Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios

Edwin J. Elton
Martin J. Gruber
Sanjiv Das
Matthew Hlavka
New York University

We investigate the informational efficiency of mutual fund performance for the period 1965–84. Results are shown to be sensitive to the measurement of performance chosen. We find that returns on S&P stocks, returns on non-S&P stocks, and returns on bonds are significant factors in performance assessment. Once we correct for the impact of non-S&P assets on mutual fund returns, we find that mutual funds do not earn returns that justify their information acquisition costs. This is consistent with results for prior periods.

The evaluation of professional money management has long been a topic of considerable interest to financial economists. Two developments have stimulated renewed interest in this topic. The first is the work of Grossman (1976) and Grossman and Stiglitz (1980). They argue that trades by informed investors take place at prices that compensate these investors for the cost of becoming informed. Under these conditions, it is

The authors would like to thank Eugene Fama, Andrew Lo, William Sharpe, and the referee (Bruce Lehmann) for helpful comments and suggestions on an earlier version of this article. We have benefited from useful suggestions received at workshop presentations at MIT, the University of Chicago, and the University of Michigan and presentation at the European Finance Association. Address correspondence to Edwin J. Elton and Martin Gruber, Stern School of Business, Department of Finance, 44 West 4th Street, New York, NY 10012.

The Review of Financial Studies 1993 Volume 6, number 1, pp. 1–22

possible for informed investment managers to pass on the cost of analysis to their clients and for their clients to be no worse off after reimbursing managers for the cost of becoming informed. Managers who charge higher fees do not necessarily earn lower returns for their clients. Higher fees may or may not be associated with lower returns, depending not on the efficiency of security markets, but rather on the efficiency of the market for money managers.

The second development stems from the work of Roll (1977) and Ross (1977) and its implication for the evaluation of the money manager. Roll's argument against the use of CAPM as a benchmark for performance together with Ross's discovery of arbitrage pricing theory has led to questions about the appropriate benchmarks against which to judge abnormal performance.

Recent work, particularly that of Lehman and Modest (1987), Admati et al. (1986), Conner and Korajczyk (1991), and Grinblatt and Titman (1989), raise further questions about the appropriate benchmark to use in evaluating performance. In particular, Lehman and Modest (1987) stress the sensitivity of performance to the benchmark chosen and the need to find a set of benchmarks that represent the common factors determining security returns. This literature should be contrasted with a set of literature that seems to indicate that a single-index model provides an adequate description of portfolio returns and that the evaluation of the returns is reasonably insensitive to the definition of the index used. For example, Roll (1971) found that three market proxies provided nearly identical performance measures for randomly selected portfolios. Copeland and Mayers (1982) found that inferences about the Value Line Enigma were unaffected by the choice of a performance benchmark.

Our purpose in this article is to show that even a very parsimonious description of the proxies generating returns on portfolios of bonds and stocks can lead to very different and superior inferences about the attributes of active portfolio management compared to a single index. We use as an example the data used (and conclusions reached) in a recent study of mutual fund performance by Ippolito (1989). The Ippolito study is important because it is the first test of the expanded (Grossman) definition of market efficiency using data on managed portfolios. In addition, Ippolito's study is interesting because his results and conclusions are different from those reached in a long series of mutual fund studies.

Ippolito has at least two findings that are important and consistent with a theory of efficiency with informed investors.

(1) Estimated risk adjusted return (α) for the mutual fund industry

is greater than zero even after accounting for transaction costs and expenses. Ippolito attributes nonnegative α to the existence of informed actions by management.

(2) There does not seem to be any evidence that turnover, management fees, or expenses are associated with inferior returns, net of management fees and expenses.

These results and conclusions are very different from those reached by earlier researchers [see Jensen (1968), Friend, Blume, and Crockett (1970), and Sharpe (1966)]. In particular, they are different from the results presented by Jensen, who found a negative α on average for a broad sample of mutual funds over the periods 1955–1964 and 1945–1964.

In this article, we show that Ippolito's results are primarily due to the difference in the performance of non-S&P 500 assets in the Ippolito period (1965–1984) compared to the performance of these assets in the Jensen period.[1] Furthermore, once we explicitly account for mutual funds holding non-S&P assets, the results of the type of analysis Ippolito performs change and are identical to those found in earlier studies.

In the first section of this article, we examine the effect of holding non-S&P equities and bonds on alphas of mutual funds. Mutual funds have historically held some stocks not in the S&P index and held some stocks in the S&P index in proportions below their weighting in the S&P index. We shall see that holding non-S&P stocks would cause negative α's for funds over the earlier period studied by Jensen and positive α's for funds over the period studied by Ippolito, *even if* management had no selection ability. In both the earlier period and the latter period, the effect of bonds being part of the portfolio had very little impact on α.[2]

Having documented in the first section of this article two influences that could account for the positive α's found by Ippolito, we then correct for these influences. This is done by introducing two new indexes into the standard one-index performance measurement model. This allows us to examine if funds have a positive α above and beyond that arising from extending holdings beyond the S&P 500.

In the final section, we examine the influence of expenses and turnover on α when we have explicitly taken into account the effect of non-S&P assets.

---

[1] In part, his results are also due to data errors that led to a particular bias. This will be explored in later sections of the article.

[2] Bonds impact α indirectly if the presence of bonds means that the aggregate portfolio has a smaller percentage of non-S&P stocks.

Table 1
α by size decile

	With S&P indexes		With combined indexes[1]	
Decile by size	1945–1964 Jensen period	1965–1984 Ippolito period	1945–1964 Jensen period	1965–1984 Ippolito period
Smallest	−5.03	12.81[2]	−6.58	15.40[2]
2	−4.39	10.64[2]	−5.26	13.15[2]
3	−3.40	9.08[2]	−4.04	11.51[2]
4	−4.42	8.62[2]	−5.11	10.96[2]
5	−3.60	6.87[2]	−4.22	8.99[2]
6	−3.02	5.88[2]	−3.69	7.89[2]
7	−1.73	4.18	−2.41	6.29
8	−1.84	3.88[2]	−2.56	5.74
9	1.14	1.35	.52	2.97
Largest	.19	−1.07	−.12	.43

All α's come from a simple regression of the annual excess return on each decile against the indicated index measured in excess return form. In this and all tables that follow, annual excess returns are used where excess return is defined as the actual annual return minus the annual return from holding 30-day Treasury bills.

