Evaluating the Performance of International Mutual Funds

Author(s): Robert E. Cumby and Jack D. Glen

Source: *The Journal of Finance*, Vol. 45, No. 2 (Jun., 1990), pp. 497-521
Published by: Blackwell Publishing for the American Finance Association
Stable URL: http://www.jstor.org/stable/2328667
Accessed: 26/11/2009 07:17

THE JOURNAL OF FINANCE • VOL. XLV, NO. 2 • JUNE 1990

Evaluating the Performance of International Mutual Funds

ROBERT E. CUMBY and JACK D. GLEN*

ABSTRACT

In this paper, we examine the performance of a sample of fifteen U.S.-based internationally diversified mutual funds between 1982 and 1988. Two performance measures are used, the Jensen measure and the positive period weighting measure proposed by Grinblatt and Titman. We find no evidence that the funds, either individually or as a whole, provide investors with performance that surpasses that of a broad, international equity index over this sample period.

IT IS WIDELY RECOGNIZED that international diversification may produce significant reductions in systematic risk. One way for investors to achieve such diversification without investing in costly information acquisition is through internationally diversified mutual funds. In addition to providing diversification, do fund managers possess information that provides superior returns for an uninformed investor? While considerable evidence on the behavior of mutual fund returns now exists, there is surprisingly little evidence about the performance of internationally diversified funds.[1] This paper attempts to remedy that deficiency by providing empirical evidence on the performance of a sample of U.S.-based internationally diversified mutual funds.

Two measures of performance are utilized in this paper. The first is the widely used Jensen (1968, 1969) measure, which uses the security market line to evaluate fund performance. The second is Grinblatt and Titman's (1989b) positive period weighting measure. While popular, the Jensen measure is subject to well known limitations, the most important of which is the possibility that errors in inference may arise when the fund manager is a market timer. In particular, the Jensen measure may indicate poor performance when the manager possesses and utilizes superior timing information. The performance measure suggested by Grinblatt and Titman (1989b) is not subject to these problems but has other limitations.

The plan of the paper is as follows. Section I briefly discusses Grinblatt and Titman's performance measure and compares it with Jensen's measure. In order to obtain valid inference with either measure, the benchmark used to evaluate fund performance must be mean-variance efficient from the point of view of the

* Stern School of Business, New York University and Wharton School, University of Pennsylvania, respectively. Jack D. Glen acknowledges financial support from the Geewax-Terker Research Program in Financial Instruments. The comments of Mark Grinblatt and David Modest are acknowledged with thanks.

[1] Particular attention should be paid to internationally diversified funds since the benchmark portfolios used in existing studies of mutual fund performance are purely domestic and may therefore be inappropriate for evaluating the performance of internationally diversified funds.

uninformed investor. In Section II, after discussing tests of benchmark efficiency, we examine the efficiency of a single-portfolio benchmark and a multiple-portfolio benchmark. We present evidence on the performance of a sample of internationally diversified mutual funds in Section III, using two benchmark portfolios. The first is the Morgan Stanley world index. The second is a two-portfolio benchmark consisting of the world index and an equally weighted portfolio of Eurocurrency deposits. Section IV examines the performance of the funds relative to the United States market and conducts tests for market timing ability. We conclude in Section V.

I. Tests of Performance

Suppose we wish to evaluate the performance of a managed portfolio with excess return r_{pt}. This portfolio is chosen from N assets with excess returns r_{jt}, $j = 1, \cdots, N$. Let r_{et} be the excess return on the portfolio that, from the point of view of the uninformed investor, is mean-variance efficient within the set of N tradable risky assets and whose orthogonal portfolio is used to compute excess returns.[2] Next, assume that the uninformed investors' expected return on this mean-variance efficient portfolio is μ_e and its variance is σ_e^2. In addition, suppose that the expected return on asset j is μ_j and that the covariance matrix of the returns $r = (r_1, r_2, \cdots, r_j, \cdots, r_N, r_e)'$ is constant given the information available to the uninformed investor. Given these definitions, it follows from Roll (1977) that

$$r_{jt} = \beta_j \, r_{et} + \varepsilon_{jt}, \qquad \beta_j = \frac{\operatorname{cov}(r_{jt}, r_{et})}{\sigma_e^2}, \tag{1}$$

where, from the point of view of the uninformed investor, β_j is constant and $E(\varepsilon_{jt}) = 0$. The portfolio manager is assumed to possess a different information set that results in time-varying expected returns. As a result, the portfolio weights will vary over time, as will the uninformed investors' expected value of r_{pt}. Since the portfolio weights change over time, β_{pt} is likely to be time varying as well.

We can write the returns on the managed portfolio as

$$r_{pt} = \beta_{pt} \, r_{et} + \varepsilon_{pt}. \tag{2}$$

The manager of the fund potentially possesses two types of superior information. The first of these is information about r_{et}. Letting Φ_t denote the time t information set of the manager, if $E(r_{et}|\Phi_t) \neq E(r_{et}) = \mu_e$, the manager is said to possess timing information. When $E(\varepsilon_{jt}|\Phi_t) \neq E(\varepsilon_{jt}) = 0$, the manager is said to possess selectivity information.

A popular measure of the performance of a managed fund is the Jensen (1968, 1969) measure, which is the intercept, α_p, of a least squares regression of r_{pt} on r_{et}. While the Jensen measure provides an accurate indication of the performance of a fund when the manager has selectivity information but no timing informa-

[2] In carrying out the tests, we will assume that investors can borrow and lend at a riskless rate of interest rather than computing the appropriate "zero-beta" rate.

tion, Jensen (1972), Admati and Ross (1985), and Dybvig and Ross (1985a) show that, when a manager has superior timing information, the Jensen measure, α_p, can be negative.

Grinblatt and Titman (1989b) examine a class of performance measures that includes the Jensen measure and show that certain members of the class do not suffer from the problems that arise with the Jensen measure.[3] The class of measures, called period weighting measures, is defined for a sample of T observations by

$$a = \sum_{t=1}^{T} w_t r_{pt}, \qquad \sum_{t=1}^{T} w_t = 1, \qquad \sum_{t=1}^{T} w_t r_{et} = 0, \tag{3}$$

where the weights, w_t, are functions of the return on the benchmark portfolio. Their main result is that, if $w_t > 0$ for all t, the performance measure, denoted α^* when $w_t > 0$, converges in probability to zero for an uninformed investor and to a positive number for an investor with selectivity information and no timing information or selectivity information and independently distributed timing information.[4]

Grinblatt and Titman (1989b) show that the Jensen performance measure is obtained by setting

$$w_t = \frac{V_e - (r_{et} - \bar{r}_e)\,\bar{r}_e}{T\ V_e},$$

where $\bar{r}_e$ and V_e are the sample mean and sample variance of the benchmark return. We then have

$$a_p = \sum_{t=1}^{T} w_t r_{pt} = \bar{r}_p - b_p \bar{r}_e, \tag{4}$$

where b_p is the estimated least squares slope coefficient from a regression of r_{pt} on r_{et}. The problem that arises with the Jensen measure when the investor has timing information can be seen by examining the weights, w_t. For large values of r_{et}, $w_t < 0$. When the investor has timing ability, r_{pt} will, on average, be large when $E(r_{et}|\Phi_t)$ is large and will therefore also be large, on average, when r_{et} is large. These large returns will then receive negative weights, making it possible that $\alpha_p < 0$ when the investor has timing information.

Grinblatt and Titman propose restricting the class of performance measures to those with $w_t > 0$ in order to overcome the problem of a negative Jensen measure arising when the manager has timing information. Following their suggestion, suppose the utility function of the uninformed investor in period t is

[3] Unlike other measures of performance that are valid when the manager has timing information, the period weighting measures do not require that the composition of the managed portfolio be observed.

