance measure even if he lacks superior information. The spurious performance generated from this will be on the order of the average k-period change in beta (as measured with an efficient benchmark) times the risk premium of the benchmark. For reasonably large samples, this number is unlikely to be large.[4]

Most of the problems described above can be addressed by regressing the dollar returns of the zero-cost portfolio on the return of various market indexes. If the average systematic risk of the zero cost

3. The extent to which performance is exhibited by managers who are exploiting serial correlation in stock returns is examined in Grinblatt, Titman, and Wermers (1992).

4. For example, if a fund's risk premium jumps from 8% per year at the beginning of a 10-year sample period to 12% at the end (a highly unlikely 50% increase in systematic risk), its yearly performance in the 10-year sample period would be increased by .4%. This would be statistically insignificant both for an individual fund and for a portfolio composed of many funds doing the same thing. It is not going to affect any of the results in our article.

portfolio is close to zero, the bias in the Portfolio Change Measure is probably fairly small. If necessary, we can modify this measure by focusing only on the intercept in this time-series regression as a measure of performance. This modification in the performance technique, while subject to benchmark biases of the type discussed by Roll (1978), will nevertheless be much less sensitive to the benchmark portfolio selected for this modification than traditional performance measures.[5]

III. Performance Estimates

A. *Construction of the Performance Measure*

This study examines the quarterly holdings of 155 mutual funds over the December 31, 1974, to December 31, 1984, period with the Portfolio Change Measure developed in the previous section. This mutual fund data, which was examined previously in Grinblatt and Titman (1989*a*), was obtained from CDA Investment Technologies, Inc.[6]

From these quarterly holdings we calculated the returns of two zero-cost portfolios corresponding to each mutual fund. The quarterly weights of the first portfolio are the portfolio weights of the mutual fund at the beginning of the quarter minus the weights at the beginning of the previous quarter. The second portfolio consists of the portfolio weights of the funds at the beginning of the quarter minus the corresponding weights 1 year earlier.

The differenced weights (weights of the "current portfolio" less the weights of some "past portfolio") were then multiplied by CRSP monthly stock returns (compounded from the returns in the daily returns file) and summed to obtain the monthly returns of the zero-cost portfolios. Unless return data are missing, the differenced weights are identical for the three consecutive months following the quarterly Securities and Exchange Commission (SEC) filing (which reports holdings at the end of March, June, September, and December).[7] The

5. This is because the modified measure computes the difference in the returns of two portfolio weight vectors, $w_t - w_{t-k}$, after taking out the market factor component of their expected returns. Hence, the bias in the modified measure due to nonmarket factors is on the order of the risk premium of the factor times the difference in the factor loadings of the two portfolios. Traditional techniques have a hidden factor bias on the order of the factor risk premium times the factor loading of the portfolio represented by the weight vector w_t.

6. These 155 funds included all funds in the data base that were examined in Grinblatt and Titman (1989*a*) less two funds that did not have holdings on December 31, 1984. Although these funds were selected to have data over the entire holding period, Grinblatt and Titman (1989*a*) demonstrated that the fund survival requirement has a negligible effect on performance inferences for the overall sample or for the subgrouping of the sample employed here. We present tests that confirm this later in the article.

7. Only weights for CRSP-listed stocks with returns available in the relevant month were used. For example, if the return for stock XYZ was missing in the month of May but existed for April and June, the March 31 holding of XYZ was deleted for May but not for April and June.

weights are then updated for the subsequent 3 months and applied to the corresponding returns and so forth until a time series of 120 (111) monthly portfolio returns is created for the 1- (4)-quarter lag measure. Both series end in March 1985.

For example, with the 1-quarter lag, the April, May, and June 1975 returns (the first three observations) are multiplied by the difference between the portfolio weights held on March 31, 1975, and the weights held on December 31, 1974. The July, August, and September 1975 returns are multiplied by the weights on June 30, 1975, less the weights on March 31, 1975. For the 4-quarter lag, the January, February, and March 1976 returns (the first three observations) are multiplied by the portfolio weights on December 31, 1975, less the weights on December 31, 1974. The April, May, and June 1976 returns are multiplied by the weights on March 31, 1976, less the weights on March 31, 1975, and so forth.[8]

