Roy D. Henriksson
University of California, Berkeley

Market Timing and Mutual Fund Performance: An Empirical Investigation

The evaluation of the performance of investment managers is a topic of considerable interest to practitioners and academics alike. Using both the parametric and non-parametric tests for the evaluation of forecasting ability presented by Henriksson and Merton, the market-timing ability of 116 open-end mutual funds is evaluated for the period 1968–80. The empirical results do not support the hypothesis that mutual fund managers are able to follow an investment strategy that successfully times the return on the market portfolio.

I. Introduction

The evaluation of investment performance is of importance for allocating investment funds efficiently and setting appropriate management fees. Because actively managed mutual funds are an important form of investment in the United States, a valid question is whether the active management has achieved a sufficient increase in returns to offset the associated costs of information and transactions, as well as the management fees charged. As a corollary, the ability to earn superior returns based on superior forecasting ability would be a violation of the efficient markets hypothesis[1] and would have far-reaching implications for the theory of finance.[2]

Henriksson and Merton (1981) present statistical techniques for testing forecasting ability with a particular emphasis on the market-timing ability of investment managers. The tests are derived from the basic model of market timing developed by Merton (1981), where the forecaster predicts when stocks will outperform riskless securities and when riskless securities will outperform stocks but does not predict the magnitude of the relative returns.

1. For an excellent discussion of the theory of market efficiency, see Fama (1970).
2. For a description of some of the previous work on the evaluation of investment performance, see Henriksson and Merton (1981).

(*Journal of Business*, 1984, vol. 57, no. 1, pt. 1)

0021-9398/84/5701-0006$01.50

Henriksson and Merton present both parametric and nonparametric tests of market-timing ability. The parametric tests require the assumption of either the capital asset pricing model (CAPM)[3] or a multifactor return structure. Based strictly on observable returns, the tests permit the identification of the separate contributions from market-timing ability and micro forecasting. The nonparametric tests do not require any specific structure of returns but do require knowledge of the actual forecasts or a good proxy for them.

This paper evaluates the market-timing performance of 116 open-end mutual funds using the parametric and nonparametric techniques presented by Henriksson and Merton (1981). In Section II, the statistical techniques developed by Henriksson and Merton are presented. The results of the parametric tests are presented in Section III and the results of the nonparametric tests, based on proxies for the forecasts, are presented in Section IV. In both cases, no evidence of market-timing ability is found.

II. Techniques for Testing Market-timing Ability

Merton (1981) developed a framework for evaluating market-timing ability that does not require knowledge of the distribution of returns on the market or any particular model of security valuation. It takes the simple form that the investment manager forecasts either that the stock market will provide a greater return than riskless securities—that is, $Z_M(t) > R(t)$—or that riskless securities will provide a greater return than stocks—that is, $R(t) > Z_M(t)$, where $Z_M(t)$ is the one-period return per dollar on the market portfolio and $R(t)$ is the one-period return per dollar on riskless securities. The forecaster does not attempt, or is not able, to predict by how much stocks will perform better or worse than riskless securities. Based on his forecast, the investment manager will adjust the relative proportions of the market portfolio and riskless securities that are held in the fund.

The model can be formally described in terms of the probabilities of a correct forecast, conditional on $Z_M(t) - R(t)$. Let $\gamma(t)$ be the manager's forecast variable where $\gamma(t) = 1$ if the forecast, made at time $t - 1$ for the period t, is that $Z_M(t) > R(t)$ and $\gamma(t) = 0$ if the forecast is that $Z_M(t) \leqq R(t)$. The probabilities for $\gamma(t)$ conditional on the realized return on the market are

$$\begin{aligned} p_1(t) &= \text{prob}[\gamma(t) = 0 \quad Z_M(t) \leqq R(t)] \\ 1 - p_1(t) &= \text{prob}[\gamma(t) = 1 \quad Z_M(t) \leqq R(t)] \end{aligned} \tag{1a}$$

3. Jensen (1972*a*) provides a comprehensive review of the CAPM.

and

$$
\begin{aligned}
p_2(t) &= \text{prob}[\gamma(t) = 1 \quad Z_M(t) > R(t)] \\
1 - p_2(t) &= \text{prob}[\gamma(t) = 0 \quad Z_M(t) > R(t)].
\end{aligned} \tag{1b}
$$

Therefore, $p_1(t)$ is the conditional probability of a correct forecast, given that $Z_M(t) \leqq R(t)$, and $p_2(t)$ is the conditional probability of a correct forecast, given that $Z_M(t) > R(t)$. Neither $p_1(t)$ nor $p_2(t)$ depends on the level or distribution of the return of the market. The probability of a correct forecast depends only on whether or not $Z_M(t) > R(t)$.

Merton showed that a necessary and sufficient condition for a forecaster's predictions to have no value is that $p_1(t) + p_2(t) = 1$. Under this condition, knowledge of the forecast would not cause an investor to change his prior estimate of the distribution of returns on the market portfolio and, therefore, would not pay anything for the information. The existence of forecasting ability will result in $p_1(t) + p_2(t) > 1$.[4]

A. A Nonparametric Test of Market Timing

Henriksson and Merton's (1981) nonparametric tests take advantage of the fact that the conditional probabilities of a correct forecast can be used to measure forecasting ability and yet they do not depend on the distribution of returns on the market or any particular model for security valuation. The tests examine the null hypothesis that the market timer has no forecasting ability; that is, $H_0 : p_1(t) + p_2(t) = 1$, where the conditional probabilities of a correct forecast, $p_1(t)$ and $p_2(t)$, are not known. We want to determine the probability, P, that a given outcome from our sample came from a population that satisfies our null hypothesis.

Henriksson and Merton show that the null hypothesis is defined by the hypergeometric distribution:

$$
P(n_1 | N_1, N, n) = \frac{\binom{N_1}{n_1}\binom{N_2}{n - n_1}}{\binom{N}{n}}, \tag{2}
$$

4. For a more thorough presentation of this framework for evaluation, see Merton (1981) and Henriksson and Merton (1981). These papers also discuss the potential problems of using the unconditional probability of a correct forecast to evaluate forecasting ability. The use of the unconditional probability of a correct forecast requires additional information regarding the frequency distribution of the returns, whereas this information is not required to evaluate forecasting ability based on the conditional probabilities of a correct forecast.

where $n_1 \equiv$ number of correct forecasts, given $Z_M \leqq R$; $n \equiv$ number of times forecast that $Z_M \leqq R$; $N_1 \equiv$ number of observations where $Z_M \leqq R$; $N_2 \equiv$ number of observations where $Z_M > R$; and $N \equiv N_1 + N_2 =$ total number of observations. The distribution is independent of both p_1 and p_2; thus, to test the null hypothesis it is unnecessary to estimate either of the conditional probabilities. So, provided that the forecasts are known, all the variables necessary for the test are directly observable. Given N_1, N_2, and n the distribution of n_1 is determined by (2) for the null hypothesis, where the feasible range for n_1 is given by

$$\underline{n}_1 \equiv \max(0, n - N_2) \leqq n_1 \leqq \min(N_1, n) \equiv \bar{n}_1. \tag{3}$$

Equations (2) and (3) can be used to establish confidence intervals for testing the hypothesis of no market-timing ability. For a standard two-tail test[5] with a confidence level of c, one could reject the null hypothesis if $n_1 \geqq \bar{x}(c)$ or if $n_1 \leqq \underline{x}(c)$ where $\bar{x}$ and $\underline{x}$ are determined from the solutions of[6]

$$\sum_{n_1=\bar{x}}^{\bar{n}_1} \binom{N_1}{n_1}\binom{N_2}{n-n_1} \Big/ \binom{N}{n} = \frac{(1-c)}{2} \tag{4a}$$

$$\sum_{n_1=\underline{n}_1}^{\underline{x}} \binom{N_1}{n_1}\binom{N_2}{n-n_1} \Big/ \binom{N}{n} = \frac{(1-c)}{2}. \tag{4b}$$

B. A Parametric Test of Market Timing

To use the nonparametric tests to evaluate forecasting ability, the predictions of the forecaster must be obtainable. However, such information is rarely available for mutual funds. Thus it is necessary either to use a proxy for the forecasts or to make additional assumptions with respect to the generating process of security returns. Henriksson and Merton present a parametric test of market-timing ability requiring only observable returns data based on the additional assumption that securities are priced according to the CAPM, although the tests are easily adaptable to a multifactor framework.