[1] This is the second index used by Ippolito (1989). It is constructed by equally weighting the S&P index and the Ibbotson–Sinquefield long-term bond index.

[2] Indicates t-value >2.1.

1. The Effect of Non-S&P Assets on Mutual Fund α's

To examine the effect of non-S&P assets on fund performance we considered two non-S&P asset categories commonly held by mutual funds: bonds and non-S&P stocks.[3] The α on a portfolio, which is the Jensen measure of portfolio performance, is a weighted average of the α's on the individual assets that compose the portfolio. Thus, to see the effect of inclusion of an asset category on the portfolio α, one only has to estimate the α for that asset category.

We wish to examine the effect on measured α's when mutual funds hold non-S&P assets and there is *no selection ability.* To do so, we will proxy the performance of non-S&P stocks and bonds by the return on several alternative passively managed portfolios.

1.1 Non-S&P stocks

The common stocks that compose Standard & Poor's index are selected to guarantee broad industry representation; however, within each industry the larger firms are generally selected. The weight placed on each stock in computing the index is proportional to the total market value of the firm's equity. The result is that the S&P index is primarily comprised of, and affected by, the large firms listed on the NYSE. Thus, indexes of the return on small (low capitalization) stocks

[3] All results in this and subsequent sections are based on annual returns. This allows direct comparison with Ippolito's results.

Table 2
***α* for alternative indexes**

Index	With S&P indexes		With combined indexes[1]	
	1945–1964	1965–1984	1945–1964	1965–1984
CRSP value-weighted	−0.26	0.57	−0.71	2.21
CRSP equal-weighted	−3.13	6.25	−3.91	8.41
CRSP small stock index	−4.04	10.06	−5.07	12.38

All returns are in annual excess return form.

[1] Second index used by Ippolito (1989). It is constructed by equally weighting the S&P index and the Ibbotson–Sinquefield long-term bond index.

should be good proxies for non-S&P stocks. We examine the impact of small stocks on α in two ways. The first uses the return indexes on deciles formed by market size constructed by the Center for Research in Security Prices (CRSP) at the University of Chicago. The CRSP return indexes are constructed by ranking all stocks listed on the NYSE by market capitalization, dividing the stocks into 10 equal-sized groups, and then value-weighting the return on individual stocks within each group.[4] Table 1 shows the α for each of the 10 deciles in the Jensen period and in the Ippolito period. A manager who had no selection ability and selected part of the portfolio from non-S&P stocks would have shown a negative α in the Jensen period and a strong positive α in the Ippolito period. The average α of the smallest five deciles in the Jensen period is −4.17 percent, while in the Ippolito period it is +9.61 percent. Similar results are obtained when the analysis is repeated employing Ippolito's combined bond stock index as the independent variable defining α's. Negative α's for the small stock index in the Jensen period do not imply that the returns on the small stock index are less than returns on the S&P index. In fact, in the Jensen period they were higher. In Table 2, we present another illustration of the same phenomenon. The small stock index is a value-weighted index of the lowest quintile of the stocks listed on the NYSE. It is the standard small stock index used in many academic studies. In the Jensen period, the small stock index exhibited a negative α of −4.04 percent, while in the Ippolito period the α is +10.06 percent.

In Table 2, we also present the α's for market-weighted and equally weighted averages of all stocks on the NYSE. These α's reinforce the effects shown by examining deciles. Examination of the α's on the market-weighted CRSP index of all NYSE stocks shows that extending holdings beyond stocks in the S&P index, even when maintaining

[4] If security α's are a function of size as in Table 1, the CRSP weighting procedure understates the α compared to equal weighting when α's increase with decreasing size and overstates the α when α decreases with increasing size.

Table 3
Regression results of potential bond proxies versus bond funds and Shearson Lehman index

	With Shearson Lehman[1]			With bond funds[2]		
Bond proxies	α	β	R^2	α	β	R^2
Intermediate govt.	1.61	0.83	.93	.51	0.85	.92
Long-term govt.	−4.61	1.46	.90	−.55	1.57	.90
Long-term corporate	−4.93	1.55	.95	−.42	1.66	.95
80:20 mix[3]	−0.30	0.97	.96	.33	1.01	.94

[1] Regressions of the bond proxies versus the Shearson Lehman Government Corporate Index were over the period 1973–1988.

[2] Regressions of the bond proxies versus the Bond Funds Index were over the period 1979–1988.

[3] The 80:20 mix is made up of 80% of the return on the intermediate-term government bond index and 20% of the return on the long-term bond index.

market weights, results in negative α's for the Jensen period and positive α's for the Ippolito period. The larger negative α's for the Jensen period and positive α's for the Ippolito period associated with the equally weighted index illustrates another influence. Holding stocks in proportions closer to equal rather than market proportions results in large positive α's in the Ippolito period and negative α's in the Jensen period, even with no selection ability.

The results shown in Tables 1 and 2 are dramatic. Given the extremely large α's, even a small part of the portfolio not being invested in S&P stocks would explain Ippolito's results on average performance (α), even with *zero* forecasting ability.[5]

1.2 Bonds

The second category of assets we examined was bonds. There were three widely used bond return indexes that existed from 1945 to 1984. These are the Ibbotson and Sinquefield (I&S) long-term corporate, intermediate-term government, and long-term government bond indexes. The first question we need to examine is which index is the best proxy for a passively managed bond fund. The most common index used by passive bond managers is the Shearson Lehman Government Corporate Index. Unfortunately, the Shearson Lehman index

[5] Another category of non-S&P assets held by mutual funds is international stocks. Academic literature on the advantages of international diversification first appeared in the late 1960s. This literature plus the availability of ADRs has led to the recent increase in the ownership of foreign assets by mutual funds. The Morgan Stanley index is the most widely used index of international securities and is the index most often matched by passive managers. The α on the Morgan Stanley index from the inception of the index in 1970 until 1984 was 6.97 percent. Thus, international diversification and random selection would have led to a positive α in Ippolito's period. Although the large size of the α associated with international stocks could account for Ippolito's results, the likelihood is that the effect is small. There are no international funds in the corrected sample. Non-U.S. stocks (generally in the form of ADRs) are only a small part of most noninternational mutual funds even today. Thus, the effect on Ippolito's α's of international diversification is likely to be very small over Ippolito's period but is likely to become increasingly important in the future.