The average returns of the two time series created with this procedure provide the 1-quarter and 4-quarter versions of the Portfolio Change Measure of performance for a fund. Although neither of the time series represent the abnormal returns of the actual mutual funds, they do represent the abnormal returns that an investor could have achieved by buying the stocks listed in the fund's quarterly report and shorting the stocks from the fund's report in the previous quarter (for the 1-quarter lag measure) or the previous year (for the 4-quarter lag measure). Under the null hypothesis that the funds have no superior information and that, on average, the systematic risk of the current portfolio is the same as the systematic risk of the past portfolio (in the current period), both measures represent average returns of zero-cost, zero systematic risk portfolios. Both measures thus have expected values of zero.[9]

If a mutual fund's superior information is fully revealed to the market within 1 quarter, then the measure using 1-quarter lags will be the better of the two measures since it captures the entire excess return of the funds with the minimum amount of noise. However, if this superior information is not fully incorporated into prices until 1 year

8. The investment strategy tested with this technique is a strategy where holdings are rebalanced monthly within the quarter. If a buy and hold strategy is used over the quarter, rather than a rebalancing strategy, performance would tend to increase (decrease) if the monthly returns of stocks exhibited first- or second-order positive (negative) serial correlation in our sample period. An empirical investigation of this issue indicated that our conclusions were unaffected by whether a buy and hold or a rebalancing strategy was used within the quarter.

9. While these procedures seem to be adequate for determining if a fund is being actively managed with superior information, they cannot tell us if an investor who buys shares in the mutual funds is earning abnormal returns. The portfolio-weighted averages of the CRSP returns on which our measure is based exclude not only the fees and expenses charged by the fund but also the transaction costs, such as brokerage commissions and bid-ask spreads that are charged for every trade.

after the stocks are purchased, then using the previous quarter's portfolio weights to calculate the benchmark returns biases the performance measure downward. In this case, using the 4-quarter lagged weights to construct the zero-cost portfolio is preferred.

This insight may be easier to understand from the perspective of the Event Study Measure. The 4-quarter (1-quarter) version of the Portfolio Change Measure is similar to an Event Study Measure with the event return being the portfolio return for the quarter and the benchmark return being the quarterly return for 4 quarters (1 quarter) ahead. Now consider a fund manager who possesses private information about a stock in his fund. If this information is equally likely to be revealed in the first quarter or the second quarter after the stock is held, the expected first-quarter return will be identical to the expected second-quarter return and the expected difference between the two will be zero. In this case, the 1-quarter lag measure will have an expectation of zero. If the information has some probability of public revelation within the first quarter after the stock is held, and a lower probability of revelation in the quarter occurring 1 year after it is held, then the quarterly return 1 year after the stock is held will, on average, be lower than the quarterly return just after the stock is held. The expected 4-quarter version of the measure will then be positive for an informed manager.

B. *Results for Mutual Funds Grouped by Investment Objective*

Table 1 displays performance estimates for equally weighted portfolios of mutual funds grouped by investment objective. Both 1-quarter and 4-quarter lagged portfolio weights are used to compute the measures. Average performance for each grouping of funds, as well as t- and F-statistics, are reported. Recognizing that portfolios with dynamically changing weights may have returns that are not independently and identically distributed normal, we also report significance levels from the nonparametric Wilcoxon test.[10] The Wilcoxon test results, as well as those from an unreported binomial sign test, are consistent with the reported t-statistics.

The average performance for the entire sample with the 1-quarter version of the measure is close to zero, as is the average performance within the different investment objective categories. In addition, while some of the individual funds show significant abnormal performance with this measure, the incidence and magnitude of the abnormal perfor-

10. If securities returns are serially uncorrelated, the central limit theorem can be applied and asymptotic z-tests and chi-square tests are valid for non-normal portfolio returns. Given the length of our time series, these asymptotic test statistics are virtually identical to the t- and F-statistics used here and have negligibly different significance levels.