In pathfinding papers, Jensen (1968, 1969) used a CAPM framework to evaluate the performance of open-end mutual funds over the period

5. If the forecasts are known, Henriksson and Merton argue that if forecasters behave rationally, then a one-tail test is more appropriate, as it should never be the case that $p_1(t) + p_2(t) < 1$. In this paper, however, the forecasts are not known. As the proxy that is used will be affected by management fees and transaction costs, it is possible that $p_1(t) + p_2(t) < 1$. Therefore, a two-tail test is more appropriate.

6. Because the hypergeometric distribution is discrete, the strict equalities of (4a) and (4b) will not usually be obtainable. Therefore, in (4a), $\bar{x}$ should be interpreted as the lowest value of x for which the summation does not exceed $(1 - c)/2$. In (4b), $\underline{x}$ should be interpreted as the highest value of x for which the summation does not exceed $(1 - c)/2$.

1945–64. He found no evidence that mutual funds were able to generate superior returns. However, Jensen did not allow for the possibility that the mutual funds were undertaking market-timing strategies.

Henriksson and Merton allow for the possibility of market-timing ability. They assume that the investment manager chooses among discretely different systematic risk levels for the fund, dependent on his forecast. In our analysis we assume that the fund has two target risk levels, one for when the forecaster predicts $Z_M > R$ and one for when he predicts $Z_M \leqq R$. The technique can be extended to multiple target risk levels, and this possibility will be discussed in Section III.

The per period return on the investment manager's fund is assumed to be of the form

$$Z_P(t) = R(t) + [b + \theta(t)]x(t) + \lambda + \epsilon_p(t) \tag{5}$$

where b is the unconditional (on the forecast) expected value of $\beta(t)$; $\theta(t)$ is the unanticipated (dependent on the forecast) component of $\beta(t)$; $x(t) \equiv Z_M(t) - R(t)$; λ is the expected excess return from microforecasting; and $\epsilon_p(t)$ has the following characteristics:

$$\begin{aligned} E[\epsilon_p(t)] &= 0 \\ E[\epsilon_p(t)|x(t)] &= 0 \\ E[\epsilon_p(t)|\epsilon_p(t-i)] &= 0 \quad i = 1, 2, 3, \ldots, \end{aligned} \tag{6}$$

In this form, $\eta_1(t) = b + \theta(t)$ is the target level of systematic risk when the forecaster predicts $Z_M(t) \leqq R(t)$ and $\eta_2(t) = b + \theta(t)$ is the target level of systematic risk corresponding to a forecast of $Z_M(t) > R(t)$.

Using the returns process described in (5), least squares regression analysis can be used to estimate the separate contributions from security analysis and market timing. This regression specification is of the form

$$Z_p(t) - R(t) = \alpha_p + \beta_1 x(t) + \beta_2 y(t) + \epsilon(t) \tag{7}$$

where $y(t) \equiv \max[0, R(t) - Z_M(t)] = \max[0, -x(t)]$.

This specification comes from the analysis of the value of market-timing ability in Merton (1981). He showed that up to an additive noise term, the returns per dollar from a portfolio involved in market timing as described here are identical to those of a partial "protective put" option investment strategy,[7] where for each dollar of investment, $[p_n\eta_2 + (1 - p_2)\eta_1]$ dollars are invested in the market portfolio; $(p_1 + p_2 - 1)(\eta_2 - \eta_1)$ put options on the market portfolio are purchased with an exercise price (per dollar of the market portfolio) of $R(t)$; and the balance is invested in riskless securities. The value of the market tim-

7. For a description of the "protective put" option investment strategy, see Merton, Scholes, and Gladstein (1982).

ing is reflected in the fact that the put options are obtained for free. The variable $y(t)$ in (7) represents the return on one such option.

Henriksson and Merton show that the large sample least squares estimates of β_1, β_2, and α_p can be written as

$$\text{plim } \hat{\beta}_1 = p_2\eta_2 + (1 - p_2)\eta_1 \quad (8)$$

$$\text{plim } \hat{\beta}_2 = (p_1 + p_2 - 1)(\eta_2 - \eta_1) \quad (9)$$

$$\text{plim } \hat{\alpha}_p = \lambda. \quad (10)$$

From (8), plim $\hat{\beta}_1 = E[\beta(t)|x(t) > 0]$. The market-timing ability of the forecaster is measured by β_2, which will equal zero if either the forecaster has no ability ($p_1(t) + p_2(t) = 1$) or does not act on his forecast ($\eta_1 = \eta_2$).

In addition, (7) has the characteristic that

$$\lim_{N \to \infty} \left[\frac{\Sigma \epsilon_p(t)}{N} \right] = 0. \quad (11)$$

Thus, the coefficients from least squares estimation of (7) provide consistent estimates of the parameters of portfolio performance. However, ordinary least squares estimation is inefficient because $\beta(t)$ is not stationary. As Henriksson and Merton show, this causes the standard deviation of ϵ_p to be an increasing function of $|x(t)|$. Thus, it is necessary to correct for heteroscedasticity to improve the efficiency of the estimates.

While this parametric procedure separates the contributions of micro forecasting and macro forecasting, the analysis is dependent on the specified return-generating process. Empirical tests of the CAPM by Black, Jensen, and Scholes (1972), Blume and Friend (1973), and Fama and MacBeth (1973) seem to show that the security market line (SML) relationship

$$Z_i(t) = R(t) + \beta_i x(t) + \epsilon \quad (12)$$

does not hold for individual securities. In particular, they found evidence of a "zero-beta" effect, where the return on low-beta securities tended to be greater than predicted by (12) and the return on high-beta securities tended to be lower than predicted by (12). If this deviation from the model is the result of a second factor that is uncorrelated with the return on the market portfolio, it will affect the estimates of α_p in (7) but will not affect the estimates of β_2. As the primary focus of this study is on market-timing ability, the existence of a zero-beta factor should not affect the results. While not done here, Henriksson and Merton's methodology can easily be adapted to take this second factor into account.

However, Roll (1977) has questioned the validity of the tests of the CAPM mentioned above because they are not based on the true market portfolio. He shows how misspecification of the market portfolio can cause these results. However, Stambaugh (1982) shows that the empirical results of the tests of the CAPM do not seem to be very sensitive to the composition of the "market" portfolio.

In addition, Roll (1977, 1978) attacks the use of the SML for portfolio evaluation. Roll argues that if the market portfolio is ex post mean-variance efficient, then all securities will lie exactly on the SML, as well as any portfolio where the securities are held for the entire sample period. Mayers and Rice (1979), however, show that this will not be the case if the investment manager is allowed to change the composition of his fund.