Table 4
Results for 80:20 bond proxy regressed against the S&P index

	α	β	t_α	R^2	Excess return[1]
1945–1964	1.39	−.057	.83	.07	.57
1965–1984	−0.42	.12	−.25	.03	−.19

Regressions were also run using the equally weighted bond stocks index as the dependent variable. Results were substantially unchanged.

[1] Excess return refers to the annual return of the 80:20 bond index over and above the annual return of 30-day Treasury bills.

has only existed since 1973 and so cannot be examined for our entire sample period. To find a proxy for it, we regressed each of the Ibbotson and Sinquefield indexes as well as a composite I&S index comprised of 80 percent intermediate government and 20 percent long-term corporate bonds against the Shearson Lehman index over the period 1973 to 1988 (the period for which we have data on the Shearson Lehman index). The 80:20 ratio was examined as a possible proxy because its duration closely matched the duration of the Shearson Lehman index.

The results are shown in Table 3. As expected, the long duration of the Ibbotson and Sinquefield long-term indexes made these indices much more volatile than the Shearson Lehman index. The 80:20 mixture closely matches the index. The β on the Shearson Lehman index is 0.97 with an α of 0.30. The composite index is a reasonable proxy for a passive bond index fund.

We did one additional test. We have data from 1979 to 1988 on all bond mutual funds that existed in 1979.[6] Although these funds may have under- or overperformed an index fund, our proxy index ought to be highly related to the average active fund. Table 3 shows the results. The β of the 80:20 mixture was 1.01 with an α of 0.33. Thus, the 80:20 mixture has a very similar pattern of return to actual bond funds, but earned in this period an excess return. In subsequent analysis, we will use the 80:20 mixture as a proxy for a passive bond fund.

As shown in Table 4, the relationship between our proxy for a passive bond fund and the S&P index is insignificant in both periods. In the 1945–1964 period, the adjusted R^2 is 0.07 for the 80:20 mixture; whereas in the 1965–1984 period, the adjusted R^2 is 0.03. Thus, it is difficult to place much confidence in the α values.

If bond funds are unrelated to the S&P index, then it is the return above the riskless rate that determines their α.[7] The return on the

[6] See Blake (1992) for a detailed description of this data.

[7] In the Jensen model, a zero correlation with the index means a zero β. Thus, α is the return above the riskless rate.

bond proxy was 0.57 percent above the riskless rate in the first period and -0.19 percent in the second. The magnitude of the α's and excess returns for bonds are much smaller than the magnitude of the α's and excess returns for other non-S&P assets. Thus, any effects of mutual funds holding debt on the mutual fund α's are likely to be swamped by other influences. However, bonds do mitigate to a limited extent the effect of other non-S&P assets. Although the performance of bonds is unlikely to affect results in the Ippolito period, the reader should be cautious about generalizing. Most studies analyze about 10 years of data. We examined all 10-year periods in our sample. The average α for bonds on the S&P index varied from -2.6 to $+1.0$. Most recent studies have factor-analyzed stock returns in order to obtain a return generating process. Unless some of the factors that are produced by factor-analyzing equity returns are highly correlated with bond returns, the average α on balanced funds could be heavily influenced by the bond α.

2. Adjusting for Other Indexes

In the prior section we showed that inclusion of non-S&P assets in a mutual fund portfolio would result in nonzero α's, even if selection of all assets was random. The magnitude of the impact is a direct function of the fraction of non-S&P assets held. Since the fraction of non-S&P assets held by a fund is correlated with many of the characteristics of funds we wish to examine, it is important to control for this effect. In this section, we will discuss how one can control for the effect of non-S&P assets on mutual fund α's.[8]

One way to view a mutual fund is as a combination of three portfolios: one containing S&P stocks, one containing non-S&P stocks, and one containing bonds. The return on the fund is a weighted average of the return on the three portfolios. Management performance is the extra return earned on the fund compared to holding a combination of three passive portfolios with the same characteristics as the overall fund. The measure of performance developed from this approach can be considered a generalization of the Jensen measure to encompass the situation where managers hold non-S&P stocks and

[8] The other way to control for the effect of the non-S&P assets is the multifactor approach of Lehman and Modest (1987), Grinblatt and Titman (1989), Conner and Korajczyk (1991), and Jagannathan and Korajczyk (1986), where the factors include bond factors. The disadvantage of this approach is that the factors do not have passive portfolios that are actively traded in the market. Thus, an investor would be unable to replicate this strategy.

bonds.[9] It could also be considered a measure of performance given a belief in a three-factor return generating process. If a passive portfolio interpretation of the model is employed, one must consider costs.

Mutual fund returns are net of management fees and trading costs. When utilizing indexes as proxies for passive portfolios, these costs are not subtracted. The question that needs to be analyzed is how much lower are the returns on passive portfolios after fees are deducted relative to the indexes they duplicate. Institutional investors have available passive S&P index portfolios and bond portfolios that almost exactly match indexes after all fees and trading costs. For individual investors, index funds such as the Vanguard funds match the index within 25 basis points for the part of the fund not in cash. A number of analysts argue that returns on small stock indexes are unattainable by actual portfolio managers because of high transaction costs [e.g., Fouse (1989) and Stoll and Whaley (1983)]. Sinquefield (1991) examined the returns on the five passive small stock index funds with any substantial history. He found that net of trading costs and management fees, the passive funds earned 2 basis points per month more than the particular small stock index they followed. The worst shortfall was 12 basis points per month. Thus, even for small stocks, index returns may well be reasonable measures of passive returns net of costs, and a deduction of 25 basis points would be a generous adjustment.

