The Investment Performance of U.S. Equity Pension Fund Managers: An Empirical Investigation

T. Daniel Coggin; Frank J. Fabozzi; Shafiqur Rahman

The Journal of Finance, Vol. 48, No. 3, Papers and Proceedings of the Fifty-Third Annual Meeting of the American Finance Association: Anaheim, California January 5-7, 1993. (Jul., 1993), pp. 1039-1055.

Stable URL:
http://links.jstor.org/sici?sici=0022-1082%28199307%2948%3A3%3C1039%3ATIPOUE%3E2.0.CO%3B2-F

The Journal of Finance is currently published by American Finance Association.

THE JOURNAL OF FINANCE • VOL. XLVIII, NO. 3. • JULY 1993

The Investment Performance of U.S. Equity Pension Fund Managers: An Empirical Investigation

T. DANIEL COGGIN, FRANK J. FABOZZI, and SHAFIQUR RAHMAN*

ABSTRACT

This paper presents an empirical examination of the selectivity and market timing performance of a sample of U.S. equity pension fund managers. Regardless of the choice of benchmark portfolio or estimation model, the average selectivity measure is positive and the average timing measure is negative. However both selectivity and timing appear to be somewhat sensitive to the choice of a benchmark when managers are classified by investment style. Meta-analysis revealed some real variation around the mean values for each measure. The 80 percent probability intervals for selectivity revealed that the best managers produced substantial risk-adjusted excess returns. We also found a negative correlation between selectivity and timing, but we argue that the observed negative correlation in our data is largely an artifact of negatively correlated sampling errors for the two estimates.

EACH YEAR *Pensions & Investments*, a leading trade newspaper for the pension management industry, profiles the top 1000 public and private U.S. pension funds. At year-end 1990, these funds had total pension assets of $1.876 trillion. Approximately $750 billion (40 percent) was invested in equities. The Investment Company Institute estimates that $250 billion was invested in open- and closed-end equity-oriented U.S. mutual funds at year-end 1990. This snapshot indicates a 3 : 1 ratio for pension fund equity investment versus mutual fund equity investment. Not only is the dollar difference large, but also the difference in the number of managers in each universe is large. The total number of pension fund managers is much larger than the number of mutual fund managers, by a ratio of approximately 10 : 1. Yet surprisingly little research has been done on the investment performance of U.S. equity pension fund managers. This paper begins to fill an important gap in the literature by providing empirical evidence on the investment performance of these managers.

The focus of this study is on equity pension fund managers who have been allocated funds by pension plan sponsors. Brinson, Hood, and Beebower

*Coggin is from the Virginia Retirement System, Fabozzi is from Frank J. Fabozzi & Associates, and Rahman is from Portland State University. We thank Jon Christopherson of the Frank Russell Company for providing the pension manager data used in this study. The paper has benefited from the comments of John E. Hunter, Richard Roll, Charles Trzcinka, and seminar participants at Mellon Capital Management and SUNY-Buffalo. The opinions and conclusions offered in this study do not necessarily represent those of the Virginia Retirement System or the Frank Russell Company.

(1986), Ippolito and Turner (1987), and Berkowtiz, Finney, and Logue (1988) examined the investment performance of a sample of large U.S. pension plans. Each plan may be composed of many fund managers in different asset categories with their own specific investment objectives and styles. In a recent study containing a wealth of information about the pension management industry, Lakonishok, Shleifer, and Vishny (1992) examined the annual returns of a sample of equity pension funds over the period 1983 to 1989. However, they made no risk adjustment, used only the S & P 500 Index as a benchmark portfolio, and did not distinguish between selectivity and market timing skill. Hence their results are not comparable to ours. To date, ours is the only study we know of that specifically examines the components of the investment performance of a sample of U.S. equity pension fund managers.

The two components we examine are security selection ability and market timing skill. Security selection involves the identification of individual securities which are under- or overvalued relative to the market in general. Within the specification of the Capital Asset Pricing Model (CAPM), the investment manager attempts to identify securities with expected returns which lie significantly off the security market line. The manager will then invest in those securities which offer an abnormally high risk premium. Market timing refers to forecasts of return on the market portfolio. If the manager believes he can forecast the market return, he will adjust his portfolio risk level accordingly.

According to the efficient market hypothesis, all active investment management activity is futile. The only rational investment choice for a plan sponsor is to invest in a passively managed market index. Hence, in an efficient market, plan sponsors would not rationally invest in (or pay active management fees for) an investment program which cannot outperform a market index. However, there exists a very large active pension fund management business in the United States. Our study begins to shed some light on the question of whether or not plan sponsors are behaving rationally to perpetuate this business. Our paper is organized as follows. Section I presents the models of selectivity and market timing used in this paper. Section II describes the data and methodology. Section III presents the empirical results. Section IV presents a meta-analysis of our results. Section V discusses the results. Section VI concludes our paper.

