Stanley J. Kon
University of Michigan

The Market-timing Performance of Mutual Fund Managers*

This paper proposes an empirical methodology for measuring the market-timing performance of an investment manager and provides evidence for a sample of mutual funds. The results indicate that at the individual fund level there is evidence of significant superior timing ability and performance. However, the multivariate tests were not inconsistent with the efficient markets hypothesis. That is, fund managers as a group have no special information regarding the formation of expectations on the returns of the market portfolio.

I. Introduction

The theory and measurement of the investment performance of managed portfolios has received a great deal of attention in the financial economics literature.[1] The topic is important to potential clients, managers evaluating the effectiveness of their strategies, and regulators that formulate policy concerning the operations of the marketplace. The Jensen (1968, 1969) model of risk-adjusted performance has been the premier methodology for addressing these issues and testing the strong form of the efficient markets hypothesis (that is, whether any market participant has monopolistic access to any information relevant for price formation).

The investment management process can be dichotomized into the activities of stock selection and market timing. Stock selection is based on forecasts of company-specific events and hence the prices of individual securities. Market timing, however, refers to forecasts of future

*I am grateful to the participants of the Finance Workshops at the University of Michigan and the University of Utah for helpful comments and discussions. This paper has also benefited from comments by an anonymous referee. This research was partially supported by a summer research grant from the University of Michigan Graduate School of Business Administration.

1. For examples, see Jensen (1968, 1969, 1972), USSEC (1971), Fama (1972), Modigliani and Pogue (1975), Grant (1977), Kon and Jen (1979), Henriksson and Merton (1981), and Merton (1981).

(*Journal of Business,* 1983, vol. 56, no. 3)

0021-9398/83/5603-0002$01.50

realizations of the market portfolio. If an investment manager believes he can make better than average forecasts of market portfolio returns, he will adjust his portfolio risk level in anticipation of market movements. If successful, he will earn abnormal returns relative to an appropriate benchmark. While there is a considerable amount of empirical evidence available on stock selectivity performance, the empirical evidence on market timing has concentrated only on the existence of the timing activity. That is, evidence of nonstationarity of the systematic risk for a mutual fund is consistent with the timing activity.[2]

Recently, Merton (1981) has proposed an equilibrium theory of value for qualitative market forecasts of whether stocks will outperform bonds or vice versa. The correspondence of returns from these predictions to an option investment strategy is used to derive the equilibrium management fees. His simulations and examples indicate that a substantial additional average rate of return can be earned with superior market-timing ability. Henriksson and Merton (1981) derive a nonparametric procedure for testing market-timing ability when the forecasts are directly observable and a parametric test for the usual case of having access only to the time series of realized returns on a managed portfolio.

Fama (1972) has proposed a theoretical measure of market-timing performance whereby the appropriate forecast is only that of the deviation of the market portfolio from its consensus expected return. The empirical implementation of this measure requires a proxy for the target risk level of the fund, a time series of varying consensus expected returns on the market portfolio and the time series of risk level decisions by the fund's management. This is particularly difficult when the only direct information available is the realized rates of returns on the mutual fund and the market portfolio.[3] The purpose of this paper is to propose an empirical methodology for measuring market-timing performance (as defined by Fama [1972]) and test this methodology on a sample of mutual funds. Examining these results will also provide evidence concerning market efficiency with respect to the formation of expectations on the returns of the market portfolio.

Section II of this paper discusses the theoretical measure of market-timing performance based on the optimal behavior of an investment manager and the econometric difficulties associated with obtaining an

2. This has been approached by detecting evidence of nonstationarity of the estimated systematic risk of mutual funds. See Treynor and Mazuy (1966), Jensen (1969), Campanella (1972), and Kon and Jen (1979).

3. It would be useful to have complete portfolio composition data which include all buy and sell decisions. But this is clearly not the case for the published quarterly data available to researchers. Significant manipulation occurs between quarterly reports in actively managed funds. The procedure in this paper estimates the true portfolio risk level for each rate of return observation regardless of any portfolio manipulation during the observation interval.

empirically tractable measure. When the timing activity is ignored, Jensen (1972) and Kon (1982) have shown that nonneutral timing performance causes a bias in the Jensen (1968) ordinary least squares estimates of risk and stock selectivity performance. Kon and Jen (1979) avoid these biases in analyzing selectivity performance by using an econometric model specification that is consistent with the timing activities of an investment manager. The methodology here extends theirs with an additional discriminant procedure to classify observations and obtain the time series of risk level decisions required for the time series of timing performance estimates. The condition required for the estimate of average timing performance over the measurement interval to be invariant to the choice of target risk level proxy is also examined.

Section III presents the empirical results on market-timing performance for three different models of consensus expected returns on the market portfolio. This section also contains a multivariate analysis of selectivity performance, the trade-off between the timing and selectivity activities, and the implication for the efficient markets hypothesis. Section IV contains a summary of results and conclusions.

II. Empirical Measures of Timing Performance

Since the market timer's forecasts are not directly observable, constructing tests of market-timing performance requires a specific assumption for the security-return-generating process. Therefore, the following methodology results in tests of the joint hypothesis of no market-timing ability (or performance) and the assumed generating process for asset returns. The separable selectivity and timing performance measures and their empirical implementation described below assume a pattern of equilibrium security returns consistent with the traditional capital asset pricing model of Sharpe (1964), Lintner (1965), and Mossin (1966). However, the timing performance measure is also consistent with the Black (1972) equilibrium model and the multifactor pricing models of Merton (1973) and Ross (1976) under the assumption that one of the factors is the market portfolio.

A. The Behavioral Model of Timing Performance

To model the timing activity of fund managers, the rate of return on the market portfolio can be expressed as

$$R_{Mt} = E_t(R_{Mt}) + \pi_{Mt}, \tag{1}$$

where $E_t(R_{Mt})$ is the consensus expected rate of return on the market portfolio at time t and π_{Mt} is the unanticipated return on the market portfolio with $E(\pi_{Mt}) = 0$ for all t and finite variances $\sigma_t^2(\pi_{Mt})$. These parameter values are those perceived by the market participants under the null hypothesis. The definitions of bull and bear market outcomes are $\pi_{Mt} > 0$ and $\pi_{Mt} < 0$, respectively. Whenever the investment manager's expectations concerning the market factor are identical to the consensus of all market participants, a target risk level, β_T, will be selected. If, however, the manager's information set at time $t - 1, \phi_{t-1}$, is different from the consensus, then $E_t(\pi_{Mt}|\phi_{t-1})$ will be nonzero for some or all observations and $0 < \sigma_t^2(\pi_{Mt}|\phi_{t-1}) \leq \sigma_t^2(\pi_{Mt})$. With this special information he is expected to adjust portfolio risk β_t in anticipation of market changes to earn abnormal returns.[4] Fama (1972) derived the following single-period timing performance measure of an investment manager relative to a passive strategy:

$$\tau_t = (\beta_t - \beta_T)\,\pi_{Mt}. \tag{2}$$

Note that performance will be positive if the sign of $\beta_t - \beta_T$ is the same as π_{Mt} and negative if they are opposite. That is, a manager who increases portfolio systematic risk above his target level in anticipation of a bull market, and is correct, will earn an additional return that is dependent on the magnitudes of the risk level shift and the unanticipated market movement. An appropriate loss in the fund's return results from the same manager's increase in risk level (bull market prediction) when the actual outcome is a bear market. The resulting gain (loss) to a correct (incorrect) bear market forecast is also explicit in this timing measure. Hence, overall timing performance for the measurement interval is

$$\tau_0 = \frac{1}{n}\sum_{t=1}^{n}(\beta_t - \beta_T)\,\pi_{Mt}. \tag{3}$$

B. Econometric Model Specification and Estimation

The primary goal of the econometric methodology is to obtain the time series of risk level decisions of the manager required to estimate the timing performance measures in equations (2) and (3). This is accomplished by incorporating the Quandt (1972) switching regression model with the Kon and Jen (1979) identifiability condition and an additional discriminant procedure.