[4] They also require that β_{pt} increases when the timing signal of the informed investor raises that investor's expected return on r_{et}. This condition will not hold when a large expected return raises risk aversion (through its effect on expected end-of-period wealth) by enough to offset the "substitution" effect of a higher expected return. They note that their results go through when the timing and selectivity information are correlated, provided that restrictions are imposed on preferences.

given by

$$U(W_t) = \frac{1}{1-\theta} W_t^{1-\theta},$$

where W_t is wealth at the end of period t. If the uninformed investor can hold the N risky assets and a riskless asset with return R_{ft}, end-of-period wealth is given by $W_t(\gamma) = 1 + \gamma R_{et} + (1-\gamma)R_{ft}$, where beginning-of-period wealth is normalized to one. Choosing γ to maximize expected utility yields, $E(W_t(\gamma^*)^{-\theta} r_{et}) = 0$, where γ^* is the optimal choice of γ and r_{et} is the excess return on the benchmark portfolio. An estimate of γ^* can be obtained by setting $\sum W_t(\hat{\gamma}^*)^{-\theta} r_{et} = 0$. Choosing the weights in α^* to be $w_t = W_t(\hat{\gamma}^*)^{-\theta}$ will then satisfy the last condition in (3) above. Normalizing the weights to sum to one implies that the positive period weighting measure has the same units as does r_{et}. When multiple portfolio benchmarks are employed, the same procedures are utilized to obtain a γ for each portfolio in the benchmark.

Choosing the weights to be marginal utilities provides an intuitive interpretation of the positive period weighting measure. If $\alpha^* > 0$, the uninformed investor with power utility will choose to hold some of the managed portfolio.

Inference using the positive period weighting measure may proceed as follows. Under the null hypothesis that the manager does not possess superior timing or selection information, the return on the managed portfolio is given by (2) with $E(\varepsilon_{pt}) = 0$ and $\beta_{pt} = \beta_p$. We then have

$$\begin{aligned} \alpha^* &= \sum_{t=1}^{T} w_t\,(\beta_p r_{et} + \varepsilon_{pt}) = \beta_p \sum_{t=1}^{T} w_t r_{et} + \sum_{t=1}^{T} w_t \varepsilon_{pt} \\ &= \sum_{t=1}^{T} w_t \varepsilon_{pt}. \end{aligned} \tag{5}$$

Assuming that the weights are distributed independently of ε_{pt} and that ε_{pt} is homoscedastic, the variance of α^* is $\sigma_\varepsilon^2 \sum w_t^2$, where σ_ε^2 can be computed using the residuals from the regression used to compute the Jensen performance measure. If ε_{pt} is heteroscedastic, or if the weights are not distributed independently of ε_{pt}, the variance of α^* is $\sum w_t^2\, \varepsilon_{pt}^2$. Joint tests that the positive period weighting measures for a group of assets are equal to zero can be performed by noting that, under the assumption of homoscedasticity, the covariance between the positive period weighting measure for assets i and j is given by $\sigma_{ij} \sum w_t^2$, where σ_{ij} is the covariance between ε_i and ε_j. With heteroscedasticity, this covariance is $\sum w_t^2\, \varepsilon_{it} \varepsilon_{jt}$.

If the benchmark return is measured with mean zero error, both performance measures will be biased upwards. The bias in the Jensen measure arises since measurement error will bias the slope parameter, b_p, in (4) toward zero. As a result, when $\beta_p > 0$, plim $\alpha_p > 0$ even if the manager does not possess superior information. The bias in the positive period weighting measure arises because, with measurement error, the weights are not chosen so that $\sum w_t r_{et} = 0$. Instead, they will be chosen so that $\sum U'[1 + \gamma(r_{et} + \eta_t) + R_{ft}]\ (r_{et} + \eta_t) = 0$, where η_t is the measurement error in period t. Since a positive realization of η_t will raise computed end-of-period wealth and therefore reduce $w_t = U'(W_t)$, plim $\sum w_t\eta_t <$ 0. With the weights chosen so that $\sum w_t(r_{et} + \eta_t) = 0$, $p$lim $\sum w_t r_{et} > 0$. As can be seen from (5), $p$lim α^* will therefore be positive even when the null hypothesis is true.

The choice of benchmark to carry out the tests of performance is relatively straightforward. Grinblatt and Titman (1989b) show that under reasonable conditions both the Jensen measure and the positive period weighting measure require only that we consider a portfolio that is mean-variance efficient from the point of view of the uninformed investor within the set of risky assets available to the manager. We thus do not need to measure the returns on a "market" portfolio that includes nontraded assets. We do need to verify, however, that the benchmark chosen is, in fact, mean-variance efficient within the set of securities available to the fund manager. The next section describes the procedures we use to test for mean-variance efficiency of the benchmark.

II. Mean-Variance Efficiency of the Benchmark Portfolio

Suppose we wish to test the mean-variance efficiency of a candidate benchmark portfolio, b, with excess returns r_{bt}. Consider a regression in which the portfolio orthogonal to b is used to compute excess returns:

$$r_{jt} = \alpha_{jb} + \beta_{jb} r_{bt} + \varepsilon_{jt}.$$

It follows from Roll (1977) that, if the portfolio b is mean-variance efficient within the set of N assets, then $\alpha_{jb} = 0$. We carry out a test of the mean-variance efficiency of a proposed benchmark portfolio by estimating the N regressions and testing whether $\alpha_{jb} = 0$ individually and jointly.

We do not have return data on the full set of N equities available to the managers of the internationally diversified mutual funds that we are examining. Instead, we have data on the returns on the national market indexes from thirteen countries that include all major equity markets worldwide.[5] The tests we carry out will then be tests of the efficiency of the proposed benchmark with respect to the thirteen national market index portfolios. If these passive national market portfolios yield positive intercepts, the apparent superior performance to these portfolios will be attributed to inefficiency of the benchmark portfolio.

How do we choose a benchmark that is likely to be mean-variance efficient? International asset pricing models predict that a world market portfolio of equities will be mean-variance efficient only under restrictive assumptions.[6] Under less restrictive assumptions, multiple portfolio benchmarks will be efficient. Adler and Dumas (1983) show that, when inflation is nonstochastic, a benchmark consisting of a world market portfolio of equities and a portfolio

[5] The data are from Morgan Stanley's *Capital International Perspective*. The countries are, in alphabetical order, Australia, Belgium, Canada, France, Germany, Hong Kong, Italy, Japan, Netherlands, Singapore-Malaysia, Switzerland, United Kingdom, and United States.

[6] Adler and Dumas (1983) and Stulz (1984) present overviews of international asset pricing models and describe the conditions under which a single international market portfolio will be efficient. Other models, such as those presented by Hodrick (1981) and Ross and Walsh (1983), are multiple β models and thus predict that a linear combination of portfolios will be efficient.

consisting of nominal riskless assets denominated in each currency, which hedges purchasing power risk, will be efficient. In addition to a single-portfolio benchmark, we also consider

$$r_{jt} = \alpha_j + \sum_{l=1}^{L} \beta_{jl} r_{lt} + \beta_{jw} r_{wt} + \varepsilon_{jt}, \tag{7}$$

where r_{lt} is the excess home currency return on an asset that is riskless in terms of currency l and r_{wt} is the excess return on the world equity portfolio. As is the case with single-portfolio benchmarks, mean-variance efficiency of the multiple-portfolio benchmark implies $\alpha_j = 0$.