The return on an active fund relative to the return on the three passive portfolios is

$$R_i - R_F = \alpha_i + \beta_{iM}(R_M - R_F) + \beta_{iS}\,(R_S - R_F) + \beta_{iD}(R_D - R_F) + e_i.$$

In this model,[10] R_M is the return on the S&P 500 index, R_S is the return on a non-S&P equity index that has been made orthogonal to the S&P index, R_D is the return on a bond return index that has been made orthogonal to both the S&P and the non-S&P equity index, β_{ij} is the sensitivity of portfolio i to the relevant index j, and e_i is the residual risk.

We will use two alternative proxies for the returns on non-S&P stocks: the small stock index and a value-weighted index of all NYSE stocks, each with the effect of the S&P removed. By orthogonalizing (removing the effect of another index from) the non-S&P indexes, we lose the interpretation of each index as the return on a passive

[9] The Jensen measure is usually stated with an equilibrium interpretation. Another interpretation is that it is the extra return earned by a manager compared to the return on a combination of an index fund and Treasury bills of the same risk.

[10] This regression model is consistent with a multifactor asset-pricing model. In independent research, Sharpe (1988) has developed a nine-index model based on asset classes to decompose investment returns.

Table 5
α's for individual funds

	No. pos.	Significant[1] (5%) Pos.	Neg.	Mean α	t-value*	Mean S&P β	Mean small stock β	Mean bond β	R^2
A. 1945–1964 (115 firms)									
1. Jensen		3	14	−1.1		.84			
B. 1965–1984 (143 firms)									
1. S&P (Ippolito uncorrected)		12	4	.81		.88			
2. S&P (Ippolito corrected)	87	9	5	.61	.56	.94			.74
3. S&P and small stock	38	0	15	−1.49	−2.11	.94	0.21		.79
4. S&P and value-weighted residual	47	1	11	−.81	−1.25	.94	2.58		.78
5. S&P, small stock, bond	34	0	21	−1.59	−2.35	.94	0.21	0.13	.83
6. S&P, value-weighted residual, bond	44	1	12	−.88	−1.46	.94	2.58	0.15	.83

[1] The t-value is calculated using the cross-sectional correlation of α's. The standard error of the mean α is $1/N^2$ times the sum of all the elements in the variance covariance matrix of the residuals. All t's in subsequent tables are calculated the same way.

fund. However, using orthogonalized rather than raw indexes does not affect the explanatory power of α and does highlight the marginal effect of non-S&P assets above that accounted for by the S&P index.[11] The portfolio used to generate a return on a passive debt portfolio is the portfolio consisting of 80 percent intermediate government bonds and 20 percent long-term corporate bonds orthogonalized to remove any effects of the other two indexes. Orthogonalization is performed by running a regression of the non-S&P index (e.g., the small stock index) against the S&P index. A new index is created by using the intercept plus the residual from this regression as the orthogonalized small stock index. This new index captures both the mean effect and the time-series impact of the effect.

Table 5 shows the average α in our sample for a number of different definitions of passive portfolios. We present results for the single-index case, the two-index case, and the three-index case. Jensen's results for the time period 1945–1964 are presented in part A of Table 5. Our new results and Ippolito's results are presented in part B. Our results on a corrected version of Ippolito's sample are presented in the second row of part B. While the corrections we make lower the average α for mutual funds and lead to estimates that are not statistically different from zero, they do not change the sign of the average α. The corrections we made in his sample are of two types. In checking his data, we found several errors. By far the most frequent error was

[11] By orthogonalization we leave the β on the S&P index unchanged and the β on other indexes captures the marginal impact of those indexes.

the omission of the negative sign on individual observations of returns with no corresponding misclassification of positive returns. Correcting individual observations changed the average α from 0.81 to 0.38. The second correction involved a slight redesign of the sample. Ippolito selected all funds that met certain criteria in 1984 and also had data or were formed from a fund that had data in 1965. Twelve of these funds did not meet Ippolito's criteria (e.g., no international funds) in 1965. We formed our sample from all firms that met Ippolito's criteria at the beginning of 1965. By coincidence the sample size was the same but 12 of Ippolito's funds were replaced by different funds. Actually drawing the sample correctly resulted in increasing the β from 0.38 to 0.61, which is still below Ippolito's estimate of 0.81. Also note the low t-value that occurs, in part because of the change in α, but also because we corrected t's for cross-sectional correlation and heteroskedasticity.[12]

Once we account for non-S&P stocks or bonds, the α's become negative and the results are similar to Jensen's results. Using any of the two-index or three-index models, the average α is negative. The α's involving the small stock index are significantly different from zero at the 1 percent level; more than two thirds of the funds have negative α's; and the number of negative α's that are significantly different from zero at the 5 percent level far exceeds the number of α's that are positive and different from zero at the 5 percent level. In computing the t-statistics for the α's, the standard error of the mean α was computed taking into account the cross-sectional correlation among α's. The variance of the average α is the vector of proportions (each equal to $1/N$) times the variance–covariance matrix of residuals times the vector of proportions. All subsequent t-statistics on α's were computed in this way. The pattern and significance of the α's are further evidence that Ippolito's results are due to the relative performance of alternative classes of securities rather than the ability of managers to select superior securities from any class.

Since in subsequent analysis we will feature the α's from the three-index model with the small stock index as one index, a few additional comments on this formulation are in order. The F-ratio shows that the addition of the index for non-S&P stocks and bonds improves the explanatory power of the equation. In addition, a substantial number of the sensitivities are significant. In the three-index model using the small stock index, 73 sensitivities are significant for the small stock index and 52 for the bond index.