Given n observations on the mutual fund's excess returns, $y_t = R'_t$, and the independent variables, $\mathbf{x}'_t = [1\ x_t]$ where $x_t = R'_{Mt}$, the excess return on the market portfolio, there exists some *true* permutation of the rows of $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ and $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)'$, so that

4. See Jensen (1972) or Kon and Jen (1979) for the equation of the target β and the magnitude of the risk level decision associated with bull and bear market forecasts. These market-timing β values depend on the uncertainty of the manager's forecast, $\sigma_t^2(\pi_{Mt}|\phi_{t-1})$, the degree of risk aversion for the fund, and the magnitude of the manager's bull or bear market forecast, $E_t(\pi_{Mt}|\phi_{t-1})$.

they may be partitioned according to the risk level decisions of the fund manager.

Without a priori information on the fund manager's decision-making process, it is reasonable to assume a multinomial prior whereby the fund manager selects risk level β_i (or regime i) for generating observations with probability (or proportion of the n decisions) λ_i, $i = 1, 2, \ldots, N$. Therefore, each observation is viewed as a drawing from the mixture distribution (conditional p.d.f. of y_t given $\mathbf{x}_t'$ and the parameter vector $\boldsymbol{\theta}$).

$$f_t\,(y_t|\mathbf{x}_t', \boldsymbol{\theta}) = \sum_{i=1}^{N} \lambda_i p(y_t|\mathbf{x}_t', \boldsymbol{\gamma}_i) \tag{4}$$

where $\boldsymbol{\gamma}_i = (\alpha_i, \beta_i, \sigma_i^2)$ is the vector containing the stock selectivity performance, risk level, and residual variance parameters, respectively, for regime i; $\boldsymbol{\theta} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_N, \lambda_1, \lambda_2, \ldots, \lambda_{N-1})$ is the parameter vector to be estimated subject to the parameter space $\Omega = \{\boldsymbol{\theta}: -\infty < \alpha_i < \infty, -\infty < \beta_i < \infty, 0 < \sigma_i^2 < \infty, i = 1, 2, \ldots, N, 0 < \lambda_i < 1, i = 1, 2, \ldots, N-1\}$; $p(y_t|\mathbf{x}_t, \boldsymbol{\gamma}_i) = \text{p.d.f.} \sim N(\alpha_i + \beta_i x_t; \sigma_i^2)$, $i = 1, 2, \ldots, N$; $\Sigma_{i=1}^{N} \lambda_i = 1$; and Identifiability Condition: $\beta_N > \beta_{N-1} > \ldots > \beta_1$.[5]

For a sample of size n, Quandt (1972) proposed maximizing the likelihood function (joint probability),

$$\ell(\boldsymbol{\theta}|\mathbf{y}, X) = \prod_{t=1}^{n} \left[\sum_{i=1}^{N} \lambda_i p(y_t|\mathbf{x}_t', \boldsymbol{\gamma}_i) \right], \tag{5}$$

or equivalently the logarithmic likelihood,

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \log f_t(y_t|\mathbf{x}_t', \boldsymbol{\theta}). \tag{6}$$

The maximum likelihood estimator, $\hat{\boldsymbol{\theta}}_n$, is the solution to the normal equations:[6]

$$\frac{\partial L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n} \frac{1}{f_t(y_t|\mathbf{x}_t', \boldsymbol{\theta})} \cdot \frac{\partial f_t(y_t|\mathbf{x}_t', \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{O}. \tag{7}$$

Details of this estimation procedure and hypothesis testing are given in Kon and Jen (1979).

This estimation procedure avoids the OLS biases of risk and selectivity performance by modeling the true nonstationary stochastic process. The procedure thus far, however, does not identify each observation with a specific risk level. Only the probability that it was generated

5. This condition ensures that the mixture in eq. (4) is uniquely determined. See Kon and Jen (1979) for a discussion and simulation of the importance of this restriction.
6. Kiefer (1978) has verified that these likelihood equations do have a consistent root.

by each risk level is estimated. Thus, timing performance as defined by equations (2) and (3) is not yet empirically tractable. However, an additional discriminant procedure given the estimated parameter vector, $\hat{\boldsymbol{\theta}}$, can be used to estimate the true partition of the data by an appropriate "most likely" classification rule. Given the information in $\hat{\boldsymbol{\theta}}$, choose the regime (or risk level) that maximizes the posterior probability density associated with it. That is, classify the tth observation as having been generated by regime i according to the rule

$$\underset{i}{\text{maximum}}\ [\hat{\lambda}_i\, p(y_t|\mathbf{x}_t', \hat{\gamma}_i)]. \tag{8}$$

In theory, if the number of observations is large then the estimates substituted for true values in equation (8) should be reasonably accurate and the classification techniques should be quite good (see Press 1972, pp. 374–75). This was verified by sampling experiments for the number of observations used in the empirical section of this paper. Once classified, a vector $\hat{\boldsymbol{\beta}}$ of $\hat{\beta}_t$'s is estimated where each element of $\hat{\boldsymbol{\beta}}$ is one of the $\hat{\beta}_i$, $i = 1, 2, \ldots, N$.

C. *Timing Performance Estimates*

Given the information in $\hat{\boldsymbol{\beta}}$ from the estimation procedure of the previous subsection and suitable proxies for the fund's target risk level and the unanticipated market portfolio returns, the estimates of timing performance (eq. [2]) for observations $t = 1, \ldots, n$ and of overall timing performance (eq. [3]) are empirically tractable. The definition of π_{Mt} is the deviation of R_{Mt} from its consensus expectation, $E_t(R_{Mt})$. Two possible proxies for $E_t(R_{Mt})$ are the sample mean, $\bar{R}_M$, and the current riskless rate plus the mean risk premium, $R_{Ft} + (\bar{R}_M - \bar{R}_F)$, for the measurement interval. Given the choice of proxy,

$$\hat{\pi}_{Mt} = R_{Mt} - E_t(R_{Mt}). \tag{9}$$

An appropriate proxy for β_T must be substituted in (2) and (3). The proxy to be employed in this study for each fund will be the average risk level of all funds in that stated objectives class. Note, however, that the proxy for β_T is invariant with respect to the overall timing performance estimate if the consensus expected returns for the market portfolio is, on average, realized during the measurement interval. That is, if $(1/n)\,(\Sigma_{t=1}^{n}\,\hat{\pi}_{Mt}) = 0$, then

$$\begin{aligned}\hat{\tau}_0 &= \frac{1}{n}\sum_{t=1}^{n}(\hat{\beta}_t - \beta_T)\,\hat{\pi}_{Mt} = \frac{1}{n}\sum_{t=1}^{n}\hat{\beta}_t\,\hat{\pi}_{Mt} - \beta_T\frac{1}{n}\sum_{t=1}^{n}\hat{\pi}_{Mt} \\ &= \frac{1}{n}\sum_{t=1}^{n}\hat{\beta}_t\,\hat{\pi}_{Mt}.\end{aligned} \tag{10}$$

Both of the proxies for $E_t(R_{Mt})$ suggested above satisfy this condition. When the invariance property of the choice of β_T fails to hold it does not mean that the timing performance estimates are biased unless there is a prior reason to believe that the choice of β_T proxy is biased. This property only emphasizes the benefit of a very large time series of observations in a market with rational expectations. Then the invariance property follows.

III. Empirical Results

A. Data

The sample consists of 37 mutual funds, each with 198 months of total rate of return data (based on net asset values including all distributions) from January 1960 to June 1976. The range of objectives include four funds defined as maximum capital gains, 10 as growth, 11 as growth/income, 9 as balanced, and three with income objectives. The proxies for the riskless rate and market portfolio are the 30-day Treasury bill rate and the monthly rate of return on the CRSP value-weighted market index (adjusted for dividends), respectively.