TABLE 1 Performance Estimates for 155 Surviving Mutual Funds Grouped by Investment Objective Categories (in % Return per Year)

		Performance Measure					
		Lagged 1 Quarter			Lagged 4 Quarters		
	No. of Funds	Mean Performance	t-statistic[a]	Wilcoxon Probability[b]	Mean Performance	t-statistic[a]	Wilcoxon Probability[b]
Total sample	155	.37	1.47	.233	2.04	3.16*	.004
Aggressive growth funds	45	.39	.98	.475	3.40	3.55*	.004
Balanced funds	10	−.48	−1.87	.057	.01	.03	.902
Growth funds	44	.66	2.01*	.017	2.41	2.94*	.009
Growth-income funds	37	.14	.61	.095	.83	1.75	.107
Income funds	13	.54	1.54	.475	1.33	2.64*	.002
Special purpose funds	3	−.10	−.16	.233	.21	.19	.711
Venture capital/special situation funds	3	1.26	1.07	.812	2.66	1.43	.035

$F1$-statistic (Abnormal performance in every category = 0)

$F = 3.1438$*

Prob $> F = .0028$[c]

$F2$-statistic (Abnormal performance across categories is equal)

$F = 3.6590$*

Prob $> F = .0014$[c]

[a] The mean over all months divided by the standard error of mean.
[b] The probability that the absolute value of the Wilcoxon-Mann-Whitney Rank z-statistic is greater than the absolute value of the observed z-statistic under the null.
[c] The probability of the F-statistic being greater than the outcome shown, under the null hypothesis (Type I error).
* Type I error < .05.

mance is not large. For this reason we devote the rest of our analysis to the measure using the 4-quarter version of the measure.

For the entire sample, average performance with the 4-quarter version of the measure is estimated to be about 2% per year, which is statistically significant. As in Grinblatt and Titman (1989*a*), the strongest evidence of positive abnormal performance is present in the aggressive growth funds, which average 3.4% per year. The superior performance of the aggressive growth funds is not surprising given that these funds have the greatest turnover and the greatest expenses. This group must also generate the best performance on their gross returns if the net returns offered to their investors are to be comparable to the returns offered by other funds. We are surprised, however, at the consistency of this performance. For an equally weighted portfolio of these funds, the time series of the zero-cost portfolio used to compute the measure is positive in 67 out of 111 months. Moreover, 41 out of 45 aggressive growth funds realize positive performance with the measure.

The growth funds and the income funds also show statistically significant performance, although at a lower level than the aggressive growth funds. The F-statistics reported in table 1 reject the hypothesis that the average performance for the different categories of funds are all equal. This indicates that there are statistically significant differences in average performance between categories.

The average performance measures for the various groups of funds are very similar to the Jensen Measures found in Grinblatt and Titman (1989*a*) using the eight-portfolio benchmark that controls for dividend policy, firm size, and past returns.[11] In contrast, these performance measures are very different from those generated by the factor analysis benchmark and the CRSP equally weighted and value-weighted indexes. This provides additional support for the Grinblatt and Titman (1989*a*) claim that the eight-portfolio benchmark is more reliable than traditional benchmarks.

A comparison between the standard errors of the Portfolio Change Measure and the standard errors reported in Grinblatt and Titman (1989*a*) suggests that the Portfolio Change Measure estimates smaller standard errors than approaches that use the securities market line. The reason for the increase in estimation precision is that the benchmark for the Portfolio Change Measure, the current returns of a fund's

11. The reason for this is found in Grinblatt and Titman (1989*b*), who show that, in large samples, covariance-based performance measures for informed investors who select stocks but cannot time the market are equivalent to Jensen Measure alphas that are derived with a mean-variance efficient benchmark. The eight-portfolio benchmark, which is designed to eliminate the dividend-yield and size anomalies, must therefore be close to the efficient frontier in order to provide results so similar to the covariance-based measure used here.

historical portfolio, is more highly correlated with the returns of the current portfolio than any benchmark portfolio.

C. Do Mutual Funds Benefit from the January Effect?

The Portfolio Change Measure of fund performance requires that expected stock returns be constant over time. However, research by Keim (1983), Reinganum (1983), and Roll (1983) indicates that, in January, small stocks outperform large stocks and past losers outperform past winners. Since our measure fails to account for this seasonal difference in expected returns, a mutual fund that took advantage of these seasonal patterns would exhibit positive performance with our measure.[12]

To analyze the effect of this January seasonal, we have broken the abnormal returns reported in table 1 into January and non-January returns. These performance estimates indicate that Januaries do not yield higher abnormal returns than the rest of the year. In fact, January performance is lower than performance during the remainder of the year for the entire sample of funds as well as for the three investment objective categories that contain a large number of funds. For example, the January performance for the entire sample averages .03%, and the January performance for the aggressive growth funds averages .11%. January performance numbers are statistically indistinguishable from zero in each of the categories as well as for the sample as a whole.