All studies that employ multiple indexes have a potential problem: by using multiple indexes, they may fail to attribute to management

[12] We made this replacement since Brown et al. (1992) have shown survivorship bias can be a serious problem in this type of study and Ippolito's sample has a potential survivorship bias.

the element of good performance associated with superior sector selection. This is true of our study as well as those of Grinblatt and Titman (1989), Lehman and Modest (1987), Jagannathan and Korajczyk (1986), and Conner and Korajczyk (1991). In our study, given the performance of the indexes, the main concern is the split between S&P equities and non-S&P equities. We will present evidence that shows that investors in 1964 had information that would allow them to determine which funds would have significant small stock exposure. Furthermore, we will show that managers did not successfully time changes in the relative performance between S&P and non-S&P equities. Since at the beginning of the period the investor could duplicate the small stock strategy of any firm with an index fund, and since, as we show, fund management did not improve performance by changes in small stock strategy within the period, we believe the intercept from the multi-index model is the relevant measure of performance.[13]

To examine whether an investor could determine the sector selection of the funds in 1964, we examined the fund descriptions and the stock holdings reported in Wiesenberger in 1964. While the policy description is subjective, it is clear that the majority of the funds with very high (positive) small stock sensitivity were funds that had a policy of investing in high technology, high growth stocks. On the other hand, the funds with low small stock sensitivity tended to be funds that had been established for a longer period of time and stated their objectives as holding larger well-established companies. We computed the percentage of the stock portfolio held in non-S&P stocks as of 1965 for the 10 funds with the largest and the 10 funds with the smallest sensitivity to small stocks. For the funds with high positive sensitivity to small stocks, 56 percent of the stock portion of their portfolio consisted of non-S&P stocks; while for the low sensitivity group, the proportion of non-S&P stock was only 17 percent. These numbers were computed from Wiesenberger reports and thus were data that were available to any investor.[14] Clearly, the investor could have known in 1965 which funds were holding non-S&P equities.

In order to examine the fund managers' ability to sector select, we compared their exposure in periods of good and poor small stock

[13] An argument could be made that, while the investor could have matched the funds with passive portfolios, fund management should get credit for showing the investor the way. This is a difficult problem and one that is more general than this study. As an analogy, should the manager of a utility fund who buys exclusively utility stocks be judged against the performance of utility stocks or against the S&P index? We believe the appropriate benchmark is utility stocks, but that gives the manager neither credit (or discredit) for choosing that sector of the market. Having expressed our opinion, we leave the final resolution of this point up to the reader.

[14] We do not repeat these numbers for later years because Wiesenberger stopped reporting portfolio composition and other sources are not available until after our period.

performance. Our sample contained five years in which small stocks underperformed large stocks. If managers are sector timers, and have forecasting ability, then the sensitivity to small stocks should decrease in the years when small stocks did poorly. To test this, we estimate a separate small stock sensitivity measure (small stock β) for the years in which small stocks did poorly. This procedure, which employed a zero-one dummy variable to estimate the new slope, is analogous to the Merton–Henriksson [Henriksson and Merton (1981)] procedure, except that the switching is between S&P stocks and small stocks rather than between stocks and Treasury bills. Our procedure is also similar to that followed by Jagannathan and Korajczyk (1986). If managers are successful timers, then the differential β for years when small stocks did poorly should be negative. In fact, for our overall sample, only 44 out of 143 were negative, indicating some slight tendency for rotation into small stocks to take place at the wrong time. It is possible that timing was only relevant for those funds that had high sensitivity to small stocks. For the 10 percent of the firms with the largest small stock β's, seven had positive differential β's and seven negative. One was significantly positive at the 5 percent level and none significantly negative. For the top 20 percent of the firms ranked by small stock β's, the numbers are 14 positive and 15 negative with 1 significantly positive. These results lead us to believe that management has shown no ability to time the small stock effect. Given the ability of investors to identify the funds with significant non-S&P exposure and the lack of timing demonstrated by these funds, we feel our multi-index approach is appropriate.

Table 6 shows the performance of managers when mutual funds are classified in various ways. These classifications were used by Ippolito and are included here because they provide additional evidence on the performance of funds and the linkage between performance and portfolio composition. Ippolito in his analysis eliminated 15 funds because the β in the first 10 years differed from the β in the second 10 years, and we did not eliminate them. We repeated the test he used to determine changing β with the corrected sample. Only two firms were eliminated using his procedure. The difference arose because the data for the other firms he eliminated generally had substantial data errors. We included these firms, although eliminating them would not affect our results.[15]

[15] We utilized a different procedure for analyzing the effect of changing β. The percentage in common stock is highly related to β. In addition, firms that change their investment policy would be expected to change their sensitivities. Thus, we formed a sample that eliminated those firms with a change in policy (the exact method is discussed later in the text when we discuss Wiesenberger classifications) or changed the percentage invested in common stock by more than 50 percent. This eliminated 44 firms. We repeated all the tests and the results were substantially unchanged. Hence, we only report the results for the full sample.

The first classification is β. The β ranges used in this table are identical to those used by Ippolito. All numbers in the table except for t-values are simple averages across the included funds. The t-values were calculated by the method described in discussing Table 5. Examining the one-index model shows that α is positively related to β. This pattern is what should be expected given that non-S&P stocks generally have higher β's. Ippolito found no relationship between α and β when he used a single-index model. The difference is due to data errors in Ippolito's sample. A sign error (which was always a negative return recorded as positive) increases the α. In addition, it lowers β. Thus, the sign errors in Ippolito's sample were concentrated in the low β category. When the three-index model is utilized, the pattern disappears. Adding the additional passive portfolios significantly increases the explanatory power. In addition, the sensitivities to the non-S&P indexes are sensible. The sensitivity to the small stock index increases as the S&P β increases. Mutual funds with high S&P β have high exposure to small stocks. Likewise, the exposure to bonds decreases as the S&P β increases.

The second categorization we utilized is percent of the funds' assets invested in common stock. We utilized the average percent over 20 years. We adopted Ippolito's categories. For the one-index model, the greater the percent in common stock, the greater the α. When the three-index model is examined, the relationship disappears. The β's on the other indexes show why. The small stock β increases with percent in common stock and the bond β decreases.

