Title
Do Benchmarks Matter? Do Measures Matter? A Study of Monthly Mutual Fund Returns

Authors
Grinblatt, Mark
Titman, Sheridan

Publication Date
1991-03-01

DO BENCHMARKS MATTER? DO MEASURES MATTER? A STUDY OF MONTHLY MUTUAL FUND RETURNS

by

Mark Grinblatt*

and

Sheridan Titman*

March 1991

*UCLA Graduate School of Management. We would like to thank Julian Franks, Bruce Lehmann, David Mayers, Rena Repetti, Jay Shanken, and seminar participants at UCLA, Berkeley, the University of British Columbia, the University of Washington, Duke, Rutgers, and Wharton for valuable comments on an earlier version of this work. We also appreciate the contributions of Jim Brandon, Pierre Hillion, Haeyon Kim, Erik Sirri, Mark Tsesarsky and Khai Kan, who provided excellent research assistance, and of Bruce Lehmann and David Modest, who supplied monthly factor returns.

ABSTRACT

Do Benchmarks Matter? Do Measures Matter?
A Study of Monthly Mutual Fund Returns

This paper empirically examines the extent to which some of the different performance measures developed in the literature provide different evaluations of performance. It also examines how robust these performance scores are to the choice of benchmark portfolio. An analysis of 109 passive portfolios and 279 mutual funds finds that the measures generally yield similar inferences when using the same benchmark and that inferences can vary, even from the same measure, when using different benchmarks.

Introduction

The development of the Capital Asset Pricing Model (CAPM) in the mid-1960's provided financial economists with a tool for adjusting returns for risk. An important application of this model, implemented by Jensen (1968, 1969),[1] is the evaluation of the performance of managed portfolios. However, this approach to evaluating portfolio performance has been the subject of a great deal of controversy.

There are two major reasons for this controversy. The first is that the CAPM approach (and analogous multi-factor approaches) to performance evaluation requires the use of a benchmark portfolio(s). As Roll (1978) and others have noted, performance evaluation with these methods is likely to be sensitive to the benchmark choice and will provide misleading inferences if the benchmark portfolio is not mean-variance efficient. Second, even if a mean-variance efficient benchmark is used, theoretical work by Jensen (1972), Admati and Ross (1985), Dybvig and Ross (1985), and Grinblatt and Titman (1989b) suggests that these methods may penalize investors with the ability to time the market. In response to the latter criticism, Grinblatt and Titman (1989b) proposed a new measure, which they called the Positive Period Weighting Measure, that is not subject to the timing-induced biases of the Jensen's Measure. An alternative to this is a measure developed here, which uses the Treynor and Mazuy (1966) quadratic regression to aggregate the effects of timing and selectivity ability. We call this the "Treynor-Mazuy Total Performance Measure."

If stock returns are normally distributed and if portfolios exhibit no

[1]An equivalent approach was developed by Treynor (1965). The issues discussed in this paper that apply to Jensen's Measure also apply to Treynor's Measure.

timing ability, the three performance measures should all generate the same inferences (see Grinblatt and Titman (1989b)). However, funds that either time the market or pick portfolios with returns that are co-skewed with the benchmark returns will exhibit different performance scores with the different measures.

This paper empirically examines (i) the extent to which these different measures provide different inferences when they use the same benchmarks and (ii) the sensitivity of the different performance measures to the choice of the benchmark. We do this by studying the performance of a sample of 109 passive portfolios constructed from securities characteristics and industry groups, as well as a sample of 279 mutual funds. Differences in the various performance measures on the passive portfolios would confirm that performance was potentially sensitive to the technique used, since the passive portfolios probably span most of the strategies that professional fund managers follow. Differences in the performance of the mutual funds would indicate that the actual set of strategies followed by investment managers was sensitive to the technique used.

For the mutual funds studied here, the Jensen and Positive Period Weighting Measures provide similar performance scores. However, differences between the Jensen Measure and the Positive Period Weighting Measure are significantly related to the coefficient on the quadratic (timing) term in the Treynor-Mazuy regression. This suggests that while few funds successfully time the market, when timing does exist, it is likely to bias the Jensen Measure. While the Treynor-Mazuy Measure is highly correlated with the other two measures when the same benchmark is used there is also some evidence that it may be more sensitive to non-normality in the return generating process than the other measures.

The benchmarks we examine include the CRSP equally-weighted index, a benchmark consisting of 10 factor portfolios constructed by Lehmann and Modest (1987), and two benchmarks developed by Grinblatt and Titman (1988) that are based on securities characteristics--an eight-portfolio benchmark (P8) and a six-portfolio benchmark (P6). In contrast to the Lehmann and Modest conclusions, we find that the 10-factor benchmark yields inferences that are very similar to those generated with the equally-weighted index. However, the P8 and P6 benchmarks, which do not exhibit the size, dividend yield, and beta biases of the other two benchmarks, generate inferences that differ substantially from those generated with the other benchmarks. This is true for the sample of mutual funds as well as for the sample of passive portfolios.