D. Evidence on Survivorship Bias

Even though our measure is designed to avoid biases due to surviving assets, the use of surviving funds could cause a bias. We do not believe that this bias is large because of Grinblatt and Titman's (1989*a*) finding that fund survivorship bias is small with this sample. To confirm this, table 2 employs the sample of funds that existed at the end of 1974. The number of funds diminishes in size with each successive year and is identical to the sample used in table 1 in the last quarter. Since the sample used in table 2 lacks the survivorship bias of the sample in table 1, and since the results in table 2 are remarkably similar to those in table 1, we do not believe that fund survivorship bias is affecting any of our conclusions.

12. This seasonal effect would be less severe with the 4-quarter measure. For example, the 1-quarter measure for the month of January for a fund manager who shifts into small-cap stocks at the end of December and who shifts out of them at the end of January would equal the average return of a zero-cost portfolio that is long in small stocks and short in large stocks. However, the corresponding zero-cost portfolio for the 4-quarter version of the measure would be both long and short small stocks. The 4-quarter measure does not, however, correct for the portion of the January effect related to past winners and losers.

TABLE 2 Performance Estimates for All 274 Mutual Funds Grouped by Investment Objective Categories (in % Return per Year)

		Performance Measure Lagged 4 Quarters		
	No. of Funds	Mean Performance	t-statistic[a]	Wilcoxon Probability[b]
Total sample	274	1.90	3.13*	.001
Aggressive growth funds	73	3.44	3.69*	.002
Balanced funds	19	.61	1.18	.902
Growth funds	81	2.06	2.82*	.009
Growth-income funds	57	.84	1.89	.035
Income funds	31	.93	2.02*	.0003
Special purpose funds	6	−.04	−.04	.536
Venture capital/special situation funds	7	1.74	.96	.536

$F1$-statistic (Abnormal performance in every category = 0)

$F = 2.6461$*

Prob $> F = .0104$[c]

$F2$-statistic (Abnormal performance across categories is equal)

$F = 3.0505$*

Prob $> F = .0059$[c]

[a] The mean over all months divided by the standard error of mean.
[b] The probability that the absolute value of the Wilcoxon-Mann-Whitney Rank z-statistic is greater than the absolute value of the observed z-statistic under the null.
[c] The probability of the F-statistic being greater than the outcome shown, under the null hypothesis (Type I error).
* Type I error < .05.

E. *Evidence on Changing Risk*

The Portfolio Change Measure may spuriously indicate positive performance for funds that select stocks in time periods in which their systematic risk is temporarily high. To examine the extent to which this bias exists, we estimated the betas (against the CRSP value-weighted index) and corresponding Jensen Measures of the zero-cost portfolios that were used to compute the 4-quarter lag version of the measure. Table 3 indicates that, for each investment objective category, the average beta of these zero-cost portfolios is positive. The mutual funds thus have a tendency to hold stocks during time periods in which they are riskier than average. While these betas are statistically significant (see the F-tests at the bottom of the table), they are not particularly large and do not have much of an effect on our inferences. Indeed, the Jensen Measures of the zero-cost portfolios (which represent a beta-adjusted version of the Portfolio Change Measure) are almost identical to the performance measure itself (the unadjusted average of the time series of portfolio returns). The cross-sectional correlation between the two measures is .993.

TABLE 3 **Beta-adjusted Measures and Betas (Equally Weighted Portfolios for Each Category of Funds)**

	Aggressive Growth	Balanced	Growth	Growth-Income	Income	Special Purpose	Venture Capital/ Special Situation
No. of funds (total = 155)	45	10	44	37	13	3	3
Beta-adjusted measure[a]:	3.1020	−.0342	2.1900	.7380	1.2984	−.0008	2.2764
t-statistic	(3.321)*	(−.079)	(2.716)*	(1.568)	(2.552)*	(−.074)	(1.232)
Prob > |t|[b]	(.001)	(.938)	(.008)	(.120)	(.012)	(.941)	(.221)
Beta:	.0519	.0082	.0372	.0153	.0061	.0508	.0660
t-statistic	(2.880)*	(.976)	(2.391)*	(1.680)	(.618)	(2.400)*	(1.858)
Prob > |t|[b]	(.005)	(.331)	(.019)	(.096)	(.538)	(.018)	(.066)

$F1$-statistic (Abnormal performance in every category = 0)

$F = 2.7927$

$\text{Prob} > F = .0071^{c*}$

$F2$-statistic (Abnormal performance across categories is equal)

$F = 3.2437$

$\text{Prob} > F = .0037^{c*}$

$F3$-statistic (Betas of zero-cost portfolios for every category = 0)

$F = 2.2140$

$\text{Prob} > F = .0313^{c*}$

[a] In % return per year.
[b] The probability of the absolute value of the *t*-statistic being greater than the absolute value of the outcome shown, under the null hypothesis (Type I error).
[c] The probability of the *F*-statistic being greater than the value shown, under the null hypothesis.
* Type I error < .05.

