MUTUAL FUND PERFORMANCE: MEASUREMENT AND PREDICTION

W. F. Sharpe

March 1965

P- 3096

MUTUAL FUND PERFORMANCE: MEASUREMENT AND PREDICTION*

W. F. Sharpe**

The RAND Corporation, Santa Monica, California

Within the last few years considerable progress has been made in two closely related areas -- the theory of portfolio selection[1] and the theory of the pricing of capital assets under conditions of risk.[2] Portfolio selection theory is primarily normative: it provides the investment manager with methods for selecting preferred portfolios for his clients; while capital theory is essentially positive: it attempts to describe the manner in which investment managers (and investors) do in fact select their portfolios and to derive the implications of such behavior for the pricing of capital assets. Both areas are relevant for investigations of the performance of mutual funds. Such funds provide the economist with data concerning the workings of the capital market, and thus allow him to test some of his theories. The investor, more interested in his own welfare, is concerned with measuring the differences in performance among funds in the past and, even more important, attempting to find some method for predicting their differences in the future. The implications of mutual fund performance for capital theory have been described elsewhere;[3] this paper is concerned with the problems facing the investor -- how to measure and predict mutual fund performance.

In a recent article,[4] Jack L. Treynor has suggested a method for rating the performance of mutual funds. His technique can be used either for measuring a fund's performance after the fact, or for predicting its performance in the future, and Treynor apparently intended that it be used for both purposes.[5] However, there is no rule that states that

---

*I am grateful to Norman H. Jones of The RAND Corporation for helpful comments and suggestions.

the appropriate measure for past performance is also the best predictor of future preformance -- in fact, only by testing the predictive ability of alternative measures can the preferred one be selected. Moreover, there are good theoretical reasons to adopt a measure that differs somewhat from Treynor's to rate the past performance of a fund. This paper provides empirical information to aid in assessing the appropriateness of the measure proposed by Treynor for the two roles of interest. Briefly, we will show that Treynor's index does serve rather well in the role of a predictor, even though a superior measure of past performance is available. The effects on performance of other attributes (expense ratios and size) will also be described, and -- in the tradition of empirical studies of mutual funds -- we will provide the usual comparisons with the performance of the securities used to compute the Dow-Jones Industrial Average.

For expository purposes, Treynor's technique will not be presented at this point, but will be introduced later, after an alternative method for measuring mutual fund performance is described. This order or presentation is not intended to detract from Treynor's contribution, only to assist readers unfamiliar with the material.

THE TASKS OF THE MUTUAL FUND

A mutual fund can usefully be considered a voluntary association of investors who have decided to pool their resources and to hire a professional management (or management service). Some of the advantages of such an arrangement are attributable primarily to pooling, others to the use of professional management. One classification distinguishes among diversification, selection of risk class, and selection of incorrectly priced securities.

Perhaps the most important task of the mutual fund is to obtain for its investors the advantages of diversification. There is considerable risk attached to any given investment -- only rarely can the over-all rate of return on a specific stock be predicted with certainty. However, if an investor spreads his funds over many diverse investments, the total risk can be reduced, since a particularly poor outcome on one holding may be offset by a particularly good outcome on another. In theory, the investor could provide such diversification himself, but the

transaction cost might be substantial. By pooling his funds with those of others he can take advantage of the lower percentage transaction cost associated with large purchases and sales of securities. Moreover, the use of professional management to select the securities to be held may prove advantageous. It is not always completely obvious whether two securities move together or not, and effective diversification requires that the returns from holdings be relatively independent. Thus management as well as size may play a role in fulfilling this first function of the mutual fund.

Although diversification can reduce risk, it can rarely eliminate it entirely, The rate of return on a well-diversified portfolio will vary from year to year, and the variability of some portfolios will be greater than that of others. Since investors dislike such variability (because it makes the outcome during any given year uncertain) the prices of securities must adjust to provide some incentive to induce people to hold the more risky assets. For this reason portfolios showing greater variability in the rate of return typically also give greater returns on the average than of those whose outcome is more certain.[6] Given this relationship, some investors will prefer to take few chances, investing their funds in a manner that will provide a relatively small but relatively assured return. Others will be willing to incur more risk, realizing that in any given year they may do rather poorly but that on the average they will earn a greater return than those holding more conservative portfolios. These differences imply that in pooling their funds, investors may wish to associate with like-minded individuals. In practice, the managers of mutual funds provide investors with statements concerning the general risk-classes of the portfolios they intend to hold. To fulfill such commitments, managers must be able to discern the riskiness of individual securities and, even more important, the total risk of various combinations of securities. Considerable skill may be required to perform this task: relatively simple selection methods are likely to provide most of the advantages of diversification but they may not suffice to insure that the fund remain in the selected risk-class.

The market price of a security can be viewed as the consensus among

investors concerning its present value. If the participants in the market are well-informed and rational there will be little opportunity to find securities for which this consensus is in error. However, if some investors are poorly informed, or irrational when acting on good information, some securities may be underpriced or overpriced, at least temporarily. The third major task of the management of a mutual fund is to find such securities, buying them if underpriced and selling them if overpriced. To perform this task successfully, the manager must be better than his colleagues in the market, since in essence he is betting against them. Moreover, the mutual fund manager is constrained regarding the extent to which he can take advantage of a particularly good discovery. By law, no more than 5 per cent of the assets of a fund can be invested in any given security, and no more than 10 per cent of the assets of a given firm can be held by the fund. Moreover, the very size of many mutual funds makes it impossible to invest even the legal maximum in a security without driving its price to a point far above the original (bargain) amount. For all these reasons, this task may be the least important of the three outlined here; although there is evidence that some managers are more successful at it than others.

The advantages of a mutual fund are considerable, but they have a cost. The professional management typically requires compensation which, together with other recurring expenses, takes from one half of one per cent to one per cent of the fund's net assets each year. And the load charge -- levied on new investment in the fund -- amounts to 8½ per cent for small amounts, ranging down to as little as one per cent for very large investments. But, as the proponents of mutual funds point out, the small investor attempting to obtain a reasonably diversified portfolio might well incur similar costs if he were to attempt to invest his funds directly.[7]

MEASURING HISTORICAL PERFORMANCE

Since a key element of investment in common stocks is risk, it is obviously undesirable to measure the performance of mutual funds over a very short period of time. At the very least, a sufficient number of years should be considered to provide reasonable estimates of both the <u>average</u> return from the fund and the <u>variability</u> of the annual

return during the period. On the other hand, if too many years are used it is likely that the results may include changes in management philosophy. Following Treynor, we select a ten-year period as a reasonable compromise. For each of the ten years, the rate of return on the fund can be found by simply determining the per cent increase (or decrease) in the wealth of an investor holding shares in the fund during the year. The formula is:

$$\text{Rate of Return} = \frac{(NAV_{end} - NAV_{beginning}) + G + D}{NAV_{beginning}} \times 100$$

where:

NAV_{end} = net asset value per share at the end of the year;

$NAV_{beginning}$ = net asset value per share at the beginning of the year;

G = capital gains distributions per share during the year;

D = dividend payments per share during the year.

Note that this rate of return is net of all costs that the fund incurs (with the exception of the costs associated with the initial purchase of shares).

Once the rate of return on a fund in each of the ten years is computed it is a simple matter to determine the <u>average rate of return</u> on the fund during the period. We will use the term A to refer to this quantity and a subscript to indicate the fund in question (A_1 thus refers to the average rate of return on Fund 1, A_i to that on Fund i). A number of possible candidates for measuring variability can be found, but tradition (both in statistics and in capital theory) suggests the standard deviation;[8] we will use the term V for this measure of variability (thus V_i represents the variability of Fund i). Generally speaking, the variability indicates the extent to which actual returns tend to diverge from the average return during the period. For example, the average return for Massachusetts Investors' Trust during the ten-year period from 1954 through 1963 was 16.2 per cent per annum, while the variability was 20.8 per cent per annum. In six of the ten years

the actual return differed from the average return by less than 20.8 per cent;[9] in four of the years it differed by more. Although such a relationship will not always hold, it is not atypical.[10] In any event, the standard deviation represents as good a measure of variability as any, and we will follow tradition by adopting it.

Investors generally prefer funds offering high average returns to those offering low average returns, but they also prefer funds offering little variability. Since the marketplace tends to arrange security prices so that portfolios with large average returns also have large variability it is difficult to compare funds without some measure that takes both factors into account. If one fund has greater variability than another it is not necessarily a poor selection; for if the fund provides a greater reward to the investor willing to incur the risk associated with its variability it may be the better buy.

When evaluating the returns from risky investments it is important to remember that it is always possible to find some investments whose returns are virtually free from variability. In 1953 it was possible to purchase a ten-year U.S. Government bond at a price that would have guaranteed the investor a return of slightly less than 3 per cent if held to maturity. Although it is not a simple matter to determine the precise value of the riskless (or "pure") interest rate, for the period from 1954 to 1963 a value of 3 per cent is probably approximately correct.