I. Models of Selectivity and Timing

It is important that portfolio managers be evaluated on both security selection ability and market timing skill. Furthermore, it has become standard practice to model selectivity and timing simultaneously. Jensen (1968, 1969) formulated a return-generating model to measure performance of managed portfolios. The model is:

$$R_{pt} = \alpha_p + \beta_p R_{mt} + u_{pt} \tag{1}$$

where R_{pt} is the excess (net of risk-free rate) return on the pth portfolio. R_{mt} is the excess (net of risk-free rate) return on the market portfolio, α_p is a measure of security selection ability, β_p measures the sensitivity of the portfolio to the market return, u_{pt} is a random error which has expected value of zero and t denotes time. This specification assumes that the risk level of the portfolio under consideration is stationary through time and ignores the market timing skill of the managers. Indeed, portfolio managers may shift the overall risk composition of their portfolio in anticipation of broad market price movements. Fama (1972) and Jensen (1972) addressed this issue and suggested a somewhat finer breakdown of performance.

Treynor and Mazuy (1966) added a quadratic term to equation (1) to test for market timing skill. They argued that if a manager can forecast market returns, he will hold a greater proportion of the market portfolio when the return on the market is high and a smaller proportion when the return on the market is low. Thus, the portfolio return will be a nonlinear function of the market return as follows:

$$R_{pt} = \alpha_p + \beta_p R_{mt} + \gamma(R_{mt})^2 + \epsilon_{pt} \tag{2}$$

A positive value of γ would imply positive market timing skill.

Jensen (1972) developed a similar model to detect selectivity and timing skill of managers. Jensen's measure of market timing performance calls for a fund manager to forecast the deviation of the market portfolio return from its consensus expected return. By assuming that the forecasted return and the actual return on the market have a joint normal distribution, Jensen shows that a market timer's forecasting skill can be measured by the correlation between the market timer's forecast and the realized return on the market. He concluded that, under the above structure, the separate contributions of selectivity and timing cannot be identified unless, for each period, the manager's forecast and the consensus expected return on the market portfolio, $E(R_m)$, are known.

Bhattacharya and Pfleiderer (1983) extended the work of Jensen (1972). By correcting an error made in Jensen (1972), they show that one can use a simple regression technique to obtain measures of timing and selection ability. Jensen assumed that the manager uses the unadjusted forecast of the market return in the timing decision. Bhattacharya and Pfleiderer assume that the manager adjusts forecasts to minimize the variance of the forecast error. They specify a relationship in terms of observable variables, which is similar to the Treynor and Mazuy model:

$$R_{pt} = \alpha_p + \theta E(R_m)(1 - \Psi)R_{mt} + \Psi\theta(R_{mt})^2 + \theta\Psi\epsilon_t R_{mt} + u_{pt} \tag{3}$$

where

α_p = security selection ability (risk-adjusted excess return),
θ = the fund manager's response to information; i.e., risk level deviation from the target risk level depending on the optimal forecast of the market return,

Ψ = the coefficient of determination between the manager's forecast and the excess return on the market, and
ϵ_t = the error of the manager's forecast.

The quadratic regression of R_{pt} on R_{mt} allows us to detect the existence of stock selection ability as revealed by α_p. The disturbance term in equation (3):

$$\omega_t = \theta \Psi \epsilon_t R_{mt} + u_{pt} \tag{4}$$

contains the information needed to quantify the manager's timing skill. We can extract this information by regressing $(\omega_t)^2$ on $(R_{mt})^2$:

$$(\omega_t)^2 = \theta^2 \Psi^2 (\sigma_\epsilon)^2 (R_{mt})^2 + \varsigma_t, \tag{5}$$

where

$$\varsigma_t = \theta^2 \Psi^2 (R_{mt})^2 \left[(\epsilon_t)^2 - (\sigma_\epsilon)^2 \right] + (u_{pt})^2 + 2\theta \Psi R_{mt} \epsilon_t u_{pt}. \tag{6}$$

The proposed regression produces a consistent estimator of $\theta^2 \Psi^2 \sigma^2 \epsilon$, where $(\sigma_\epsilon)^2$ is the variance of the manager's forecast error. Using the consistent estimator of $\theta \Psi$, which we recover from equation (3), we obtain $(\sigma_\epsilon)^2$. This coupled with knowledge about $(\sigma_\pi)^2$, the variance of excess return on the market, allows us to estimate $\Psi = (\sigma_\pi)^2 / [(\sigma_\pi)^2 + (\sigma_\epsilon)^2] = \rho^2$, where ρ is the correlation between the manager's forecast and excess return on the market. Finally, we calculate ρ, which truly measures the quality of the manager's timing information.