The third section of the table examines performance by objective and policy as reported by Wiesenberger. Four categories were used. First, firms' investment policy was used to divide funds into groups that either invest only in common stock or as a policy invest in bonds and common stock (balanced). Within the common stock category, we further classified firms according to their stated objective. A number of firms changed their policy or objective over the 20 years. If the firm changed its policy from common stock to balanced or the reverse, it was excluded from this part of Table 6.[16] Further, if it changed its objective more than just to an adjacent category or was not in the same category for at least 14 of the 20 years, it was excluded. These procedures led to 27 firms being excluded.

When examining the one-index model, the α declines as the category becomes less aggressive. When three indexes are used, the pattern is reversed for the common stock funds. The pattern of sensitivities is as expected. Sensitivity to the S&P index declines uni-

[16] Three firms that should have been discarded by this criteria were not. They were listed in a different category their first two years. However, the percent in equity was of the same magnitude in these years so that we judged the policy was not changed but that the category change was to reflect practice more accurately.

Table 6
Performance of mutual funds for various criteria

Classification criteria	One-index model[1] No. of funds	α	t_α[4]	$\beta_{S\&P}$	R^2_{adj}	Three-index model[2] α	t_α	$\beta_{S\&P}$	β_{SS}	β_B	R^2_{adj}	Ippolito[3] α
β												
$\beta <$ .77	30	−0.31	−.25	0.66	.69	−1.92	−3.76	0.66	.10	.72	.86	0.86
.77 $< \beta <$ 1.02	72	0.59	.80	0.91	.81	−0.77	−1.44	0.91	.13	.12	.86	0.85
$\beta >$ 1.02	41	1.33	.56	1.19	.65	−2.80	−1.78	1.19	.44	−.29	.77	0.75
Stock share												
<65%	32	0.36	.30	0.70	.67	−1.41	−2.77	0.70	.12	.65	.85	0.44
65–90%	75	0.54	.41	0.98	.75	−1.92	−2.21	0.98	.25	.01	.82	0.99
>90%	36	1.00	.81	1.06	.78	−1.07	−1.16	1.06	.21	−.09	.84	0.88
Wiesenberger type												
1. Common stock												
a. Max capital gain	12	1.63	.42	1.34	.54	−4.59	−1.87	1.34	.69	−.81	.77	0.82
b. Growth	33	0.82	.54	1.05	.73	−1.55	−1.23	1.05	.24	.06	.79	1.28
c. Growth and income	40	0.66	.98	0.92	.86	−0.68	−1.65	0.92	.12	.16	.89	0.45
2. Balanced	31	0.34	.30	0.71	.73	−1.27	−2.73	0.71	.11	.66	.89	0.62

[1] The one-index model regresses excess returns of funds on the excess returns on the S&P.

[2] The three-index model regresses excess fund returns on the excess returns on the S&P, the excess returns on the small-stock index orthogonalized on the S&P, and the 80:20 bond index orthogonalized on the S&P and small stock index returns.

[3] The α's shown in this column are those reported by Ippolito (1989).

[4] t_α refers to the t-statistic on α.

formly as the fund's objective becomes less aggressive. The small stock β declines and the bond β increases as the fund's objective becomes less aggressive. Thus, the Wiesenberger objective seems to describe fund behavior accurately on average.

3. Market Efficiency Turnover and Expenses

The basic idea underlying the Grossman (1976) and Grossman and Stiglitz (1980) view of efficiency is that informed investors earn a sufficient amount to just compensate for the cost of information gathering. High expenses do not necessarily imply inferior performance, even after expenses have been deducted. Ippolito's article has received much attention because he found no relationship between turnover or expenses and performance net of expenses. This was different than most prior studies and lent strong support to the Grossman view of efficiency.

In this section, we reexamine this question after adjusting for the effect of non-S&P assets. We will examine the relationship between performance and three cost variables: expense ratios, turnover, and load costs.

Recall that mutual fund returns are measured after the deduction of expenses and costs associated with turnover, but before any costs to the investor associated with load charges. Because of the way returns are calculated, load costs should have a different impact on performance statistics than the other costs. Let us start by examining expense ratios and turnover. Expenses are incurred or portfolios changed in an attempt to increase performance. If mutual funds invest money efficiently, there should be no relationship between performance (α) and either expense ratios or turnover. Table 7 shows the relationship between performance and both turnover and expenses. Part A of Table 7 divides the funds into quintiles by expense ratios. Equally sized quintiles are formed on the basis of average expense ratios for each fund and the α for each quintile is displayed.[17] In addition, we ran a cross-sectional regression between the α on each fund in the sample and their average expense ratio. When a single-index model is employed, there is no statistically significant relationship between performance and expense ratios. When we examine the three-index model, the results change. There is a statistically

[17] Ippolito treated each year's expense ratio separately. There are two reasons why we emphasize average rather than year-to-year data. First, there is very little year-to-year variation in expenses. The high expense firms are high throughout and the low are low throughout. Second, there is a fair amount of latitude concerning when expenses are charged and mutual funds time expenses to smooth performance. For example, management fees can and are deferred in years of relatively poor performance. An extra firm was included in each of the three middle quintiles.

Table 7
Effect of average expenses and average turnover on α

A. Quintiles by expense ratios

Single-index[1]

Group	Average expenses	α	t_α[2]	$\beta_{S\&P}$	R^2_{adj}
High	.912 < E < 2.020	0.06	0.06	1.01	.60
2	.753 < E < 0.912	1.22	0.90	0.93	.69
3	.680 < E < 0.753	1.10	1.08	0.91	.73
4	.590 < E < 0.680	0.28	0.34	0.92	.79
Low	E < 0.590	0.33	0.47	0.92	.86

Three-index[3]

Group	Average expenses	α	t_α	$\beta_{S\&P}$	β_{SS}	β_{B}	R^2_{adj}
High	.912 < E < 2.020	−3.87	−3.56	1.01	.40	−.07	.76
2	.753 < E < 0.912	−1.68	−2.34	0.93	.27	.23	.83
3	.680 < E < 0.753	−0.69	−0.86	0.91	.17	.16	.82
4	.590 < E < 0.680	−1.19	−1.93	0.92	.13	.18	.86
Low	E < 0.590	−0.59	−0.89	0.92	.08	.13	.89