If a riskless investment provides an average return equal to the pure interest rate (p), then the reward offered by a risky investment for bearing the risk associated with its variability can best be measured by the average return it pays <u>over and above</u> the pure interest rate. Putting it another way, the average return on an investment can be considered to be made up of two components: the pure interest rate and the reward for bearing risk. Letting R_i represent the reward for Fund i:

$$R_i = A_i - p .$$

If a fund offers a reward of R_i for bearing an amount of risk equal to V_i, its desirability can be measured by the reward-to-variability

ratio:

$$R/V = \frac{R_i}{V_i}$$

or, in terms of the original measures:

$$R/V = \frac{A_i - p}{V_i} .$$

We propose this ratio as a measure of the performance of mutual funds. The greater the reward-to-variability ratio, the better the fund's performance. Intuitively it seems to be a useful measure, and many may be willing to accept it on this basis. However, a stronger case can be made for its adoption, as we show next.

Figure 1 shows the average return and variability of each of 34 open-end mutual funds during the period 1954-1963.[11] The fund with the highest R/V ratio was the Boston Fund, with an average return of 12.4 per cent and a variability of 12.07 per cent. Using 3 per cent for the pure interest rate:

$$R/V = \frac{12.4 - 3.0}{12.07} = .78 .$$

The performance of the Boston Fund is indicated by point Y in Fig. 1; the line PYZ connects point Y with point P (which represents the result obtained by investing funds in a riskless security). Notice that the distance from P to X measures the reward $(A_i - p)$ from the fund and that the distance from X to Y measures the variability (V_i). It follows that the steepness of the line represents the reward-to-variability ratio:[12] the steeper the line, the poorer the ratio. Thus using the R/V ratio to measure the performance of funds is the same as rating them by the steepness of the lines connecting their (A_i, V_i) points with the point representing the pure interest rate. Boston Fund receives the best rating, since the line PYZ has the smallest slope. And Incorporated Investors receives the poorest rating, since its line (PQ) has the greatest slope

Lines such as those shown in Fig. 1 are important in their own right.

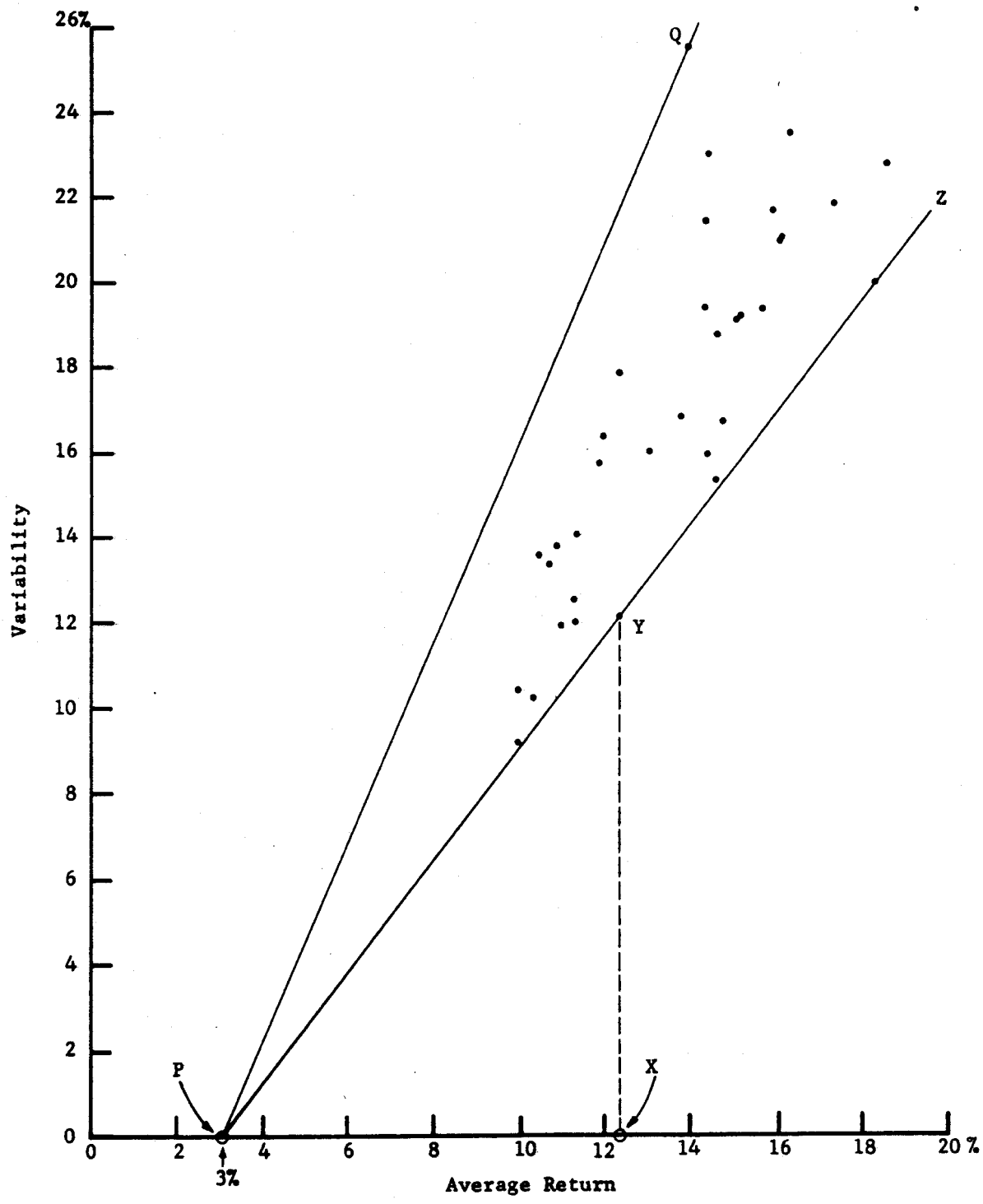


Fig. 1 -- Average Return and Variability: Thirty-four Open-end Mutual Funds, 1954-1963

Think of a relatively conservative investor who would have been unwilling to incur the risk associated with Boston Fund (V = 12.07%). American Business Shares, offering a risk of 9.2 per cent and an average return of 10.0 per cent might seem to have offered an attractive alternative. But the investor could have obtained the same variability and a higher average return[13] if he had put some of his money in shares of the Boston Fund and the remainder in a riskless investment at 3 per cent per annum. In fact, by altering the proportions, any combination of variability and average return lying along line PY could have been obtained using only the Boston Fund and investment at the pure interest rate. And points lying along the line beyond point Y could have been obtained by borrowing funds at 3 per cent to invest even more in the Boston Fund. Thus Boston Fund, plus either borrowing or investing in riskless securities could have provided a greater average return for <u>any</u> desired degree of risk than any other fund or combination of funds. The line connecting the (A_i,V_i) point for a fund with that for the pure interest rate thus represents <u>all</u> the (A,V) combinations obtainable by investing in that fund plus borrowing or lending at the pure interest rate. And the smaller the slope of the line, the better are those combinations. This is the most convincing argument for using the line -- or the R/V ratio, which is the same thing -- as a measure of fund performance.

Table 1 shows the performance of the 34 funds during the period 1954-1963; both the average return and variability are given, along with the R/V ratios based on a pure interest rate of 3 per cent. The performance ratios vary considerably -- from almost .78 (the Boston Fund) to slightly over .43 (Incorporated Investors). Thus historically, funds do show considerable differences in performance. But the investor will gain little by knowing which fund was best historically unless such differences persist over time. The more interesting questions concern the future. To what extent can performance be predicted in advance? To find out we must consider only the information available to the investor at the end of 1953; given that information and various rules for using it to select funds, how would he have done in subsequent years?

Table 1

PERFORMANCE OF 34 MUTUAL FUNDS; 1954-1963

Mutual Fund	Average Annual Return (%)	Variability of Annual Return (%)	Reward-to-Variability Ratio (R/V)
Affiliated Fund	14.6	15.3	.75896
American Business Shares	10.0	9.2	.75876
Axe-Houghton, Fund A	10.5	13.5	.55551
Axe-Houghton, Fund B	12.0	16.3	.55183
Axe-Houghton, Stock Fund	11.9	15.6	.56991
Boston Fund	12.4	12.1	.77842
Broad Street Investing	14.8	16.8	.70329
Bullock Fund	15.7	19.3	.65845
Commonwealth Investment Company	10.9	13.7	.57841
Deleware Fund	14.4	21.4	.53253
Dividend Shares	14.4	15.9	.71807
Eaton and Howard, Balanced Fund	11.0	11.9	.67399
Eaton and Howard, Stock Fund	15.2	19.2	.63486
Equity Fund	14.6	18.7	.61902
Fidelity Fund	16.4	23.5	.57020
Financial Industrial Fund	14.5	23.0	.49971
Fundamental Investors	16.0	21.7	.59894
Group Securities, Common Stock Fund	15.1	19.1	.63316
Group Securities, Fully Administered Fund	11.4	14.1	.59490
Incorporated Investors	14.0	25.5	.43116
Investment Company of America	17.4	21.8	.66169
Investors Mutual	11.3	12.5	.66451
Loomis-Sayles Mutual Fund	10.0	10.4	.67358
Massachusetts Investors Trust	16.2	20.8	.63398
Massachusetts Investors-Growth Stock	18.6	22.7	.68687
National Investors Corporation	18.3	19.9	.76798
National Securities - Income Series	12.4	17.8	.52950
New England Fund	10.4	10.2	.72703
Putnam Fund of Boston	13.1	16.0	.63222
Scudder, Stevens and Clark Balanced Fund	10.7	13.3	.57893
Selected American Shares	14.4	19.4	.58788
United Funds - Income Fund	16.1	20.9	.62698
Wellington Fund	11.3	12.0	.69057
Wisconsin Fund	13.8	16.9	.64091

NOTE: R/V Ratio = (Average return - 3.0%)/variability. The ratios shown were computed from original data and thus differ slightly from the ratios obtained from the rounded data shown in the table.