The Bhattacharya and Pfleiderer model of equation (3) is a refinement of the Treynor and Mazuy model. It focuses on the coefficient of the squared excess market return as an indication of timing skill. It is the first model that analyzes the error term to identify a manager's forecasting skill. Such a refinement should make the model more useful than previous ones. Further detail and econometric issues relating to the Bhattacharya and Pfleiderer model are discussed in Section II. In the empirical tests reported in Section III, we employed both the Treynor and Mazuy and the Bhattacharya and Pfleiderer models. This allows us to examine the sensitivity of results to alternative model specifications.

There are other models in the literature that permit identification and separation of selectivity and timing skills of portfolio managers; i.e., models by Grinblatt and Titman (1989b), Henriksson and Merton (1981), and an alternative to the Henriksson and Merton model proposed by Kon and Jen (1978, 1979). The Grinblatt and Titman model requires the historical sequence of portfolio weights (i.e., the amount invested in each stock) for the manager. Unfortunately, data on portfolio weights are very costly, time-consuming, and not often available. The Henriksson and Merton model provides no significant advantage over the Bhattacharya and Pfleiderer model. The weakness of the Henriksson and Merton model is that it only tests whether the manager has special information, and does not test whether the manager uses the information correctly (Dybvig and Ross (1985)). The forecasters in this model are less sophisticated than those of the Bhattacharya and Pfleiderer model, where they do forecast how much better the superior investment will perform. Henriksson and Merton assume that managers have a coarse information structure in which dichotomous signals are only predictive of the sign of the excess return of the market relative to the risk-free rate. In their model, the probability of receiving an "up" or a "down" signal in no way depends upon how far the market will be "up" or "down."

II. Data and Methodology

The data for this study consist of monthly returns for the period January 1983 through December 1990 (96 months) for a random sample of 71 U.S. equity pension fund managers with complete data for the entire period. The 71 managers were chosen from the Frank Russell pension manager database by a stratified random drawing so as to reflect as close as possible the actual distribution of managers (by investment style) in the database. Returns include dividends and are before expenses and management fees. An earlier version of this paper incorrectly stated that the returns are net of fees. The data include returns on tax-free, fully discretionary equity portfolios that are at least $5 million in size. These portfolios are managed by banks, insurance companies, and investment advisors who have been allocated funds by pension plan sponsors. The identities of the fund managers and sponsors are not included. The managers invest exclusively in the U.S. equity market. The random sample of pension fund managers was provided by the Frank Russell Company of Tacoma, Washington. Among other services, the Frank Russell Company evaluates the performance of the managers of a number of pension funds throughout the United States. The Frank Russell Company segregates equity managers into four basic investment styles on the basis of managers' portfolio characteristics. The styles are: (1) Earnings Growth, (2) Market-Oriented, (3) Price-Driven, and (4) Small Capitalization. Our sample consists of 18 Earnings Growth, 19 Market-Oriented, 18 Price-Driven, and 16 Small Capitalization managers. An appendix is available that more fully describes these four investment styles and their benchmark portfolios. Monthly returns on the 91-day Treasury bill were used as a proxy for the risk-free rate.

Our study uses several alternative equity benchmark portfolios. Two of these are the S & P 500 Index and the Russell 3000 Index. The Russell 3000 Index is a broad equity market index like the S & P 500. The idea of investment "style management" is becoming increasingly important to both academic studies and professional investment management (see, e.g., Tierney and Winston (1991)). Therefore, in addition to the two broad equity market indices, we also use four style indices as benchmarks. To be more specific, we use separate benchmarks for four different investment styles. The style indices are the Russell 1000 Index (for Market-Oriented managers), the Russell 2000 Index (for Small Cap managers), the Russell Price-Driven Index (for Price-Driven managers), and the Russell Earnings Growth Index (for Earnings Growth managers). The use of several alternative indices allows us to examine the sensitivity of pension fund manager's performance to alternative benchmarks. An estimate of the variance of the excess return on the market, $(\sigma_\pi)^2$, was derived from observed returns for each benchmark following the procedure of Lee and Rahman (1990).

In the empirical test, it is necessary to correct for heteroscedasticity in both the Treynor and Mazuy model and the Bhattacharya and Pfleiderer model. In the Treynor and Mazuy model, the error term will exhibit conditional heteroscedasticity because of the fund manager's attempt to time the market, even though security returns are assumed to be independent and identically distributed through time. To correct this, following Breen, Jagannathan, and Ofer (1986) and Lehmann and Modest (1987), we use heteroscedasticity-consistent standard errors proposed by White (1980), Hansen (1982), and Hsieh (1983). The significance tests used in Section IV are based on heteroscedasticity-adjusted t-statistics.