IV. Tests for Differential Performance within Categories

Given that the average performance of funds with aggressive growth, growth, and income objectives is significantly positive, it is natural to ask whether some funds in these categories are better able to generate performance than others. This section presents two types of tests to examine this issue. The first type calculates *F*-statistics that test the equality of the performance measures of the individual funds within each investment objective category. The second type examines whether funds that perform well in the first half of the sample period also perform well in the last half of the sample period.

A. F-*tests of Differential Performance within Categories*

Following Grinblatt and Titman (1989*a*), table 4 calculates *F*-statistics for each investment objective category to test the joint hypotheses that the abnormal returns of all individual funds within a particular category are equal to zero (*F*1) or are equal to each other (*F*2). The *F*-statistics are calculated for (i) the Portfolio Change Measure, which implicitly assumes that the betas of the zero-cost portfolios used to create our measure are equal to zero (the "restricted beta" case), and (ii) the Jensen Measures of these same portfolios (the beta-adjusted Portfolio Change Measure), which use the value-weighted CRSP benchmark.

The test results indicate that there is positive as well as differential performance in the aggressive growth, growth, growth-income, and venture capital/special situation categories, but not in the balanced, income, or special purpose categories. In comparison, Grinblatt and Titman (1989*a*) found differential performance with their eight-portfolio benchmark only within the aggressive growth and growth categories.

The results in table 4 indicate that the tests are not sensitive to whether the value-weighted portfolio is used in calculating performance. Table 4 also provides *F*-statistics to test the hypothesis that the betas of the zero-cost portfolios corresponding to individual funds in a given category are jointly zero. We reject this hypothesis for five categories: the aggressive growth, growth, growth-income, income, and special purpose fund categories. It should be noted that, although the betas for the zero-cost portfolios associated with some of the funds are reliably different from zero, even the highest betas are still quite small and have little effect on the performance measures. For example, in the aggressive growth category, which has the widest variation in performance and beta, all betas fall between −.08 and .32, with only four aggressive growth funds out of 45 having a beta above .15. This suggests that differences in the betas of the zero-cost portfolios used to measure performance do not account for differences in performance between funds.

TABLE 4 **Mutual Fund Performance *F*-tests**

	Aggressive Growth	Balanced	Growth	Growth-Income	Income	Special Purpose	Venture Capital/ Special Situation
No. of funds (total = 155)	45	10	44	37	13	3	3
*F*1-statistic (all funds in the category have zero abnormal performance):							
Unadjusted measure	2.5939*	.8910	2.0001*	1.9768*	1.3856	.6977	4.5705*
	(.0001)	(.5410)	(.0001)	(.0004)	(.1587)	(.5540)	(.0038)
Beta-adjusted measure	2.7019*	.8576	1.9635*	2.0597*	1.3496	.5176	4.4188*
	(.0001)	(.5730)	(.0002)	(.0002)	(.1772)	(.6704)	(.0046)
*F*2-statistic (all funds in the category have equal abnormal performance):							
Unadjusted measure	1.7744*	.9827	1.9600*	1.9920*	1.4962	.4989	6.8488*
	(.0013)	(.4525)	(.0002)	(.0004)	(.1186)	(.6076)	(.0012)
Beta-adjusted measure	1.9194*	.9510	1.9383*	2.0922*	1.4550	.4038	6.6229*
	(.0003)	(.4796)	(.0002)	(.0002)	(.1346)	(.6681)	(.0015)
*F*3-statistic (zero-cost portfolios of funds in the category have beta = 0)	3.2632*	1.1581	2.2612*	2.2757*	2.2656*	4.1608*	1.3477
	(.0001)	(.3154)	(.0001)	(.0001)	(.0060)	(.0065)	(.2588)

NOTE.—The probability of the *F*-statistic being greater than the outcome shown, under the null hypothesis (Type I error), is given in parentheses.
* Type I error $< .05$.