Roll also questions the meaning of the estimates of α_p when the true market portfolio is not known. This issue is not addressed here, as the focus of this paper is on the market-timing ability of mutual funds, and the measures of forecasting ability used are not dependent on α_p. As long as the portfolio used as a proxy for the true market portfolio is highly correlated with the true market portfolio,[8] then $\hat{\beta}_2$ will be a reasonable measure of timing ability with respect to the true market.[9] Intuitively, this is because small errors in the proxy for the realized return on the market portfolio will usually not change the correctness of a forecast as the forecaster is assumed to predict only direction and not magnitude.

In addition, investment managers may not be attempting to forecast the returns on the true market portfolio. Instead, they may attempt timing with respect to the universe of securities that they tend to invest

8. The correlation will almost certainly be high, as the stocks of companies traded on the NYSE must account for a substantial portion of the total market, considering the magnitude of the dollar value of the stocks.

9. By combining the riskless asset with the true market portfolio, it is possible to form a portfolio that is perfectly correlated with the market portfolio and has the same expected return as the market proxy. The excess return on the market proxy will be $bx(t) + \epsilon(t)$ where $x(t)$ is the excess return on the true market portfolio, b is the proportion of the new portfolio that is made up of the market portfolio, and $\epsilon(t)$ is a variable with the characteristics that $E[\epsilon(t)] = 0$, $E[\epsilon(t)|x(t)] = 0$, and $E[\epsilon(t)^2] = \sigma^2$. As modeled, the correlation between the true market portfolio and the market proxy will be a decreasing function of σ. To consider the effect of misspecification, we will examine the large sample least squares estimate of β_2 which can be written as

$$\text{plim } \hat{\beta}_2 = (\rho_{py} - \rho_{px}\rho_{xy})\sigma_p/(1 - \rho_{xy}^2)\sigma_y. \quad \text{(F1)}$$

If our investment manager is only interested in forecasting with respect to the market portfolio, then the difference in the expected estimate of β_2 because a proxy is used instead of the true market portfolio will only be a function of σ^2 as the variables in (F1) will differ only by a function of σ^2. If σ^2 is small (relative to σ_x^2), then the estimation error will be small. If the investment manager also attempts to forecast with respect to $\epsilon(t)$, then β_2, as estimated, will be a combined measure of forecasting ability with respect to both $x(t)$ and $\epsilon(t)$. Of course, if it were possible to determine the values of $\epsilon(t)$, then it could be treated as a second factor and evaluated separately. In either case, if the manager is attempting to forecast $\epsilon(t)$, then this forecasting should be taken into account in the evaluation.

in. Inspection of the holdings of the mutual funds in the sample show that in most cases, this is a trade-off between equities and high-grade fixed income securities. Predominately, the equities were those traded on the New York Stock Exchange (NYSE); thus the NYSE Composite seems to be a reasonable proxy for the market portfolio. Precisely, in this paper timing is tested with respect to a portfolio replicating the securities traded on the NYSE.

III. Empirical Results: Parametric Tests

As it was not possible to obtain the actual market-timing forecasts of the mutual fund managers, it was necessary either to use a proxy for the forecast or to assume a specific return-generating process. In this section, the parametric tests described in Section II are run on the assumption that the CAPM holds.

Both the parametric and nonparametric tests examined the performance of 116 open-end mutual funds using monthly data from February 1968 to June 1980. The returns data include all dividends paid by the fund and are net of all management costs and fees. The returns data were obtained from Standard & Poor's *Over-the-Counter Daily Stock Price Record* and Wiesenberger Investment Companies Service (1975, 1980). A list of the funds in the sample, including the objective of the individual funds, is presented in the Appendix.

Monthly returns (including dividends on the NYSE Index) are used for the returns on the market portfolio. This index is a value-weighted portfolio of all stocks traded on the NYSE. The returns on Treasury bills are used for the riskless asset and are obtained from Ibbotson and Sinquefield (1979).[10] The Treasury-bill return used is the 1-month holding period return on the shortest maturity bill with at least a 30-day maturity.

Using the returns process described in (5), weighted least squares regression analysis, with a correction for heteroscedasticity, is used to obtain the separate contributions from micro forecasting and market timing. The regression specification is as shown in (7)

$$Z_p(t) - R(t) = \alpha_p + \beta_1 x(t) + \beta_2 y(t) + \epsilon(t). \tag{7}$$

The correction for heteroscedasticity in the parametric tests is of the following form. Least squares estimation is run for each fund using the regression specification found in (7). Then the absolute value of the residuals from this estimation, $|\epsilon(t)|$, are used as the dependent variable in the regression of

$$|\epsilon_i(t)| = \phi_i + \Omega_{1i} x_1(t) + \Omega_{2i} x_2(t) + \xi_i, \tag{13}$$

10. Treasury bill returns for 1979 and 1980 were calculated using data from the *Wall Street Journal*.

TABLE 1 **Parametric Tests: $Z_p(t) - R(t) = \hat{\alpha}_p + \hat{\beta}_1 x(t) + \hat{\beta}_2 y(t)$; 116 Open-End Mutual Funds, Sample Split by Time**

Sample	1968:2–1980:6	1968:2–1974:4	1974:5–1980:6
Mean (SD):			
$\hat{\alpha}$ (σ_α)	.0007 (.0041)	−.0010 (.0053)	.0022 (.0057)
$\hat{\beta}_1(\sigma_{\beta_1})$	.92 (.21)	1.01 (.27)	.86 (.20)
$\hat{\beta}_2(\sigma_{\beta_2})$	−.07 (.15)	−.02 (.21)	−.08 (.18)
Number of funds:			
Reject $\hat{\alpha}$ = 0 at 5%*	11+ 8−	6+ 13−	21+ 5−
Reject $\hat{\alpha}$ = 0 at 1%	6+ 4−	4+ 4−	10+ 1−
Reject $\hat{\beta}_2$ = 0 at 5%	3+ 9−	1+ 4−	2+ 3−
Reject $\hat{\beta}_2$ = 0 at 1%	1+ 1−	0+ 1−	1+ 1−
$\hat{\alpha} > 0$	59	32	67
$\hat{\beta}_2 > 0$	44	64	46

Note.—Only one fund had a significantly positive $\hat{\alpha}$ in both periods. Only one fund had a significantly positive $\hat{\beta}_2$ in both periods. No funds had significantly negative $\hat{\alpha}$ or $\hat{\beta}_2$ in both periods. *F*-test: number of funds that reject hypothesis that coefficients are equal in both periods are 45 at 5%, 26 at 1%.

* + Denotes number of funds with significantly positive estimates.

− Denotes number of funds with significantly negative estimates.

where $x_1 \equiv \min[0, x(t)]$ and $x_2 \equiv \max[0, x(t)]$, to estimate the degree of heteroscedasticity with respect to the realized excess return on the market. The variables in (7), including the constant, are then divided by $[\hat{\phi}_i + \hat{\Omega}_{1i}x_1(t) + \hat{\Omega}_{2i}x_2(t)]$ and the coefficients for α_p, β_1, and β_2 are reestimated.[11]

The tests are run for the entire period, February 1968–June 1980, as well as for the subperiods February 1968–April 1974 and May 1974–June 1980. The results are summarized in table 1.[12]

The results show little evidence of market-timing ability. In fact, 62% of the funds had negative estimates of market timing, as shown by $\hat{\beta}_2$. Using the assumption that the returns from the mutual funds and the market portfolio follow a joint-normal distribution, only three of the funds exhibited positive estimates for β_2 with 95% confidence.[13]

11. The specification of the heteroscedasticity correction directly follows from the specification of each period's estimation error, as shown in (26) in Henriksson and Merton (1981). While it would be technically more correct to run a number of iterations of the correction described above, there was virtually no change in the results when this was done for a number of the funds. In fact, there was little difference in the results from using ordinary least squares, without the heteroscedasticity correction, and the results from using the weighted least squares estimation described above.