In spite of the restrictive assumptions needed to prove efficiency of the world portfolio of equities, a world index portfolio is a useful starting point in searching for an efficient benchmark. We therefore begin by considering as a candidate benchmark portfolio the Morgan Stanley *Capital International Perspective* world index. This is a U.S. dollar-denominated index based on the share prices for 1450 companies listed on the stock exchanges in 20 countries. The companies used to compute the index represent approximately 60 percent of the market value of stocks in the 20 countries. Importantly, 99 percent of non-U.S. stocks covered in the index are readily purchasable by non-nationals, indicating that the components of this benchmark are indeed available to the managers of internationally diversified mutual funds based in the United States.[7] We also assume that the U.S. Treasury bill rate is viewed as riskless to the potential purchasers of the funds we analyze and thus compute excess returns by subtracting the return on Treasury bills.[8]

The results of tests of the mean-variance efficiency of the Morgan Stanley world index are found in Table I, where we report point estimates of the intercept and its t-ratio obtained from regressions of Morgan Stanley country index returns on Morgan Stanley world index returns. Eight of the thirteen estimated intercepts are positive, with one (Belgium) statistically significant at the ten percent level. We also estimate the thirteen regressions as a system and carry out a Wald test that the intercepts are jointly zero. The test statistic, reported in Table I along with its marginal significance level, indicates that the null hypothesis that the intercepts are jointly zero cannot be rejected at standard significance levels. While we find that the intercepts are generally not statistically significant, a few of them are quite large. The intercept for the Belgian index exceeds one percent per month, while another (Japan) is in the neighborhood of .75 percent per month. Intercepts of this size, while not statistically significantly greater than zero, raise questions about the power of the test of mean-variance efficiency.

Previous work using U.S. stock returns has found inefficiencies in standard market indexes arising due to seasonality, firm size, and dividend yield.[9] Since

[7] The remaining one percent is due to restricted Swiss stocks and to Swedish bank stocks.

[8] Adler and Dumas (1983) point out that this amounts to an implicit assumption that inflation rates are nonstochastic or that the variance of inflation is so much lower than that of security returns that inflation may be treated approximately as nonstochastic.

[9] Banz (1981) and Keim (1983) present evidence on the size effect and seasonality in U. S. stock returns. Litzenberger and Ramaswamy (1982) present evidence that high dividend yields are associated with positive abnormal returns in U.S. stocks. International evidence on stock return seasonality is found in Gultekin and Gultekin (1983). Hawawini (1988) surveys international evidence on stock return anomalies.

Table I

Jensen and Positive Period Weighting Tests of the Efficiency of the Capital International World Index

The sample period is January 1982 to June 1988. r_{wt} is the return on the Capital International World Index, and the r_{jt} are the returns on the Capital International country indexes. The w_t are the marginal utilities of an investor with power utility. All returns are measured in percent per month. t-Ratios are in parentheses. Significantly positive Jensen or positive period weighting measures indicate superior performance of the passive country indexes and therefore suggest benchmark inefficiency.

	Jensen Test ($r_{jt} = \alpha_j + \beta_j r_{wt} + \varepsilon_{jt}$)	Positive Period Weighting Test ($PPW_j = \sum w_t r_{jt}$)
Country Index	α_j	PPW_j
Australia	−0.121	−0.505
	(−0.125)	(−0.529)
Belgium	1.050	1.021
	(1.754)	(1.719)
Canada	−0.716	−0.743
	(−1.507)	(−1.576)
France	0.351	0.350
	(0.532)	(0.534)
Germany	0.213	0.118
	(0.307)	(0.172)
Hong Kong	0.027	−0.424
	(0.022)	(−0.352)
Italy	0.186	0.224
	(0.222)	(0.270)
Japan	0.777	0.904
	(1.323)	(1.551)
Netherlands	0.380	0.329
	(0.815)	(0.710)
Singapore	−1.060	−1.376
	(−1.239)	(−1.622)
Switzerland	−0.228	−0.290
	(−0.466)	(−0.599)
United Kimgdom	0.029	0.047
	(0.054)	(0.088)
United States	−0.303	−0.352
	(−0.913)	(−1.070)
Joint Test	χ^2 (13) = 12.652	χ^2 (13) = 13.222
[Probability Value]	[0.475]	[0.431]

the national market indexes we use above are portfolios selected using other criteria (such as currency or national economic policies), it is possible that we have not grouped assets in a way that will reveal inefficiencies in the benchmark. While the use of country indexes is useful if inefficiencies in the world index arise due to market segmentation or due to currency or purchasing power risk, this grouping of equities may not reveal the effects of size, dividend yield, or other characteristics that may lead us to reject the efficiency of the proposed benchmark. For that reason, we carry out tests by sorting the national market indexes according to several criteria: country size (GNP), size of market (market capitalization in U.S. dollars), turnover in the market, liquidity (turnover/

capitalization), concentration (percentage of market capitalization accounted for by the largest 10 firms), and dividend yield. For each criterion, we construct four portfolios. The first portfolio consists of the three national market indexes that ranked highest according to the criterion at the beginning of our sample, and so on. We then regress each of these four equally weighted portfolios on the Morgan Stanley world index and test whether the intercepts are zero, both individually and jointly.

The results of the tests using the rank-based portfolios are found in Table II, where the point estimates of the intercepts are reported along with their t-ratios. There is, on the whole, little evidence against the null hypothesis that the benchmark is efficient. Only the portfolios formed by the largest quartile of indexes ranked by capitalization, concentration, and turnover have intercepts with t-ratios in excess of 1.0. If small firms tended to outperform large firms, we might expect to find small capitalization indexes and low concentration indexes exhibiting superior performance. This is clearly not the case in Table II. There is also no evidence that national indexes with high dividend yields have systematically higher returns than national indexes with low dividend yields.[10]

Tables I and II also contain the positive period weighting performance measures (along with their t-ratios) computed for the country indexes and the portfolios formed from the country indexes.[11] If we find significantly positive measures using the country indexes, we have reason to suspect the reliability of any evidence of superior performance found in the following section of the paper when we use the positive period weighting measure to examine the performance of our sample of mutual funds. Fortunately, no evidence of superior performance is found. Instead, the results of the positive period weighting tests are remarkably similar to the results of the Jensen measures.[12] This similarity is reassuring as both are valid performance measures in the absence of market timing information on the part of an investment manager. Clearly, market timing can play no role in the passive portfolios we examine.

Next we turn to tests of the efficiency of the benchmarks that include a currency risk hedge portfolio along with the world equity portfolio. We consider three specifications of the hedge portfolio. First, we construct an equally weighted portfolio of Eurocurrency deposits denominated in seven currencies, thus constraining $\sum \beta_{jl} r_{lt}$ to be $(\beta_{jc}/L) \sum r_{lt}$, where β_{jc} is the "beta" of the currency risk hedge portfolio and L is the number of Eurocurrency deposits in the portfolio.[13] Next, we use relative import shares to compute weights for a portfolio of Eurocurrency deposits denominated in the same seven currencies and constrain $\sum \beta_{jl} r_{lt}$ to be $\beta_{jc} \sum \lambda_l r_{lt}$, where λ_l is the import share of country l. Finally, we

[10] This is not surprising since the tax treatment of dividends differs across countries.

[11] The weights are the marginal utilities of an investor with power utility. The coefficient of relative risk aversion, θ, is assumed to be 6. The results do not appear to be sensitive to the choice of θ.

[12] The similarity is not surprising since the Jensen weights and the Grinblatt-Titman weights have a sample correlation coefficient of .936.

[13] The currencies are the Canadian dollar, the French franc, the Deutsche mark, the Japanese yen, the Dutch guilder, the Swiss franc, and the pound Sterling. We consider the three constrained portfolios since evidence presented in Adler and Dumas (1983) and Jorion (1985) shows that portfolios formed with unconstrained weights generally involve implausibly large short positions in some currencies that are offset by implausibly large long positions in other currencies.

restrict L to be two and consider an unconstrained hedge portfolio consisting only of Euro-DM and Euro-yen deposits.

Table III contains the results of tests of the efficiency of the two-portfolio benchmark consisting of the Morgan Stanley world index and the equally weighted portfolio of Eurocurrency deposits. The results obtained using the other two Eurocurrency portfolios described above are nearly identical to those in Table III and are not reported. While the currency portfolios are jointly significant at any reasonable level, there are no meaningful differences between the single-benchmark results reported in Table I and those reported in Table III. Little evidence of benchmark inefficiency is found in either table.