In the Bhattacharya and Pfleiderer model, the procedure discussed in Section I does not produce the most efficient estimates of the parameters since the disturbance terms in equations (3) and (5) are heteroscedastic. We used a generalized least squares (GLS) procedure to obtain efficient estimates of parameters. This methodology is more fully described in Lee and Rahman (1990).

As noted in Coggin and Hunter (1993), one weakness of the Treynor and Mazuy and the Bhattacharya and Pfleiderer models is that they ignore negative or inferior market timing. We modify these models to allow negative timing skill. We hypothesize that managers can exhibit negative ex post timing skill. In the Treynor and Mazuy model, this means the managers hold a smaller portion of the market portfolio when the market return is high. In the Bhattacharya and Pfleiderer model, this is indicative of a negative correlation between the manager's beta and the market return. Such results in both models could be due to the inability of managers to correctly forecast the expected return on the market portfolio. Hence these managers would forecast the market return to be high when it is actually low and vice versa. In the Treynor and Mazuy model of equation (2), a negative value of γ would be indicative of poor market timing skill.

For the Bhattacharya and Pfleiderer model, we examine the sign of the coefficient of $(R_{mt})^2$ in equation (3). Intuitively, in the spirit of the Treynor and Mazuy model, the sign of this coefficient will be indicative of the direction of timing skill. If the estimated value of this coefficient is negative, we desginate timing skill (given by ρ) to be poor (negative). This modification makes these models more realistic. A similar adjustment of the Bhattacharya and Pfleiderer model was implicitly introduced in Jagannathan and Korajczyk (1986, p. 229).

III. Empirical Results

A. Mean Values of Performance Measures

Table I presents the means of the selectivity and timing values for all managers and for the subsets of managers classified by investment style. For the entire sample (All Managers), both models show a positive mean selectivity value for all three alternative benchmarks. These values are significant at the 0.05 level for two of the three benchmarks. For timing skill, the results are just the opposite. For the entire sample, both models show a negative mean timing value for all three alternative benchmarks. However, for only one of the three benchmarks (the S & P 500), the mean timing value is significant at the 0.05 level for both models. Hence the results using the S & P 500 Index as a benchmark contrast with the results obtained using the Russell 3000 Index and the Style Index as benchmarks. The latter two indices are more representative of the managers' investment universes (i.e., true investment opportunities) than the former and, as such, are more appropriate benchmarks.

The results in Table I suggest that pension fund managers are on average better stock pickers than market timers. Our results relating to selection ability are consistent with those of Lee and Rahman (1990), who found some evidence of superior selection ability on the part of mutual fund managers. They also found evidence of superior market timing skill for several managers. However, it should be pointed out that Lee and Rahman (1990) ignored negative market timing skill in their model, while we allow negative market timing here. Our market timing results are consistent with those of previous studies on mutual fund performance (see Kon (1983), Chang and Lewellen (1984), Henriksson (1984), Lehmann and Modest (1988), Cumby and Glen (1990), and Connor and Korajczyk (1991)). These studies found more evidence of negative market timing than positive. These studies also found some evidence of negative selection ability for mutual funds.

B. Investment Style Results

There are differences in the portfolio characteristics and investment styles among the Earnings Growth, Market-Oriented, Price-Driven, and Small Capitalization managers. It is therefore useful to examine performance measures for each investment style separately. Table I presents mean values of the performance measures for each style of manager. It also provides the aggregated rank of each group. These ranks do not vary between the models for a given benchmark. However, they do vary somewhat across benchmarks for a given model.

The period 1983 to 1990 was a period in which the overall stock market was up substantially. For the eight years, the Russell 3000 grew at an annualized rate of 14.17 percent, and the S & P 500 grew at a 15.60 percent rate. For the majority of this period (up until the end of 1988) the "value" investment style was favored by the market relative to other investment

Table I

Mean Values of Performance Measures across Models and Benchmarks

This table presents the mean values of the selectivity and timing values across models and benchmarks for the entire sample period. The numbers in parentheses indicate the rank among investment styles for each measure.

	Russell 3000		Style Index		S & P 500	
	Selectivity	Timing	Selectivity	Timing	Selectivity	Timing
Bhattacharya and Pfleiderer Model						
All Managers	0.0008**	−0.0092	0.0016**	−0.0100	0.0003	−0.0470**
Earnings Growth	−0.0003(3)	0.0538**(1)	0.0007*(4)	0.0217(2)	−0.0008*(4)	0.0072(1)
Market Oriented	0.0022**(1)	0.0012(2)	0.0020**(2)	−0.0274(3)	0.0017**(1)	−0.0451 (2)
Price Driven	0.0015**(2)	−0.0234(3)	0.0009*(3)	−0.0843**(4)	0.0009*(2)	−0.0539**(3)
Small Capitalization	−0.0005 (4)	−0.0765**(4)	0.0031**(1)	0.0586*(1)	−0.0007(3)	−0.1024**(4)
Treynor and Mazuy Model						
All Managers	0.0008**	−0.0828	0.0016**	−0.0706	0.0004	−0.2799**
Earnings Growth	−0.0003(3)	0.2014**(1)	0.0008*(4)	−0.0011(2)	−0.0008*(4)	0.0108(1)
Market Oriented	0.0021**(1)	−0.0278(2)	0.0019**(2)	−0.1261(3)	0.0017**(1)	−0.1799(2)
Price Driven	0.0016**(2)	−0.0877 (3)	0.0009*(3)	−0.3286**(4)	0.0011**(2)	−0.2293*(3)
Small Capitalization	−0.0004(4)	−0.4625**(4)	0.0032**(1)	0.2074*(1)	−0.0004(3)	−0.7828**(4)