I. Measures of Performance

This section describes the measures that are used to evaluate the 109 passive portfolios and 279 mutual funds. Each measure calculates performance relative to a benchmark, which is a portfolio or a group of portfolios.

The Jensen Measure is the intercept in a regression of the time series of excess returns (above the one month Treasury Bill rate) of the evaluated portfolio against the time series of excess returns of the benchmark portfolio(s). This is the traditional measure used in most previous studies of fund performance.

The Positive Period Weighting Measure, developed in Grinblatt and Titman (1989b), is obtained in two steps. First, one selects a set of nonnegative weights that make the weighted sum of the excess returns of the benchmark portfolio(s) sum to zero. These weights are then applied to the excess returns of the mutual fund to compute a performance score. Obviously, there are many sets of weights with the properties mentioned above. The weights employed in this study can be interpreted as the marginal utilities of an uninformed investor with power utility. Given this interpretation, uninformed investors with power utility increase their expected utility by adding a small amount of any mutual fund return with a measure that is positive to their unconditionally optimal portfolio (the unconditional mean-variance efficient portfolio). Grinblatt and Titman (1989b) show that under fairly standard conditions, positive values for these measures imply that the mutual fund manager has superior information. This measure is discussed in more detail in Appendix A.

The Treynor-Mazuy (1966) quadratic regression is similar to the Jensen Measure regression. Here, however, there are two explanatory variables: the excess return of the benchmark portfolio and the square of that excess return. The intercept in this regression provides an estimate of selectivity ability; the product of the quadratic term slope coefficient and the variance of the benchmark return provides an estimate of timing ability. We call the sum of the timing and selectivity terms the Treynor-Mazuy measure of total performance. This measure is discussed in more detail in Appendix B.

II. Benchmark Portfolios

Our first benchmark is the monthly rebalanced equally-weighted index computed from all CRSP securities. The second benchmark is a factor portfolio benchmark, created from factor portfolio weights provided by Bruce Lehmann and David Modest. These were derived from a 10-factor maximum likelihood factor analysis over the 1978-1982 period.[2] The portfolios contain 750 securities in the 1978-1982 period and slightly fewer in the 1975-1977 and 1983-1984 periods since some of the securities from the middle period did not exist in the early and later periods. Although this method of forming factor portfolios can potentially create survivorship bias, (unreported) comparisons with the equally-weighted index suggest that this bias is not large.

Past research indicates that neither of these two traditional benchmarks are mean-variance efficient. In particular, they generate biased performance scores that relate to size (Banz (1981), Reinganum (1981)), dividend yield (Litzenberger and Ramaswamy (1979, 1982)), past returns (DeBondt and Thaler (1985), Jegadeesh (1990)), and beta (Black, Jensen and Scholes (1972)). In the 1975-1984 sample period, Grinblatt and Titman (1988) also found size, dividend yield, past returns, and beta-related biases with the two benchmarks.

The third benchmark, the P8 benchmark developed in Grinblatt and Titman (1988) and used in Grinblatt and Titman (1989a), is not subject to any of these biases. The formation of this benchmark is premised on the idea that various firm characteristics are correlated with their stocks' factor loadings. As a result, characteristic-based portfolios can be used as proxies for the factors. The P8 benchmark consists of four size-based portfolios, three dividend-yield-based portfolios, and the lowest past returns portfolio: The smallest 8 1/3% of firms comprised the first size-based portfolio; the average of the second and third smallest size portfolios (out of 12) comprised the second portfolio; the average of the fourth through ninth smallest size portfolios comprised the third portfolio; and the average of the three largest

---

[2]See Lehmann and Modest (1985b) for a complete description of these factor portfolios.

size portfolios comprised the fourth. An equal weighting of the two lowest dividend-yield portfolios (out of 12), the fifth and sixth lowest dividend-yield portfolios, and the tenth and eleventh dividend-yield portfolios are the fourth through sevenths portfolios in the benchmark. The lowest past returns portfolio (out of 12) is the eighth portfolio in the benchmark.

To test the sensitivity of performance inferences to minor variations in the procedure for forming benchmarks based on securities characteristics, we also employed the P6 benchmark from Grinblatt and Titman (1988). This benchmark consists of three size-based portfolios, two dividend yield portfolios, and a portfolio based on past returns. An equal weighting of the smallest and second smallest size portfolios (out of 12) is the first portfolio in the benchmark. An equal weighting of the sixth and seventh smallest size portfolios comprised the second portfolio; an equal weighting of the two largest size portfolios is the third. The fourth and fifth portfolios in the benchmark are comprised of an equal weighting of the two lowest dividend-yield portfolios (out of 12) and the tenth and eleventh lowest dividend yield portfolios (out of 12) respectively. The lowest past returns portfolio (out of 12) is the sixth portfolio in the benchmark.

III. The Data

Stock returns were obtained from the CRSP Daily Returns File. The daily returns were compounded to calculate the monthly portfolio returns used to form and test the benchmark portfolios. Mutual fund data were obtained from CDA Investment Technologies, Inc. of Silver Springs, Maryland. The data consists of monthly cash-distribution-adjusted returns and investment goals for 279 funds that existed from December 31, 1974 to December 31, 1984. The data was spot checked with data collected by hand and found to be accurate.