Since our tests do not reject efficiency of the proposed benchmarks, and since no superior alternative has been found in other work, we will examine the performance of our sample of U.S.-based internationally diversified mutual funds first using the world index as the benchmark and then using the world index and the equally weighted Eurocurrency portfolio as a two-portfolio benchmark.[14]

III. Evidence on Mutual Fund Performance

In this section we examine the performance of a sample of U.S.-based internationally diversified mutual funds. From a list of mutual funds obtained from CDA Investment Technologies, Inc., a total of 59 funds with the description "international" were identified as of June 1988. Unfortunately, most of these 59 funds began in the mid-1980s and thus do not have a sufficiently long history of returns to allow reliable inference on their performance. Our choice of sample period is made in an attempt to trade off the number of funds available against the length of time used to evaluate them. In choosing to examine the funds from January 1982 to June 1988, we are able to obtain a reasonably long sample of monthly returns and are able to analyze fifteen funds. This sample period has the advantage of including substantial periods of both appreciation and depreciation of the U.S. dollar. Unfortunately, it has the disadvantage of consisting primarily of a period of rising equity prices.

In all studies of this type, selection bias is a potential problem. For the group of funds analyzed here, however, it is unlikely to be a serious one. As the discussion above suggests, in recent years there has been strong growth in international equity funds, and market exit is unlikely to have been significant.

Table IV contains the Jensen measures of the performance of the fifteen funds obtained with the Morgan Stanley world index used as the benchmark. As Lehmann and Modest (1987) point out, since the composition of the managed funds may be changing over time as a result of the private information of the fund manager, the returns on these funds may not be homoscedastic even though the returns on individual securities are assumed to have a constant distribution.[15]

[14] Our failure to detect statistically significant evidence of inefficiency of the proposed benchmark may simply reflect data limitations. It is possible that inefficiency could be detected if we had return data for individual equities. Korajczyk and Viallet (1989) and Cho, Eun, and Senbet (1986) have found some anomalous pricing evidence in individual stock data using single- and multi-factor benchmarks.

[15] Lagrange multiplier tests of homoscedasticity suggested by White (1980) are carried out but not reported. These tests provide strong evidence against the null hypothesis of homoscedasticity.

Table II

Jensen and Positive Period Weighting Tests of the Efficiency of the Capital International World Index Using Portfolios of Capital International Country Indexes

The sample period is January 1982 to June 1988. r_{wt} is the return on the Capital International World Index, and the r_{jt} are the returns on the portfolios formed by sorting the returns on the Capital International country indexes according to the criteria listed. The w_t are the marginal utilities of an investor with power utility. All returns are measured in percent per month. t-ratios are in parentheses. Marginal significance levels are in brackets. Significantly positive Jensen or positive period weighting measures indicate superior performance of the passive country indexes and therefore suggest benchmark inefficiency.

A. Jensen Tests: $r_{jt} = \alpha_j + \beta_j r_{wt} + \varepsilon_{jt}$

	α_j for Quartile				$\alpha_j = 0$ Jointly
Criterion	1	2	3	4	χ^2 (4)
Capitalization	0.168	−0.208	−0.139	0.421	2.556
	(1.062)	(−0.468)	(−0.419)	(0.703)	[0.633]
GDP	0.229	0.189	−0.171	0.006	1.264
	(0.958)	(0.417)	(−0.448)	(0.009)	[0.867]
Turnover	0.168	−0.051	0.096	0.059	1.166
	(1.062)	(−0.130)	(0.148)	(0.131)	[0.884]
Liquidity	0.229	−0.109	0.173	−0.021	1.332
	(0.958)	(−0.294)	(0.433)	(−0.032)	[0.856]
Concentration	0.401	0.142	−0.094	−0.222	2.586
	(1.056)	(0.227)	(−0.229)	(−0.765)	[0.629]
Dividend Yield	0.476	0.096	−0.258	−0.032	3.022
	(0.837)	(0.212)	(−0.849)	(−0.082)	[0.554]

Table II—*Continued*

B. Positive Period Weighting Tests: $PPW_j = \sum w_t r_{jt}$

Criterion	PPW_j for Quartile 1	2	3	4	$PPW_j = 0$ Jointly χ^2 (4)
Capitalization	0.168	−0.377	−0.247	0.274	2.943
	(1.074)	(−0.854)	(−0.750)	(0.461)	[0.567]
GDP	0.223	0.176	−0.302	−0.260	1.076
	(0.942)	(0.391)	(−0.799)	(−0.422)	[0.898]
Turnover	0.168	−0.092	−0.200	−0.044	1.328
	(1.074)	(−0.237)	(−0.312)	(−0.098)	[0.857]
Liquidity	0.223	−0.233	0.167	−0.325	1.647
	(0.942)	(−0.629)	(0.422)	(−0.491)	[0.800]
Concentration	0.353	−0.272	−0.256	−0.248	3.057
	(0.938)	(−0.044)	(−0.632)	(−0.861)	[0.548]
Dividend Yield	0.316	0.074	−0.317	−0.083	2.978
	(0.560)	(0.165)	(−1.049)	(−0.212)	[0.562]

Table III

Jensen and Positive Period Weighting Tests of the Efficiency of a Two-Portfolio Benchmark Formed with the Capital International World Index and an Equally Weighted Portfolio of Eurocurrency Deposits

The sample period is January 1982 to June 1988. r_{wt} is the return on the Capital International World Index, and r_{et} is the return on an equally weighted portfolio of Eurocurrency deposits. The r_{jt} are the returns on the Capital International country indexes. The w_t are the marginal utilities of an investor with power utility. All returns are measured in percent per month. t-Ratios are in parentheses. Significantly positive Jensen or positive period weighting measures indicate superior performance of the passive country indexes and therefore suggest benchmark inefficiency.

	Jensen Test ($r_{jt} = \alpha_j + \beta_j r_{wt} + \beta_{jc} r_{et} + \varepsilon_{jt}$)	Positive Period Weighting Test ($PPW_j = \sum w_t r_{jt}$)
Country Index	α_j	PPW_j
Australia	−0.120	−0.454
	(0.124)	(−0.476)
Belgium	1.047	0.982
	(1.810)	(1.725)
Canada	−0.715	−0.725
	(−1.507)	(−1.553)
France	0.347	0.298
	(0.545)	(0.475)
Germany	0.210	0.077
	(0.308)	(0.116)
Hong Kong	0.030	−0.330
	(0.025)	(−0.277)
Italy	0.182	0.162
	(0.222)	(0.201)
Japan	0.772	0.818
	(1.402)	(1.508)
Netherlands	0.379	0.314
	(0.812)	(0.683)
Singapore	−1.056	−1.277
	(−1.253)	(−1.539)
Switzerland	−0.231	−0.333
	(−0.502)	(−0.734)
United Kingdom	0.026	−0.071
	(0.050)	(−0.137)
United States	−0.298	−0.281
	(−1.104)	(−1.058)
Joint Test	χ^2 (13) = 12.775	χ^2 (13) = 13.009
[Probability Value]	[0.465]	[0.447]

We therefore report t-ratios computed under the assumption that returns are heteroscedastic as well as under the assumption that they are homoscedastic. The heteroscedastic t-ratios are computed using the covariance matrix estimator proposed by White (1980). The Jensen measures provide no evidence that the managed funds outperform the Morgan Stanley world index. Only three of fifteen Jensen measures are positive, and none statistically significantly so. The largest point estimate is only .2 percent per month. While the majority of point estimates

Table IV

Jensen and Positive Period Weighting Tests of Mutual Fund Performance with the Capital International World Index as the Benchmark

The sample period is January 1982 to June 1988. r_{wt} is the return on the Capital International World Index, and the r_{jt} are the returns on the mutual funds. In computing the PPW_j, the w_t are the marginal utilities of an investor with power utility. All returns are measured in percent per month. Homoscedastic t-ratios are in parentheses. Heteroscedastic t-ratios are in brackets. The joint test statistic is distributed as chi-squared (15) and tests the hypothesis that all measures are jointly equal to zero. Significantly positive Jensen or positive period weighting measures indicate superior performance of the managed funds.