12. The results for the individual funds for all of the tests run are available from the author on request.

13. The confidence interval mentioned is that which would be used to evaluate each fund in isolation. It does not take into account the fact that there are 116 funds in my sample, which are certainly not independent of one another. To test for the market-timing ability for the sample as a whole, the "seemingly unrelated regression model" of Zellner (1962) can be used to test the significance of the timing variable with respect to the entire sample. However, since the explanatory variables are the same for all of the individual funds in the sample, the point estimates of the individual coefficients will not change. It is therefore virtually certain that the results from this technique would not show significantly positive market-timing ability.

TABLE 2 Parametric Tests: $Z_p(t) - R(t) = \hat{\alpha}_p + \hat{\beta}_1 x(t)$; Results Assuming No Market-timing Activity

Sample	1968:2–1980:6	1968:2–1974:4	1974:5–1980:6
Mean (SD):			
$\hat{\alpha}$ (σ_{α})	−.0002 (.0027)	−.0013 (.0036)	.0011 (.0038)
$\hat{\beta}_1(\sigma_{\beta_1})$	.96 (.23)	1.02 (.25)	.90 (.21)
Number of funds:			
Reject $\hat{\alpha}$ = 0 at 5%	5+ 13−	4+ 18−	20+ 6−
Reject $\hat{\alpha}$ = 0 at 1%	2+ 8−	1+ 12−	7+ 2−
$\hat{\alpha} > 0$	52	38	68

NOTE.—Only one fund had a significantly positive $\hat{\alpha}$ in both periods. Only two funds had a significantly negative $\hat{\alpha}$ in both periods. *F*-test: number of funds that reject hypothesis that coefficients are equal in both periods: 50 at 5%; 33 at 1%.

And only one of the three exhibited significantly positive estimates of β_2 in both subperiods.

While there were many more significantly positive estimates of α for the overall period and the two subperiods, again only one of the funds exhibited significantly positive estimates in both periods. None of the funds had significantly negative estimates of α or β_2 in both periods.

For the funds in the sample, the correlation between the estimate of α in the first period and the second period was .15. For $\hat{\beta}_2$, the correlation between the two periods was .34. In fact, using a 2 × 2 test of independence, it is not possible to reject the hypothesis that the estimates of α for each fund are independent for the two periods or the hypothesis that the estimates of β_2 for each fund are independent for the two periods.

Further evidence of the instability of the parameters can be found in the results of tests of the hypothesis that the coefficients in (7) are equal in both subperiods. Using an *F*-test,[14] 45 of the funds reject the hypothesis with 95% probability and 26 reject with 99% probability.

These results are similar to those from the regression specification used by Jensen (1968, 1969):

$$Z_p(t) - R(t) = \alpha_p + \beta_p[Z_M(t) - R(t)] + \epsilon_p(t), \qquad (14)$$

where possible market-timing activity is ignored. This is as would be expected, since there does not appear to be any evidence of market-timing ability.

The results from regressions using the specification in (14) are summarized in table 2. Estimates of α tend to be slightly lower when market-timing activity is ignored, reflecting the fact that the coefficient for the omitted variable, $\hat{\beta}_2$, was on average slightly negative for the sample when the regression specification as shown in (7) was used.

14. For a description of the *F*-test run, see Fisher (1970).

TABLE 3 Parametric Tests: $Z_p(t) - R(t) = \hat{\alpha}_p + \hat{\beta}_1 x(t) + \hat{\beta}_2 y(t)$; without Heteroscedasticity Correction

Sample	1968:2–1980:6	1968:2–1974:4	1974:5–1980:6
Mean (SD):			
$\hat{\alpha}$ (σ_α)	.0008 (.0044)	−.0010 (.0054)	.0023 (.0062)
$\hat{\beta}_1(\sigma_{\beta_1})$	.90 (.19)	1.00 (.25)	.85 (.19)
$\hat{\beta}_2(\sigma_{\beta_2})$	−.07 (.18)	−.02 (.23)	−.08 (.21)
Number of funds:			
Reject $\hat{\alpha}$ = 0 at 5%	8+ 5−	5+ 10−	18+ 6−
Reject $\hat{\alpha}$ = 0 at 1%	5+ 2−	2+ 2−	8+ 2−
Reject $\hat{\beta}_2$ = 0 at 5%	3+ 13−	3+ 6−	3+ 13−
Reject $\hat{\beta}_2$ = 0 at 1%	1+ 8−	0+ 1−	1+ 7−
$\hat{\alpha} > 0$	55	31	65
$\hat{\beta}_2 > 0$	47	64	54

NOTE.—Only one fund had a significantly positive $\hat{\alpha}$ in both periods. Only one fund had a significantly negative $\hat{\alpha}$ in both periods. Only one fund had a significantly positive $\hat{\beta}_2$ in both periods. Only two funds had significantly negative $\hat{\beta}_2 s$ in both periods.

When market-timing activity is ignored, only one fund exhibited significantly positive estimates of α in both subperiods. This was the same fund that had significantly positive estimates of α in both subperiods when market-timing activity was taken into account. In addition, two funds had significantly negative estimates of α in both subperiods.

Tests of the hypothesis that the coefficients in (14) are equal in both subperiods again provides evidence that the parameters are not stationary. When an *F*-test is used, 50 of the funds reject the hypothesis with 95% probability and 33 reject with 99% probability.

The results for regressions as specified in (7), but without the heteroscedasticity correction, are summarized in table 3. As can be seen, the results are not qualitatively different than those with the correction, as shown in table 1. The major difference is that when heteroscedasticity is not taken into account, a greater number of "significant" negative estimates of β_2 were found and fewer "significant" estimates of α were found.

While the possible existence of heteroscedasticity did not seem to have much effect on the results, strong evidence was found that it indeed does exist. Examining the regressions run using (13) in the correction for heteroscedasticity shows that 73 funds had significantly negative estimates of Ω_1 and 57 funds had significantly positive estimates of Ω_2 at the 95% probability level. Of these, 55 of the negative estimates of Ω_1 were significant with 99% probability, as were 45 of the positive estimates of Ω_2.

As the values of $x_1(t)$ will always be either zero or negative and the values of $x_2(t)$ will always be either zero or positive, these results imply that the absolute size of the residuals is increasing with the absolute

TABLE 4 **Parametric Tests: $Z_p(t) - R(t) = \hat{\alpha}_p + \hat{\beta}_1 x(t)\ \hat{\beta}_2 y(t)$; Sample Split by Magnitude of $|x(t)|$, 1968:2–1980:6**

Sample	Small Magnitudes	Large Magnitudes
Mean (SD)		
$\hat{\alpha}$ (σ_α)	.0017 (.0057)	.0034 (.0075)
$\hat{\beta}_1(\sigma_{\beta_1})$	.91 (.45)	.87 (.20)
$\hat{\beta}_2(\sigma_{\beta_2})$	−.23 (.54)	−.14 (.26)
Number of funds:		
Reject $\hat{\alpha}$ = 0 at 5%	9+ 5−	4+ 2−
Reject $\hat{\alpha}$ = 0 at 1%	3+ 2−	1+ 1−
Reject $\hat{\beta}_2$ = 0 at 5%	2+ 8−	3+ 2−
Reject $\hat{\beta}_2$ = 0 at 1%	0+ 2−	1+ 1−
$\hat{\alpha} > 0$	73	78
$\hat{\beta}_2 > 0$	35	39

NOTE.—Two funds had significantly positive $\hat{\alpha}$'s for both groups. *F*-Test: number of funds that reject hypothesis that coefficients are equal for both groups are 10 at 5%; one at 1%.

magnitude of the market return. Henriksson and Merton show that these results can be caused by imperfect market-timing activity.