As with most mutual fund studies, the mutual fund return data is subject to survivorship bias. Since CDA's non-academic clients are not interested in mutual funds that no longer exist, funds that went out of business prior to December 31, 1984 were excluded from the CDA data set. Grinblatt and Titman (1989a) estimated the survivorship bias in this sample and it does not appear to be large, on the order of .5% per year.

In addition to the sample of mutual funds, we also evaluate the performance of 109 passive strategies. Since these strategies do not use private information, in theory, they should generate zero performance with properly designed measures and benchmarks. The 109 portfolios include 72 portfolios that were formed on the basis of six characteristics that are related to CAPM and APT "anomalies" and 37 industry portfolios. The 72 characteristic portfolios were formed by ranking the stocks on the basis of the different characteristics and then dividing them into 12 equally-weighted portfolios based on their rankings. For a given characteristic, portfolio 1 represents the portfolio formed from firms with the lowest rankings of that characteristic. (For example, portfolio 1 of the size portfolios consists of firms with sizes among the lowest 8 1/3%.) Specifically, the six characteristics are:

i) firm size, determined by the most recent capitalization available on the CRSP daily master file prior to the month of the observed return;

ii) dividend yield, calculated from the CRSP master file using the calendar year prior to the observed return;

iii) past returns, computed from the CRSP daily returns file using the three calendar years prior to the observed return;

iv) interest rate sensitivity, as measured by the slope coefficient on an equally-weighted index of 16-21 year government bonds in an excess return regression using this bond portfolio and the equally-weighted portfolio of all CRSP securities. The time series uses the three calendar years prior to the observed return;

v) co-skewness, as measured by the slope coefficient on the "squared term" in a regression using the excess return and squared excess return of the equally-weighted portfolio. The time series uses the three calendar years prior to the observed return; and

vi) beta, as computed against the equally-weighted portfolio in a monthly excess return regression using the three calendar years prior to the observed return.

We also looked at industry portfolios. Firms were divided into portfolios based on their two-digit SIC codes at the beginning of the sample period. All "two-digit" industries with at least 20 firms were included in excess return regressions. This yielded 37 passive industry portfolios, which combined with the 72 portfolios described above gives us 109 portfolios.

IV. The Measured Performance of Passive Investment Strategies

Tables 1-3 analyze the sensitivity of performance to benchmarks and measures using the sample of 109 passive portfolios. Table 1 presents the means and standard deviations of the performance measures with the four benchmarks. Table 2 presents correlation matrices that examine the extent to which the different benchmarks generate different performance scores. Table 3 presents correlation matrices that determine the extent to which the different measures generate different performance scores.

Table 1 indicates that the means of the different measures are almost all close to zero and their standard deviations are generally very similar to each other. The similarity of the means and standard deviations across measures and benchmarks indicates that if the various measures and benchmarks generate performance scores that are highly correlated across portfolios, the magnitude of the individual scores must be very similar.

The only case for which average performance is not close to zero is when performance is measured with the Treynor-Mazuy quadratic regression with the

P8 benchmark. With the P8 benchmark, almost all of the passive portfolios have a negative coefficient on the quadratic term in the Treynor-Mazuy quadratic regression, causing the total performance measure to be generally negative. This anomaly is due to January 1975, which was an exceptional month for most stocks, but much more so for small stocks. Our procedure for computing the Treynor-Mazuy coefficients with a multiple portfolio benchmark uses the ex-post efficient combination of the portfolios that comprise the benchmarks as the regressor. Given the outstanding performance of small firms in our sample period, this combination is heavily weighted towards the smallest size portfolio and as a result of this greatly outperforms most other portfolios in January 1975. The extremely high January 1975 return for the combined benchmark implies that in a simple regression, the passive portfolios would achieve strongly negative residuals for this month. However, with a quadratic regression, we have an extra degree of freedom that allows it to better fit this observation. In order to do this, the Treynor-Mazuy coefficients on the quadratic term are consistently negative across the passive funds and the mutual funds. The ex-post efficient combination of the P6 benchmark, which contains a size portfolio of the smallest 16 2/3 % of firms rather than one with the smallest 8 1/3 % of firms, is not as heavily weighted towards its small size portfolio as the P8 benchmark. It does not outperform the passive portfolios and the mutual funds to the same degree as the combination used for the P8 benchmark in January 1975 and thus has a Treynor-Mazuy regression that is much less affected by this outlier.