** Significant at the 0.05 level.
* Significant at the 0.10 level.

styles. Our benchmark for this style is the Price-Driven index which grew at an annualized rate of 15.53 percent. This compares to the "growth" investment style (represented by the Earnings Growth index) which grew at a 13.72 percent rate, and the Small Capitalization style (represented by the Russell 2000 index) which grew at a 7.38 percent rate. In Table I we see that, using the broad stock market indices as benchmarks, a negative mean selectivity value is consistently observed for the Earnings Growth and Small Capitalization managers. This is consistent with the preference of the stock market for the period. However, if we use the appropriate Style Index as a benchmark, we see that these managers (as well as all other styles) have *positive* selectivity values. Thus it does appear to matter which benchmark portfolio is used (and, perhaps, which time period) when we examine the results for the four investment styles.

C. *Importance of Timing*

Finally, one needs to be somewhat concerned about the size of the timing values. At a purely statistical level, one can assess the significance of the timing values by looking at the t-values. However, in the Treynor-Mazuy model the impact of timing on portfolio return is, in effect, measured by multiplying a rather small decimal fraction, γ, by a squared decimal fraction, $(R_{mt})^2$. Thus, at the level of actual portfolio returns, there is a relatively small reward/penalty to this activity in our data. Further research in the area of the measurement and assessment of market timing would help clarify this issue.

IV. Meta-Analysis of Results

Meta-analysis is a parametric statistical technique for the cumulation of results across studies or units of analysis. The contribution of meta-analysis is to offer a statistical technique to produce direct estimates of the mean and standard deviation of population values. Thus meta-analysis allows more statistically powerful inferences from data than are possible using more traditional disaggregated analyses. Recent comprehensive texts on meta-analysis include Hedges and Olkin (1985) and Hunter and Schmidt (1990).

There are a number of "study artifacts" which can cause the results of one study to appear different or even contradictory to those of another. Among the more prominent artifacts are sampling error, error of measurement, and restriction of range on the dependent variable. These artifacts are discussed in detail in Hunter and Schmidt (1990, chapters 2 and 3). In this paper, we focus on sampling error in the regression values for selectivity and market timing across managers. Meta-analysis has been primarily developed for correlational data. However, the time series regressions performed in our paper have identical specifications (by performance measurement model) across the sample of pension fund managers. Thus, for the purpose of meta-analysis, we can consider each of the 71 managers as a "study," cumulate the results, and apply meta-analysis. A more complete discussion of the meta-analysis procedures used here is given in Coggin and Hunter (1993).

As discussed in Coggin and Hunter (1993), the standard meta-analysis formulas must be adjusted for the effect of correlated regression residuals. The average correlations, r, between the time series residuals were used in the adjusted formulas to calculate the meta-analysis results given in Tables II and III.

Table II presents the results of the meta-analysis of the selectivity and timing coefficients based on three benchmark portfolios and using heteroscedasticity-adjusted t-values. The first row of this table gives the frequency-weighted mean of the observed values for each parameter, $\bar{b}$; the second row gives estimates of the standard deviation of the observed values, s_b; the third row gives estimates of the standard deviation of the population values, s_β; the fourth row gives estimates of the frequency-weighted average squared deviation of the observed values, s_b^2; the fifth row gives estimates of the variance of the population values, s_β^2 ; the sixth row gives estimates of the sampling error variance, s_e^2; the seventh row gives the chi-square value for the ratio of the observed variance to the sampling error variance; and the last

Table II

Meta-Analysis Results

This table presents the meta-analysis results for the selectivity and timing values based on the three benchmark portfolios and using heteroscedasticity-consistent standard errors, for the entire period (N = 71 managers).