B. Model Specification Tests

In order to empirically define the number of distinct risk levels selected during the performance measurement interval by each fund manager, the maximum likelihood estimation procedure described in Section II, subsection *B* was employed for the alternative hypotheses of two and then three regression models having generated the data.[7] Table 1 lists each fund, its stated objectives and the likelihood ratio tests of model specification. The statistic in column 4, $-2 \log [\ell (\boldsymbol{\theta}|\mathbf{y}, X, N = 1)/\ell (\boldsymbol{\theta}|\mathbf{y}, X, N = 2)]$ is a test of the standard linear model (single regime) null hypothesis versus the alternative two-regime model. The statistic has an asymptotic χ^2 distribution with 4 degrees of freedom. Out of the 37 funds, 35 rejected the null hypothesis at the .05 level of significance (exceeded 9.49). However, 10 of these funds rejected the null hypothesis of two regimes against the three-regime alternative (col. 6) where the test statistic $-2 \log [\ell (\boldsymbol{\theta}|\mathbf{y}, X, N = 2)/\ell (\boldsymbol{\theta}|\mathbf{y}, X, N = 3)]$ also has an asymptotic χ^2 distribution with 4 degrees of freedom. The statistic in column 5, $-2 \log [\ell (\boldsymbol{\theta}|\mathbf{y}, X, N = 1)/\ell (\boldsymbol{\theta}|\mathbf{y}, X, N = 3)]$, has an asymptotic χ^2 distribution with 8 degrees of freedom. It is used to test the null hypothesis of the standard linear model against the three-regime alternative. Of the two cases where the single-regime model could not be rejected by the two-regime alternative, both were rejected

7. Equation (6) was maximized by the modified quadratic hill-climbing algorithm described in Goldfeld and Quandt (1972).

TABLE 1 Mutual Funds, Objectives, and Likelihood Ratio Specifications Tests

Mutual Fund	Objective	ID Number	$-2\Delta L_{12}$	$-2\Delta L_{13}$	$-2\Delta L_{23}$
Affiliated Fund	Growth/income	1	45.773	48.335	2.562
Amer. Business Shs.	Balanced	2	26.224	36.645	10.421
American Mutual	Growth/income	3	36.176	N.A.	N.A.
Anchor Income	Income	4	61.631	N.A.	N.A.
Axe-Houghton Fund B	Balanced	5	48.114	N.A.	N.A.
Axe-Houghton Stock	Max. capital gain	6	47.472	N.A.	N.A.
Bullock Fund	Growth/income	7	71.986	76.558	4.572
Chase Fund of Boston	Max. capital gain	8	24.435	N.A.	N.A.
Chemical Fund	Growth	9	22.273	22.273	15.120
Colonial Fund	Growth/income	10	105.838	N.A.	N.A.
Delaware Fund	Growth	11	38.687	74.171	35.484
Dreyfus Fund	Growth	12	37.132	N.A.	N.A.
Eaton & Howard Bal.	Balanced	13	39.216	N.A.	N.A.
Eaton & Howard Stk.	Growth/income	14	72.021	87.887	15.866
Fidelity Growth	Growth	15	13.455	18.065	4.610
Fidelity Fund	Growth/income	16	40.998	N.A.	N.A.
Fidelity Trend	Growth	17	50.237	N.A.	N.A.

George Putnam Fund	Balanced	18	31.896	37.222	5.236
Invest. Co. of Amer.	Growth/income	19	70.418	N.A.	N.A.
Keystone K-2	Growth	20	12.505	N.A.	N.A.
Keystone S-2	Growth/income	21	31.571	N.A.	N.A.
Loomis-Sayles Mutual	Balanced	22	38.005	47.595	9.590
Massachusetts Fund	Balanced	23	54.618	68.111	13.493
Mass. Investors Gr. Stk.	Growth	24	18.366	25.382	7.016
Mass. Investors Trust	Growth/income	25	89.326	N.A.	N.A.
Nation-Wide Sec.	Balanced	26	65.980	72.649	6.669
One William Street	Growth/income	27	76.352	98.489	22.137
Oppenheimer Fund	Max. capital gains	28	13.264	23.830	10.566
Pioneer Fund	Growth/income	29	9.854	14.007	4.153
Puritan Fund	Income	30	N.A.	18.668	N.A.
Putnam Growth Fund	Growth	31	N.A.	29.906	N.A.
Stein R&F Bal.	Balanced	32	19.649	47.540	27.891
T. Rowe Price Gr. St.	Growth	33	10.578	N.A.	N.A.
Value-Line Income	Income	34	13.270	N.A.	N.A.
Value-Line Spec. Sit.	Max. capital gain	35	15.418	N.A.	N.A.
Wellington Fund	Balanced	36	121.116	130.950	9.834
Windsor Fund	Growth	37	15.252	N.A.	N.A.

NOTE.—N.A. refers to an unsuccessful attempt at reaching an interior optimum.

by the three-regime alternative at the .05 level of significance (exceeded 15.51). Therefore, the sample is assumed to consist of 12 mutual funds where rates of return are generated by three regression equations and 25 mutual funds having returns generated by the two-regime model.

C. *Parameter Estimates*

Tables 2 and 3 contain the parameter estimates and their differences between regimes for the funds classified as two-regime and three-regime models, respectively. A test for the existence of timing activities is whether the risk level (beta) decisions between regimes are significantly different. Of the 25 two-regime funds, 15 had t-values that reject the equality of beta at the .05 level of significance and 12 at the .01 level. All of the 12 three-regime funds had at least one pair of beta estimates that were significantly different at the .01 level. Hence, the sample exhibits a considerable amount of market-timing activities.[8] In many cases, the portfolio decisions that changed portfolio risk also significantly affected selectivity performance (differences in $\hat{\alpha}$'s) and portfolio diversification (differences in $\hat{\sigma}^2$s).

D. *Timing Performance*

In order to implement the timing performance estimates in equations (2) and (3) we require (*a*) the time series of beta estimates for each fund, (*b*) a proxy for the target beta of the fund, and (*c*) a proxy for the consensus expected return on the market portfolio at each observation to calculate the time series of unanticipated market returns.

The time series of beta estimates are obtained by the most likely classification rule in equation (8). The proxy used for the target beta of each fund is the average risk level of all funds in its stated objectives class. Finally, three alternatives for a model of consensus expected market portfolio returns at the beginning of each month are used: (1) the sample mean of the market portfolio, $\bar{R}_M$, over the measurement interval (January 1960–June 1976); (2) the current riskless rate plus the sample mean excess market returns, $R_{Ft} + (\bar{R}_M - \bar{R}_F)$, over the measurement interval; and (3) $R_{Ft} + (\bar{R}_M - \bar{R}_F)$ where the estimate of the

8. Although beta nonstationarity can arise from a passive portfolio due to ex post market movements and/or nonstationarity of the underlying securities, it is still attributed to the timing activity of the fund manager. The reason is that as a decision variable, if a fund manager is not forecasting market movements, he should rebalance portfolio composition to achieve the target risk level when acting in the best interests of shareholders. Hence, not fully rebalancing can be a least-cost method of implementng an active timing decision. Furthermore, any transactions cost bounds on rebalancing would not result in the magnitude or statistical significance of the risk level differences found in tables 2 and 3.

excess rate of return on the market portfolio, $(\bar{R}_M - \bar{R}_F)$, is .7666% per month (9.2% annual rate over the 1926–76 period reported by Ibbotson and Sinquefield [1977], p. 19).[9] Alternative 3 is a model including information prior to as well as during the measurement interval for the purpose of forming expectations. Alternative 3 is particularly important here since the sample mean excess market returns over the measurement interval of .2746% per month was considerably less than the 51-year mean of .7666%. Hence, alternatives 1 and 2 may be downward biased estimates of consensus expected return on the market portfolio if realizations on this portfolio during the measurement interval were on average less than the consensus expected. If alternative 3 is the most appropriate model for the formation of consensus expected returns, then the measurement interval provides an exceptional opportunity for those managers who are particularly good at forecasting bear markets to exhibit superior timing performance.

Since the empirical results for alternatives 1 and 2 are essentially the same, only the results for alternatives 2 and 3 are reported here.[10] Assuming that consensus expectations on the market portfolio returns are time varying and equal to the current riskless rate plus the ex post sample mean excess market return over the measurement interval, the time series of timing performance estimates (eq. [2]) were calculated for each fund. Table 4 reports the estimate of overall, or average, timing performance (eq. [3]) for each fund. Of the 37 mutual funds, 14 had overall timing performance estimates that were positive but none had a t-statistic indicating that performance was significantly greater than zero at a reasonable probability level (i.e., $< .10$). Hence, at the individual fund level it seems that there is no evidence of significant superior timing performance.