	Jensen Test ($r_{jt} = \alpha_j + \beta_j r_{wt} + \varepsilon_{jt}$)	Positive Period Weighting Test ($PPW_j = \sum w_t r_{jt}$)
Mutual Fund	α_j	PPW_j
Alliance	−0.191	−0.338
	(−0.476)	(−0.850)
	[−0.476]	[−0.746]
GT Pacific	−0.281	−0.425
	(−0.536)	(−0.818)
	[−0.531]	[−0.763]
Kemper	−0.182	−0.344
	(−0.471)	(−0.895)
	[−0.422]	[−0.682]
Keystone	−0.389	−0.436
	(−1.420)	(−1.603)
	[−1.491]	[−1.618]
Merrill	0.042	−0.108
	(−0.088)	(−0.229)
	[−0.084]	[−0.196]
Oppenheimer	−0.486	−0.606
	(−1.240)	(−1.558)
	[−1.159]	[−1.297]
Price	−0.083	−0.184
	(−0.259)	(−0.580)
	[−0.265]	[−0.535]
Putnam	−0.030	−0.108
	(−0.110)	(−0.402)
	[−0.116]	[−0.390]
Scudder	−0.150	−0.290
	(−0.453)	(−0.886)
	[−0.413]	[−0.679]
Sogen	0.198	0.105
	(0.849)	(0.455)
	[0.685]	[0.306]
Templeton Global	−0.054	−0.212
	(−0.152)	(−0.595)
	[−0.124]	[−0.401]
Templeton World	−0.093	−0.175
	(−0.300)	(−0.570)
	[−0.284]	[−0.480]

Table IV—*Continued*

Mutual Fund	Jensen Test ($r_{jt} = \alpha_j + \beta_j r_{wt} + \varepsilon_{jt}$) α_j	Positive Period Weighting Test ($PPW_j = \sum w_t r_{jt}$) PPW_j
Transatlantic	−0.543	−0.652
	(−1.378)	(−1.669)
	[−1.352]	[−1.528]
United	0.095	0.016
	(0.419)	(0.073)
	[0.377]	[0.058]
Vanguard	0.281	0.211
	(0.756)	(0.574)
	[0.826]	[0.608]
Joint Test	χ^2 (15) = 24.659	χ^2 (15) = 25.946
[Probability Value]	[0.055]	[0.039]

are negative, none are measured sufficiently precisely to reject the null hypothesis that they are zero. While the hypothesis that the Jensen measures are jointly zero is rejected at significance levels just above five percent, this rejection is probably due to the preponderance of negative estimates.[16]

Table V contains the Jensen measures of performance for the fifteen funds obtained with the two-portfolio benchmark comprised of the Morgan Stanley world index and the equally weighted portfolio of Eurocurrency deposits. Again, the Jensen measures are mostly negative, and in no case is there evidence of statistically significant superior performance.[17]

Why are so many of the Jensen measures negative? Since error in measuring the benchmark return biases the Jensen measures upward, measurement error cannot explain the results. Could it be due to the use of an inefficient benchmark? Dybvig and Ross (1985b) show that, with a riskless rate and in the absence of any timing information on the part of the manager, when an inefficient benchmark is used, any efficient portfolio (or any portfolio efficient relative to the benchmark) will have a positive Jensen measure. Thus, while positive Jensen measures are not reliable evidence of superior performance, negative Jensen measures are reliable evidence that the manager of the fund does not provide superior performance. The results cannot be explained by the possibility of low power of the tests of benchmark efficiency carried out in the previous section.

Another possibility is that the negative Jensen measures arise due to market timing ability on the part of the fund managers. As discussed in Section I, unlike the Jensen measure, the positive period weighting measure proposed by Grinblatt and Titman (1989b) will provide valid inference when the fund manager is a market timer.

Tables IV and V also contain the positive period weighting measures for the

[16] Eun, Kolodny, and Resnick (1988) have independently examined the performance of a somewhat different sample of internationally diversified mutual funds.

[17] The similarity of the results in Table V with those in Table IV is not due to a lack of statistical significance of the currency portfolio. A test that the coefficients on the currency portfolios are jointly zero yields a χ^2 (13) statistic of 67.610.

Table V

Jensen and Positive Period Weighting Tests of Mutual Fund Performance with the Benchmark Formed by the Capital International World Index and an Equally Weighted Portfolio of Eurocurrency Deposits

The sample period is January 1982 to June 1988. r_{wt} is the return on the Capital International World Index, r_{et} is the return on the equally weighted portfolio of Eurocurrency deposits, and the r_{jt} are the returns on the mutual funds. In computing the PPW_j, the w_t are the marginal utilities of an investor with power utility. All returns are measured in percent per month. Homoscedastic t-ratios are in parentheses. Heteroscedastic t-ratios are in brackets. The joint test statistic is distributed as chi-squared (15) and tests the hypothesis that all measures are jointly equal to zero. Significantly positive Jensen or positive period weighting measures indicate superior performance of the managed funds.

	Jensen Test ($r_{jt} = \alpha_j + \beta_j r_{wt} + \gamma_j r_{et}$)	Positive Period Weighting Test ($PPW_j = \sum w_t r_{jt}$)
Mutual Fund	α_j	PPW_j
Alliance	−0.192	−0.342
	(−0.480)	(−0.871)
	[−0.465]	[−0.742]
GT Pacific	−0.282	−0.428
	(−0.536)	(−0.827)
	[−0.527]	[−0.765]
Kemper	−0.184	−0.354
	(−0.486)	(−0.951)
	[−0.405]	[−0.674]
Keystone	−0.390	−0.455
	(−1.465)	(−1.737)
	[−1.500]	[−1.649]
Merrill	0.040	−0.119
	(0.085)	(−0.259)
	[0.076]	[−0.207]
Oppenheimer	−0.486	−0.586
	(−1.232)	(−1.511)
	[−1.167]	[−1.309]
Price	−0.085	−0.204
	(−0.277)	(−0.677)
	[−0.259]	[−0.552]
Putnam	−0.030	−0.109
	(−0.112)	(−0.410)
	[−0.116]	[−0.390]
Scudder	−0.151	−0.397
	(−0.468)	(−0.934)
	[−0.394]	[−0.664]
Sogen	0.199	0.136
	(0.869)	(0.600)
	[0.750]	[0.460]
Templeton Global	−0.050	−0.134
	(−0.164)	(−0.444)
	[−0.154]	[−0.372]
Templeton World	−0.090	−0.119
	(−0.332)	(−0.447)
	[−0.350]	[−0.453]

Table V—*Continued*

	Jensen Test ($r_{jt} = \alpha_j + \beta_j r_{wt} + \gamma_j r_{et}$)	Positive Period Weighting Test ($PPW_j = \sum w_t r_{jt}$)
Mutual Fund	α_j	PPW_j
Transatlantic	−0.544	−0.681
	(−1.446)	(−1.834)
	[−1.320]	[−1.496]
United	0.094	0.023
	(0.416)	(0.103)
	[0.374]	[0.083]
Vanguard	0.278	0.176
	(0.782)	(0.504)
	[0.834]	[0.494]
Joint Test	χ^2 (15) = 24.659	χ^2 (15) = 25.655
[Probability Value]	[0.055]	[0.042]

fifteen funds, along with t-ratios computed under the assumptions of homoscedastic as well as heteroscedastic returns. The positive period weighting measures are remarkably similar to the Jensen measures, both in magnitude and in statistical significance.[18] In no instance does a fund with a negative Jensen measure have a positive period weighting measure of the opposite sign, suggesting that the negative Jensen measures are not due to market timing ability on the part of the fund managers. As is the case with the Jensen measures, the positive period weighting measures are measured imprecisely, and none are significantly different from zero. The relatively large standard errors encountered with both measures may indicate that both have relatively low power. Fortunately, the imprecision does not present problems for detecting superior performance in this case since nearly all of the point estimates are negative.