The tests of market-timing ability depend on the assumption that the conditional probabilities of correct forecasts are uniform over the entire range of outcomes where $Z_M > R$ and $Z_M \leqq R$. This assumption will be violated if the forecaster is able to predict periods of extreme market movement better than other periods. To test for this, the sample data was split in half by the magnitude of the excess market return, $|x(t)|$. Periods where the absolute value of $x(t)$ are greater than the sample median are split from those below the median. For the sample period, the median value of $|x(t)|$ is 3.1% per month.

A summary of the results for the 116 funds, with the sample split by the magnitude of $|x(t)|$, is shown in table 4. Only two of the funds (including the only fund when the sample was split by time) had significantly positive estimates of α in both samples. None of the funds had significantly positive estimates of β_2 in both samples or significantly negative estimates of α or β_2 in both samples.

The evidence does not support the hypothesis that investment managers are able to forecast large changes in the market or that it is necessary to model more than two levels of systematic risk. While three funds (out of 116) had significantly positive estimates of market-timing ability in the sample of large magnitude returns, all three had large negative (significant in two of the cases) estimates of α that more than offset the gains from market timing.

Ten of the funds did reject (at the 5% level) the hypothesis that the coefficients are equal for both the high-magnitude and low-magnitude samples. Those 10, however, did not exhibit any market-timing ability. In fact, the 10 appeared to do quite poorly with regard to market timing. By segmenting the sample by magnitude, estimates of the level of systematic risk are obtained for four different ranges of market returns. It is possible to compare the ranking of the level of systematic risk (from low to high) for each range of returns with the ranking of the market return for each range. For the 10 funds that reject the hypothesis that the coefficients are equal for both high-magnitude returns and low-magnitude returns, the correlation between the ranking of the level of systematic risk and the ranking of the market return for each range was $-.55$. The 10 funds seemed to be moving in the wrong direction in their market-timing activities.

Further examination of the estimates for the individual funds, using (7) with the heteroscedasticity correction, shows the existence of a strong negative correlation between $\hat{\alpha}$ and $\hat{\beta}_2$. For the total period, 49 of 59 funds with positive estimates of α had negative estimates of β_2. Of the 57 funds with negative estimates of α, 34 had positive estimates of β_2. In all but two of the cases where either $\hat{\alpha}$ or $\hat{\beta}_2$ are significantly different from zero, the two coefficients have different signs. In the two exceptions, in both cases, the estimate of α is significantly negative and the estimate of β_2 is also negative but small and not significant. While this negative correlation can be partially explained by measurement error in the estimates, as the covariance around the *true* values of α and β_2 will be negative, the results are so strong that this is not likely to be the entire explanation.[15]

The negative correlation was strong in both subperiods as well. As the market portfolio performed worse than riskless securities in the first period and performed better than riskless securities in the second period, the negative correlation between α and β_2 does not seem to depend on the market return for the period being examined.[16]

The negative correlation between estimates of α and β_2 seems to imply that funds that earn superior returns from stock selection also seem to have negative market-timing ability and performance. This is quite disturbing, as Treynor and Black (1973) showed that investment managers can effectively separate their stock-selection activities from their decision regarding the market risk of their fund. As this result has also been found in two previous studies using completely different

15. As the measurement error is relative to the true values of α and β_2 for the sample period, the results seem to imply that the true values of α and β_2 must either be opposite in sign or very close to zero in most cases. For the sample period, many of the funds had returns that exceeded the returns from the market portfolio and from riskless securities. As the average excess return on the market portfolio was approximately zero, the superior return must show up in the estimates of α or β_2 or both. Yet the negative correlation was the strongest for the funds that had the highest returns.

16. The negative correlation also existed when the tests were run without the heteroscedasticity correction and in both samples when the data were split by the magnitude of the market return. There were no qualitative changes in the results when investment periods longer than a month were used. In fact, the negative correlation was found even when the tests were run using random portfolios.

methodologies over different sample periods,[17] the possibility of misspecification of the return-generating process must be considered.

One potential source of error is misspecification of the market portfolio. This results from the fact that the proxy used for the market portfolio, the NYSE Index, does not include all risky assets. While it seems unlikely that the universe of investment opportunities relevant to the mutual funds in the sample has sufficiently different characteristics from the assets of the NYSE Index to cause the negative correlation of $\hat{\alpha}$ and $\hat{\beta}_2$, it is possible, especially for a particular sample period.

Another potential source of error is omission of relevant factors in addition to the return on the market portfolio from the return-generating process. If the omitted factor can be identified, then the return-generating process can be modified to take into account the omitted factor, and a multifactor version of the parametric test presented by Henriksson and Merton can be used. However, as the identity of the omitted variable, if it exists, is not known, a different procedure is used to focus on the potential biases.

To take into account the influence of these potential sources of error on the returns of mutual funds, the excess return on an equally weighted portfolio of all 116 funds in the sample, net of the influence of the return on the market portfolio, is added to (7) as a second factor. Formally, this factor is defined as

$$w(t) \equiv Z_{EW}(t) - R(t) - \beta_{EW}x(t), \tag{15}$$

where $Z_{EW}(t)$ is the return, in period t, from the equally weighted portfolio of all 116 funds, and β_{EW} is the least squares regression coefficient using the following specification:

$$Z_{EW}(t) - R(t) = \alpha_{EW} + \beta_{EW}x(t) + \epsilon(t). \tag{16}$$

For the total sample period, 1968:2–1980:6, the estimate of β_{EW} = .942 with a standard error of .013 and the estimate of $\alpha_{EW} = -.0002$ (i.e., $-.02\%$ per month) with a standard error of .0006.

As defined, $w(t)$ serves as a proxy for both assets omitted from the market proxy and relevant factors omitted from the returns-generating process as specified in (5).

The expanded regression specification for portfolio returns, taking into account the mutual fund factor, $w(t)$, is

$$Z_p(t) - R(t) = \alpha_p + \beta_1 x(t) + \beta_2 y(t) + \delta_1 w(t) + \delta_2 v(t) + \epsilon(t), \tag{17}$$

17. In an unpublished paper, Rex Thompson (1973) used a quadratic term as described in Jensen (1972*b*) to test for market timing using data for the period 1960–69. Stanley Kon (1981) used the switching regression methodology to test for timing during the period 1960–76.

TABLE 5 Parametric Tests: $Z_p(t) - R(t) = \hat{\alpha}_p + \hat{\beta}_1 x(t) + \hat{\beta}_2 y(t) + \hat{\delta}_1 w(t) + \hat{\delta}_2 v(t)$; with Mutual Fund Factor, 1968:2–1980:6

Sample	Mean	SD
$\hat{\alpha}$	−.0004	.0037
$\hat{\beta}_1$	.94	.22
$\hat{\beta}_2$	.00	.15
$\hat{\delta}_1$	.98	.96
$\hat{\delta}_2$	.05	.69
Number of funds:		
Reject at $\hat{\alpha} = 0$ at 5%	3+ 10−	
Reject at $\hat{\alpha} = 0$ at 1%	1+ 2−	
Reject at $\hat{\beta}_2 = 0$ at 5%	3+ 6−	
Reject at $\hat{\beta}_2 = 0$ at 1%	1+ 1−	
Reject at $\hat{\delta}_1 = 0$ at 5%	63+ 1−	
Reject at $\hat{\delta}_1 = 0$ at 1%	44+ 1−	
Reject at $\hat{\delta}_2 = 0$ at 5%	1+ 1−	
Reject at $\hat{\delta}_2 = 0$ at 1%	0+ 0−	
$\hat{\alpha} > 0$	44	
$\hat{\beta}_2 > 0$	70	
$\hat{\delta}_1 > 0$	110	
$\hat{\delta}_2 > 0$	58	

where $v(t) \equiv \max[0, -w(t)]$. This specification allows for possible timing of the mutual fund factor as well as the market portfolio.