However, when substituting the estimates of β_t and π_{Mt} in equation (2) the distribution of $\hat{\tau}_t$ may not be normal as required for the validity of the t-statistic. For the 198 timing performance observations on each fund a studentized range statistic greater than 7.03 when sampling from a normal distribution is expected to occur only once in 200 replications (David, Hartley, and Pearson 1954, p. 491), .005 probability level.[11] Therefore, the evidence in table 4 indicates that the normality assump-

9. The models in alternatives 2 and 3 are the state of the art for estimating the expected rate of return on the market portfolio. We would have also used the estimator proposed by Merton (1980), but its use requires a specification for the *true* variance changing process for the market portfolio that is currently not known. Although investigating this specification is beyond the scope of this paper, it should be noted that it is an important area for future empirical research.

10. See Kon (1982) for the empirical evidence given the constant consensus expectations on the market portfolio alternative 1.

11. The studentized range statistic is the difference between the maximum and minimum observations in the sample divided by the sample standard deviation.

TABLE 2 Two-Regime Parameter Estimates, Differences and t-Values

ID Number	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\lambda}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\lambda}_2$	$\hat{\alpha}_2 - \hat{\alpha}_1$	$\hat{\beta}_2 - \hat{\beta}_1$	$\hat{\sigma}_2^2 - \hat{\sigma}_1^2$
	1.3500	.7454	.1905	− .1090	.9171	.8095	−1.4590	.1717	− 5.1691
1	(2.0196)	(6.6588)	(2.7108)	(−1.1517)	(41.1998)	(11.5214)	(−2.1143)	(1.4584)	(− 2.9414)
	.6504	.7794	.4238	− .1019	.8603	.5762	− .7524	.0809	− 8.3788
3	(1.7246)	(9.0059)	(4.5699)	(− .7324)	(28.0305)	(6.2144)	(−1.7745)	(.8205)	(− 4.4070)
	.4103	.3454	.1693	− .1243	.7170	.8307	− .5346	.3716	− 5.8043
4	(.7681)	(3.5355)	(2.7542)	(−1.3354)	(25.9683)	(13.5093)	(− .9629)	(3.7777)	(− 2.6282)
	− .0536	.6828	.7588	.4325	.7170	.2412	.4862	.0342	13.9268
5	(− .3968)	(14.0788)	(9.3396)	(.6756)	(6.0052)	(2.9686)	(.7090)	(.2332)	(3.0600)
	− .0202	.7647	.8433	2.2192	1.6965	.1567	2.2394	.9318	17.3307
6	(− .1052)	(17.5839)	(13.9580)	(1.8927)	(5.0702)	(2.5933)	(1.8586)	(2.7634)	(2.3522)
	.2702	.8014	.3539	− .0843	.9874	.6461	− .3545	.1860	−11.5635
7	(.6099)	(7.5730)	(5.1918)	(− .8220)	(38.9351)	(9.4791)	(− .7541)	(1.6806)	(− 4.4010)
	− .5560	1.1574	.7519	1.2241	1.6915	.2481	1.7801	.5341	20.1860
8	(−1.6254)	(15.0000)	(4.5898)	(1.2261)	(6.7235)	(1.5146)	(1.7291)	(2.0783)	(1.9337)
	− .0627	.7198	.1817	.0095	.8802	.8183	.0722	.1604	−18.6654
10	(− .0809)	(5.7355)	(3.4549)	(.1000)	(33.9466)	(15.5566)	(.0908)	(1.2243)	(− 3.1185)
	.6779	.7810	.2763	− .1419	1.0391	.7237	− .8198	.2580	− 6.3186
12	(1.2364)	(7.1516)	(2.4439)	(−1.1364)	(38.8588)	(6.4018)	(−1.3688)	(2.3016)	(− 3.3652)
	− .0931	.5935	.7624	.0029	.6525	.2376	.0960	.0590	5.4215
13	(−1.1138)	(25.7290)	(10.6659)	(.0073)	(7.6708)	(3.3246)	(.2276)	(.6221)	(3.0763)
	− .4241	.9740	.5079	.7278	1.3497	.4921	1.1518	.3757	6.8513
15	(−1.5100)	(12.7557)	(3.1262)	(1.5593)	(12.6003)	(3.0289)	(2.0198)	(3.0120)	(3.2367)
	.9034	.6519	.2083	− .0751	1.0000	.7917	− .9785	.3481	− 1.2971

16	(2.3689)	(10.3402)	(2.7185)	(− .8171)	(42.5358)	(10.3325)	(−2.4362)	(5.5684)	(− 2.0357)
	− .2069	1.0986	.8810	5.7121	1.8056	.1190	5.9190	.7070	6.0219
17	(−1.0051)	(25.4066)	(15.6480)	(3.0621)	(7.0904)	(2.1139)	(3.2673)	(2.6645)	(.9184)
	.2351	.5703	.3664	− .0432	.8591	.6336	− .2783	.2889	− 1.9367
18	(.9868)	(8.0684)	(2.8229)	(− .4601)	(27.6481)	(4.8821)	(−1.0266)	(4.5043)	(− 3.3531)
	.6208	.7999	.4249	− .0280	1.0100	.5751	− .6488	.2100	− 8.1795
19	(1.7752)	(10.0463)	(5.8455)	(− .3032)	(44.6896)	(7.9117)	(−1.7339)	(2.4868)	(− 5.0202)
	.4829	.8193	.1542	− .1124	1.1797	.8458	− .5953	.3603	−12.8098
20	(.4332)	(2.5606)	(1.2720)	(− .5485)	(27.0091)	(6.9793)	(− .5001)	(1.0992)	(− 1.8460)
	.2093	.6264	.2443	− .0774	1.0660	.7557	− .2866	.4396	− 2.6104
21	(.5274)	(5.2649)	(2.8062)	(− .7184)	(41.4131)	(8.6817)	(− .6555)	(3.7770)	(− 2.5587)
	− .1642	.9378	.3979	.0410	1.1069	.6021	.2053	.1691	− 6.2710
24	(− .4287)	(10.2406)	(2.7346)	(.2429)	(23.5300)	(4.1388)	(.4327)	(1.4615)	(− 3.3515)
	− .0450	.9038	.6877	− .0087	.9476	.3123	.0364	.0438	7.9087
25	(− .6148)	(41.0917)	(9.9704)	(− .0227)	(10.9234)	(4.5270)	(.0910)	(.4573)	(4.0300)
	− .7262	.5511	.2056	− .0952	.6663	.7944	.6309	.1153	− 9.1946
26	(−1.2648)	(4.2815)	(3.3372)	(−1.0904)	(33.0677)	(12.8978)	(1.0570)	(.8602)	(− 3.1857)
	.5109	.6949	.5669	.0576	1.0207	.4331	− .4533	.3257	− 1.5007
29	(1.8941)	(8.8775)	(2.9892)	(.2382)	(15.2274)	(2.2835)	(−1.0826)	(4.3128)	(− 1.6387)
	− .0374	.6229	.2297	− .1137	1.1393	.7703	− .0763	.5165	− .3955
33	(− .0946)	(2.8571)	(1.3375)	(− .7352)	(17.7147)	(4.4853)	(− .1637)	(2.9506)	(− .2943)
	1.4688	.7296	.3655	− .7226	.8211	.6345	−2.1914	.0915	− 2.2144
34	(1.5645)	(9.6931)	(1.7248)	(−2.2629)	(17.8849)	(2.9939)	(−2.5709)	(.9025)	(− 1.5161)
	−1.9976	.9247	.2490	.5182	1.6834	.7510	2.5159	.7587	19.5787
35	(−2.7180)	(8.6341)	(2.3594)	(1.0911)	(12.4674)	(7.1162)	(2.9131)	(4.6593)	(5.7826)
	.6940	.9012	.7270	−1.1041	1.2745	.2730	−1.7980	.3733	− 1.0665
37	(2.9483)	(17.6988)	(5.6809)	(−2.0262)	(16.6223)	(2.1333)	(−3.2251)	(4.6326)	(− 1.0636)

NOTE.—The numbers in parentheses are t-statistics.