In October 1987 all major world markets declined substantially. There was, however, significant variation in the declines experienced by individual markets.[19] If mutual funds merely provide investors with diversification services, then their performance during this month should mimic that of the overall market. Conversely, if portfolio managers have superior information and market timing ability, then this should be reflected at this time as well. One approach to measuring the impact of the 1987 crash on mutual fund performance is to include an October 1987 dummy variable in a regression of mutual fund returns on the world market return. This dummy variable will then capture the response of each fund to the decline in the value of the world market. Similar regressions using the returns on the country indexes summarize the relative performance of the national market indexes in October 1987.

Table VI contains the results of regressions of the country index returns on a constant, the dummy variable, and the return on the world index. While several

[18] Since the weights are scaled to sum to one, the positive period weight measures, like the Jensen measures, are expressed in percentage per month. The similarity is not surprising given the correlation between the weights used in the two performance measures.

[19] Roll (1988) documents the effect of the October 1987 "crash" on individual markets and investigates the ability of institutional features to explain the cross-sectional variation.

Table VI

Jensen Tests of the Efficiency of the Capital International World Index with October 1987 Dummies

The sample period is January 1982 to June 1988. r_{wt} is the return on the Capital International World Index, D_t is a dummy variable equal to one in October 1987 and zero otherwise, and the r_{jt} are the returns on the Capital International country indexes. All returns are measured in percent per month. t-Ratios are in parentheses. The joint test statistic is distributed as chi-squared (15) and tests the hypothesis that all measures are jointly equal to zero. A significantly positive Jensen measure indicates superior performance of the passive country indexes after excluding October 1987 from the sample and therefore suggests benchmark inefficiency. The performance of the country index relative to the Capital International World Index in October 1987 is measured by δ_j.

	$r_{jt} = \alpha_j + \beta_j r_{wt} + \delta_j D_t$	
Country Index	α_j	δ_j
Australia	0.910	−36.157
	(1.002)	(−4.179)
Belgium	1.143	−3.259
	(1.829)	(−0.547)
Canada	−0.598	−4.153
	(−1.209)	(−0.882)
France	0.374	−0.786
	(0.541)	(−0.119)
Germany	0.379	−5.849
	(0.526)	(−0.851)
Hong Kong	1.170	−40.010
	(0.998)	(−3.591)
Italy	0.087	3.459
	(0.100)	(0.416)
Japan	0.379	13.974
	(0.641)	(2.483)
Netherlands	0.450	−2.449
	(0.924)	(−0.527)
Singapore	−0.166	−31.383
	(−0.205)	(−4.065)
Switzerland	0.084	−5.041
	(0.166)	(−1.043)
United Kingdom	0.219	−6.666
	(0.395)	(−1.260)
United States	−0.136	−5.845
	(−0.401)	(−1.807)
Joint Tests	χ^2 (15) $\alpha_j = 0$, 12.622,	$\delta_j = 0$, 42.987
	[0.477]	[0.000]

small markets performed significantly worse than the world index, only one market, Japan, significantly outperformed the world index in October 1987.

Table VII presents Jensen measures of fund performance for the fifteen mutual funds obtained with the world index as the benchmark, along with the coefficients on the dummy variable. In fourteen of fifteen cases, the funds performed significantly worse than the world index during the month of October 1987. The negative coefficients on the dummy variables indicate that most of the funds under investigation held a smaller proportion of Japanese stocks than were in the world index at the time of the crash. When the dummy variable is included, only three of the intercepts are negative, and one is significantly positive.[20] Taken together, the results suggest that active management of the funds does not yield superior performance, but much of the poor performance is associated with relatively large holdings of assets that did relatively poorly in October 1987.

IV. Further Evidence on Mutual Fund Performance

In this section we examine two additional pieces of evidence on the returns on our sample of internationally diversified mutual funds. First, we consider the effect of using the U.S. national market index as the benchmark when carrying out the tests of performance. Next, we carry out tests of market timing suggested by Treynor and Mazuy (1966).

How do the funds perform relative to the U.S. national market index? Table VIII contains the results of the Jensen and positive period weighting measures obtained using the U.S. index as a benchmark. The picture that emerges is substantially different from the picture that emerges from Tables IV and V. Both measures are positive for all of the funds, with three funds showing statistically significant superior performance at the ten percent level using both measures. Two more funds have t-ratios exceeding 1.4 with the Jensen measure.

While one may carry out the tests of performance using the U.S. index as a benchmark, it is difficult to evaluate the results since the U.S. index is unlikely to be mean-variance efficient. Dybvig and Ross (1985b) show that arbitrarily inefficient portfolios may yield positive Jensen measures if an inefficient benchmark portfolio is used. To test the efficiency of the U.S. index as a benchmark, regressions of country index returns on the return on the U.S. index are estimated (but not reported). Ten of the twelve intercepts are positive, with three statistically significantly so at the ten percent level. Two more intercepts have t-ratios exceeding 1.4. The intercepts are jointly significantly different from zero at the ten percent level. The evidence clearly rejects the hypothesis that the U.S. benchmark is efficient. The results in Table VI cannot be used to infer superior performance by the managers of the funds since they incorrectly ascribe to the fund managers some of the benefits derived from holding a passive and efficient internationally diversified portfolio such as the world index.[21]

Treynor and Mazuy (1966) propose a test for market timing ability based on the coefficient β_2 from the regression:

$$r_{pt} = \alpha + \beta_1 r_{et} + \beta_2 r_{et}^2 + \varepsilon_{pt}.$$

The intuition behind the test is as follows. A fund manager with market timing ability is likely on average to increase the equity portion of the managed portfolio

[20] The hypothesis that the intercepts are jointly zero is rejected at the five percent level. Since three intercepts remain negative, this rejection does not necessarily indicate superior performance.

[21] Logue and Rogalski (1979) discuss the pitfalls of using a national index to evaluate the performance of internationally diversified funds and provide intuition on why it is misleading to interpret the results in Table VIII as evidence of superior performance.

Table VII

Jensen Tests of Mutual Fund Performance with the Capital International World Index as the Benchmark and October 1987 Dummies

The sample period is January 1982 to June 1988. r_{wt} is the return on the Capital International World Index, D_t is a dummy variable equal to one in October 1987 and zero otherwise, and the r_{jt} are the returns on the mutual funds. All returns are measured in percent per month. Homoscedastic t-ratios are in parentheses. Heteroscedastic t-ratios are in brackets. The joint test statistic is distributed as chi-squared (15) and tests the hypothesis that all measures are jointly equal to zero. A significantly positive Jensen measure indicates superior performance of the managed fund after excluding October 1987 from the sample. The performance of the fund relative to the Capital International World Index in October 1987 is measured by δ_j.