Table 2 reports correlations that illustrate the sensitivity of the different performance measures to the choice of the benchmark.[3] Of particular note are the correlations between the performance scores of characteristic-based benchmarks and the two other benchmarks, which are not very large. While the EW and 10-factor benchmarks also generate different inferences, the correlations are much higher. For example, with the Jensen Measure, the correlation between EW and 10-factor performance is .64, which is about twice as large as the correlation between P8 (P6) performance and performance with the EW and 10-factor benchmarks.[4]

Table 3 indicates that the Jensen and Positive Period Weighting Measures provide very similar inferences. The lowest correlation between a cross-section of Positive Period Weighting Measures and a corresponding set of Jensen Measures is .95. These high correlations also imply that the Positive Period Weighting Measure is robust to our specification of the period weights.[5] The Treynor-Mazuy Measure, however, generates performance scores that differ somewhat from the performance generated with the other measures. This is probably due to the fact that some of the passive portfolios are designed to have high co-skewness with the equally-weighted index, and

---

[3]Significance levels for the correlations are not reported. Because the vectors used to compute the correlations contain correlated random elements, the standard significance tests are highly biased. This omission does not affect the conclusion we reach because the numbers are dramatic enough to speak for themselves.

[4]We also constructed 3 alternative characteristic-based factor portfolios, varying both the number of portfolios in the benchmark and characteristics. These benchmarks provide almost the same inferences as the P8 benchmark used here.

[5]This is because the Jensen Measure and the Treynor-Mazuy Total Performance Measure are Period Weighting Measures (without the nonnegativity constraint) that are based on utility functions that differ substantially from the power utility function used to calculate the weights in this study. For further discussion of this see Grinblatt and Titman (1989b).

therefore appear to be timing this index.

V. Mutual Fund Performance

The performance statistics presented in Tables 1-3 for the passive portfolios are repeated in Tables 4-6 for the sample of 279 mutual funds. The cross-sectional means and standard deviations for each measure and each benchmark are presented in Table 4. The standard deviations are similar for all of the measures and benchmarks, but the means, which are all negative, differ substantially. The means are strongly negative for the Treynor-Mazuy Measure with each of the benchmarks as well as for the Jensen and Positive Period Weighting Measures with the equally-weighted and factor-based benchmarks. However, the Jensen and Positive Period Weighting Measures that are calculated with the P8 or P6 benchmark are close to zero.

The magnitude of the average performance scores using the equally-weighted and factor analysis benchmarks, (about -3.5 to -4% per year), are too large to be explained by the transaction costs and the expenses of the funds.[6] This suggests that the negative performance must be either due to the funds systematically picking stocks that do poorly, or to the benchmarks being inefficient. Given the known size-related bias of these benchmarks, and the fact that mutual funds tend to invest in larger than average firms, the latter possibility is the more plausible. The P8 and P6 benchmarks are not subject to this size-related bias and do not generate significant negative performance. As a result, conclusions that one would draw about the overall performance of the mutual fund industry would be strongly influenced by the choice of benchmarks.

---

[6]See Grinblatt and Titman (1989a) for estimates of these costs.

The average performance of the mutual funds with the Treynor-Mazuy Measure is negative with the P8 benchmark and is due to a negative coefficient on the quadratic term. The downward shift in performance due to the quadratic term (about .5%) is virtually identical to the downward shift for the passive portfolios and is remarkably stable across mutual funds and passive portfolios. As was the case with the passive portfolios, this negative performance is due to the January 1975 outlier.

Correlations between the performance scores generated with the different benchmarks are presented in Table 5. This table suggests that the choice of a benchmark can have a large effect on the relative rankings of the mutual funds. The correlations between the Jensen Measures calculated with the P8 benchmark and the measures calculated with either the equally-weighted or factor-based benchmarks range between .37 and .63. However, the correlations between the measures calculated with the equally-weighted and factor-based benchmarks (.86) are fairly high, as are the correlations between the P8 and P6 benchmarks (.93). These correlations, along with the average performance numbers generated with the different benchmarks show that factor analysis does not improve on the inferences generated by the equally-weighted index, while the characteristic-based benchmarks do provide different inferences and demonstrate robustness with respect to minor deviations in the characteristic portfolios used.

Table 6 examines the extent to which the different measures provide different inferences. The most striking observation here is the similarity of the inferences generated by the Jensen and Positive Period Weighting Measures when the same benchmarks are employed. For any given benchmark, the correlation between the performance scores exceed .97. The Treynor-Mazuy Total

Performance Measure is also highly correlated with these two measures, but it appears to be shifted downward for the P8 benchmark, as indicated by the mean in Table 4.

A subset of the performance measures and rankings for a sample of 11 funds, ranked first, 29th, 57th, etc. by the Jensen Measure using the equally-weighted benchmark portfolio, is presented in Table 7. This table documents that the superior performance of the Fidelity Magellan fund is quite robust with respect to the benchmark choice as well as the measure. However, the United Vanguard fund that has approximately zero performance with the equally-weighted and factor analysis benchmarks achieves significant positive performance with all of the measures only when the P8 benchmark is utilized. In all cases, the performance scores for the different measures using the same benchmark are very similar.

It is important to stress here that we are not claiming that the choice of measures never matters. As we saw earlier, passive portfolios exhibited scores that differed between the Treynor-Mazuy Total Measure and the others. Moreover, since differences between the Jensen Measure and either the Positive Period Weighting or the Treynor-Mazuy Total Performance Measure can be attributed to timing ability, performance inferences will differ for mutual funds that time the market. Although Table 6 indicates that these differences are generally small for most mutual funds, or exist for only a small number of funds, Table 8 indicates that they do exist.