TABLE 3 Three-Regime Parameter Estimates, Differences and *t*-Values

ID Number	$\hat{\alpha}_1$	$\hat{\beta}_1$	$\hat{\lambda}_1$	$\hat{\alpha}_2$	$\hat{\beta}_2$	$\hat{\lambda}_2$	$\hat{\alpha}_3$	$\hat{\beta}_3$	$\hat{\lambda}_3$
	– .0987	.5348	.6953	1.1718	.5718	.2159	–1.2505	1.0007	.0888
2	(– .4443)	(15.0080)	(2.3481)	(.9668)	(9.0337)	(.7402)	(– 4.2391)	(27.1790)	(2.4101)
	.5990	.2896	.1030	.0450	1.0074	.8377	3.1948	1.1948	.0593
9	(2.3962)	(7.9411)	(2.6378)	(.2611)	(24.1383)	(8.4903)	(.7076)	(2.9944)	(.6322)
	.7199	.9208	.5062	– .3990	1.1195	.4720	– .5282	3.2131	.0218
11	(2.3951)	(15.7306)	(5.6647)	(–3.0322)	(34.4677)	(5.2815)	(– 3.2700)	(84.1057)	1.7773)
	(– .3169)	.7004	.1121	– .0485	.8946	.6114	– .3414	1.2049	.2765
14	(– .3890)	(3.4942)	(2.6490)	(– .4864)	(34.9739)	(6.7157)	(– 1.5403)	(26.5316)	(2.9518)
	.5239	.3829	.0931	– .0456	.7215	.7124	– .1048	.9746	.1945
22	(1.3960)	(5.3818)	(1.7340)	(– .5506)	(26.8111)	(6.9109)	(– .3218	(13.5546)	(2.1328)
	.0794	.4832	.1555	– .1109	.6303	.3301	.0306	.7836	.5144
23	(.1557)	(3.9609)	(2.3231)	(–1.2131)	(31.8478)	(3.1047)	(.2289)	(23.7630)	(4.5771)
	1.5023	.2493	.0328	– .0175	.9664	.4015	– .2065	1.0366	.5657
27	(32.1189)	(41.9550)	(2.2048)	(– .0763)	(20.6751)	(3.3277)	(– 2.4010)	(47.0993)	(4.6771)
	–1.9291	.5792	.0554	– .5063	.9685	.2652	.6256	1.2814	.6794
28	– .7163)	(3.4846)	(.9565)	(–2.5146)	(28.0965)	(3.4221)	(2.0937)	(13.3602)	(7.1806)
	2.0040	.6194	.1542	– .1211	.7240	.7527	– .3470	1.2200	.0931
30	(.7529)	(6.5084)	(.6222)	(– .5965)	(18.6591)	(3.0193)	(– .9040)	(15.7399)	(1.6732)
	1.7513	.3057	.0912	– .1825	.9989	.6943	.8462	1.5652	.2145
31	(2.3199)	(3.4599)	(1.7636)	(– .7771)	(24.3428)	(6.3447)	(1.8956)	(13.0203)	(2.3218)
	.2914	.0917	.0584	– .0585	.6487	.2890	.0244	.9083	.6525
32	(2.0988)	(2.8180)	(2.3971)	(– .7872)	(31.8116)	(4.5786)	(.2026)	(31.5965)	(9.7776)
	1.3466	.3640	.0769	– .1780	.7032	.8995	.6578	1.7280	.0235
36	(1.5542)	(2.7126)	(2.1259)	(–2.6547)	(38.7414)	(22.7294)	(.4990)	(7.8031)	(1.4563)

	$\hat{\alpha}_2 - \hat{\alpha}_1$	$\hat{\beta}_2 - \hat{\beta}_1$	$\hat{\sigma}_2^2 - \hat{\sigma}_1^2$	$\hat{\alpha}_3 - \hat{\alpha}_2$	$\hat{\beta}_3 - \hat{\beta}_2$	$\hat{\sigma}_3^2 - \hat{\sigma}_2^2$	$\hat{\alpha}_3 - \hat{\alpha}_1$	$\hat{\beta}_3 - \hat{\beta}_1$	$\hat{\sigma}_3^2 - \hat{\sigma}_1^2$
2	1.2705	.0370	1.3541	−2.4224	.4288	−1.7647	−1.1519	.4659	−.4106
	(1.1614)	(.4547)	(1.4582)	(−1.9296)	(5.6481)	(−1.8337)	(−3.1301)	(9.2454)	(−.8685)
9	−.5540	.7178	3.7884	3.1498	.1305	10.2835	2.5958	.8484	13.0719
	(−1.8185)	(13.2940)	(3.9600)	(.7016)	(.3391)	(1.1623)	(.5738)	(2.2116)	(1.5007)
11	−1.1189	.1987	4.9520	−.1292	2.0935	−.6067	−1.2481	2.2923	−5.5588
	(−3.2835)	(2.8827)	(5.3550)	(−.6224)	(41.7627)	(−2.6162)	(−3.6834)	(32.8749)	(−5.7760)
14	.2685	.1942	−10.7638	−.2929	.3103	.1812	−.0245	.5045	−10.5826
	(.3283)	(.9698)	(−2.5507)	(−1.0719)	(6.9388)	(.4757)	(−.0276)	(2.5837)	(−2.5106)
22	−.5695	.3386	−.2544	−.0592	.2531	2.1204	−.6287	.5917	1.8660
	(−1.4801)	(4.7685)	(−.5946)	(−.1692)	(3.5257)	(2.1666)	(−1.2419)	(5.9083)	(1.6910)
23	−.1902	.1472	−4.9086	.1414	.1533	.6087	−.0488	.3005	−4.2999
	(−.3739)	(1.1762)	(−2.8657)	(.7708)	(4.3170)	(2.5160)	(−.0856)	(2.2780)	(−2.5547)
27	−1.5198	.7171	3.1493	−.1890	.0702	−2.7810	−1.7088	.7873	.3683
	(−6.4709)	(15.2250)	(3.9461)	(−.7073)	(1.2568)	(−3.8386)	(−17.6500)	(34.6235)	(2.5784)
28	1.4228	.3893	−4.8335	1.1318	.3129	7.3185	2.5547	.7022	2.4849
	(.5284)	(2.3138)	(−.7540)	(3.1022)	(3.0900)	(6.2713)	(.9230)	(3.7099)	(.3800)
30	−2.1251	.2046	−1.3255	−.2259	.4960	−.9098	−2.3510	.6006	−2.2352
	(−.8452)	(.9140)	(−.4449)	(−.4736)	(6.3717)	(−2.0841)	(−.8598)	(4.5905)	(−.7842)
31	−1.9338	.6932	1.1615	1.0287	.5663	.8832	−.9051	1.2595	2.0447
	(−2.2510)	(6.7561)	(.6733)	(1.9711)	(4.7201)	(.7107)	(−.9319)	(8.7483)	(.9473)
32	−.3500	.5570	.0121	.0829	.2596	1.3529	−.2671	.8166	1.3649
	(−2.2236)	(14.5356)	(.1457)	(.5544)	(7.6090)	(6.2233)	(−1.4527)	(19.1652)	(6.0632)
36	−1.5246	.3392	−5.9834	.8358	1.0248	.1947	−.6888	1.3641	−5.7887
	(−1.7545)	(2.5537)	(−2.0129)	(.6329)	(4.6204)	(.2442)	(−.4343)	(5.3109)	(−1.8737)

NOTE.—The numbers in parentheses are t-statistics.

TABLE 4 **Market-timing Performance**
$E_t(R_{Mt}) = R_{Ft} + [\bar{R}_M - \bar{R}_F]$ January 1960–June 1976