B. Quintiles by turnover

Single-index

Group	Average turnover	α	t_α[2]	$\beta_{S\&P}$	R^2_{adj}
High	72% < T < 162%	.75	.46	1.05	.69
2	51% < T < 72%	.75	.49	0.98	.68
3	34% < T < 51%	.35	.30	0.93	.74
4	22% < T < 34%	.64	.69	0.87	.76
Low	T < 22%	.58	.93	0.86	.81

Three-index

Group	Average turnover	α	t_α	$\beta_{S\&P}$	β_{SS}	β_{B}	R^2_{adj}
High	72% < T < 162%	−2.21	−1.94	1.05	.30	−.03	.80
2	51% < T < 72%	−1.87	−1.60	0.98	.26	−.01	.79
3	34% < T < 51%	−2.17	−3.39	0.93	.24	.14	.85
4	22% < T < 34%	−1.11	−2.11	0.87	.15	.34	.86
Low	T < 22%	−0.58	−1.49	0.86	.10	.20	.88

[1] The one-index model regresses excess returns of funds on the excess return on the S&P.

[2] t_α refers to the t-statistic on α.

[3] The three-index model regresses excess fund returns on the excess returns on the S&P, the excess returns on the small-stock index orthogonalized on the S&P, and the 80:20 bond index orthogonalized on the S&P and small stock index returns.

significant relationship between α and expenses (t-values of −7.12) and a relationship that can clearly be seen by examining the quintiles. Higher expenses are associated with poorer performance. Management does not increase performance by an amount sufficient to justify higher fees. Examination of the individual sensitivities to each of the three indexes shows why the results are different for the single-index model. High expense funds tend to be more sensitive to (place more money in) small stocks and less in bonds than do low expense managers. However, they underperform what an investor could earn by allocating his or her funds to three passive portfolios. When a single-index model is used, this relationship between performance and

Table 8
Performance of load and no-load funds

	Number	From one-index model[1]				From three-index model[2]					
		α	t_α	$\beta_{S\&P}$	R^2_{adj}	α	t_α	$\beta_{S\&P}$	β_{ss}	β_{B}	R^2_{adj}
Load	90	.48	.49	.92	.74	−1.55	−2.38	.92	.19	.19	.84
No load	19	.86	.78	.96	.72	−0.84	−0.89	.96	.17	.01	.81
Switch	33	.84	.56	.97	.72	−2.17	−3.30	.97	.30	.05	.83

One fund, State Street, was not classified by Wiesenberger.

[1] t_α refers to the t-statistic on α.

[2] The one-index model regresses excess returns of funds on the excess returns on the S&P.

expenses is obscured by differences in the investment philosophy of funds. This is the same phenomenon we saw earlier when we examined the relationship between α and β in the single-index model.

When we examine turnover, we find the same type of results, although the relationships are not nearly as strong. With the one-index model, performance appears to be very weakly positively related to turnover. With the three-index model, the relationship between performance and turnover becomes negative and significant at the 5 percent level using a two-tailed test. Management does not earn enough excess return to compensate for the full cost of increased turnover.

Our principal test relating performance to either expenses or turnover used the average values for expenses and turnover because there was little variation in year-to-year values. Although we do not have a long history on turnover, we have 20 years of expense ratios. Since Ippolito examined yearly values, we performed a test on yearly values for expenses to be more comparable with his tests. Each year we regressed α plus the residual for each firm on the yearly expenses for the firm. Of the 20 slopes, 15 were negative and 10 significantly so. Estimating the standard deviation of the mean of the slopes, using the 20 observed slopes, and testing for the difference of the average slope from zero, produced a t-value of −3.28, which is highly significant.

The last cost we examine is the load charge on funds. Since the load cost of purchasing funds is not deducted in calculating returns, load funds need a higher α in order to be an attractive investment. In fact, as seen in Table 8, load funds have a lower α whether the single-index or three-index model is used. There is no evidence in the data that mutual funds that charge a load compensate investors for the added cost.

Having examined the relationship between performance and average expenses, average turnover and load, we now examine whether funds change these in response to performance. One view of mutual

Table 9
α in two 10-year subperiods and percent change in expenses from earlier subperiod to later (deciles sorted by α during 1965–1974)

Deciles	1965–1974 α	1975–1984 α	% change in expenses
Low	−5.22	−4.37	34.1
2	−2.71	−2.68	34.9
3	−2.01	−2.59	34.2
4	−1.40	−2.02	22.2
5	−1.03	0.18	16.6
6	−0.56	−1.98	27.6
7	0.05	−0.48	16.4
8	0.64	−0.98	17.2
9	1.54	−1.26	17.4
High	3.97	−1.68	9.4

fund expenses is that mutual fund managers charge what their performance justifies. A large portion of expenses are under the control of managers. That is, they are expenses to a fund's shareholder but they are set by and controlled by management. Successful managers can raise their fees and unsuccessful managers may have to lower them. To examine this question, we divided the sample into two 10-year periods (Table 9). In the first 10-year period, we calculate α from our three-index model, rank the firms by α, and divide the sample into 10 equally sized groups. The last column shows the percentage change in average yearly expenses from the first 10 years to the last 10 years for each of our groups. The percentage increase in expenses declines almost linearly, with an increase in performance. Thus, the table does not provide evidence of expenses being changed in relationship to past performance.[18]

Another way to test the hypothesized interrelationship between expenses and performance is to examine performance over time. If expenses are just sufficient to eliminate the excess return due to being informed, then performance should be unpredictable. To test this, we used the same deciles ranked by α in the first 10 years. For these deciles, we calculated the average α in the second 10 years. This is shown in the second column of Table 9. A rank correlation test is highly significant. In addition, a regression of α in the second period on α in the first period is significant at the 5 percent level. Thus, historical performance of managers is somewhat predictive of future performance and superior performance is not simply a reward for the cost of obtaining information.[19]

[18] The results are consistent with expenses having a fixed and variable component and, thus, average expenses increasing for firms with relatively poor performance.