B. Persistence of Performance

To test for persistence, we first rank the performance of the funds over the first half of our sample period. The funds are then formed into portfolios based on these rankings, and the performance of these portfolios is measured over the second half of the sample period. If some fund managers possess superior investment talent, then funds that performed well in the past would be expected to perform well in the future. In addition, since the returns we examine are before expenses, fees, and commissions, we would not expect to find persistently poor performers, as found, for example, in Hendricks, Patel, and Zeckhauser (1991) and Grinblatt and Titman (1992).

Table 5 presents persistence tests for the entire sample of funds. The results suggest that there is a strong association between performance in the first 56 months of the sample and performance in the last 55 months. The second subperiod performance of the funds ranked in the top decile in the first subperiod is about 3.5% per year, which is statistically significant.[13]

Some of the persistence found in table 5 could be due to differences in performance among funds with different investment objectives. If, for example, the aggressive growth funds had superior performance in both the first and second half of the sample period, then the preceding test would find evidence of persistence in the entire sample. For this reason, it is of interest to test for persistence of performance within the different categories of investment objectives.

Table 6 presents persistence results for the funds within the aggressive growth, growth, and growth-income categories. The remaining four categories do not contain enough funds to make persistence tests meaningful. For each category, a portfolio is formed which consists of a long position in the top $x\%$ of first-half performers and a short position in the bottom $x\%$, where x is 10, 20, or 50. Given the smaller sample sizes for the individual categories, these tests have larger standard errors than the persistence tests for the entire sample. However, the results do suggest that there is weak evidence of persistence within these categories. In each of the categories, the funds that do relatively well in the first half of the sample period do better in the second half of the sample period than the funds that do poorly in the first half.

This issue is clearly illustrated by examining the only significant category for persistence: the growth-income category. The two persistence portfolios that are long in the 20% and 50% best growth-income funds and respectively short the 20% and 50% worst growth-income funds display significant performance. However, an (unreported) F-test indicates that this significant performance is not significantly dif-

13. We performed similar persistence tests on our 274 fund sample that does not require survival and found very similar results.

TABLE 5 Persistence Test on Mutual Funds (Lagged 4 Quarters)

	First 56 Months		Second 55 Months	
	Average Fund Size	Mean Performance	Mean Performance	t-statistic
Equal-weight portfolio—top 10%	3,495,343	9.0835	3.5182	2.12
Equal-weight portfolio—top 20%	6,810,028	7.2721	2.8169	1.79
Equal-weight portfolio—top 50%	9,437,952	4.8813	2.7774	2.04
Equal-weight portfolio—bottom 10%	8,716,274	−1.5632	.6018	.80
Equal-weight portfolio—bottom 20%	9,445,020	−.9945	.2081	.32
Equal-weight portfolio—bottom 50%	14,256,427	.0783	.3449	.50
Equal-weight portfolio—all 155	11,831,646	2.4953	1.5690	1.62
Long top 10%—short total portfolio	. . .	6.5882	1.9492	1.79
Long top 10%—short bottom 10%	. . .	10.6467	2.9164	1.75
Long top 50%—short bottom 50%	. . .	4.8029	2.4325	2.53
Long top 20%—short bottom 20%	. . .	8.2667	2.6088	1.78

TABLE 6 Persistence Test on Mutual Funds by Investment Objective (Lagged 4 Quarters)