	Selectivity			Timing		
	S & P 500	Russell 3000	Style Index	S & P 500	Russell 3000	Style Index
Panel A: Bhattacharya and Pfleiderer Model						
$\bar{b}$	0.000339	0.000769	0.001624	−0.046979	−0.009194	−0.010003
s_b	0.002646	0.002646	0.002450	0.105010	0.114070	0.123984
s_β	0.001827	0.001868	0.001807	0.054236	0.058340	0.075100
s_b^2	0.000007	0.000007	0.000006	0.011027	0.013012	0.015372
s_β^2	0.000003	0.000003	0.000003	0.002941	0.003404	0.005640
s_e^2	0.000005	0.000004	0.000003	0.010501	0.010489	0.010476
χ^2(df = 70)	130.66**	136.27**	145.78**	96.83**	96.15**	112.15**
$(1 - r)s_e^2/s_b^2$	0.5434	0.5210	0.4870	0.7333	0.7384	0.6331
Panel B: Treynor and Mazuy Model						
$\bar{b}$	0.000422	0.000796	0.001645	−0.279925	−0.082756	−0.070593
s_b	0.002646	0.002646	0.002450	0.635032	0.598273	0.593544
s_β	0.001922	0.001883	0.001927	0.588197	0.550055	0.552251
s_b^2	0.000007	0.000007	0.000006	0.403266	0.357931	0.352295
s_β^2	0.000004	0.000003	0.000004	0.345976	0.302560	0.304981
s_e^2	0.000004	0.000004	0.000003	0.074403	0.060449	0.050930
χ^2 (df = 70)	145.24**	140.50**	165.14**	499.77**	458.96**	528.66**
$(1 - r)s_e^2/s_b^2$	0.4889	0.5053	0.4299	0.1421	0.1547	0.1343

** Significant at the 0.05 level or less.

Table III

80 Percent Probability Intervals for Observed and Population Selectivity and Market Timing Values

This table presents the 80 percent probability intervals for the observed and population values of selectivity and market timing using all managers for the entire period. The observed values are bounded by $b \pm 1.28(s_b)$, and the population values are bounded by $b \pm 1.28(s_\beta)$.

	Observed Values				Population Values			
	Selectivity		Market Timing		Selectivity		Market Timing	
Benchmark	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Panel A: Bhattacharya and Pfleiderer Model								
S & P 500	−0.003121	0.003800	−0.181391	0.087434	−0.001999	0.002678	−0.116401	0.022443
Russell 3000	−0.002687	0.004255	−0.155203	0.136814	−0.001623	0.003161	−0.083870	0.065481
Style Index	−0.001606	0.004854	−0.168703	0.148697	−0.000689	0.003937	−0.106131	0.086126
Panel B: Treynor and Mazuy Model								
S & P 500	−0.003019	0.003864	−1.092766	0.532916	−0.002038	0.002883	−1.032817	0.472967
Russell 3000	−0.002632	0.004223	−0.848546	0.683034	−0.001615	0.003206	−0.786826	0.621314
Style Index	−0.001622	0.004912	−0.830330	0.689144	−0.000822	0.004112	−0.777474	0.636288

row gives estimates of the proportion of total observed variance accounted for by sampling error, $(1 - r)s_e^2/s_b^2$.

A. Selectivity

For selectivity, the mean monthly values in Table II are positive in every case but rather small. However, on an annualized basis, these numbers become more meaningful. For the Bhattacharya and Pfleiderer model, the annualized mean selectivity values are 0.41 percent (S & P 500), 0.93 percent (Russell 3000), and 1.97 percent (Style Index). For the Treynor and Mazuy model, the annualized mean selectivity values are 0.51 percent (S & P 500), 0.96 percent (Russell 3000), and 1.99 percent (Style Index). Hence we see that for both models, as suggested in Table I, managers do better on average relative to their *own* style index as compared to the broader market indices. This result is instructive, since much of the common investment wisdom implies that investment managers "can't beat the market." This result suggests that such a comment begs an important question regarding which benchmark should be used in evaluating a manager. We remind the reader that these returns do not include investment management fees.

The chi-square values are significant at the 0.05 level or less for the selectivity values using all three benchmarks for both models. This implies that there is real variation (in excess of that attributable to sampling error) around the mean selectivity value in each case.

B. Timing

For market timing, the mean values in Table II are negative in each case. This result is consistent with the results of Kon (1983), Chang and Lewellen (1984), Henriksson (1984), Grinblatt and Titman (1988), Lehmann and Modest (1988), Cumby and Glen (1990), Coggin and Hunter (1993), and Connor and Korajczyk (1991) who examined mutual fund returns. Furthermore, the chi-square values are significant at the 0.05 level or less in each case. Thus in every case there is evidence of real variation around the negative mean timing value.