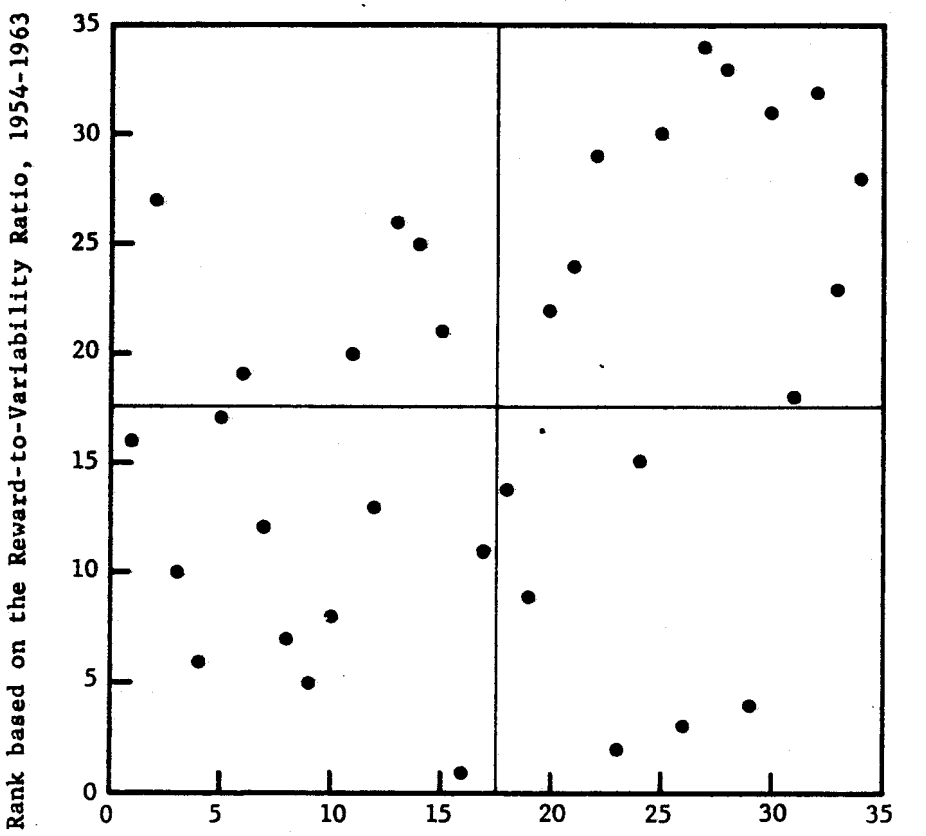
PREDICTIONS BASED ON THE R/V RATIO

Do differences in fund performance persist over time? An obvious test is to compare the R/V ratios in two separate periods. To do this the performance of each of the funds during the ten-year period from 1944 to 1953 was computed[14] and the funds ranked from 1 -- the fund with the highest (best) R/V ratio, to 34 -- the fund with the lowest (worst) R/V ratio. If differences in performance are relatively stable over time a fund's ranking based on its performance (R/V ratio) for the period 1954-1963 should be similar to that based on the ratio in the earlier period.

Figure 2 shows the rankings of the funds in the two periods on a scatter diagram. Each point represents one fund; the horizontal axis shows its ranking based on the R/V ratio in the first period and the vertical axis its ranking in the second period. Although the points do not fall along a 45° line from the origin (representing equal rankings in the two periods) there is a general upward trend, suggesting that funds ranking low in the early period tend to rank low in the later period, while those ranking high in the early period tend to rank high in the later period.

There are a number of ways in which the information in a scatter diagram of the type shown in Fig. 2 can be summarized. We will use two.

First, the diagram can be divided into four quadrants by drawing a vertical line between the 17th and 18th ranks on the horizontal axis and then drawing a horizontal line between the 17th and 18th ranks on the vertical axis. The number of funds falling into each of the four quadrants is shown in the smaller diagram in Fig. 2. Its interpretation is relatively straightforward. Of the 17 funds with the best performance in the early period (ranks 1 through 17), 6 were in the poorer half (ranks 18 through 34) in the second period, and 11 were in the better half. In other words, at the end of 1953 if an investor had chosen one of the 17 best funds based on past performance the odds would have been 11 to 6 that he would have held one of the 17 best during the subsequent ten years. Conversely, if he had chosen one of the 17 worst funds during the first period, the odds would have been 11 to 6 that he would end up with one of the 17 worst funds in the second period.



Rank correlation coefficient = .360

	best	worst
worst	6	11
best	11	6

Fig. 2 -- Predictions Based on the Reward-to-Variability Ratio

While this measure is rather crude, it does provide some indication of the predictive ability of the R/V ratio.

The second method for summarizing the relationship uses a rank correlation coefficient.[15] If there were no relationship at all between the rankings in the two periods this coefficient would be zero; if the rankings were precisely the same (i.e., all points fell on the 45^{o} line) it would be equal to one. Intermediate values indicate the extent to which the two rankings are related -- the higher the coefficient the better the relationship. The value in this case is .360, indicating a significant relationship.[16] As we will see, however, other measures of performance show even more significant correlation and are thus better predictors of future performance.

THE TREYNOR INDEX

In a perfect capital market no securities would be incorrectly priced. Thus the third function of the mutual fund (finding such securities) would be eliminated, leaving only the tasks of diversification and selecting the appropriate risk class. Moreover, under such conditions it has been shown[17] that all truly diversified portfolios will move with the over-all market, giving high returns when the market in general provides high returns and low returns when the market provides low returns. The data bear out this hypothesis. During the period from 1954 to 1963, almost 90 per cent of the variance[18] of the return on the typical fund in our sample was due to its comovement with the return on the 30 securities used to compute the Dow-Jones Industrial Average; moreover, as shown in Fig. 3, the percentage was quite similar for most of the 34 funds. Treynor has taken advantage of this relationship by using the volatility of a fund as a measure of its risk instead of the total variability suggested here. Since the returns on all diversified portfolios move with the market, the extent to which changes in the market are reflected in changes in a fund's rate of return can stand as a good measure of the total variability of the fund's return over time. By observing this relationship over some past period, a reasonably good estimate of volatility -- the change in the rate of return on a fund associated with a 1 per cent change in the rate of return on, say, the Dow-Jones portfolio -- can be obtained. We will use B_i to represent

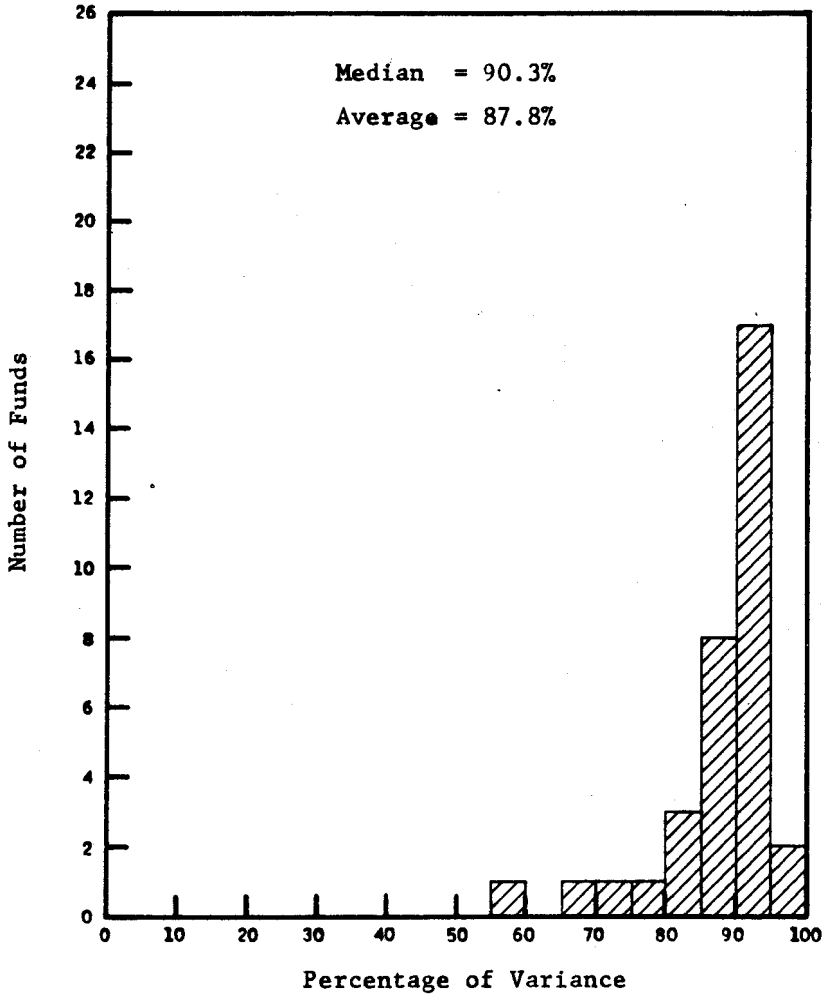


Fig. 3 -- Percentage of Variance Due to Comovement with the Dow-Jones Industrial Fund, Thirty-four Open-end Mutual Funds, 1954-1963

this value for the i^{th} fund.[19]