Table 8 reports a statistically reliable relation between the coefficient of the Treynor-Mazuy quadratic term and the difference between the Jensen and Positive Period Weighting Measures across mutual funds. The t-statistics reported for the four cross-sectional regressions in Table 8 were calculated with a time-series procedure described in Appendix C. The need to calculate the t-statistics with this time series approach arises because the usual t-statistics derived from cross-sectional regressions like these are biased if fund returns are correlated cross-sectionally.

VIII. Conclusion

This study examined the sensitivity of performance inferences to the selection of a benchmark and to the measure used. It found that measures do not seem to matter as much as benchmarks. The different measures of total performance that were examined in the paper, the Jensen Measure, the Treynor-Mazuy Total Performance Measure, and the Positive Period Weighting Measure, generated similar performance scores on average and displayed high cross-sectional correlations when they used the same benchmarks. This suggests that the concerns of Jensen (1972), Admati and Ross (1985), Dybvig and Ross (1985), and Grinblatt and Titman (1989b) about a timing-related problem in the Jensen Measure may not be important in practice since measures that eliminate this problem yield almost identical inferences as those generated by the Jensen Measure.

The choice of a benchmark, by contrast, can have a large effect on inferences about performance. For instance, the large size and dividend-related biases of the equally-weighted benchmark and the factor benchmark appear to carry over to the mutual funds. The mutual funds displayed strong negative performance, on average, with both the equally-weighted index and the 10-factor benchmark. This negative bias is likely to be due to the size and dividend-related biases of these benchmarks. The P8 and P6 benchmarks that were designed to eliminate these and other well-known biases generated performance scores that were close to zero on average.

Although the performance scores generated with the equally-weighted index and the 10-factor benchmark are not highly correlated with either the scores generated with the P8 or P6 benchmark, the scores generated with the equally-equally-weighted and 10-factor benchmark are highly correlated with each other as are those generated by the P6 and P8 benchmarks. The high correlation between the scores generated with the equally-weighted index and the 10-factor benchmark indicates that benchmarks formed with statistical factor analysis do not correct for the biases in the equally-weighted index. The high correlation between the scores generated with the P8 and P6 benchmark, however, is somewhat reassuring, since it suggests that minor variations in the procedure for forming benchmark portfolios based on securities characteristics will have little effect on performance infererences. While these results are not conclusive, they suggest that forming factor portfolios using securities characteristics rather than relying solely on return covariances is likely to be a better way of implementing multi-factor risk-adjustment models.

Finally, it appears that when a benchmark loads too heavily on a single factor, the Treynor-Mazuy Measure may be sensitive to outliers. Theoretically, this could have been corrected by implementing the measure on the eight portfolios in the P8 benchmark directly, rather than on the ex-post efficient combination of them. However, the resulting number of regressors, 72, would have been too large to allow reasonable inferences. Further research is needed on how to best implement the Treynor-Mazuy measure for a multiple portfolio benchmark before this measure can be of much practical use.

APPENDIX A: THE POSITIVE PERIOD WEIGHTING MEASURE

Let r_{It} denote the period t return above the risk-free rate for the Ith index portfolio in a benchmark. The Positive Period Weighting Measure is calculated by specifying a set of nonnegative weights, w_t, that satisfy the weighted excess return condition

$$\sum_{t=1}^{T} w_t r_{It} = 0$$

for each of the portfolios in a benchmark. Grinblatt and Titman (1989b) demonstrated that the Jensen Measure can be characterized as a Period Weighting Measure that does not satisfy the nonnegativity constraint. With the Jensen Measure, the period weights are proportional to the marginal utilities of a quadratic utility investor who holds some combination of the benchmark portfolios. The nonnegativity constraint required by the Positive Period Weighting Measure is not satisfied by the Jensen Measure because quadratic utility exhibits satiation and hence, negative marginal utility above certain wealth levels.

The period weights used in this study can be interpreted as the marginal utilities of an investor with power utility. Since this utility function does not exhibit satiation, period weights derived from it satisfy the nonnegativity constraint. A risk aversion parameter of eight was chosen because it generated an optimal portfolio that required almost no holdings of the risk-free asset. To calculate a set of weights for each of the four benchmarks we

(i) Applied an algorithm that searched for the utility optimal combination of the portfolios in the benchmark and the risk-free asset.

(ii) Calculated the time series of returns of the optimal portfolio.

(iii) Interpreted the gross returns as wealth levels (i.e. WLOG set initial wealth to one for each observation) and calculated the marginal utility of this wealth level with the power function.

(iv) Used the time series of marginal utilities as positive period weights.

Since the first order condition for utility maximization requires that $E(U'(\ )r_I) = 0$ for the return, r_I, of each portfolio in the benchmark, the four step procedure described above derives weights that approximately satisfy the weighted excess return condition. These weights are then scaled to sum to one so that observed Positive Period Weighting Measures can be interpreted as average monthly excess returns.