If there were no real variation around the observed mean value, then the observed mean would be the true value for each of the 71 managers. However, in our case, there is evidence of real variation in every set of selectivity and market timing values. To put these results in perspective, we can look at the last row of Table II for each model and examine the proportion of total observed variance accounted for by sampling error. For the Bhattacharya and Pfleiderer model, the percentage of observed variance in selectivity accounted for by sampling error goes from 54 to 52 to 49 percent across benchmarks; while the percentage of variance in timing accounted for by sampling error goes from 73 to 74 to 63 percent across benchmarks. For the Treynor and Mazuy model, the percentages for selectivity go from 49 to 50 to 43 percent across benchmarks; while the timing percentages go from 14 to 15 to 13 percent across benchmarks. We should note that, as discussed in Hunter and

Schmidt (1990), these percentages of variance attributable to sampling error may well contain other unaccounted for study artifacts (such as measurement error).

C. The 80 Percent Probability Intervals for Selectivity and Timing

Assuming selectivity and market timing to be normally distributed, we can also examine the 80 percent probability intervals (i.e., the lower and upper 90 percent probability values) for the spread of the observed and population values presented in Table III. The probability intervals in Table III clearly show the amount of variation in both the observed and the population values for selectivity and market timing. As noted above, there is real variation in the selectivity and timing values in every case. The 80 percent probability intervals for selectivity are all shifted towards positive values, while the 80 percent probability intervals for timing are all shifted towards negative values.

Using the 80 percent probability intervals for the population selectivity values in Table III, we can examine the true spread in pension manager excess returns for the two models across benchmarks. The return for the top 10 percent of managers is obtained by annualizing the appropriate upper bound return in Table III, and the return for the bottom 10 percent of managers is obtained by annualizing the appropriate lower bound return in Table III. For the Bhattacharya and Pfleiderer model using the S & P 500 benchmark, the true annualized spread in returns is 5.63 percent (top 10 percent = 3.26 percent, bottom 10 percent = −2.37 percent); using the Russell 3000, the true spread is 5.79 percent (top 10 percent = 3.86 percent, bottom 10 percent = −1.93 percent); and using the Style Index, the true spread is 5.65 percent (top 10 percent = 4.83 percent, bottom 10 percent = −0.82 percent). For the Treynor and Mazuy model, the true annualized spread in returns using the S & P 500 benchmark is 5.93 percent (top 10 percent = 3.51 percent, bottom 10 percent = −2.41 percent); using the Russell 3000, the true spread is 5.84 percent (top 10 percent = 3.92 percent, bottom 10 percent = −1.92 percent); and using the Style Index, the true spread is 6.03 percent (top 10 percent = 5.05 percent, bottom 10 percent = −0.98 percent). Hence there is evidence in our data that the best equity pension fund managers delivered substantial risk-adjusted excess returns, no matter which model or benchmark we use. This complements the results of Grinblatt and Titman (1989a), Ippolito (1989), Lee and Rahman (1990), and Coggin and Hunter (1993) who found some evidence of superior performance in their studies of mutual funds.

V. Discussion

A. Sensitivity of Results to Benchmarks and Models

Our general finding is that selectivity is positive and timing is negative on average across all models and benchmarks. However, we did observe some sensitivity of results to the choice of a benchmark when we divided up the managers by investment style in Table I.

Our basic results differ from those of Lehmann and Modest (1987) and Grinblatt and Titman (1989a), who found that performance results varied across both models and benchmark portfolios. It should be pointed out that there is a problem in the Lehmann and Modest (1987) analysis. They examined selectivity in the context of a Jensen-like measure using the CAPM and APT models. Market timing and factor timing activities were not included in their analysis. Market timing was also ignored by Grinblatt and Titman (1989a). Grant (1977) explained how market timing actions will affect the results of empirical tests that focus only on selection ability. He showed that market timing skill will cause the observed regression estimate of selectivity to be downwardly biased. The results of Lee and Rahman (1990) are consistent with Grant's (1977) contention. A similar conclusion was drawn by Chang and Lewellen (1984) and Henriksson (1984). Moreover, as Jensen (1972), Admati and Ross (1985), Dybvig and Ross (1985), and Grinblatt and Titman (1989b) have shown, the Jensen-like measure may penalize the performance of market timers.

B. Negative Correlation between Selectivity and Timing

Table IV presents the correlations between selectivity and timing. In Table IV we observe a strong negative correlation between selectivity and market timing in our data. Furthermore, this is consistent with the results of several other studies. The literature on investment management contains a number of studies documenting the negative market timing skill of mutual fund managers (see Chua and Woodward (1986) for a summary and extension of

Table IV

Correlation between Selectivity and Timing

Each model was estimated for all managers for the entire period using each of the three benchmark portfolios. This table presents the Pearson and Spearman correlations between the selectivity and timing values for each model for each benchmark.