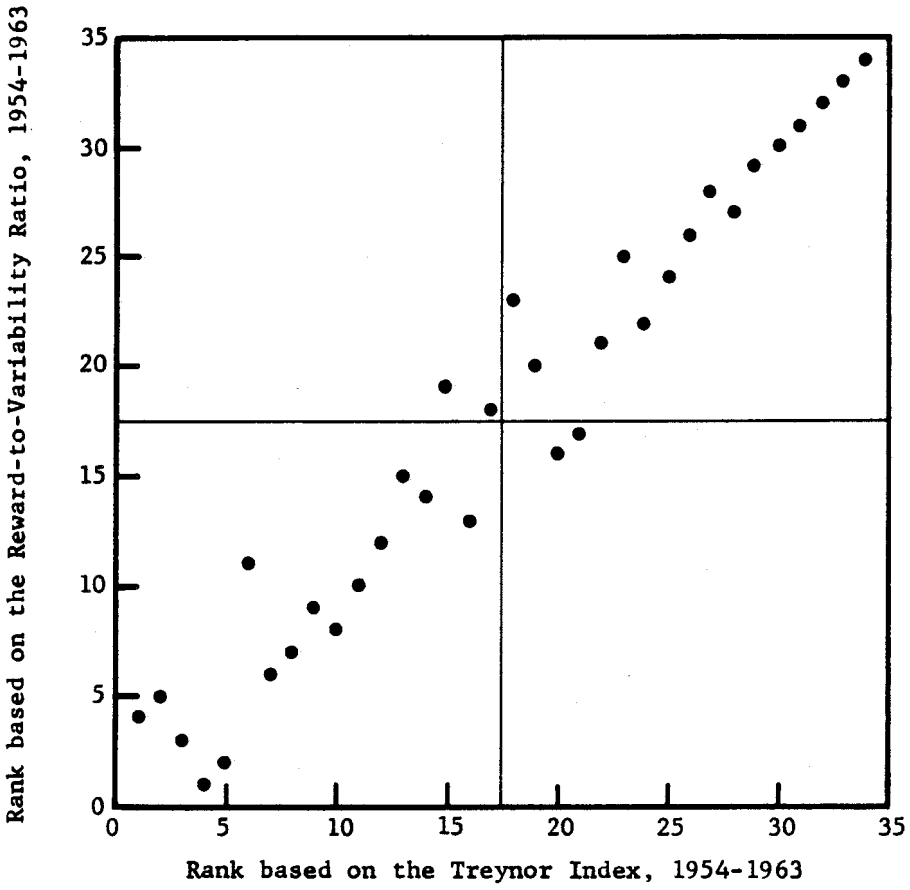
The measure that we will term the Treynor Index can be obtained by simply substituting volatility for variability in the formula for the R/V ratio:

$$TI = \frac{R_i}{B_i} = \frac{A_i - p}{B_i} .$$

Stated in this manner, the relationship between the two measures is clear. And the extent of the contribution of volatility to over-all variability makes the ranking of funds on the basis of the Treynor Index very close to that based on the R/V ratio. Figure 4 shows the rankings using the two measures for the period 1954-1963. Since the mutual funds in our sample all hold highly diversified portfolios, the similarity of the rankings is not surprising. And the cost of using the Treynor Index as a measure of past performance is relatively slight, since it mirrors the R/V ratio quite well. However, if some relatively undiversified funds (or more likely, privately held portfolios) had been included the results could have been significantly different, since the Treynor Index cannot capture the portion of variability that is due to lack of diversification. For this reason it is an inferior measure of <u>past</u> performance. But for this reason it may be a superior measure for predicting <u>future</u> performance.

If mutual funds hold well-diversified portfolios, any major discrepancies between the variability of their returns and that portion due to movements in the market are likely to be due to transitory effects. By concentrating on the systematic part of a fund's variability -- that is, its volatility -- we can avoid paying attention to these transitory effects and concern ourselves with the more permanent relationships. Thus, given some reasonable assurance that a fund will perform its diversification function well, the Treynor Index may provide better predictions of future performance than the R/V ratio.

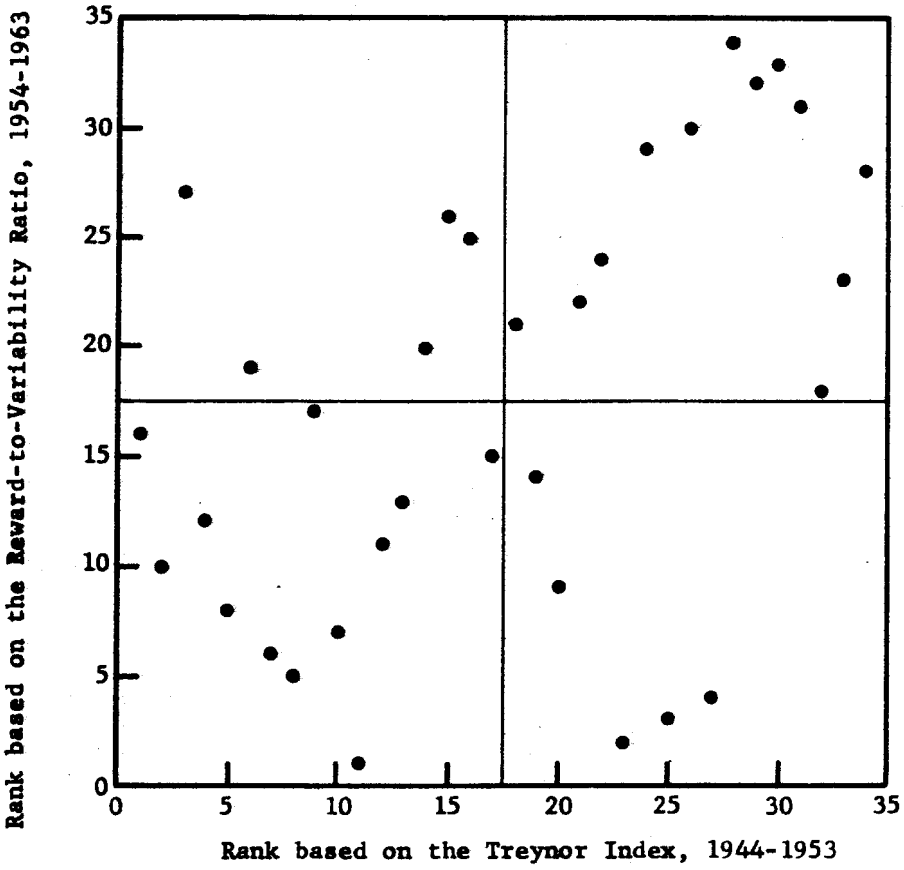
As Fig. 5 shows, the data bear out this suspicion. Using the rankings of funds based on the Treynor Index computed from 1944-1953 data to predict rankings based on performance (measured by the R/V ratio) in the subsequent ten years gives somewhat better results than those



Rank correlation coefficient = .974

	best	worst
worst	2	15
best	15	2

Fig. 4 -- Correlation between Ranks: Reward-to-Variability Ratio versus the Treynor Index



Rank correlation coefficient = .454

	best	worst
worst	5	12
best	12	5

Fig. 5 -- Predictions Based on the Treynor Index

obtained before. The odds of remaining in the selected half are now 12-to-5 (instead of 11-to-6), and the rank correlation coefficient is substantially higher -- .454 instead of .360. While the differences are far from overwhelming, the fact remains that the Treynor Index proved superior to the R/V ratio in this role. Unless other data show different results, it should be accepted as the better predictor.

Before turning to other methods for predicting mutual fund performance, the relationship between the measure we termed the Treynor Index and that suggested by Treynor needs to be clarified. Our measure (TI) is similar in form to the R/V ratio. One measure proposed by Treynor[20] was simply the negative of ours:

$$\text{Slope-angle} = -\text{TI} .$$

This form has a major advantage, since it can be transformed into a rather different quantity -- the rate of return on a market portfolio (e.g., the Dow-Jones portfolio) that would cause the fund in question to provide a rate of return equal to the pure interest rate.[21] This substitute measure has considerable intuitive appeal. Moreover, if predictions are to be made subjectively (using some judgment) rather than objectively (solely on the basis of historical data) the advantages of this alternative index are substantial. For the purposes of this paper, however, such considerations are relatively unimportant, since we are dealing only with objective methods of prediction. Accordingly, the Treynor Index will be stated as before, in the form directly comparable to the R/V ratio.

OTHER MEASURES FOR PREDICTING PERFORMANCE

We have shown that past performance (whether measured directly by R/V ratios or, better yet, indirectly with the Treynor Index) provides some aid to the investor attempting to predict the future performance of mutual funds. But other measures may be useful as well. Since mutual funds return income and capital gains to their investors only after deducting the costs of management, administration and brokerage fees, it is entirely possible that differences in performance are due more to differences in the amount spent on such services. A cynical view holds

that the funds spending the least will show the best (net) performance. The more charitable view holds that funds devoting more resources to management will gain enough to more than offset the increased expenditure and thus show better net performance. Intimately related with such considerations is the impact of size. A fund with substantial assets can obtain a given level of security analysis by spending a smaller per cent of its income than can a smaller fund; alternatively, by spending the same per cent it can obtain more (and/or better) analysis. On the other hand, more analysis may be required for a large fund than for a small one. In any event, both influences should be considered.