ID Number	Overall Timing	t-Statistic	Student Range	Sign Statistic	Signed Rank Statistic
4	.0256	.3976	11.3359	.9239	.7048
30	.0249	.5332	11.4527	.4975	.6131
34	−.0010	−.0365	8.5997	.6396	.0192
5	−.0083	−1.4799	8.5079	−.9239	−1.5793
13	.0027	.1035	6.6284	−.6396	−.3029
18	−.0098	−.2118	7.1915	.2132	.0508
22	.0069	.1217	10.4909	1.6345	1.2826
2	.0049	.0895	10.1280	−.4975	−.5066
23	−.0010	−.0279	9.9886	−.3553	−.4069
26	.0088	.4130	9.6893	−.6396	−.4075
32	−.1009	−1.4737	7.3701	−1.9188	−1.7775
36	−.0524	−.7267	12.9681	1.0660	.8900
1	.0188	.7929	11.4159	1.2081	.8404
3	−.0108	−.4153	8.2017	−.6396	−.5017
7	.0015	.0529	7.2868	1.0660	.5636
10	.0370	1.0294	10.2291	−.6396	.2168
14	−.0556	−.9269	10.0729	−2.2031	−1.7224
16	.0015	.0300	9.1981	.2132	.3697
19	−.0478	−1.4864	6.5012	−1.0660	−1.2238
21	−.0598	−1.0130	7.3547	.0711	−.6460
25	−.0027	−.3246	8.9785	−.9239	−.6936
27	.0094	.1501	14.7966	.9239	.6825
29	−.0492	−.9408	7.9793	−2.3452	−1.5242
9	−.0542	−.6022	12.6208	−.0711	.3803
11	−.0771	−.8866	16.6114	−2.3452	−1.9664
12	−.0180	−.4297	10.8976	−.6396	−.3313
15	−.0016	−.0240	9.4329	.7817	.7036
17	.0468	.7454	11.5405	.6396	.7494
20	−.0036	−.0750	8.0601	.2132	.0359
24	.0056	.2177	6.4792	.4975	.5846
33	−.0202	−.2992	8.2753	.0711	−.3431
31	−.0026	−.0252	10.1770	1.3503	1.5025
37	−.0606	−1.1535	7.9804	−1.7767	−1.4387
6	−.0435	−.3249	6.3700	.2132	−.3803
8	.0403	.5781	9.8024	1.2081	1.5174
28	−.0411	−.5547	14.4682	.0711	−.3264
35	−.1244	−.9365	6.6749	−.7817	−.8107

Multivariate Tests

F	Sign	Signed Rank
.6477	1.7052	38.8675

tion is rejected for each of 32 funds at the .005 level and individually all reject it at the .05 level. Hence, better inferences may be available with nonparametric methods. In particular, the sign and Wilcoxon signed rank test for location.

These nonparametric tests also allow us to separate the manager's ability to correctly predict unanticipated market movements and shift the fund's risk level accordingly (timing ability) from the combined effects of the size of the market turn and the magnitude of the risk level shifts (timing performance). This distinction is made in the ordinary sign test and the signed rank test, respectively.[12] The sign test, however, should be interpreted primarily as a check that significant positive overall timing performance was not a result of one or two exceptionally large market movements. For this model, however, there was no significant positive overall timing performance reported. Of the 37 funds, 20 did have positive sign test statistics of which two were significant at the .10 probability level and none were significant at the .05 level. Therefore, even the distribution free sign test could not detect very much significant superior forecasting ability at the individual fund level. Of the 20 funds with positive sign timing *ability*, four had negative *performance* when considering the magnitude effects in the signed rank test. Although 18 funds had positive performance in the signed rank test, only three funds had significant performance at the .10 probability level and no individual fund could reject a zero location value at the .05 significance level. The implication for funds with negative timing performance, and in particular the many with significant negative timing performance, is that the allocation of effort and expense of timing activities should be reevaluated.

Any cross-sectional inferences such as for the efficient markets hypothesis requires accounting for the interdependence among the fund decisions and market movements. That is, if fund managers as a group have no special information regarding consensus unanticipated market movements, the vector of overall (average) timing performance of all 37 funds in the sample should not be significantly different from a vector of zeros. The multivariate analogue to a univariate t-statistic is the Hotelling T^2 statistic.

Under the null hypothesis, $F = (N - P)T^2/P(N - 1)$ has an F distribution with degrees of freedom P and $N - P$. For this sample of funds the F-statistic is .6477 with $P = 37$ and $N - P = 198 - 37 = 161$ degrees of freedom. This is not inconsistent with the null hypothesis.

12. The large sample univariate sign test statistic is $Z = (S - 0.5 - 0.5\,n)/0.5\sqrt{n}$. S is either S_+ or S_- where S_+ is the number of plus signs observed and S_- is the number of minus signs observed. The total number of observations is n, and the second term in the numerator (-0.5) is the continuity correction. The mean is $0.5n$ and the standard deviation is $0.5\sqrt{n}$. In the tables of the text the Z value associated with the larger of S_+ or S_- is reported. If $S_- > S_+$, the Z statistic has a minus sign in front. For the sample sizes in this study the sampling distribution of Z is the standard normal. The large sample univariate Wilcoxon signed rank statistic is $Z = \{T - .5 - [n(n + 1)/4]\}/\sqrt{n(n + 1)(2n + 1)/24}$. T is either T_+ or T_- where T_+ is the sum of positive ranks and T_- is the sum of negative ranks. .5 in the numerator is the continuity correction. $n(n + 1)/4$ is the mean and the denominator is the standard deviation. The Z-value for the larger of T_+ or T_- is reported in the text where minus signs in front of the Z-statistic are used to indicate that $T_- > T_+$. The sampling distribution of the Wilcoxon signed rank Z-statistic is also well approximated by the standard normal.

TABLE 5 Market-timing Performance
$E_t(R_{Mt}) = R_{Ft} + [\bar{R}_M - \bar{R}_F]$, 1926–1976

ID Number	Overall Timing	t-Statistic	Student Range	Sign Statistic	Signed Rank Statistic
4	.0501	.7742	11.2526	−.3553	−.3406
30	−.0002	−.0045	11.4178	−.2132	−.2818
34	−.0403	−1.5039	8.5837	−.9239	−1.5143
5	−.0073	−1.2891	8.3891	.2132	−.7259
13	.0450	1.7302	6.6209	.9239	1.2913
18	−.0404	−.8655	7.1540	−.2132	−.5778
22	−.0158	−.2754	10.0432	.0711	.0526
2	.0571	1.0500	10.1094	.0711	.6621
23	.0021	.0564	9.8366	−.6396	−.4948
26	.0336	1.5651	9.6060	.7817	1.0795
32	−.1469	−2.1646	7.7961	−1.9188	−2.5925
36	−.0522	−.6852	12.8976	−.2132	−.3053
1	.0282	1.1725	11.2424	−.0711	.1034
3	.0271	1.0475	8.2204	.9239	1.0064
7	−.0062	−.2192	7.2414	1.0660	.1802
10	.0640	1.7558	10.0992	.7817	1.5632
14	−.0828	−1.3569	9.9091	−.6396	−.9798
16	−.0098	−.2015	9.1481	−.4975	−.4341
19	−.0550	−1.7051	6.4708	−.7817	−1.4560
21	−.0888	−1.5062	7.1048	−.6396	−1.4090
25	−.0089	−1.0545	8.9205	.0711	−.2093
27	−.0300	−.4703	14.6137	−.3553	−.8757
29	−.0171	−.3253	7.9431	−1.4924	−1.0683
9	−.0056	−.0634	12.7514	1.0660	1.5143
11	−.0814	−.9120	16.1810	−2.6295	−1.7899
12	.0165	.3940	10.9008	.7817	1.0095
15	−.0540	−.8235	9.3843	1.0660	.4193
17	−.0224	−.3606	11.6847	−.9239	−.9488
20	−.0463	−.9570	8.0311	−.7817	−.9946
24	.0068	.2611	6.2078	−.2132	.3716
33	−.0138	−.2041	8.2527	−1.3503	−1.2449
31	−.0107	−.1043	10.1417	1.9188	1.9484
37	−.0442	−.8306	7.8759	−.9239	−.9971
6	.0884	.6553	6.3305	.7817	.6169
8	−.0083	−.1194	9.8134	1.3503	1.6251
28	−.0332	−.4507	14.5564	.0711	−.7890
35	−.2606	−1.9559	6.6529	−.7817	−1.6586

Multivariate Tests

F	Sign	Signed Rank
.6721	1.7278	34.9024

That is, an efficient market with respect to the formation of expectations on the return of the market portfolio.

The Hotelling T^2 test requires that the timing performance estimates be multivariate normally distributed. Because this is contrary to the studentized range evidence, the nonparametric multivariate sign and sign rank statistics were calculated. The multivariate sign statistic of 1.7052 and multivariate signed rank statistic of 38.8675 both have asymptotic χ^2 distributions with $P = 37$ degrees of freedom.[13] Neither test, however, could reject the efficient markets null hypothesis.