The Positive Period Weighting Measure, like the Jensen Measure, is a linear weighting of returns. The standard t-statistic can thus be used to test whether it is significantly different from zero, when conditioned on the excess returns of the portfolios in the benchmark. The test statistic, which has a t-distribution[7] with T-K-1 degrees of freedom if there are T returns and K benchmark portfolios is

$$PW/\sqrt{s^2 \Sigma w_t^2},$$

where s is the standard error of the excess return regression used to compute the Jensen Measure for the portfolio and PW is the Positive Period Weighting Measure.

APPENDIX B: THE TREYNOR-MAZUY QUADRATIC REGRESSION

The Treynor-Mazuy quadratic regression is

$$R_p = \alpha_p + \beta_{1p} r_I + \beta_{2p} r_I^2 + \epsilon_p,$$

---

[7]This assumes that the benchmark portfolios add up to a point on the efficient frontier of the 109 test portfolios and that the test portfolios have normally distributed residuals.

where r_I is the excess return on the benchmark portfolio and R_p is the excess return on the portfolio being measured. Jensen (1972) and Admati, Bhattacharya, Pfleiderer, and Ross (1986) have analyzed the asymptotic properties of the two slope coefficients in the regression when the portfolio strategy has linear risk adjustments to timing signals. They noted that the second slope coefficient, which measures co-skewness with the benchmark portfolio, is related to timing ability. In this special case, it can easily be shown that the contribution of timing information to the excess return of the portfolio is proportional to the coefficient on the quadratic term in large samples. The Treynor-Mazuy Total Performance Measure is defined to be

$$TM = \alpha_p + \beta_{2p} var(r_I),$$

which is easily demonstrated to be the added return from superior information under the Admati et al. (1986) assumptions (i.e. exponential utility and multivariate normality).

In principal, it may be possible to use the Treynor-Mazuy regression to analyze performance with multiple portfolio benchmarks. This would require cross-product terms and would generate a very large number of regressors for the 8 and 10 portfolio benchmarks used earlier. For this reason, we employ the returns of the ex-post efficient combination of the portfolios used as the input to the Treynor-Mazuy regression.

APPENDIX C: TIME SERIES TESTS OF CROSS-SECTIONAL RELATIONSHIPS

The significance of the cross-sectional coefficients in Table 8 is an extension of a technique developed in Fama and MacBeth (1973). The slope coefficient in a cross-sectional regression of the difference between the Positive Period Weighting Measure and the Jensen Measure on the Treynor-Mazuy quadratic coefficient is a linear "weighting" of the dependent variable, where the weights sum to zero. The weight attached to fund i's observation is proportional to the difference between fund i's Treynor-Mazuy coefficient and the sample mean of all the Treynor-Mazuy coefficients. It necessarily follows that one can interpret the slope coefficient in the regression as the difference between the Positive Period Weighting Measure and the Jensen Measure for a self-financing portfolio of mutual funds, where the portfolio weight on fund i is given above.

Since the performance of a portfolio of funds is the portfolio-weighted average of each fund's performance, the time series of the returns of this self-financing portfolio can be used to compute its Jensen and Positive Period Weighting Measures. From Grinblatt and Titman (1989b), we know that the Jensen Measure is a weighted average of the time series of excess returns, where the weights are linear in the returns of the benchmark (or the ex-post efficient combination of the portfolios in the benchmark). These weights, denoted w_{Jt} for the return in period t, are the marginal utilities of a quadratic utility investor who invests in the benchmark. The weights for the Positive Period Weighting Measure, described in Appendix A and denoted w_{Pt} for period t, represent the marginal utilities of a power utility investor. The difference between the two measures for the self-financing portfolio of mutual funds is then

$$\sum_t (w_{Pt} - w_{Jt}) r_t$$

where r_t is the excess return of the self-financing portfolio in period t. If the returns are conditionally i.i.d. normal, an unbiased t-statistic for the difference in the measures is

$$(PW - JM) / \sqrt{s^2 \sum_t (w_{Pt} - w_{Jt})^2} \quad ,$$

where s is the standard error of the excess return regression used to compute the Jensen Measure for the self-financing portfolio, PW is its Positive Period Weighting Measure and JM is its Jensen Measure. These are the t-statistics reported in Table 8.

REFERENCES

Admati, Anat, Sudipto Bhattacharya, Paul Pfleiderer, and Stephen Ross (1986), "On Timing and Selectivity," Journal of Finance, 46, 715-730.

Banz, Rolf (1981), "The Relationship Between Return and Market Value of Common Stocks," Journal of Financial Economics, 9, 3-18.

Black, Fischer, Michael Jensen, and Myron Scholes (1972), "The Capital Asset Pricing Model: Some Empirical Tests," in Jensen (ed.) Studies in the Theory o Capital Markets, New York, Praeger.

DeBondt, Werner and Richard Thaler (1985), "Does the Stock Market Overreact?" Journal of Finance, 40, 793-805.

Dybvig, Philip and Stephen Ross (1985), "Differential Information and Performance Measurement Using A Security Market Line," Journal of Finance, 40, 282-99.

Fama, Eugene (1976), Foundations of Finance, New York, Basic Books.