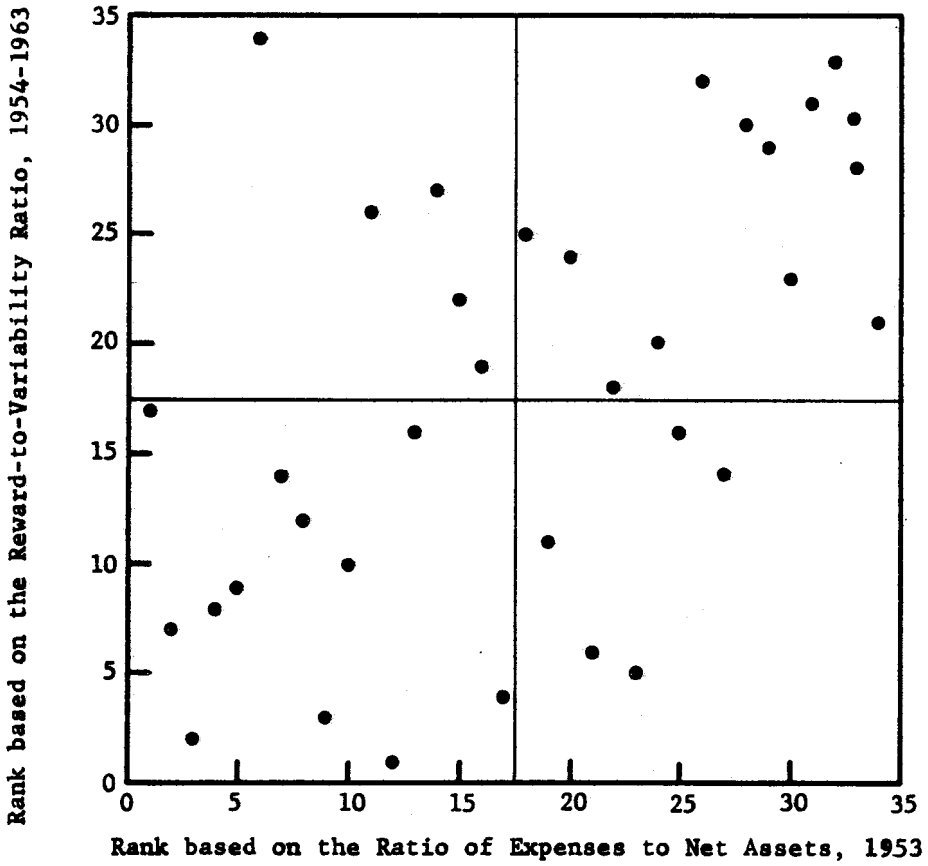
Figure 6 shows the relationship between the expenses[22] incurred by the 34 funds and their performance. Since we are concerned with prediction, the funds were ranked on the basis of the ratio of expenses to net assets during 1953 (needless to say, expense ratios changed somewhat during the subsequent ten years). Expenses ranged from .27 per cent of net assets (rank 1) to 1.49 per cent of net assets (rank 34).

The results tend to support the cynics. One of our summary measures (the correlation coefficient)[23] suggests that expense ratios provide somewhat better predictions than the Treynor Index; the other suggests that the two are equally good. Thus selecting a fund with a low ratio of expense to net assets may not be as foolish as some have suggested.

Figure 7 provides information concerning the predictive ability of the amount of a fund's assets. Size, measured by net asset value at the end of 1953, ranged from $522 million (rank 1) to $5.26 million (rank 34). Although the data show some correlation -- with the larger funds exhibiting somewhat better performance -- the relationship is marginal at best. Clearly the other measures we have considered provide more useful predictions.

A COMPOSITE MEASURE FOR PREDICTING PERFORMANCE

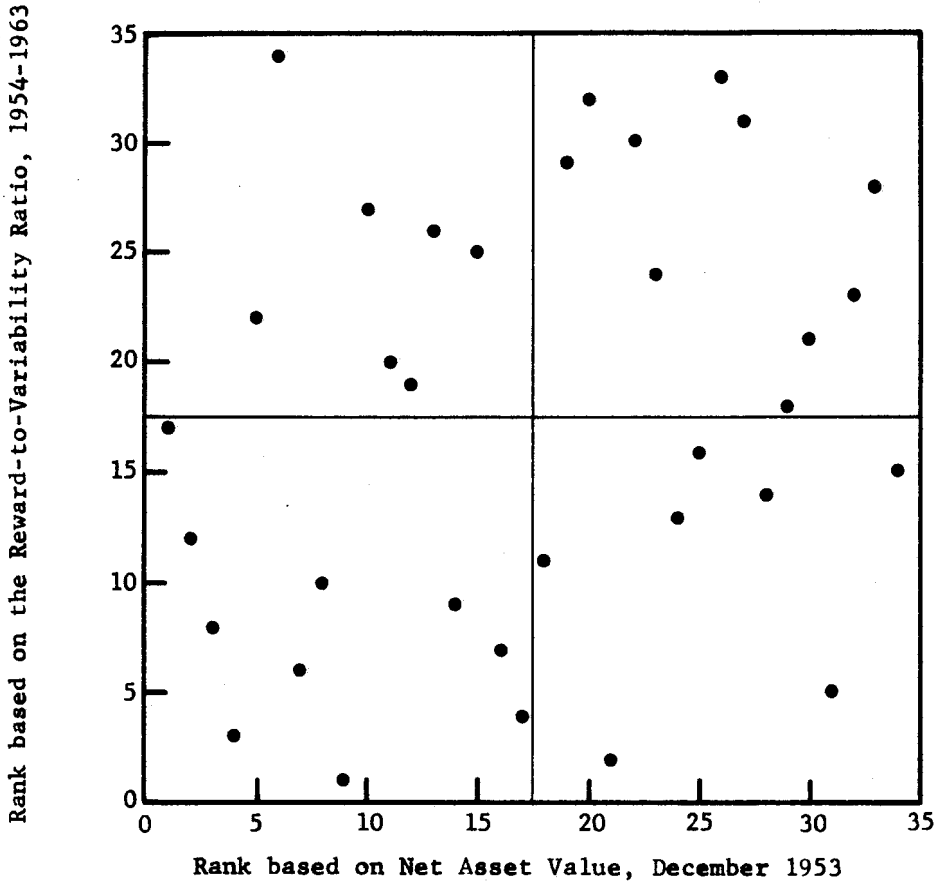
Our results indicate that two of the four measures considered can reasonably be discarded. Although the R/V ratios for past periods provide relatively good predictions of future performance, the Treynor Index values appear to be superior, at least for funds holding diversified portfolios. And the size of a fund is apparently of little importance



Rank correlation coefficient = .505

	lowest	highest
worst	5	12
best	12	5

Fig. 6 -- Predictions Based on the Ratio of Expenses to Net Assets



Rank correlation coefficient = .234

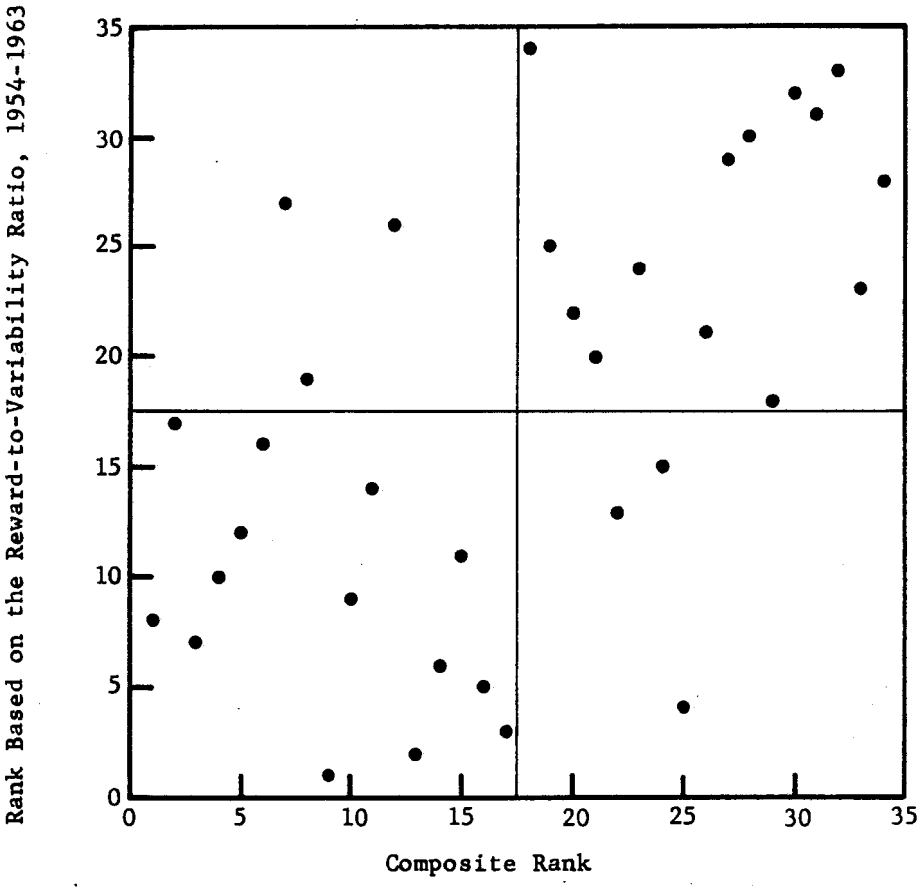
	largest	smallest
worst	7	10
best	10	7

Fig. 7 -- Predictions Based on Size

in determining its performance. However, both of the remaining measures -- the Treynor Index and the expense ratio -- appear to predict reasonably well. To some extent this may be due to a relatively high correspondence between them. Funds with low expense ratios tend to perform better than those with high expense ratios; thus funds ranking well on the basis of the Treynor Index (reflecting relatively good performance in the past) are likely to also rank well on the basis of expense ratios (i.e., have low ratios). It is at least possible that all the differences in performance can be attributed to differences in expenses. However, additional tests suggest that such is not the case: both expense ratios and past performance are relevant for predicting future performance.[24] The calculations behind this assertion are given in the Appendix, and will not be described here. However, one simple predictive method incorporating both measures will be shown.