The results from (17) are summarized in table 5. For the sample period, the mutual fund factor is predominately positively correlated with the performance of the mutual funds and appears to play an important role in explaining the behavior of returns for many of the funds. For the sample, 64 funds have estimates of δ_1 significantly different from zero at the 95% confidence level. For all but one of the 64 funds, the estimate of δ_1 is positive. At the 99% confidence level, 44 of the funds have estimates of δ_1 significantly different from zero, with only one having a significantly negative estimate. For the entire sample of 116 funds, only six had negative point estimates of δ_1.[18]

The funds did not demonstrate any significant timing ability with respect to the mutual fund factor, as only one fund had a significantly positive estimate of δ_2 and one fund had an estimate of δ_2 that was significantly negative. In both cases, the point estimates were significantly different from zero at the 95% confidence level but not at the 99% confidence level.

The inclusion of the mutual fund factor did change the estimates of the other coefficients, as the number of funds with positive estimates of

18. After running this test, I discovered that in an unpublished paper Black, Farrell, and Scholes (1972) used a similar approach to test for a mutual fund factor. They also found that this factor's coefficient was significantly positive for many of the funds in their sample. Their estimation period was 1960–69.

α fell from 59 to 44 while the number of funds with positive estimates of market-timing ability increased from 44 to 70. It did not change the number of significant positive estimates of market-timing ability, and the only fund to have a positive estimate of market-timing ability with 99% confidence was also the only fund to have an estimate of β_2 significant at that level for the entire sample period in table 1. In fact, it was also the only fund to exhibit significant market-timing ability in both time subperiods examined.

The mutual fund factor did not explain the negative correlation between $\hat{\alpha}$ and $\hat{\beta}_2$. For the 45 funds with positive estimates of α, 28 also had negative estimates of $\hat{\beta}_2$, while 54 of the 71 funds with negative estimates of α had positive estimates of β_2.

While the number of funds with significant coefficients for the mutual fund factor is consistent with either the misspecification of the market portfolio or the omission of other relevant factors in (5), the results are also consistent with the hypothesis that mutual funds tend to follow similar investment strategies. This is sometimes referred to as the "herd" effect.

IV. Nonparametric Tests

Because of the negative correlation between α and β_2, the specification used for the parametric tests must be questioned. The nonparametric test described in Section II does not require a specified model of returns and therefore avoids this problem. However, the nonparametric test requires that the forecaster's predictions are known or that a proxy for the forecasts can be found. As the market-timing forecasts of mutual funds are not available, it is necessary to use a proxy for the forecasts.

Henriksson (1980) used changes in the proportion of equities held in the portfolios of mutual funds as a proxy for the market-timing forecasts. Using quarterly data from 1973–80 for 186 mutual funds, he found no evidence of market-timing ability. However, his proxy will be measured with error if the fund manager's forecast intervals do not correspond to the quarterly periods for which data are available or if the funds follow an adjustment process for the level of market-related risk more complex than those modeled by Henriksson. Such measurement error will bias the results against detecting superior forecasting, even if the forecaster does have market-timing ability. Also, the limited availability of the data makes it difficult to find significant forecasting ability.

To avoid measurement errors in the proxy, the actual returns of the mutual funds are used as the measure of performance and compared with the performance of a feasible passive strategy. The returns of the mutual funds and the passive strategy are compared using the 2×2

TABLE 6 **Nonparametric Tests: 116 Open-End Funds, Return on Fund versus Passive Strategy Reflecting Fund, Null Hypothesis: $p_1 + p_2 = 1$**

	1968:2–1980:6	1968:2–1974:4	1974:5–1980:6
Average estimated $(p_1 + p_2)$	.984	.947	1.021
Standard deviation	.115	.148	.168
Number of funds:			
Reject null at 5%	4+	4+	7+
Reject null at 1%	1+	2+	3+
Estimate $(p_1 + p_2) > 1$	54	39	65

NOTE.—Correlation (period 1, period 2) = .05. Only one fund exhibited positive forecasting ability in both periods.

test of independence described in Section II. The returns of the passive strategy are segmented around $R(t)$ and the null hypothesis, $H_0: p_1(t) + p_2(t) = 1$, reflects the probabilities of the mutual fund's outperforming the passive strategy, conditional on whether or not the return on the passive strategy exceeds the return on riskless securities. In this form, the test examines the total performance of the fund and not just the market-timing ability. It examines whether or not active portfolio management can generate returns in excess of those earned by a feasible passive strategy.

The passive strategy used is a portfolio consisting of the market portfolio and Treasury bills with the proportion of the portfolio invested in the market portfolio equal to the β of the fund where β is measured using

$$Z_p(t) - R(t) = \alpha + \beta x(t) + \epsilon(t). \tag{18}$$

The tests did not seem to be sensitive to the passive strategy chosen, as others were also used with little change in the results.

The results of the nonparametric test for the individual funds are summarized in table 6 and figure 1. On average, the funds appear to do slightly worse than the passive strategy, which is consistent with the hypothesis of no forecasting ability and the use of returns that are net of management costs and fees.

Four funds were able to reject the null hypothesis of no forecasting ability at the 95% confidence level for the entire sample. However, only one fund was able to reject the null hypothesis in both subperiods, the same fund that also had significantly positive estimates for α for both subperiods in the parametric tests. In fact, in every case where the null hypothesis is rejected, both for the total period and the two subperiods, the same fund had a positive estimate of α and a negative estimate of β_2 in the parametric test for the same period. All but two of these funds had estimates of α that were significantly different from zero. In every case where a fund had a significantly positive estimate of market-timing ability in the parametric tests, that same fund had a