__________ and James MacBeth (1973), "Risk, Return and Equilibrium: Empirical Tests", Journal of Political Economy, 607-636.

Grinblatt, Mark and Sheridan Titman (1988), "Mutual Fund Performance: An Analysis of Monthly Returns," UCLA Working Paper.

Grinblatt, Mark and Sheridan Titman (1989a), "The Evaluation of Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings," Journal of Business, 62, 394-415.

Grinblatt, Mark and Sheridan Titman (1989b), "Portfolio Performance Evaluation: Old Issues and New Insights," Review of Financial Studies, 2, 396-422.

Jegadeesh, Narasimhan (1990), "Evidence of Predictable Behavior of Security Returns," Journal of Finance,45, 881-898.

Jensen, Michael (1968) "The Performance of Mutual Funds in the Period 1945-1964," Journal of Finance, 23, 389-416.

__________ (1969), "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios," Journal of Business, 42, 167-247.

__________ (1972), "Optimal Utilization of Market Forecasts and the Evaluation of Investment Portfolio Performance," in Szego/Shell (eds.), Mathematical Methods in Investment and Finance, Amsterdam, North Holland.

Lehmann, Bruce and David Modest (1985), "The Empirical Foundations of the Arbitrage Pricing Theory II: The Optimal Construction of Basis Portfolios," Columbia University Working Paper.

__________ (1987), "Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons," Journal of Finance, 42, 233-265.

Litzenberger, Robert and Krishna Ramaswamy (1979), "The Effects of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical Evidence," Journal of Financial Economics, 7, 163-195.

__________ (1982), "The Effects of Dividends on Common Stock Returns: Tax Effects or Information Effects?" Journal of Finance, 37, 429-443.

Reinganum, Marc (1981), "Misspecification of Capital Asset Pricing: Empirical Anomalies Based on Earnings' Yields and Market Values," Journal of Financial Economics, 9, 19-46.

Roll, Richard (1978), "Ambiguity When Performance is Measured by the Securities Market Line," Journal of Finance, 33, 1051-1069.

Treynor, Jack (1965), "How to Rate Management of Investment Funds", Harvard Business Review, 44, 131-136.

__________ and F. Mazuy (1966) "Can Mutual Funds Outguess the Market?" Harvard Business Review, 45, 131-136.

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR 3 MEASURES WITH 4 BENCHMARKS
USING 109 PASSIVE TEST PORTFOLIOS, EACH WITH 120 MONTHLY OBSERVATIONS

BENCHMARK:	EW INDEX	10 FACTORS	P8	P6
JENSEN	.0002	.0004	.0001	.0003
	(.0023)	(.0021)	(.0023)	(.0024)
POS.PER.WTG.	.0002	.0007	.0001	.0003
	(.0024)	(.0023)	(.0022)	(.0024)
TMTOT	.0003	.0003	-.0022	.0002
	(.0025)	(.0025)	(.0023)	(.0023)

----

*Standard Deviations below in parentheses

TABLE 2

CORRELATIONS BETWEEN BENCHMARKS USING 109 PASSIVE TEST PORTFOLIOS EACH WITH 120 MONTHLY OBSERVATIONS

Matrix A: Jensen Measure

BENCHMARK:	10 FACT	P8	P6
EW	.64	.35	.35
10 FACT		.32	.38
P8			.91

Matrix B: Positive Period Weighting Measure

BENCHMARK:	10 FACT	P8	P6
EW	.69	.19	.25
10 FACT		.21	.31
P8			.93

Matrix C: Treynor-Mazuy Total Measure

BENCHMARK:	10 FACT	P8	P6
EW	.71	.37	.31
10 FACT		.41	.39
P8			.92

TABLE 3

CORRELATIONS BETWEEN MEASURES USING 109 PASSIVE TEST PORTFOLIOS
EACH WITH 120 MONTHLY OBSERVATIONS

Matrix A: Equal-Weighted Benchmark

MEASURE:	PER.WT.	TMTOT
JENSEN	.97	.58
PER.WT.		.64

Matrix B: Factor-Based Benchmark

MEASURE:	PER.WT.	TMTOT
JENSEN	.95	.44
PER.WT.		.51

Matrix C: 8 Portfolio Benchmark

MEASURE:	PER.WT.	TMTOT
JENSEN	.95	.88
PER.WT.		.88

Matrix D: 6 Portfolio Benchmark

MEASURE:	PER.WT.	TMTOT
JENSEN	.96	.90
PER.WT.		.89

TABLE 4

MEANS AND STANDARD DEVIATIONS FOR 3 MEASURES WITH 4 BENCHMARKS
USING 279 MUTUAL FUNDS, EACH WITH 120 MONTHLY OBSERVATIONS

BENCHMARK:	EW INDEX	10 FACTORS	P8	P6
JENSEN	-.0028	-.0033	-.0004	.0000
	(.0030)	(.0032)	(.0032)	(.0034)
POS.PER.WTG.	-.0035	-.0037	-.0001	.0002
	(.0030)	(.0032)	(.0036)	(.0036)
TMTOT	-.0037	-.0043	-.0025	-.0004
	(.0031)	(.0034)	(.0032)	(.0034)