	First 56 Months		Second 55 Months	
	Average Fund Size	Mean Performance	Mean Performance	*t*-statistic
Aggressive-growth funds only:				
Equal-weight portfolio—top 10%	1,432,355	10.6248	2.3437	1.10
Equal-weight portfolio—top 20%	2,453,670	8.8592	3.1646	1.74
Equal-weight portfolio—top 50%	6,134,124	6.5350	2.3653	1.37
Equal-weight portfolio—bottom 50%	6,787,993	2.2336	2.3410	1.80
Equal-weight portfolio—bottom 20%	5,823,872	.3688	.3544	.17
Equal-weight portfolio—bottom 10%	2,964,488	−.5818	−.9899	−.50
Equal-weight portfolio—all 45	6,453,793	4.4321	2.3534	1.65
Long top 10%—short bottom 10%	...	11.2066	3.3337	1.24
Long top 20%—short bottom 20% (beta measure = −.034)	...	8.4904	2.8102	1.61
Long top 50%—short bottom 50%	...	4.3015	.0243	.02
Growth funds only:				
Equal-weight portfolio—top 10%	3,432,059	9.2992	2.0883	1.51
Equal-weight portfolio—top 20%	12,284,392	6.9628	3.1307	1.86
Equal-weight portfolio—top 50%	12,942,966	4.8921	3.3548	2.12
Equal-weight portfolio—bottom 50%	17,252,403	.2924	1.0775	.97
Equal-weight portfolio—bottom 20%	10,264,107	−.6262	1.6244	1.23
Equal-weight portfolio—bottom 10%	2,319,086	−1.1889	2.7474	1.54
Equal-weight portfolio—all 44	15,097,684	2.5923	2.2162	1.79
Long top 10%—short bottom 10%	...	10.4881	−.6591	−.29
Long top 20%—short bottom 20% (beta measure = .032)	...	7.5890	1.5063	.83
Long top 50%—short bottom 50%	...	4.5997	2.2773	1.96
Growth-income funds only:				
Equal-weight portfolio—top 10%	11,829,988	4.7714	3.3032	1.76
Equal-weight portfolio—top 20%	13,642,745	3.8454	3.3116	2.11
Equal-weight portfolio—top 50%	14,178,291	2.2581	1.0549	1.32
Equal-weight portfolio—bottom 50%	18,844,922	−.2612	−.2513	−.43
Equal-weight portfolio—bottom 20%	16,551,708	−1.1829	.1954	.31
Equal-weight portfolio—bottom 10%	20,650,917	−1.5433	.1651	.18
Equal-weight portfolio—all 37	16,448,544	1.0325	.6167	.81
Long top 10%—short bottom 10%	...	6.3147	3.1380	1.76
Long top 20%—short bottom 20% (beta measure = .028)	...	5.0283	3.1161	2.28
Long top 50%—short bottom 50%	...	2.5193	1.3062	2.19

ferent from the comparable portfolios in the other two categories. Moreover, the magnitude of performance with the 50% growth-income persistence portfolio is lower than four of the seven remaining insignificant persistence portfolios. Obviously, higher volatility, due to both the number of funds and the composition of the funds' assets, is largely responsible for the failure of the non-growth-income persistence portfolios to achieve significance.

Irrespective of this, the performance numbers in table 6 tells us that the persistence observed in table 5 is not strictly an aggressive growth phenomenon. There is a fair degree of homogeneity observed in the performance numbers across the three investment objective categories. For example, the three top 20% past performers in the three investment objective categories all have abnormal returns outside of the ranking period of about 3% per year. This suggests that past performance per se is a valid indicator of future performance and is not a variable that is confounded with investment objective.

Table 6 also presents estimated betas from the second half of the sample period for the three 20% portfolios to test whether there are systematic differences between the betas of the top quintile first-half performers and those of the bottom quintile first-half performers. The betas of the three zero-cost portfolios should be zero if there are no differences in the betas of the average mutual fund in each of the two quintiles. Table 6 indicates that the three beta differences are quite small, and an F-test of the significance of these three beta differences does not reject, at the 5% level, that they are jointly zero. Therefore, the persistent performance apparent in these three portfolios of funds is not due to differences in systematic risk between mutual funds.

V. Conclusion

The evidence in this article indicates that the CRSP-listed quarterly holdings of mutual fund portfolios, on average, achieved positive abnormal investment performance from January 1, 1976, to March 31, 1985. Consistent with Grinblatt and Titman (1989*a*), the strongest evidence of abnormal performance was found in the aggressive growth category of funds. However, in contrast to this earlier work, the measure of performance used in this study required the observation of portfolio holdings, and under no circumstances could the results be attributed to an inefficient benchmark portfolio. No benchmark was required!

Not all mutual fund managers achieved superior performance. Those who achieved superior performance, on average, exhibited this ability persistently. That is, a fund that did well in the first half of the sample period continued to do well in the second half, suggesting that, to some extent, superior performance is predictable. Funds that exhibited negative performance in the first half of our sample period continued to lag behind the rest in the second half by exhibiting second-half performance that was indistinguishable from zero.