	$r_{jt} = \alpha_j + \beta_j r_{wt} + \delta_j D_t$	
Mutual Fund	α_j	δ_j
Alliance	0.123	−10.991
	(0.308)	(−2.902)
	[0.345]	[−4.492]
GT Pacific	0.005	−10.017
	(0.009)	(−1.976)
	[0.009]	[−3.224]
Kemper	0.208	−13.689
	(0.563)	(−3.889)
	[0.585]	[−6.119]
Keystone	−0.280	−3.831
	(−0.990)	(−1.422)
	[−1.063]	[−2.634]
Merrill	0.407	−12.811
	(0.862)	(−2.847)
	[0.902]	[−5.715]
Oppenheimer	−0.164	−11.308
	(−0.424)	(−3.070)
	[−0.452]	[−6.286]
Price	0.132	−7.540
	(0.410)	(−2.458)
	[0.454]	[−4.084]
Putnam	0.129	−5.555
	(0.467)	(−2.116)
	[0.536]	[−3.848]
Scudder	0.192	−11.981
	(0.611)	(−4.008)
	(0.694)	[−7.370]
Sogen	0.484	−10.040
	(2.285)	(−4.969)
	[2.507]	[−11.338]
Templeton Global	0.392	−15.663
	(1.211)	(−5.077)
	[1.363]	[−11.140]
Templeton World	0.157	−8.780
	(0.514)	(−3.011)
	[0.578]	[−6.986]
Transatlantic	−0.308	−8.235
	(−0.771)	(−2.162)
	[−0.770]	[−3.643]

Table VII—*Continued*

	$r_{jt} = \alpha_j + \beta_j r_{wt} + \delta_j D_t$	
Mutual Fund	α_j	δ_j
United	0.297	−7.101
	(1.351)	(−3.390)
	[1.365]	[−5.921]
Vanguard	0.387	−3.743
	(1.005)	(−1.018)
	[1.087]	[−1.569]
Joint Tests	χ^2 (15) $\alpha_i = 0$, 26.446,	$\delta_i = 0$, 153.170
	[0.034]	[0.000]

Table VIII

Jensen and Positive Period Weighting Tests of Mutual Fund Performance with the U.S. Index as the Benchmark

The sample period is January 1982 to June 1988. r_{ust} is the return on the Capital International U.S. Index, and the r_{jt} are the returns on the mutual funds. In computing the PPW_j, the w_t are the marginal utilities of an investor with power utility. All returns are measured in percent per month. Homoscedastic t-ratios are in parentheses. Heteroscedastic t-ratios are in brackets. The joint test statistic is distributed as chi-squared (15) and tests the hypothesis that all measures are jointly equal to zero. Significantly positive Jensen or positive period weighting measures indicate superior performance of the managed funds relative to the U.S. market index.

	Jensen Test ($r_{jt} = \alpha_j + \beta_j rus_t$)	Positive Period Weighting Test ($PPW_j = \sum w_t r_{jt}$)
Mutual Fund	α_j	PPW_j
Alliance	0.465	0.395
	(0.901)	(0.773)
	[0.852]	[0.690]
GT Pacific	0.445	0.369
	(0.694)	(0.583)
	[0.644]	[0.514]
Kemper	0.486	0.415
	(0.940)	(0.813)
	[0.831]	[0.667]
Keystone	0.294	0.267
	(0.654)	(0.601)
	[0.619]	[0.547]
Merrill	1.007	0.926
	(1.456)	(1.356)
	[1.310]	[1.147]
Oppenheimer	0.132	0.087
	(0.284)	(0.189)
	[0.269]	[0.170]
Price	0.611	0.556
	(1.269)	(1.169)
	[1.189]	[1.040]
Putnam	0.513	0.479
	(1.423)	(1.342)
	[1.421]	[1.292]

Table VIII—*Continued*

	Jensen Test ($r_{jt} = \alpha_j + \beta_j rus_t$)	Positive Period Weighting Test ($PPW_j = \sum w_t r_{jt}$)
Mutual Fund	α_j	PPW_j
Scudder	0.499	0.431
	(1.070)	(0.935)
	[0.971]	[0.791]
Sogen	0.536	0.510
	(2.285)	(2.201)
	[1.938]	[1.737]
Templeton Global	0.279	0.236
	(1.026)	(0.879)
	[0.878]	[0.692]
Templeton World	0.203	0.192
	(1.080)	(1.034)
	[1.001]	[0.920]
Transatlantic	0.216	0.161
	(0.388)	(0.293)
	[0.360]	[0.259]
United	0.578	0.547
	(1.785)	(1.710)
	[1.652]	[1.498]
Vanguard	0.975	0.934
	(1.887)	(1.830)
	[1.854]	[1.738]
Joint Test	χ^2 (15) = 26.456	χ^2 (15) = 26.277
[Probability Value]	[0.033]	[0.035]

when equity returns are high and reduce it when equity returns are low. One would therefore expect to find $\beta_2 > 0$ if the manager possesses and utilizes superior information on the returns on market indexes. A formal treatment of the test is found in Pfleiderer and Bhattacharya (1983) and Admati, Bhattacharya, Pfleiderer, and Ross (1986), where it is shown that a positive β_2 indicates market timing ability.[22] Table IX reports the Treynor-Mazuy timing measures for our sample of fifteen funds, along with both homoscedastic and heteroscadistic t-ratios. All 15 funds have estimates of β_2 that are significantly negative at the five percent level. The apparently perverse timing ability suggested by these results is consistent with the earlier findings of Chang and Lewellen (1984), Henriksson (1984), Lehmann and Modest (1988), Connor and Korajczyk (1988), and Grinblatt and Titman (1989a). Jagannathan and Korajczyk (1986) and Connor and Korajczyk (1988) offer explanations for apparent perverse timing involving possible options-like characteristics of fund returns. Lehmann and Modest (1988) suggest that the negative measures may partially be illusory, a product of the small sample distribution of returns. Following their suggestion, we present estimates of t-ratios obtained using Effron's (1982) bootstrap technique. These t-ratios are obtained with 2,000 artificial samples drawn by sampling

[22] The estimates of β_2 cannot be used to rank the quality of the market timing information of the managers since it combines both the quality of the manager's information and the manager's response to that information.

Table IX

Treynor-Mazuy Tests of Market Timing Ability with the Capital International World Index as the Benchmark

The sample period is January 1982 to June 1988. The r_{wt} is the return on the Capital International World Index, and r_{wt} is its square. The r_{jt} are the returns on the funds. The bootstrap t-ratios are based on 2,000 replications. A significantly positive β_2 suggests that the fund managers exhibit market timing ability.

$$r_{jt} = \alpha_j + \beta_{1j} r_{wt} + \beta_{2j} r_{wt}^2$$

Mutual Fund	β_2	Homoscedastic t-Ratio	Heteroscedastic t-Ratio	Bootstrap t-Ratio
Alliance	−3.877	−4.635	−3.944	−1.916
GT Pacific	−3.960	−3.434	−3.472	−1.552
Kemper	−4.151	−5.314	−6.547	−3.114
Keystone	−1.301	−2.062	−2.383	−1.230
Merrill	−3.710	−3.567	−5.356	−2.575
Oppenheimer	−2.983	−3.460	−5.248	−2.690
Price	−2.743	−3.987	−3.980	−1.856
Putnam	−2.101	−5.536	−3.805	−1.683
Scudder	−3.585	−5.401	−7.441	−3.427
Sogen	−1.961	−3.884	−3.457	−1.379
Templeton Global	−3.483	−4.656	−5.665	−2.220
Templeton World	−1.768	−2.511	−3.344	−1.391
Transatlantic	−3.010	−3.478	−4.112	−1.750
United	−1.938	−3.994	−6.066	−2.688
Vanguard	−2.174	−2.581	−2.034	−0.958

with replacement from the existing data and dividing the estimate of β_2 reported in Table IX by the standard deviation of the β_2 estimates across the 2,000 artificial samples. Using the bootstrap t-ratios, only five of the fifteen funds have timing measures that are significantly negative at the five percent level. The results are to some extent consistent with Lehmann and Modest's finding that negative timing measures are largely illusory. However, ten of the fifteen measures are significant at the ten percent level. Table X contains the results of the Treynor-Mazuy tests when October 1987 is excluded from the sample. Thirteen of fifteen estimates are negative, with nine statistically significantly negative at the five percent level when asymptotic t-ratios are used. When bootstrap t-ratios are used, only five of fifteen are significantly negative at the five percent level. The evidence of perverse timing is thus reduced substantially when October 1987 is excluded and when bootstrap t-ratios are used. On the other hand, the preponderance of negative estimates remains a puzzle.