----

*Standard Deviations below in parentheses

TABLE 5

CORRELATIONS BETWEEN BENCHMARKS USING 279 MUTUAL FUNDS
EACH WITH 120 MONTHLY OBSERVATIONS

Matrix A: Jensen Measure

BENCHMARK:	10 FACT	P8	P6
EW	.86	.60	.63
10 FACT		.42	.47
P8			.93

Matrix B: Positive Period Weighting Measure

BENCHMARK:	10 FACT	P8	P6
EW	.89	.48	.53
10 FACT		.37	.44
P8			.96

Matrix C: Treynor-Mazuy Total Measure

BENCHMARK:	10 FACT	P8	P6
EW	.87	.57	.56
10 FACT		.43	.43
P8			.95

TABLE 6

CORRELATIONS BETWEEN MEASURES USING 279 MUTUAL FUNDS
EACH WITH 120 MONTHLY OBSERVATIONS

Matrix A: Equal-Weighted Benchmark

MEASURE:	PER.WT.	TMTOT
JENSEN	.99	.98
PER.WT.		1.00*

Matrix B: Factor-Based Benchmark

MEASURE:	PER.WT.	TMTOT
JENSEN	.99	.98
PER.WT.		.99

Matrix C: 8 Portfolio Benchmark

MEASURE:	PER.WT.	TMTOT
JENSEN	.98	.97
PER.WT.		.97

Matrix D: 6 Portfolio Benchmark

MEASURE:	PER.WT.	TMTOT
JENSEN	.97	.99
PER.WT.		.99

* Rounded to 1.00

TABLE 7

MEASURES* OF PERFORMANCE AND RANKS** FOR A SAMPLE OF ELEVEN FUNDS

Measure:	Jensen			Positive Period Wtg			Treynor-Mazuy Total		
Benchmark:	EW	Factor	P8	EW	Factor	P8	EW	Factor	P8
Funds									
Fidelity Magellan	.0067	.0062	.0080	.0054	.0055	.0084	.0049	.0046	.0052
	(1)	(2)	(3)	(2)	(2)	(5)	(2)	(2)	(6)
United Vanguard	.0003	-.0018	.0063	-.0002	-.0020	.0061	-.0004	-.0029	.0038
	(29)	(89)	(7)	(29)	(80)	(15)	(29)	(103)	(9)
Guardian Mutual	-.0007	-.0004	.0010	-.0012	-.0002	.0018	-.0014	-.0006	-.0007
	(57)	(40)	(83)	(51)	(27)	(76)	(52)	(34)	(68)
Sigma Capital	-.0013	-.0047	-.0001	-.0016	-.0041	.0004	-.0018	-.0050	-.0023
	(84)	(193)	(127)	(68)	(179)	(117)	(64)	(180)	(123)
IDS New Dimensions	-.0020	-.0046	.0032	-.0031	-.0050	.0039	-.0036	-.0061	.0010
	(112)	(188)	(35)	(133)	(190)	(33)	(152)	(200)	(31)
ISI Trust	-.0027	-.0039	-.0034	-.0033	-.0043	-.0033	-.0034	-.0041	-.0043
	(140)	(175)	(245)	(145)	(182)	(238)	(138)	(149)	(209)
Meeschaert Cap. Accum.	-.0033	-.0027	-.0050	-.0034	-.0028	-.0043	-.0034	-.0028	-.0057
	(168)	(132)	(263)	(156)	(118)	(258)	(143)	(96)	(247)
Sentry	-.0042	-.0071	-.0011	-.0050	-.0073	-.0013	-.0054	-.0084	-.0039
	(196)	(243)	(154)	(203)	(240)	(169)	(202)	(247)	(190)
BLC Growth	-.0052	-.0061	-.0017	-.0061	-.0071	-.0012	-.0066	-.0080	-.0045
	(223)	(224)	(180)	(230)	(236)	(166)	(231)	(238)	(219)
Axe-Houghton Stock	-.0065	-.0081	-.0019	-.0071	-.0076	.0006	-.0076	-.0090	-.0028
	(251)	(259)	(194)	(249)	(250)	(107)	(253)	(255)	(141)
Steadman American Ind.	-.0124	-.0121	-.0076	-.0138	-.0139	-.0075	-.0142	-.0148	-.0106
	(279)	(277)	(271)	(279)	(279)	(272)	(279)	(279)	(274)

---

* Significance level for performance measures is not reported

** Ranks are below performance measures in parentheses.

TABLE 8

TIME SERIES t-STATISTICS FOR SLOPE COEFFICIENTS IN CROSS-SECTIONAL REGRESSIONS OF THE DIFFERENCE BETWEEN THE POSITIVE PERIOD WEIGHTING AND JENSEN MEASURES ON THE TREYNOR-MAZUY TIMING MEASURE USING 279 MUTUAL FUNDS

Benchmark:	EW	10 Factor	P8	P6
Time Series t-statistic:	5.15++	3.45++	2.22+	3.49++

+ Significant at .05 level
++ Significant at .01 level