In table 5 the model of consensus expected market portfolio returns uses information in the data before as well as during the measurement interval. For this model of the formation of consensus expectations, there is considerable evidence of *superior* market-timing performance at the individual fund level with both parametric and nonparametric tests. Recall that the studentized range value indicates a higher sample standard deviation than would be expected from a normal distribution. Hence, the parametric t-statistic may be biased downward. In spite of this potential bias, there were three funds that reject the no-market-timing-performance null hypothesis at the .10 probability level, and two of these funds were significantly positive at the .05 level. Their respective sign tests were all positive, indicating that their significant positive timing performance was not due to a few outliers. The sign tests for the entire sample indicated that two funds exhibited significantly positive timing ability at the .10 probability level of which one was significant at the .05 level. The most impressive superior market-timing performance result was registered by the Wilcoxon signed rank test. Five funds had tests that rejected the null hypothesis in favor of the superior performance alternative at the .10 probability level, and two of these funds had tests that were significant at the .05 level.

Although there is evidence of superior market-timing ability and performance at the individual fund level, none of the multivariate tests could reject the efficient markets hypothesis.

In sum, the timing ability and performance measures of mutual fund managers were sensitive to the model of consensus expectations on the market portfolio. However, for the most rational model the parametric and nonparametric tests were able to detect significant positive timing ability and performance at the individual fund level but not at the level required for rejection of the efficient markets hypothesis. The power of these tests to detect significant positive and negative timing ability and performance at the individual fund level is important for the allocation of resources to the timing activity versus the selectivity activity or a low-cost passive investment strategy.

E. Selectivity Performance

The ability of an investment manager to select undervalued securities for their portfolio is generally measured relative to a naive strategy of

13. The general nonparametric multivariate statistic is $S_n = 1/n\,(T_n V_n^{-1} T_n')$, where $T_n^j = \Sigma_{t=1}^n E_t^j C_{jt}$, $j = 1, 2, \ldots, P$; $T_n = [T_n^1, \ldots, T_n^P]'$; $C_{jt} = +1$ or -1 if $\tau_{jt} > 0$ or < 0, respectively; $V_n = [(\nu_{njk})]$, $j, k = 1, \ldots, P$, and $\nu_{njk} = 1/n\,\Sigma_{t=1}^n E_t^j E_t^k C_{jt} C_{kt}$. If the general score $E_t^j = 1$ for all $t = 1, \ldots, n$ and $j = 1, \ldots, P$, then S_n is the multivariate sign statistic. Alternatively, if $E_t^j = r_{jt}/(n + 1)$ where r_{jt} is the rank of $|\tau_{jt}|$ among all $|\tau_{jt}|$ $t = 1, \ldots, n$, then S_n leads to the multivariate generalization of the one-sample Wilcoxon signed rank statistic. See Puri and Sen (1971), chap. 4.

a combined investment in the market portfolio and riskless asset with the identical level of systematic risk. Hence, the Jensen (1968) risk-adjusted selectivity performance estimate for a fund at time t is

$$\begin{aligned} \hat{\delta}_t &= R'_t - E(R'_t|R'_{Mt}, \hat{\beta}_t) \\ &= R'_t - \hat{\beta}_t R'_{Mt}, \end{aligned} \tag{11}$$

and overall, or average, selectivity performance is

$$\hat{\delta}_O = \frac{1}{n} \sum_{t=1}^{n} \hat{\delta}_t. \tag{12}$$

Table 6 reports that 23 out of the 37 funds had positive overall selectivity performance. Evidence of superior performance at the individual fund level is the five funds that had significantly positive performance at the .05 level. In order to determine whether this is more than by random chance, the interdependence of manager's decisions (e.g., same general management and stocks generally in favor) can be accounted for by the Hotelling T^2 test. The associated F-statistic of 2.2290 with 37 and 161 degrees of freedom is significant at the .005 probability level. Hence, fund managers as a group have superior ability in forecasting individual security prices. This evidence appears to clearly reject the strong form of the efficient markets hypothesis with respect to stock selection.

The studentized range tests, however, generally reject the normality assumption required by the Hotelling T^2 test. Both the nonparametric multivariate sign and signed rank tests also reject the zero location vector hypothesis at the .005 level. But the reason for rejection is quite different. The rejection of normality is actually due to the manipulation of portfolio composition resulting in the mixture of normal generating model. That is,

$$\delta_O = \frac{1}{n} \sum_{t=1}^{n} \delta_t = \frac{1}{n} \sum_{t=1}^{n} (\alpha_t + u_t), \tag{13}$$

where the δ_ts are not identically distributed. Tables 2 and 3 indicate that many funds had significantly different residual variances among regression regimes. This is the cause of the observed "fat tail" distribution relative to the normal for δ_t detected by the studentized range statistics. Furthermore, the significantly different α_ts for a fund implies that the mean of the time series of δ_ts is nonstationary. The resulting positive skewness (median < mean) is the reason for the many changes in sign between the individual t and sign statistics in tables 6 and 7. In the Hotelling T^2 test, it is the positive selectivity performance that is rejecting the null hypothesis, while in the multivariate sign test it is the negative performance. Therefore, the interpretation of these results

TABLE 6 Stock Selectivity Performance

ID Number	Overall Selectivity	*t*-Statistic	Student Range	Sign Statistic	Signed Rank Statistic
4	−.0539	− .5492	10.4974	−1.3503	−1.1619
30	.1713	1.7471	5.8189	− .2132	.8126
34	.0804	.5816	6.1729	−1.4924	− .2676
5	.0705	.4406	9.4381	− .0711	.0762
13	−.0708	− .7016	8.5475	− .4975	−1.0207
18	.0431	.5517	8.0488	− .9239	− .5041
22	−.0408	− .6575	10.5493	.3553	− .6720
2	.0920	1.0650	5.8103	.2132	.3759
23	.0047	.0623	9.8503	− .9239	.0892
26	−.2198	−1.8437	8.6058	−1.3503	−1.7490
32	.0786	1.1540	7.2405	.4975	1.0058
36	−.0643	− .8100	8.3774	−2.2031	−2.4265
1	.1634	1.5334	7.0488	.6396	.5716
3	.2238	1.4435	6.5908	.2132	.8974
7	.0487	.3087	7.9599	.0711	− .1133
10	−.0208	− .1392	12.7086	.4975	− .2044
14	−.1469	−1.6229	9.8594	−2.7716	−2.6631
16	.1014	1.3285	6.1564	.4975	.3159
19	.2729	1.8999	7.2873	.2132	.8150
21	−.0031	− .0357	8.9155	− .0711	− .4211
25	−.0318	− .2616	8.9851	− .6396	− .7655
27	−.0828	− .9349	6.9683	−1.6345	−1.4158
29	.3535	3.5077	6.3649	2.3452	3.1920
9	.2877	2.1119	7.9745	1.4924	1.7787
11	.1987	1.5055	7.0202	−2.0609	− .0285
12	.1031	.8335	8.2326	− .7817	.0161
15	.1215	.6996	7.2462	− .0711	− .2372
17	.4414	2.0414	7.1400	.2132	1.0058
20	−.0438	− .2520	7.3018	− .0711	− .7073
24	−.0593	− .4209	7.2679	−1.0660	− .5314
33	−.1013	−1.0198	6.1993	−1.6345	−1.1284
31	.1549	1.5417	6.6393	1.6345	1.6065
37	.2682	2.0350	5.8527	.3553	1.4152
6	.3347	1.7858	8.8532	.7817	.9897
8	−.1834	− .7887	9.2054	−2.4873	−1.8468
28	.2173	1.2570	6.8501	− .9239	.1307
35	.0176	.0583	7.5812	−1.7767	−1.2665

Multivariate Tests

F	Sign	Signed Rank
2.2290	73.1267	74.8932

depends on whether the mean or median or a combination of the two (i.e., including skewness) is the preferred measure of selectivity ability or performance for this asymmetric nonnormal distribution. Note that of all the tests presented only the sign test is not sensitive to skewness. Hence, any studies concerned with average selectivity performance over the *entire* measurement interval that rely on individual *t*-statistics, the multivariate Hotelling T^2 statistic or any nonparametric statistic that is sensitive to asymmetry may be drawing inappropriate inferences due to invalid distributional assumptions.

F. Overall Investment Performance

An examination of market-timing performance in table 5 and stock selectivity performance in table 6 reveals that there are six funds with positive performance in both timing and selectivity and that 22 funds exhibited a trade-off between the two activities. Although this sample of funds produced better selectivity (23 positive) than timing (11 positive) performance, there were five funds with positive timing and negative selectivity performance. Hence, the sample exhibited a considerable amount of activity specialization.