When ranking funds on the basis of the Treynor Index we assign Rank 1 to the fund with the highest (best) index; when ranking on the basis of expense, we assign Rank 1 to the fund with the lowest (best) ratio. Since both factors appear to be relevant, a ranking that takes into account a fund's position in both rankings should provide even better predictions than either of the two separate rankings. An obvious possibility is to accord each factor equal weight. The average rank of each fund (i.e., the value midway between its rank based on the Treynor Index and that based on the expense ratio) accomplishes this. The resulting values can be used to assign composite ranks -- from 1 (the fund with the smallest average rank) to 34 (the fund with the largest average rank); ties can be broken arbitrarily.[25]

Figure 8 shows the results obtained when predictions are based on such a composite rank. Both the measures used to summarize the relationship show that the technique gives better predictions than either of the two components used alone. This result is in general agreement with the assertion that both past performance and expense ratios are useful for predicting future performance. And although the predictions are far from perfect, they appear to be preferable to selecting a fund at random.



Rank correlation coefficient = .545

	lowest	highest
worst	3	14
best	14	3

Fig. 8 -- Predictions Based on Composite Ranks

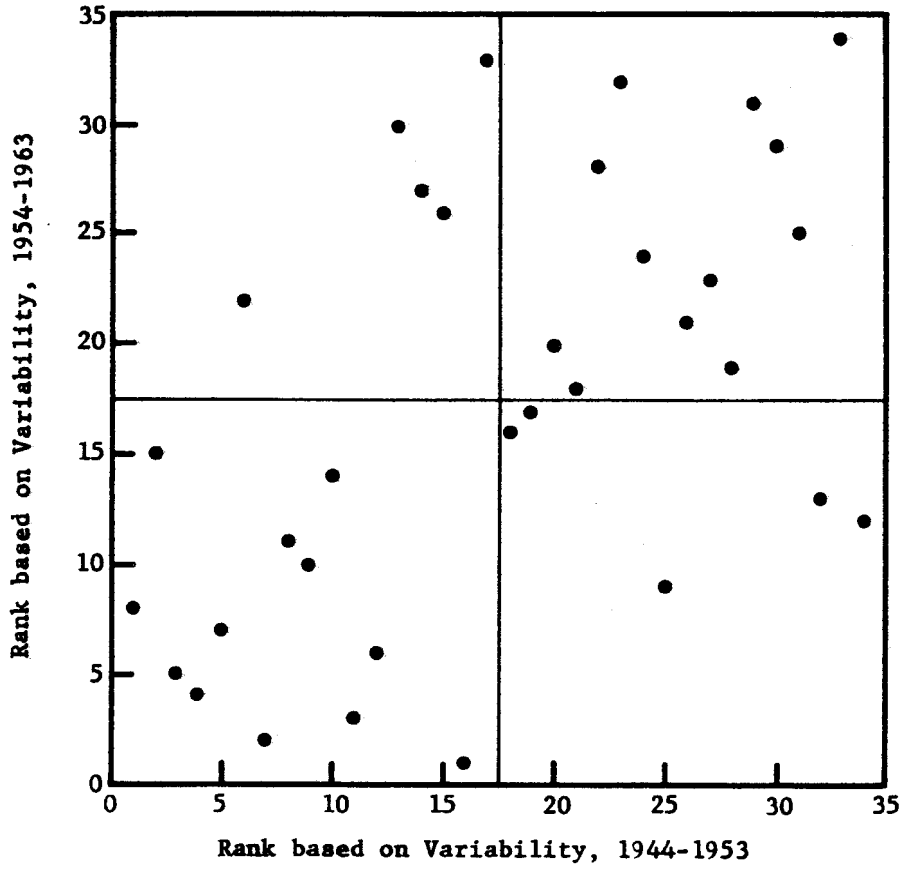
THE RELATIVE RISK OF MUTUAL FUNDS

All the calculations presented thus far deal with a measure of performance (the R/V ratio) that disregards risk per se, concentrating instead on the relationship between the reward obtained from the fund and the risk actually experienced. As shown earlier, the investor should be most concerned with this relationship, for he can arrange his other commitments (either by borrowing additional funds or by investing in some riskless security) to complement the risk inherent in a particular fund in any manner he desires. However, to do this he must have some idea of the variability the fund will actually experience. If mutual fund managers do not perform the second of the three tasks we have outlined, investors will find it very difficult to arrange their over-all holdings in the most desirable manner. Holders of mutual fund shares have a right to expect that funds will show reasonable consistency over time with regard to the variability of returns.

Figure 9 provides some evidence on this point. The funds were ranked in each of the two periods studied (1944-1953 and 1954-1963) on the basis of variability -- Rank 1 indicating the smallest amount and Rank 34 the largest. A reasonable amount of consistency between periods is evident, but a number of major shifts appear. In some cases this may have been due to announced changes in management philosophy, in others it was presumably inadvertent. Whatever the cause, the prevalence of such shifts in our sample is likely to disappoint some investors. On the other hand, there is no well-defined standard against which the results can be compared; given the difficulties involved,[26] one might reasonably argue that the data show that mutual fund managers fulfill remarkably well the obligation to stay within their selected risk classes.

MUTUAL FUNDS VERSUS THE DOW-JONES INDUSTRIAL AVERAGE

We have dealt at length with comparisons among mutual funds but have not considered an alternative strategy -- investing directly in a reasonably diversified group of securities. To investigate such an alternative we must specify the portfolio to be held; following tradition, the 30 securities used to compute the Dow-Jones Industrial Average will be used.



Rank correlation coefficient = .528

largest	5	12
smallest	12	5
	smallest	largest

Fig. 9 -- Consistency of Risk-class

When calculating the returns from the Dow-Jones portfolio, no costs (brokerage, management, or administrative) are deducted. To some extent this overstates the performance available from such a direct investment. On the other hand, the initial selling (load) charge is not deducted when determining the returns from mutual funds; thus the results from both types of investments are overstated. The magnitudes of the differences between the measures we use and those relevant for a particular investor depend on a number of factors,[27] but for most investors the comparison made here should give results similar to those obtained if all the relevant costs had been considered.

Figure 10 shows the distribution of the R/V ratios for the 34 mutual funds (based on their performance during the period 1954-1963). The vertical line represents the R/V ratio for the Dow-Jones portfolio -- its return averaged 16.3 per cent during the period with a variability of 19.94 per cent, giving an R/V ratio of .667. The average R/V ratio for the funds in our sample was .633 -- considerably smaller than that of the Dow-Jones Average. Although another group of mutual funds would give different results, the odds are greater than 100-to-1 against the possibility that the average mutual fund did as well as the Dow-Jones portfolio from 1954 to 1963.[28] In this group, only 11 funds did better than the Dow-Jones portfolio, while 23 did worse.

From the standpoint of the investor the comparison shown in Fig. 10 is the most relevant. But to account for the relatively poor performance of most mutual funds it is instructive to compare <u>gross</u> performance, (i.e., before deducting expenses) with that of the Dow-Jones portfolio. Such a comparison[29] shows that 19 funds did better than the Dow-Jones portfolio, and only 15 did worse. And although another group of funds would give different results, the odds are greater than 3-to-1 against the possibility that the gross performance of the average mutual fund was worse than that of the Dow-Jones portfolio from 1954 to 1963.[30]

Although it may be dangerous to generalize from the results found during one ten-year period, it appears that the average mutual fund manager selects a portfolio slightly better than the Dow-Jones Industrials, but that the results actually obtained by the holder of mutual fund shares (after the costs associated with the operation of the fund