V. Concluding Remarks

In this paper, we examine the performance of a sample of fifteen U.S.-based internationally diversified mutual funds between 1982 and 1988. Two performance measures are used, the Jensen (1968, 1969) measure and the positive period weighting measure proposed by Grinblatt and Titman (1989b). We find no evidence that the funds, either individually or as a whole, provide investors with

Table X

Treynor-Mazuy Tests of Market Timing Ability with the Capital International World Index as the Benchmark and Excluding October 1987

The sample period is January 1982 to June 1988. The r_{wt} is the return on the Capital International World Index, and r_{wt} is its square. The r_{jt} are the returns on the funds. The bootstrap t-ratios are based on 2,000 replications. A significantly positive β_2 suggests that the fund managers exhibit market timing ability after excluding October 1987.

$$r_{pt} = \alpha + \beta_1 r_{et} + \beta_2 r_{et}^2 + \varepsilon_{pt}$$

Mutual Fund	β_2	Homoscedastic t-Ratio	Heteroscedastic t-Ratio	Bootstrap t-Ratio
Alliance	−6.721	−3.989	−3.427	−3.132
GT Pacific	−7.918	−3.411	−4.023	−3.345
Kemper	−5.520	−3.445	−3.890	−3.366
Keystone	−2.129	−1.640	−1.633	−1.356
Merrill	−4.433	−2.066	−2.581	−2.165
Oppenheimer	−2.684	−1.509	−1.848	−1.571
Price	−4.962	−3.571	−3.941	−3.408
Putnam	−3.993	−3.325	−3.909	−1.618
Scudder	−4.629	−3.395	−3.971	−1.526
Sogen	0.510	0.515	0.658	0.479
Templeton Global	−0.990	−0.656	−1.095	−0.883
Templeton World	0.220	0.154	0.242	0.234
Transatlantic	−5.478	−3.119	−3.956	−1.372
United	−1.956	−1.953	−1.929	−0.986
Vanguard	−5.880	−3.525	−3.263	−1.492

performance that surpasses that of a broad, international equity index over this sample period.

We also examine the behavior of the funds during October 1987 and find that the funds uniformly underperformed the world index during that month. Examination of the returns on individual national market indexes suggests that the behavior of fund returns in October 1987 may be due to the funds underinvesting in the Japanese market. When October 1987 is omitted from the sample, the performance of the funds improves. Twelve of fifteen regression intercepts are positive. Taken together, the results suggest that active management of the funds does not yield superior performance, but much of the poor performance is associated with relatively large holdings of assets that did relatively poorly in October 1987.

We also examine the performance of the funds relative to the Morgan Stanley index for the United States and find some evidence that the funds outperform the U.S. index. When these results are compared to our other results, it seems clear that the performance relative to the U.S. index is most likely due to the benefits of international diversification rather than superior performance by fund managers. Finally, we carry out the tests of market timing ability suggested by Treynor and Mazuy (1966). We find evidence of apparent perverse market timing by fund managers. This evidence is considerably weaker when bootstrap t-ratios are used to perform hypothesis tests than when asymptotic t-ratios are used and when October 1987 is excluded from the sample.

REFERENCES

Adler, Michael and Bernard Dumas, 1983, International portfolio selection and corporation finance: A synthesis, *Journal of Finance* 38, 925–984.

Admati, Anat R. and Stephen A. Ross, 1985, Measuring investment performance in a rational expectations model, *Journal of Business* 58, 1–26.

——— Sudipto Bhattacharya, Paul Pfleiderer, and Stephen A. Ross, 1986, On timing and selectivity, *Journal of Finance* 41, 715–730.

Banz, Rolf W., 1981, The relationship between return and market value of common stock, *Journal of Financial Economics* 9, 3–18.

Chang, Eric C. and Wilbur C. Lewellen, 1984, Market timing and mutual fund performance, *Journal of Business* 57, 57–72.

Cho, D. C., Cheol S. Eun, and Lemma W. Senbet, 1986, International arbitrage pricing theory: An empirical investigation, *Journal of Finance* 41, 313–329.

Connor, Gregory and Robert A. Korajczyk, 1988, The attributes, behavior and performance of U.S. mutual funds, Working Paper, University of California, Berkeley.

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

——— and Stephen A. Ross, 1985b, The analytics of performance measurement using a security market line, *Journal of Finance* 40, 401–416.

Effron, Bradley, 1982, *The Jackknife, the Bootstrap, and Other Resampling Plans* (Society for Industrial and Applied Mathematics, Philadelphia, PA).

Eun, Cheol S., Richard Kolodny, and Bruce G. Resnick, 1988, U.S. based international mutual funds: A performance evaluation, Working Paper, University of Maryland.

Grinblatt, Mark and Sheridan Titman, 1989a, A comparison of measures of mutual fund performance on a sample of monthly mutual fund returns, *Journal of Business* 62, 383–416.

——— and Sheridan Titman, 1989b, Portfolio performance evaluation: Old issues and new insights, *Review of Financial Studies* 2, 393–421.

Gultekin, Mustafa N. and N. Bulent Gultekin, 1983, Stock market seasonality: International evidence, *Journal of Financial Economics* 12, 469–482.

Hawawini, Gabriel, 1988, Market efficiency and equity pricing: International evidence and implications for global investing, in Paul H. Aron, ed.: *Global Investing: A Guide to International Security Analysis and Portfolio Management* (John Wiley and Sons, New York).

Henriksson, Roy D., 1984, Market timing and mutual fund performance: An empirical investigation, *Journal of Business* 57, 73–96.

Hodrick, Robert J., 1981, International asset pricing with time varying risk premia, *Journal of International Economics* 11, 573–587.

Jagannathan, Ravi and Robert A. Korajczyk, 1986, Assessing the market timing performance of managed portfolios, *Journal of Business* 57, 217–235.

Jensen, Michael C., 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 Giorgio P. Szego and Karl Shell, eds.: *Mathematical Methods in Investment and Finance* (North Holland, Amsterdam).

Jorion, Philippe, 1985, International diversification with estimation risk, *Journal of Business* 58, 259–278.

Keim, Donald B., 1983, Size-related anomalies and stock return seasonality: Further empirical evidence, *Journal of Financial Economics* 12, 13–32.

Korajczyk, Robert A. and Claude J. Viallet, 1989, An empirical investigation of international asset pricing, Working Paper, Kellogg Graduate School of Management, Northwestern University.

Lehmann, Bruce N. and David M. Modest, 1987, Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons, *Journal of Finance* 42, 233–265.

——— and David M. Modest, 1988, Empirical issues in the measurement of market timing ability, Working Paper, University of California, Berkeley.

Litzenberger, Robert H. and Krishna Ramaswamy, 1982, The effects of dividends on common stock prices: Tax effects or information effects?, *Journal of Finance* 37, 429–443.

Logue, Dennis E. and Richard J. Rogalski, 1979, Offshore alphas: Should diversification begin at home?, *Journal of Portfolio Management* 5, 5–10.

Pfleiderer, Paul and Sudipto Bhattacharya, 1983, A note on performance evaluation, Working Paper, Stanford University.

Roll, Richard, 1977, A critique of the asset pricing theory's tests, Part I: On past and potential testability of the theory, *Journal of Financial Economics* 4, 129–176.

———, 1988, The international crash of October 1987, *Financial Analysts Journal* 44, 19–35.

Ross, Stephen A. and Michael Walsh, 1983, A simple approach to the pricing of risky assets with uncertain exchange rates, in R. Hawkins, R. Levich, and C. Wihlborg, eds.: *The Internationalization of Financial Markets and National Economic Policy* (JAI Press, Greenwich, CT).

Stulz, René M., 1984, Pricing capital assets in an international setting: An introduction, *Journal of International Business Studies* 15, 55–74.

Treynor, Jack L. and F. Mazuy, 1966, Can mutual funds outguess the market?, *Harvard Business Review* 44, 131–136.

White, Halbert, 1980, A heteroscedasticity-consistent covariance estimator and a direct test for heteroscedasticity, *Econometrica* 48, 817–838.