Table 7 provides the bottom line overall investment performance estimates from engaging in both market-timing and stock selectivity activities for each fund.[14] Out of the 37 funds in the sample, 22 had positive overall investment performance of which five had significant t-statistics at the .05 probability level. Only one fund had significant negative performance. The signed rank tests, however, had two significantly positive and three significantly negative. Again there is evidence of superior performance at the individual fund level. Of the 15 funds with negative overall investment performance, six could have had positive performance by specializing in one activity and taking a passive strategy with respect to the negative performance activity. The remaining funds with negative performance in both activities might consider a totally passive management strategy and just provide a diversification service to their clients.

Finally, the multivariate tests on overall investment performance all reject the zero location vector null hypothesis at the .025 probability level. This result is primarily that of the selectivity performance contribution whose interpretation was discussed in the previous subsection.

IV. Summary and Conclusions

This paper discusses the optimal behavior and performance measurement of an investment manager simultaneously engaged in market-timing and stock selectivity activities. The importance of an econometric model specification consistent with the risk level decisions of investment managers is evident from the resulting biases of ordinary least squares estimates of risk and selectivity performance when the timing activity is ignored. The methodology in this paper avoids these

14. The investment performance estimate for time t, $\hat{I}_t = \hat{\tau}_t + \hat{\delta}_t$, is the sum of the timing and selectivity performance estimates (eqq. [2] and [11]). Then the estimate of overall investment performance is $\hat{I}_0 = 1/n \, \Sigma_{t=1}^{n} \hat{I}_t$.

TABLE 7 Overall Investment Performance

ID Number	Overall Performance	*t*-Statistic	Student Range	Sign Statistic	Signed Rank Statistic
4	−.0037	− .0310	9.2806	− .7817	− .7859
30	.1711	1.5356	5.3174	.6396	1.1049
34	.0400	.2835	6.1960	−1.7767	− .4670
5	.0632	.3953	9.4808	− .2132	.0656
13	−.0259	− .2476	8.3139	.0711	− .3406
18	.0027	.0282	6.4991	− .9239	− .3908
22	−.0566	− .6204	7.9309	.0711	− .7073
2	.1491	1.4396	6.3383	− .4975	.8441
23	.0068	.0782	9.2594	− .3553	− .0954
26	−.1862	−1.5441	8.4236	−1.3503	−1.5818
32	−.0683	− .7013	6.8040	− .4975	− .8714
36	−.1164	−1.0588	12.1935	−2.2031	−2.5454
1	.1916	1.7120	6.9510	.4975	.7816
3	.2510	1.5921	6.6109	.3553	1.1544
7	.0425	.2641	7.6595	− .4975	− .3654
10	.0432	.2803	11.8690	.0711	− .0155
14	−.2297	−2.0220	8.0785	−1.6345	−2.4488
16	.0915	.9591	7.8274	−1.2081	− .0415
19	.2178	1.4647	6.9882	− .3553	.4484
21	−.0919	− .8187	7.5916	− .4975	−1.1990
25	−.0407	− .3338	9.2849	− .6396	− .7531
27	−.1127	−1.0105	8.3142	−1.9188	−2.0834
29	.3364	2.6597	5.7927	.6396	2.0382
9	.2820	1.7205	8.1545	.2132	1.1866
11	.1173	.7211	8.9993	− .7817	− .0248
12	.1196	.8904	7.2581	− .7817	− .0124
15	.0675	.3522	7.6395	.3553	.2143
17	.4190	1.8972	7.4523	.4975	.8918
20	−.0902	− .4834	6.9312	− .2132	− .7469
24	−.0525	− .3611	6.5654	− .6396	− .5103
33	−.1151	− .8741	6.6188	−2.0609	−1.2312
31	.1442	.9488	7.1548	.7817	.8398
37	.2240	1.4974	6.7113	.9239	1.2894
6	.4230	1.8002	8.3442	1.4924	1.6313
8	−.1917	− .7637	8.4242	−2.3452	−1.3359
28	.1841	.9545	7.3424	1.0660	.8609
35	−.2430	− .7175	6.7293	−2.0609	−1.3303

Multivariate Tests

		F		Sign	Signed Rank
		1.6995		52.5097	60.3835

biases and provides an empirically tractable separable measure of market-timing performance.

The empirical results indicate that individual mutual funds did exhibit significant positive timing ability and/or performance. However, the multivariate tests were not inconsistent with the efficient markets null hypothesis. That is, fund managers as a group have no special information regarding the unanticipated market portfolio returns.

An analysis of selectivity performance given the timing decisions of fund managers indicates that false inferences are likely with parametric and skewness sensitive nonparametric methods. For example, the distributional assumptions of the Student t and Hotelling T^2 tests on stock selectivity performance are violated severely due to the portfolio manipulation of fund managers.

The implications for investment managers are that most could improve overall investment performance considerably by reallocating resources to their more productive activity.

References

Black, F. 1972. Capital market equilibrium with restricted borrowing. *Journal of Business* 45 (3): 444–54.

Campanella, F. B. 1972. *The Measurement of Portfolio Risk Exposure*. Lexington, Mass.: Lexington.

David, H. A.; Hartley, H. O.; and Pearson, E. S. 1954. The distribution of the ratio, in a single normal sample, of range for standard deviation. *Biometrika* 61 (3): 482–93.

Fama, E. 1972. Components of investment performance. *Journal of Finance* 27 (2): 551–67.

Goldfeld, S. M., and Quandt, R. E. 1972. *Nonlinear Methods in Econometrics*. Amsterdam: North-Holland.

Grant, D. 1977. Portfolio performance and the cost of timing decisions. *Journal of Finance* 32 (3): 837–46.

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 (4): 513–33.

Ibbotson R. G., and Sinquefield, R. A. 1977. *Stocks, Bonds, Bills, and Inflation: The Past (1926–1976) and the Future (1977–2000)*. Charlottesville, Va.: Financial Analysts Research Foundation.

Jensen, M. C. 1968. The performance of mutual funds in the period 1945–1964. *Journal of Finance* 23 (2): 389–416.

Jensen, M. C. 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business* 42 (2): 167–247.

Jensen, M. C. 1972. Optimal utilization of market forecasts and the evaluation of investment performance. In Giorgio Szego and Karl Shell (eds.), *Mathematical Methods in Investment and Finance*. Amsterdam: North-Holland.

Kiefer, N. M. 1978. Discrete parameter variation: efficient estimation of a switching regression model. *Econometrica* 46 (2): 427–34.

Kon, S. J. 1982. The market timing performance of mutual fund managers. Working Paper Series no. 256, Salomon Brothers Center for the Study of Financial Institutions, New York University, March.

Kon, S. J., and Jen, F. C. 1979. The investment performance of mutual funds: an empirical investigation of timing, selectivity, and market efficiency. *Journal of Business* 52 (2): 263–89.

Lintner, J. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 41 (1): 13–37.

Merton, R. C. 1973. An intertemporal capital asset pricing model. *Econometrica* 41 (5): 867–87.

Merton, R. C. 1980. On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics* 8 (4): 323–61.

Merton, R. C. 1981. On market timing and investment performance. I. An equilibrium theory of value for market forecasts. *Journal of Business* 54 (3): 363–406.

Mossin, J. 1966. Equilibrium in a capital asset market. *Econometrica* 34 (4): 768–83.

Modigliani, F., and Pogue, G. 1975. Alternative investment performance fee arrangements and implications for regulatory policy. *Bell Journal of Economics* 6 (1): 127–60.

Press, S. J. 1972. *Applied Multivariate Analysis*. New York: Holt, Rinehart & Winston.

Puri, M. L., and Sen, P. K. 1971. *Nonparametric Methods in Multivariate Analysis*. New York: Wiley.

Quandt, R. E. 1972. A new approach to estimating switching regressions. *Journal of the American Statistical Association* 67:306–10.

Ross, S. 1976. The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 3 (December): 343–62.

Sharpe, W. F. 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. *Journal of Finance* 19 (4): 425–42.

Treynor, J. L., and Mazuy, K. K. 1966. Can mutual funds outguess the market? *Harvard Business Review* 44 (4): 131–36.

U.S. Securities and Exchange Commission. 1971. Institutional investor study report of the Securities and Exchange Commission. 92d Cong., 1st sess., House Document, No. 92-64. Washington, D.C.: Government Printing Office.