[19] The significance is heavily dependent on the poor α for the first category so that the reader should be somewhat cautious concerning these results. The results may indicate simply that there is consistency in a set of inferior managers. The predictability might also simply reflect differences in expenses across categories.

Table 10

Time	Unique return switches[1] minus average unique return nonswitches
−7	+1.22
−6	−1.91
−5	−3.17
−4	−1.88
−3	−0.61
−2	0.78
−1	0.72
+1	−2.54
+2	−0.42
+3	−4.00
+4	0.66
+5	2.22
+6	−3.38
+7	1.60

[1] The data is lined up in event time and is arranged such that the year in which the switch occurred is taken to be +1. The unique return is non-index-related return; that is, in the regression $y = \alpha + \beta x + \epsilon$, the unique return is $\alpha + \epsilon$.

As a final test, we examined the performance of funds around the time they changed from load to no-load funds. Since load fees are a sales charge and unrelated to the costs of security analysis, we would not expect to see this change affect performance. Table 10 shows the non-index-related unique return for all funds that switched from load to no-load compared to the average non-index-related return for the funds that did not switch. Data is arranged in event time. To obtain the entry for −1, we calculate for each switching fund the difference between its unique return in the year before the switch and the average unique return for all nonswitching funds in the same year. The figure in the table is the average of this number across all 33 funds that switched. The data for +1 is the year in which the switch took place. The data around the switch do not present a clear pattern. There is a slight tendency for funds that switched to do slightly better than the average fund in the two years before the switch and to do worse in the few years after the switch. As shown in Table 9, funds that charge a load fee and funds that switched did worse for investors than funds that have been no load over a long period of time.

4. Conclusion

Market efficiency is one of the major paradigms of financial economics. Modern theories of efficiency argue that informed investors in an efficient market will earn just enough to compensate for the cost of obtaining the information. Mutual fund managers are commonly viewed as the prototype of informed investors. Ippolito (1989) found that mutual funds earned a positive α before load charges and that fund performance was unrelated to expenses and turnover as predicted by efficiency arguments. In this article, we show that Ippolito's results are due to using a metric that did not appropriately account for the performance of non-S&P assets. More specifically, when the performance of non-S&P assets is appropriately accounted for, Ippolito's findings are reversed and the results are consistent with prior literature. Mutual fund managers underperform passive portfolios. Furthermore, funds with higher fees and turnover underperform those with lower fees and turnover. Finally, funds do not adjust expenses over time to reflect their performance.

Ippolito was incorrect in stating that mutual fund performance provides evidence in support of the Grossman view of efficient markets.

References

Admati, A., S. Bhattacharya, P. Pfleiderer, and S. Ross, 1986, "On Timing and Selectivity," *Journal of Finance*, 41, 715–731.

Blake, C., 1992, "Bond Fund Performance," unpublished dissertation, New York University.

Brown, S. J., W. Goetzmann, R. G. Ibbotson, and S. A. Ross, 1992, "Survivorship Bias in Performance Studies," *Review of Financial Studies*, 5, 553–580.

Conner, G., and R. Korajczyk, 1991, "The Attributes, Behavior, and Performance of U.S. Mutual Funds," *Review of Quantitative Finance and Accounting*, 1, 4–25.

Copeland, T., and D. Mayers, 1982, "The Value Line Enigma (1965–1978): A Case Study of Performance Evaluation Issues," *Journal of Financial Economics*, 10, 289–322.

Fouse, W., 1989, "The Small Stock Hoax," *Financial Analyst Journal*, July/August, 4–22.

Friend, I., M. Blume, and J. Crockett, 1970, *Mutual Funds and Other Institutional Investors*, McGraw-Hill, New York.

Grinblatt, M., and S. Titman, 1989, "Portfolio Performance Evaluation: Old Issues and Insights," *Review of Financial Studies*, 2, 393–421.

Grossman, S., 1976, "On the Efficiency of Competitive Stock Markets When Traders Have Diverse Information," *Journal of Finance*, 31, 573–585.

Grossman, S., and J. Stiglitz, 1980, "On the Impossibility of Informationally Efficient Markets," *American Economic Review*, 70, 393–408.

Henriksson, R., and R. Merton, 1981, "On Market Timing and Investment Performance II: Statistical Procedures for Evaluating Forecasting Skills," *Journal of Business*, 54, 513–534.

Ippolito, R., 1989, "Efficiency with Costly Information: A Study of Mutual Fund Performance," *Quarterly Journal of Economics*, 104, 1–23.

Jagannathan, R., and R. Korajczyk, 1986, "Assessing the Market Timing Performance of Managed Portfolios," *Journal of Business*, 59, 217–235.

Jensen, M., 1968, "The Performance of Mutual Funds in the Period 1945–1964," *Journal of Finance*, 23, 389–416.

Lehman, B., and D. Modest, 1987, "Mutual Fund Performance Evaluation: A Comparison of Benchmarks and Benchmark Comparisons," *Journal of Finance*, 42, 233–265.

Roll, R., 1971, "Sensitivity of Performance Measurement to Index Choice Commonly Used Indices," working paper, University of California, Los Angeles.

Roll, R., 1977, "A Critique of the Asset Pricing Theory Tests: Part 1 on Past and Potential Testability of the Theory," *Journal of Financial Economics*, 4, 129–176.

Ross, S., 1977, "Risk, Return, and Arbitrage," in I. Friend and J. Bicksler (eds.), *Risk and Return in Finance*, Ballinger, Cambridge, Mass.

Sharpe, W., 1966, "Mutual Fund Performance," *Journal of Business*, 39, 119–138.

Sharpe, W., 1988, "Determining a Fund's Effective Asset Mix," *Investment Management Review*, June, 5–15.

Sinquefield, R., 1991, "Are Small Stock Returns Achievable?" *Financial Analyst Journal*, Jan/Feb, 45–50.

Stoll, H., and R. Whaley, 1983, "Transactions Costs and the Small Firm Effect," *Journal of Financial Economics*, 12, 57–79.

Wiesenberger, A., 1964, *Investment Companies*, Arthur Wiesenberger and Co., New York.