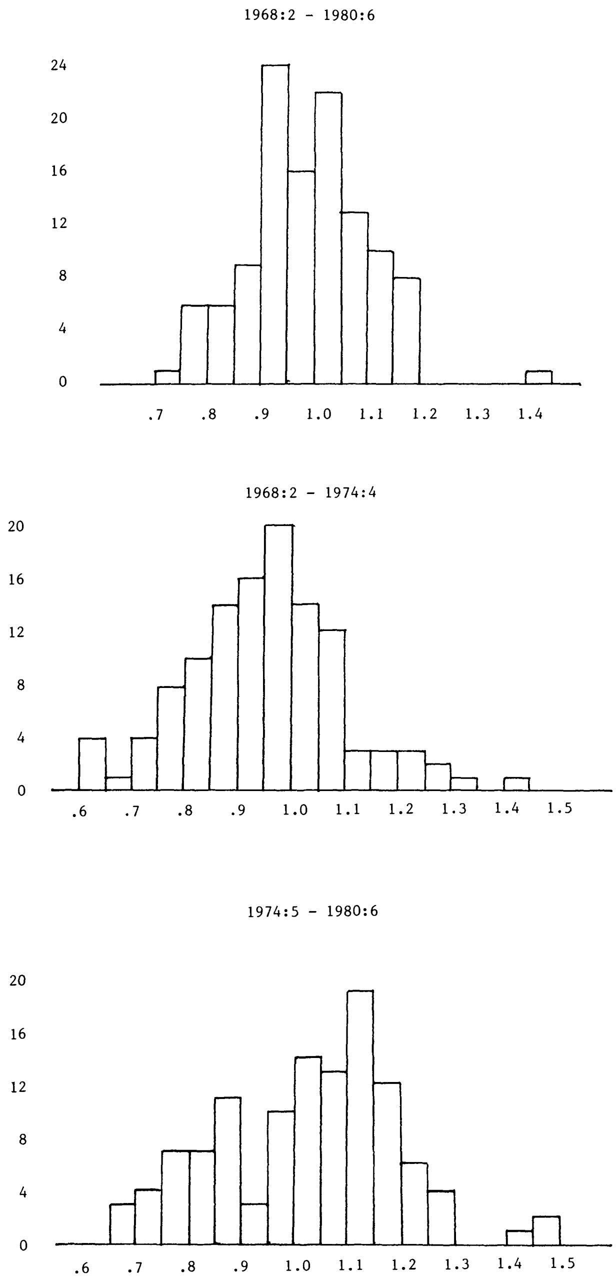

FIG. 1.—Mutual fund results. Nonparametric test fund: fund return versus passive strategy, $E(p_1 + p_2)$.

TABLE 7 **Nonparametric Tests: Return on Fund versus Passive Strategy, Sample Split by Magnitude of $|x(t)|$, 1968:2–1980:6***

Magnitude	Less than Median	Greater than Median
Averaged estimated $(p_1 + p_2)$	.986	.990
Standard deviation	.161	.123
Number of funds:		
Reject null at 5%	8	1
Reject null at 1%	2	1
Estimated $(p_1 + p_2) > 1$	55	52

NOTE.—None of the funds exhibited positive forecasting ability for both samples.
* Median $|x(t)|$ = .0308.

negative estimate of overall performance in the nonparametric tests, that is, est$(p_1 + p_2) < 1$.

As in the parametric tests, there appeared to be very little relationship between the performance of the first subperiod and the second subperiod for the individual funds. The correlation between the estimates of $p_1 + p_2$ for the same fund in the two subperiods was .05. Overall, the funds tended to do better in the second subperiod than the first.

The sample is also split by the magnitude of the absolute value of the excess return on the passive strategy to examine the assumption that the forecaster is no better at predicting large magnitude changes than small ones. The results for the individual funds are summarized in table 7. None of the funds exhibited significantly superior performance in both samples, and the results provide no evidence that the forecasters are better able to forecast large magnitude changes than small. For the sample of 116 funds, 56 had a higher estimate of $p_1 + p_2$ for the sample of large-magnitude returns than for the sample of small-magnitude returns.

V. Mutual Fund Returns

As a final comparison, the actual sample period returns from the investment of $1.00 at the beginning of the period for each of the 116 mutual funds are examined and compared with the results of the parametric and nonparametric tests. The returns for the individual funds are summarized in table 8 and figure 2.

Even though the average level of systematic risk for the mutual funds was less than that of the market portfolio, the funds on average did worse than the market portfolio in the first subperiod where the excess return on the market portfolio was negative, and the funds performed better than the market portfolio in the second subperiod when the excess return on the market was positive.

Of the eight funds that finished the total period with over $3.00

TABLE 8 **Total Returns: 116 Open-End Mutual Funds, Return on Initial Investment of $1.00**

	1968:2–1980:6	1968:2–1974:4	1974:5–1980:6
Average for 116 funds	$2.09	$1.07	$1.95
Standard deviation	.96	.31	.56
Market portfolio	$2.09	$1.14	$1.84
Riskless securities	$2.19	$1.41	$1.55
Number of funds:			
Return greater than market portfolio	45	45	55
Return greater than riskless securities	42	8	95

(starting with the $1.00 investment), all had positive estimates of α in the parametric tests and only one had a positive estimate of β_2. The only fund to reject the null hypothesis for both subperiods in the nonparametric tests and to have a significantly positive estimate of α in both subperiods in the parametric tests also had the greatest return over the sample period.[19] The fund earned over $3.00 in both subperiods and earned $9.58 overall, exceeding the next highest fund by $4.40.

The four funds that rejected the null hypothesis for the total period in the nonparametric tests all had a return greater than the average for the 116 funds, with two of them ranking first and second in total returns. The four funds had returns of $9.58, $5.18, $2.65, and $2.43. The returns for the three funds that had significantly positive estimates of β_2 in the parametric tests were $1.94, $1.95, and $2.21, with the fund that had significant estimates in both periods earning $1.94.

VI. Conclusion

The empirical results obtained using techniques developed by Henriksson and Merton (1981) do not support the hypothesis that mutual fund managers are able to follow an investment strategy that successfully times the return on the market portfolio. This is observed in both the parametric and nonparametric tests. Only three funds had significantly positive estimates of market-timing ability in the parametric tests for the period from 1968:2 to 1980:6, and only one fund had significant estimates in both subperiods when the sample was split in half. All three had negative overall estimates of performance in the nonparametric tests and total returns for the period very close to the average of all funds in the sample.

Of the four funds that exhibited superior performance in the non-

19. This fund was Templeton Growth.

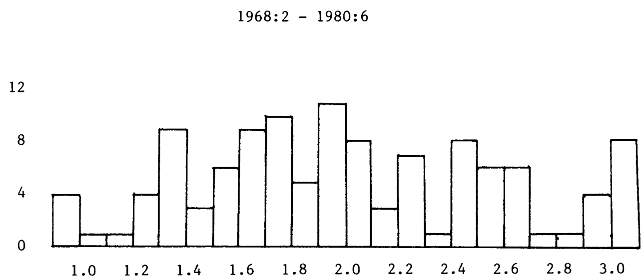

FIG. 2.—Mutual fund returns: return on initial investment of $1.00, 1968:2–1980:6 distribution of funds ($).

parametric tests, only one did so in both subperiods, and all four had positive estimates of α and negative estimates of β_2 in the parametric tests.

Strong evidence of nonstationarity in the performance parameters was found in both the parametric and nonparametric tests. In addition, no evidence was found that forecasters are more successful in their market-timing activity with respect to predicting large changes in the value of the market portfolio relative to smaller changes. The absolute magnitude of the returns on the market portfolio did not seem to have an influence on the measures of performance evaluation.

The specification used in the parametric tests must be questioned because of the persistence of a negative correlation between $\hat{\alpha}$ and $\hat{\beta}_2$, which raises new questions regarding the validity of using the CAPM to evaluate portfolio performance when the possibility of market timing is allowed for. Although it does not explain the negative correlation, a mutual fund factor was added to the specification used and was found to be significant for 64 of the 116 funds in the sample. One possible explanation of this result is the existence of a factor omitted in the return-generating process modeled.