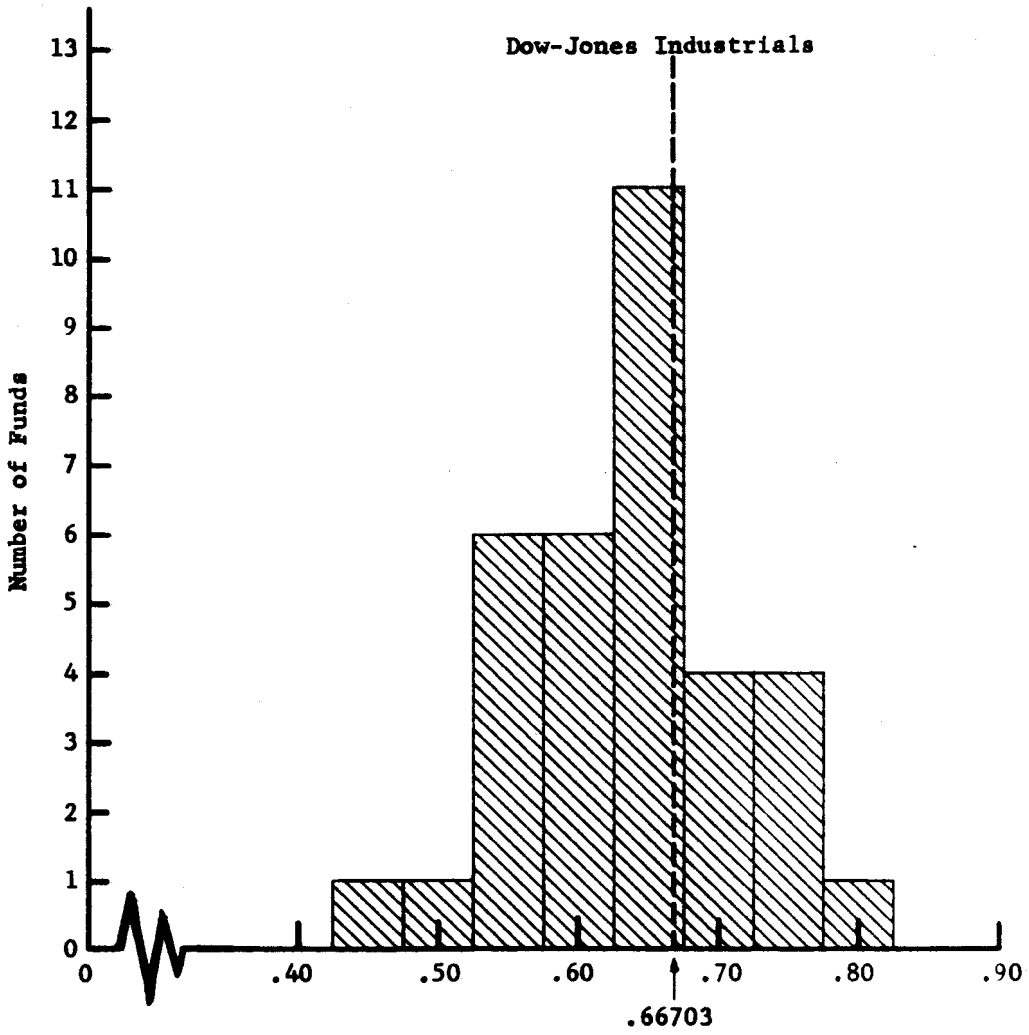


Fig. 10 -- Mutual Fund Performance versus the Dow-Jones Industrials, 1954-1963

have been deducted) fall somewhat short of those from the Dow-Jones portfolio. This is consistent with our earlier conclusion that, all other things being equal, the smaller a fund's expense ratio, the better the results obtained by its stockholders. However, mutual funds provide the investor with several services; even if their performance falls slightly below that of the Dow-Jones Average, they probably represent a preferred form of investment for many people. Moreover, while the average fund may provide performance somewhat below that of the Dow-Jones Average, many funds do better (in our sample, 11 did during the period 1954-1963). And, as we have shown, it may be possible to predict performance well enough to increase the odds that the fund selected will in fact be better than average.

PREDICTING MUTUAL FUND PERFORMANCE FROM 1964 TO 1973

We have shown that rankings based on values of the Treynor Index, expense ratios or, better yet, combinations of the two would have helped the investor attempting to predict the performance of mutual funds during the ten-year period from 1954 to 1963. But can they help the investor attempting to predict performance from 1964 to 1973? Only time can tell.

Table 2 provides the predictions. The funds are listed in order of their composite ranks (calculated in the manner described earlier)[31] -- from best to worst, with all ties indicated. The table shows the expense ratio of each fund during 1963 as reported by Wiesenberger,[32] as well as its volatility (B_i)[33] during the period 1954-1963. Finally, the table gives the ranks of the funds based on the Treynor Index[34] (Rank 1 indicates the best performance, Rank 34 the worst) and variability (Rank 1 indicates the smallest amount, Rank 34 the largest).

No claim is made for the usefulness of Table 2. The academic observer will wait until 1974 to test its value. The investor may, at hiw own peril, attempt to test it in a more direct manner now.[35]

Table 2

RANKINGS BASED ON THE COMPOSITE INDEX: 1954-1963

Mutual Fund	Expense Ratio	Yola-tility	Rank Based On Treynor Index	Rank Based On Vari-ability
National Investors Corporation	0.23	0.916	5	25
Affiliated Fund	0.38	0.679	3	12
Broad Street Investing	0.23	0.807	8	17
Wellington Fund	0.36	0.576	10	5
Massachusetts Investors - Growth Fund	0.45	1.073	9	31
Boston Fund	0.55	0.552	4	6
Dividend Shares	0.50	0.741	7	14
Bullock Fund	0.43	0.926	14	23
Investors Mutual	0.48	0.592	12	7
Massachusetts Investors Trust	0.18	1.015	21	26
Putnam Fund of Boston	0.46	0.742	15	15
Eaton and Howard, Balanced Fund	0.56	0.556	11	4
Loomis-Sayles Mutual Fund*	0.63	0.430	6	3
New England Fund*	0.72	0.427	2	2
American Business Shares	0.74	0.353	1	1
United Funds - Income Fund	0.56	0.993	19	27
Investment Company of America	0.59	1.063	16	30
Wisconsin Fund	0.66	0.780	13	18
Eaton and Howard - Stock Fund	0.58	0.927	20	22
Fidelity Fund	0.53	1.144	28	33
Scudder, Stevens and Clark - Balanced Fund	0.58	0.627	23	8
Commonwealth Investment Company	0.58	0.658	26	10
Fundamental Investors	0.59	1.061	24	29
Selected American Shares	0.62	0.937	25	24
Incorporated Investors	0.55	1.252	34	34
Group Securities, Common Stock Fund	0.75	0.909	17	21
Group Securities, Fully Administered Fund	0.75	0.634	18	11
Deleware Fund*	0.70	1.021	31	28
Equity Fund*	1.27	0.893	22	20
Financial Industrial Fund	0.66	1.135	33	32
Axe-Houghton, Fund B	0.74	0.783	30	16
National Securities - Income Series*	0.74	0.859	32	19
Axe-Houghton, Stock Fund*	1.05	0.754	27	13
Axe-Houghton, Fund A	0.98	0.649	29	9

*Adjacent funds have equal composite rankings.

APPENDIX

Since this paper is concerned primarily with predicting mutual fund performance, and since the investor is presumably interested in finding the fund that will provide the best performance, all the tests given in the text are based on rankings rather than actual values of the various measures investigated. However, for other purposes it may be preferable to measure the relationships in the more usual manner; this is especially true when attempting to determine the relative importance of alternative explanatory variables (e.g., the Treynor Index and the expense ratio).

A series of simple regression analyses using actual values rather then rankings gave results similar to those reported in the text: the Treynor Index proved superior to the R/V ratio as a predictor and size was a relatively poor measure; the sign of the coefficient from the regression using size indicates that larger funds did better than the smaller ones. However, the low t-value (which measures the significance of the slope coefficient) for the regression with size suggests that this relationship is not very significant. One minor difference should be noted: in this case the Treynor Index gave slightly better results than the expense ratio. The results were:

DEPENDENT VARIABLE: R/V RATIO, 1954-1963

Independent Variable	Correlation Coefficient	t-Value for Slope Coefficient
Treynor Index, 1944-53	.4008	+ 2.47
Expense/NAV, 1953	.3746	- 2.29
R/V ratio, 1944-53	.3157	+ 1.88
Size (NAV), Dec. 1953	.1523	+ 0.87

Multiple correlations, using these variables, gave the following results:

DEPENDENT VARIABLE: R/V RATIO, 1953-1963

Multiple Correlation Coefficients	t-Values for Slope Coefficients		
	Treynor Index 1944-1953	Expense/ NAV 1953	Size (NAV) Dec. 1953
.4633	+ 1.67	- 1.43	- 0.46
.4572	+ 1.64	- 1.38	--
.4015	+ 2.26	--	+ 0.15
.3767	--	- 2.07	- 0.24

The extremely small t-values for the slope coefficients relating R/V ratios to size, plus the fact that the signs differ from case to case, support the assertion that size per se is an unimportant factor in predicting future performance. The t-values for the Treynor Index and the expense ratio support the assertion that both are useful for such predictions (although neither is highly significant).

A simple regression using the standard deviation of return in the second period (1954-1963) as the dependent variable and the standard deviation of return in the first period (1944-1953) as the independent variable gave results similar to those found earlier. The correlation coefficient was .4518; the t-value for the slope coefficient was + 2.86.

FOOTNOTES

1. The original work in the field was that of Markowitz; see his article, "Portfolio Selection," The Journal of Finance, XII, March, 1952, pp. 71-91, or the subsequent expanded version, Portfolio Selection, Efficient Diversification of Investments, New York: John Wiley and Sons, Inc., 1959. For extensions, see W. F. Sharpe, "A Simplified Model for Portfolio Analysis," Management Science, Vol. 9, No. 2, January 1963, pp. 277-293, and Eugene F. Fama, "Portfolio Analysis in a Stable Paretian Market," Management Science, Vol. II, No. 3, January 1963, pp. 404-419.

2. See, for example, W. F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," The Journal of Finance, Vol. XIX, No. 3, Sept. 1964, pp. 425-442.