It is important to emphasize that the abnormal performance documented here does not imply that individual investors can achieve abnormal returns by investing in mutual funds. Indeed, Grinblatt and Titman (1989*a*) indicated that, on average, transaction costs and fund expenses dissipate the abnormal investment performance so that the net performance of these funds is close to zero on average. However, as Grinblatt and Titman (1989*a*) suggested, there is an efficient markets anomaly here since the investments of these fund managers are public within weeks of their quarterly SEC filing. Investors who mimicked the funds could have achieved the returns described in this article, less commissions and bid-ask spreads (without paying the fees and expenses charged to the holders of open-end mutual funds). Given the 2%–3.5% gross abnormal returns of the funds, it is plausible that the net abnormal returns to mimicking investors would still have been positive.

Grinblatt and Titman (1989*a*) asserted that this efficient markets anomaly was not due to the type of benchmark bias that Roll (1978) suggested was possible. We can be more certain of this assertion now, given that the same anomaly has been documented here with a measure that was not based on the securities market line and thus did not require a benchmark portfolio. However, there is now a new and more puzzling twist to this anomaly that becomes apparent when we try to reconcile the insignificance of the 1-quarter lag measure and the significance of the 4-quarter lag measure. This difference could have occurred only if the assets purchased (sold) by fund managers had higher (lower) than normal returns long after the quarterly holdings were reported.

In comparing the performance evaluation measure developed in this article with more traditional measures developed from asset-pricing theories, we can identify only three areas of concern that might inhibit someone from adopting our approach. First, the new technique is more costly to implement, both in terms of data collected and computer time. With the increased availability of data and computing power, we do not see this as a major problem. Second, because data on holdings are only available externally on a quarterly basis, outside evaluators are not measuring the true performance of the fund but only the performance of some hypothetical portfolio that is correlated with the fund. This is a nonissue for internal evaluations when this technique can be used with transaction data. For outside evaluators, the use of quarterly holdings as opposed to holdings that correspond to each transaction adds noise to true performance and thus biases the measure toward finding no performance. However, the finding of significant performance within our sample suggests that the frequency of the holdings data need not be a major concern for outside evaluators who are attempting to identify informed managers. A third potential concern, that managers might game the measure by selecting securities when they are riskier than usual, can be addressed by combining traditional evaluation techniques with the evaluation procedure developed here, as was done in table 3. Combining the new approach with the traditional approach results in a performance measure that is still much less sensitive to the benchmark issues raised by Roll (1978) than traditional evaluation techniques alone. The combined approach is also no more subject to gaming, based on publicly known changes in the risk premia of securities, than the more traditional evaluation techniques.

References

Copeland, Thomas, and Mayers, David. 1982. The value line enigma (1965–1978): A case study of performance evaluation issues. *Journal of Financial Economics* 10, no. 3 (November): 289–321.

Cornell, Bradford. 1979. Asymmetric information and portfolio performance measurement. *Journal of Financial Economics* 7 (December): 381–90.

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

Grinblatt, Mark, and Titman, Sheridan. 1989*b*. Portfolio performance evaluation: Old issues and new insights. *Review of Financial Studies* 2, no. 3:393–421.

Grinblatt, Mark, and Titman, Sheridan. 1992, in press. The persistence of mutual fund performance. *Journal of Finance*, vol. 47 (December).

Grinblatt, Mark; Titman, Sheridan; and Wermers, Russell. 1992. Momentum investment strategies and portfolio performance: A study of mutual fund behavior. Working paper. Los Angeles: University of California.

Hendricks, Darryl; Patel, Jayendu; and Zeckhauser, Richard. 1991. Hot hands in mutual funds: The persistence of performance, 1974–88. Working paper. Boston: Harvard University Business School.

Jensen, Michael. 1972. Optimal utilization of market forecasts and the evaluation of investment portfolio performance. In G. Szego and K. Shell (eds.), *Mathematical Methods in Investment and Finance*. Amsterdam: North-Holland.

Keim, Donald. 1983. Size-related anomalies and stock return seasonality. *Journal of Financial Economics* 12 (June): 13–32.

Lehmann, Bruce, and Modest, David. 1987. Mutual fund performance èvaluation: A comparison of benchmarks and benchmark comparisons. *Journal of Finance* 42 (June): 233–65.

Reinganum, Marc. 1983. The anomalous stock market behavior of small firms in January. *Journal of Financial Economics* 12 (June): 89–104.

Roll, Richard. 1978. Ambiguity when performance is measured by the securities market line. *Journal of Finance* 33 (September): 1051–69.

Roll, Richard. 1983. Vas ist das? The turn-of-the-year effect and the return premia of small firms. *Journal of Portfolio Management* 9 (Winter): 18–28.