Appendix

Mutual Fund Sample

Fund No.	Fund Name	Objective*
1	Axe-Houghton Stock	G
2	Boston Foundation Fund	S-G-I
3	Broad Street Investment Corp.	G-I
4	Bullock Fund	G-I
5	Canadien Fund	G
6	Century Shares Trust	G
7	Chase Fund of Boston	MCG
8	Chemical Fund	G
9	Colonial Fund	G-I

10	Colonial Growth Shares	G
11	Commerce Income Shares	G-I
12	Composite Fund	G-I-S
13	Composite Bond & Stock Fund	I-S-G
14	Common Stock Fund of Stage Bond & Mortgage Co.	G
15	Corporate Leaders Trust B	G-I
16	Decatur Income Fund	I
17	Delaware Fund	G-I-S
18	DeVegh Mutual Fund	G
19	Dividend Shares	G-I
20	Dreyfus Fund	G
21	Eaton & Howard Balance Fund	I-G-S
22	Eaton & Howard Stock Fund	G-I
23	Energy Fund	G
24	Fairfield Fund	MCG
25	Fidelity Fund	G-I
26	Fidelity Trend Fund	G
27	Financial Dynamics Fund	MCG
28	Financial Industrial Fund	G-I
29	First Investors Fund	I
30	Fund of America	MCG
31	Founders Growth Fund	MCG
32	Founders Mutual Fund	G-I
33	Franklin Custodian Fund—Income Series	I
34	Fundamental Investors	G-I
35	Guardian Mutual Fund	G-I
36	Investment Co. of America	G-I
37	Investors Mutual	I-S-G
38	Investors Stock Fund	G-I
39	Investors Variable Payment Fund	G
40	Istel Fund	G-I
41	Investment Trust of Boston	G-I
42	Ivest Fund	MCG
43	Johnston Capital Appreciation Fund	G
44	Keystone Income Fund (K1)	I
45	Keystone Growth Fund (K2)	G
46	Keystone High-Grade Common Stock Fund (S1)	G-I
47	Keystone Growth Common Stock Fund (S3)	G
48	Keystone Speculative Common Stock Fund (S4)	MCG
49	Lexington Research Fund	G
50	Life Insurance Investors Fund	G
51	Loomis-Sayles Capital Development Fund	G
52	Loomis-Sayles Mutual Fund	G-I-S
53	Manhattan Fund	G
54	Massachusetts Fund	S-I-G
55	Midamerica Mutual Fund	G
56	Mutual Investing Foundation—MIF Fund	I
57	Mutual Investing Foundation—MIF Growth Fund	G
58	Massachusetts Investors Growth Stock Fund	G
59	Massachusetts Investors Trust	G-I
60	Mutual Shares Corp.	MCG
61	National Investors Corp.	G
62	Nation-Wide Securities	I-S-G
63	New World Fund	G-I
64	Northeast Investors Trust	I
65	National Dividend Fund	I
66	National Stock Fund	G-I

67	One William Street Fund	G-I
68	Oppenheimer Fund	MCG
69	Penn Square Mutual Fund	G
70	Philadelphia Fund	G-I
71	Pilgrim Fund	G
72	Pilot Fund	MCG
73	Pine Street Fund	G-I
74	Provident Fund for Income	I
75	Puritan Fund	I
76	Putnam Fund of Boston	S-I-G
77	Putnam International Equities Fund	MCG
78	Putnam Growth Fund	G
79	Putnam Income Fund	I
80	Putnam Investors Fund	G
81	Research Equity Fund	MCG
82	Revere Fund	MCG
83	SAFECO Equity Fund	G-I
84	Salem Fund	G
85	Scudder Common Stock Fund	G
86	Scudder Income Fund	I-S
87	Scudder International Fund	G
88	Scudder Special Fund	MCG
89	Security Equity Fund	G
90	Security Investment Fund	I
91	Selected American Shares	G-I
92	Sentinel Balanced Fund	I-G-S
93	Sentinel Common Stock Fund	G-I
94	Shareholders' Trust of Boston	S-I-G
95	Kemper Growth Fund	MCG
96	Stein, Roe & Farnham Capital Opportunities Fund	G
97	Stein, Roe & Farnham Stock Fund	G
98	St. Paul Capital Fund	G-I
99	St. Paul Growth Fund	G
100	Technology Fund	G-I
101	Templeton Growth Fund	G
102	Twentieth Century Growth Investors	MCG
103	Union Income Fund	I
104	United Accumulative Fund	G
105	United Income Fund	I
106	United Science and Energy Fund	G
107	Value Line Income Fund	I
108	Value Line Fund	G
109	Value Line Special Situations Fund	MCG
110	Varied Industry Plan	G-I
111	Vance Sanders Common Stock Fund	G
112	Vance Sanders Investors Fund	G-I-S
113	Wall Street Fund	G-I-S
114	Washington Mutual Investors Fund	G-I
115	Wellington Fund	S-I-G
116	Windsor Fund	G

* Primary objective according to Wiesenberger (1980): G ≡ growth; I ≡ income; MCG ≡ maximum capital gain; and S ≡ stability.

References

Black, F.; Farrell, J. L.; and Scholes, M. 1972. Common factors in mutual fund performance and their influence on performance evaluation. Working Paper. Chicago: University of Chicago, Graduate School of Business.

Black, F.; Jensen, M. C.; and Scholes, M. 1972. The capital asset pricing model: Some empirical tests. In M. C. Jensen (ed.), *Studies in the Theory of Capital Markets*. New York: Praeger.

Blume, M. E., and Friend, I. 1973. A new look at the capital asset pricing model. *Journal of Finance* 28 (March): 19–34.

Fama, E. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25 (May): 383–417.

Fama, E., and MacBeth, J. D. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81:607–36.

Fisher, F. M. 1970. Tests of equality between sets of coefficients in two linear regressions: An expository note. *Econometrica* 38 (March): 361–66.

Henriksson, R. D. 1980. Tests of market timing and mutual fund performance. Working Paper no. 1136-80. Cambridge, Mass.: MIT, Sloan School of Management.

Henriksson, R. D., and Merton, R. C. 1981. On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business* 54 (October): 513–33.

Ibbotson, R., and Sinquefield, R. 1979. *Stocks, Bonds, Bills and Inflation: Historical Returns (1965–1978)*. Charlottesville, Va.: Financial Analysts Research Foundation.

Jensen, M. C. 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance* 23 (May): 389–416.

Jensen, M. C. 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business* 42 (April): 167–247.

Jensen, M. C. 1972*a*. Capital markets: Theory and evidence. *Bell Journal of Economics and Management Science* 3 (Fall): 357–98.

Jensen, M. C. 1972*b*. Optimal utilization of market forecasts and the evaluation of investment performance. In G. P. Szego and K. Shell (eds.), *Mathematical Methods of Investment and Finance*. Amsterdam: North Holland.

Kon, S. J. 1981. The market timing performance of mutual fund managers. Working paper. New York: New York University.

Mayers, D., and Rice, E. M. 1979. Measuring portfolio performance and the empirical content of asset pricing models. *Journal of Financial Economics* 7 (March): 3–28.

Merton, R. C. 1981. On market timing and investment performance. I. An equilibrium theory of value for market forecasts. *Journal of Business* 54 (July): 363–406.

Merton, R. C.; Scholes, M.; and Gladstein, M. L. 1982. The returns and risks of alternative put-option portfolio investment strategies. *Journal of Business* 55 (January): 1–55.

Roll, R. 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 (March): 129–76.

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

Stambaugh, R. F. 1982. On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. *Journal of Financial Economics* 10 (November): 237–68.

Thompson, R. 1973. The effects of the optimal utilization of market forecasts on mutual fund performance in the period 1960–1969. Working Paper. Rochester, N.Y.: University of Rochester, Graduate School of Business.

Treynor, J., and Black, F. 1973. How to use security analysis to improve portfolio selection. *Journal of Business* 46 (January): 66–86.

Wiesenberger Investment Companies Service. Various years. *Investment Companies*. New York: Warren, Gorham & Lamont.

Zellner, A. 1962. An efficient method for estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association* 57 (June): 348–68.