3. In W. F. Sharpe, "Risk-Aversion in the Stock Market: Some Empirical Evidence, P-3084, The RAND Corporation, Santa Monica, March 1965.

4. Jack L. Treynor, "How to Rate Management of Investment Funds," Harvard Business Review, Vol. 43, No. 1, Jan-Feb. 1965, pp. 63-75.

5. "... when one talks about the historical performance pattern of a fund, he is looking at the past; but when he considers the preferences of individual investors and their choices among funds, he is talking about their appraisal of the future. We shall continue to talk about the performance of funds in terms of historical performance patterns, even though actual investor choices among funds are necessarily based on expectations regarding future performance patterns. The implication is that a good historical performance pattern is one which, if continued into the future, would cause investors to prefer it to others." Treynor, op. cit., p..67.

6. For a model leading to this conclusion (as well as several others mentioned in this paper), see W. F. Sharpe, "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," The Journal of Finance, Vol. XIX, No. 3, Sept. 1964, pp. 425-442.

7. The load charge applies only to the initial investment in the mutual fund, thus it must be compared with the transaction cost associated with the purchase plus the sale of securities. Weisenberger (Investment Companies, 1962 edition) gives two examples for comparison. Assume that an individual has $4,000 to invest; if he selects ten securities at a price of $40 each, he will spend $235.30 in total commissions -- about 5.9 per cent of his original investment. If he selects 20 securities at $20 each, he will spend $297.80 -- over 7.4 per cent of his investment. Needless to say, mutual funds pay some transactions costs; these come out of the annual returns are are thus expenses over and above the load charge. Thus the net performance of the fund should be compared to the total performance of, say, the Dow-Jones securities in rating fund management. Such a comparison is made in a later section of this paper.

8. To compute the standard deviation of the rate of return during a ten-year period, determine the difference between the actual rate of return in each year and the average (A_i), then square each of these differences and sum them. Dividing the sum by 9 gives the variance. The square root of the variance is the standard deviation.

9. That is, it was between - 4.6 per cent (16.2 per cent - 20.8 per cent) and 37.0 per cent (16.2 per cent + 20.8 per cent).

10. If the distributions of returns followed the normal curve, two-thirds of the results would diverge from the average by less than the standard deviation. However, stock market results tend to have fatter-tailed distributions than that of the normal curve, as Mandelbrot has shown ("The variation of certain Speculative Prices," The Journal of Business, Vol. 37, No. 4, Oct. 1963, pp 394-419), thus less than two-thirds of the yearly results are likely to fall within the range ($A_i - V_i$) to ($A_i + V_i$).

11. The funds used for this and all subsequent analyses were those for which annual rates of return were given by Weisenberger for at least the last 20 years. All data are from Arthur Weisenberger and Co., Investment Companies, 1953, 1962, and 1964 editions.

12. The cotangent of the angle made by the line with the horizontal axis is equal to the R/V ratio. Putting it another way, the reciprocal of the slope of the line equals the R/V ratio.

13. Slightly less than 10.2 per cent per annum.

14. Since the long-term interest rate during this period was somewhat lower than that prevailing in the later period, a pure interest rate of 2.5 per cent was used in the calculations (this was approximately the yield on a ten-year U.S. Government bond in 1943). Although different assumptions regarding the pure interest rates in the two periods would substantially affect the rankings of individual funds, the relative predictive ability of the measures to be considered in this and subsequent sections is not significantly affected by altering the assumed rates over a considerable range.

15. Spearman's rank correlation coefficient.

16. The standard error for the sample size used in this and all subsequent calculations is .174 ($1/\sqrt{34 - 1}$).

17. In Sharpe, "Capital Asset Prices," op. cit.

18. This is not the standard deviation (which we have termed variability) but its square, as indicated in an earlier footnote.

19. To be consistent with the notation used in Sharpe, "A Simplified Model for Portfolio Analysis," op. cit.

20. Described on p. 69 in Treynor, op. cit.: "tangent $\alpha = (\mu - \mu^*)/\sigma$."

21. The original index (SA -- for slope-angle) is:

$$SA_i = - TI_i = -\left(\frac{A_i - p}{B_i}\right) = \frac{p - A_i}{B_i} .$$

If the relationship between the rate of return on the fund (r_i) and that on the Dow-Jones portfolio (D) is:

$$r_i = a_i + B_i D ,$$

then the average return on the fund is related to that on the Dow-Jones portfolio ($\bar{D}$) by:

$$A_i = a_i + B_i \bar{D} .$$

Thus:

$$SA_i = \frac{p - A_i}{B_i} = \frac{p - (a_i + B_i \bar{D})}{B_i} = \left(\frac{p - a_i}{B_i}\right) - \bar{D} .$$

D_i^* -- the return on the Dow-Jones portfolio required to make the return on the fund equal to the pure interest rate (p) is:

$$p = r_i = a_i + B_i D_i^*$$

$$D_i^* = \frac{p - a_i}{B_i}$$

Therefore:

$$SA_i = D_i^* - \bar{D} .$$

Clearly, if the slope-angle for one fund is higher than that of another, the return on the Dow-Jones required to make it yield the pure interest rate will be also. Since D_i^* is simply the slope-angle plus a constant ($\bar{D}$), both measures will give the same ranking. The advantage of using D_i^* lies in its intuitive appeal and the fact that if future values are

being estimated directly, no prediction about market performance is required. The insensitivity of Treynor's rankings (using either of the two measures) to predictions about market behavior follows directly from his assumption that variability is due only to the response of the fund's return to changes in the market. If there is another source of variability the rankings of funds (using the R/V ratio) will depend to some extent on the behavior of the market.

22. Brokerage charges are not included in the expenses reported explicitly. It is conceivable that the predictive ability of expense ratios would be improved if they were.

23. Some of the funds had equal expense ratios (to the accuracy of two decimal places, as computed by Weisenberger). In such cases ranks were assigned alphabetically. For this reason the measures of correlation (particularly the correlation coefficient) are not as trustworthy as those computed for the other comparisons.

24. Although expense ratios as normally computed do not explain all the persisting differences among funds, it is possible that if all expenses were included, no improvement could be obtained by also considering past performance. Since brokerage fees are not counted as expenses per se, it is entirely possible that the differences in past performance that seem to be useful for predicting future performance simply reflect differences in total expense ratios.

25. In our calculations, funds with equal average ranks were ranked in alphabetic order.

26. Mandelbrot, op. cit., has shown that predicting the variability of the changes in security prices is very difficult indeed. Presumably the law of large numbers cannot be relied upon to eliminate enough of the difficulty to make predictions of the return on portfolios relatively simple.

27. For example: the amount to be invested, the number of years the portfolio is to be held, and the extent to which dividends are to be reinvested.

28. The standard deviation of the R/V values for the 34 funds was .08067. If the population of mutual funds had a mean of .667 and a standard deviation of .08067, the distribution of sample means for groups of 34 would have a standard deviation of .01383 ($= .08067/\sqrt{34}$) and be roughly normally distributed. The observed mean of .633 is 2.46 standard deviations below the assumed mean of .667; the odds are 144-to-1 that under the hypothesized conditions a sample of 34 funds would have an average R/V value as low as .633.

29. The comparison was made by assuming that each fund maintained its 1953 ratio of expenses to net assets throughout the subsequent ten years. Under these conditions the only change required to compute the R/V ratios for gross performance was to add each fund's expense ratio to its average return before the R/V ratio was computed.

30. The standard deviation of the gross R/V values for the 34 funds was .08304. If the population of mutual funds had a mean of .667 and a standard deviation of .08304, the distribution of sample means for groups of 34 would have a standard deviation of .01424 (= $.08304/\sqrt{34}$) and be roughly normally distributed. The observed mean of .677 is .74 standard deviations above the assumed mean of .667; the odds are 3.36-to-1 that under the hypothesized conditions a sample of 34 funds would have an average gross R/V value as high as .677.

31. With one difference. In the previous calculations each fund was assigned a unique rank with regard to expense ratio as well as the composite rank. This approach was desirable since the goal was merely testing relationships between values for two periods of time. However, Table 2 may be used for the selection of mutual funds, and all ties need to be taken into account explicitly. Thus when assigning ranks on the basis of expense ratios, all funds with the same ratio were assigned the same rank. Initially the funds were arrayed in order of increasing expense ratio, with tying funds listed alphabetically, and ranks from 1 to 34 assigned as before. Then all funds with the same ratio were assigned a rank equal to the average initial rank for the group. These adjusted ranks were used when the composite ranks were computed. As indicated in the text, all ties with regard to composite rank are indicated in Table 2.

32. Weisenberger, Investment Companies, 1964 edition, pp. 160-165.

33. Recall that volatility is an estimate of the per cent change in the return on the fund accompanying a change of 1 per cent in the return on the Dow-Jones portfolio. A value less than one indicates that the fund is less volatile than the Dow-Jones Industrials; a value over one indicates that it is more volatile. The values shown in Table 2 are based on regression analyses using standard techniques.

34. Based on a pure interest rate of 3 per cent.

35. Needless to say, the investor will want to take into account differences in load charges, accumulation plans, etc., when evaluating the funds. The difference between a load charge of 8 per cent and none at all can compensate an investor for at least some difference in performance. Thus Equity Fund (3 1/2 per cent maximum charge), Loomis-Sayles (no load charge) and Scudder, Stevens and Clark Balanced Fund (no load charge) deserve more attention than their relative positions in the table suggest.