	Bhattacharya and Pfleiderer Model		Treynor and Mazuy Model	
Benchmark	Pearson	Spearman	Pearson	Spearman
Russell 3000	−0.447	−0.488	−0.485	−0.427[a]
Style Index	−0.359[d]	−0.315[c]	−0.399[b]	−0.359[d]
S & P 500	−0.487	−0.504	−0.467	−0.387[c]

[a] Significant at the 0.0002 level.
[b] Significant at the 0.0006 level.
[c] Significant at the 0.0008 level.
[d] Significant at the 0.0021 level.
[e] Significant at the 0.0075 level.
All other correlations are significant at the 0.0001 level.

these studies). Ours is the first study we know of which documents this finding for pension fund managers.

The negative correlation between selectivity and timing presents a problem of interpretation. Hunter, Coggin, and Rahman (1992) show that, for the regression models used in this study, the correlation between the estimates of selectivity and timing will *necessarily* be negative. They show that this is because the sampling errors for the two estimates are *negatively* correlated. The magnitude of the negative correlation between the two estimates is the same as the magnitude of the negative correlation between the two sampling errors. Thus, once we account for the effect of negatively correlated sampling errors, selectivity and timing are largely *uncorrelated* in our data.

The correlation between selectivity and market timing is currently an unsettled question in the literature. Coggin and Hunter (1993) calculated a corrected correlation of -0.62 in the Lee and Rahman (1990) data, but noted that this was also an artifact of correlated sampling errors for the two estimates. They also calculated an observed correlation of 0.04 ($N = 37$ mutual funds) between "Overall Selectivity" and "Overall Timing" in Kon (1983, Table 5), who used a different model of market timing. Using the Henrikkson and Merton (1981) model, Henrikkson (1984) and Connor and Korajczyk (1991) report a negative correlation. Lehmann and Modest (1987, footnote 33) report basically no "substantive correlation" between the two. Jagannathan and Korajczyk (1986) have presented an argument that the observed negative correlation between selectivity and timing could also be the result of some other phenomenon, such as changes in firms' debt/equity ratios in relationship to that of the benchmark portfolio. Finally, Grinblatt and Titman (1989b) have shown that many of the desirable properties of a performance measurement model which seeks to estimate both selectivity and market timing skill are not present if selectivity and timing are correlated.

C. Survivorship Bias

The issue of survivorship bias is well known in studies of investment performance. A recent study by Brown, Goetzmann, Ibbotson, and Ross (1992) highlights this issue with regard to performance measurement. We do not know the true extent of this bias in our results, but the results in Grinblatt and Titman (1989a) suggest that it may not be large.

VI. Summary and Conclusion

Our major findings are as follows. Regardless of the choice of benchmark portfolio or estimation model, the selectivity measure is positive on average and the timing measure is negative on average. However, both selectivity and timing do appear to be somewhat sensitive to the choice of a benchmark (and, possibly, the time period) when managers are classified by investment style. A meta-analysis of the regression results was performed to quantify the effect of sampling error. In every case, meta-analysis revealed some real variation (in excess of that attributable to sampling error) around the mean values for each measure. An examination of the 80 percent probability intervals for selectivity revealed that the best equity pension fund managers delivered substantial risk-adjusted excess returns, while the worst did not. Consistent with previous studies of equity mutual fund performance, we also found a negative correlation between selectivity and timing. However, we argue that the observed negative correlation in our data is largely an artifact of negatively correlated sampling errors for the two estimates.

Much work remains to be done in this area. While active equity managers are currently losing ground to passively managed index funds, actively managed equities still represent the largest fraction of the equity component of corporate pension funds. We still do not know why some active managers are able to provide substantial risk-adjusted performance, while many cannot. Identifying the characteristics of successful money mangers should be one focus of future research. Furthermore, while there are some interesting statistical explanations, we still do not have a satisfactory substantive model of the relationship between the security selection and market timing skill of active equity managers. This is another fertile area for study.

REFERENCES

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

Berkowitz, S. A., L. D. Finney, and D. E. Logue, 1988, *The Investment Performance of Corporate Pension Plans* (Quorum Books, New York).

Bhattacharya, S., and P. Pfleiderer, 1983, A Note on performance evaluation, Technical Report 714, Stanford University.

Breen, W., R. Jagannathan, and A. R. Ofer, 1986, Correcting for heteroscedasticity in tests for market timing ability, *Journal of Business* 59, 585–598.

Brinson, G. P., L. R. Hood, and G. L. Beebower, 1986, Determinants of portfolio performance, *Financial Analysts Journal* 42, 39–44.

Brown, S. J., W. N. Goetzmann, R. G. Ibbotson, and S. A. Ross, 1992, Survivorship bias in performance studies, *Review of Financial Studies* 5, 553–580.

Chang, E. C., and W. G. Lewellen, 1984, Market timing and mutual fund investment performance, *Journal of Business* 57, 57–72.

Chua, J. H., and R. S. Woodward, 1986, Gains from market timing, Monograph No. 1986–2, New York University.

Coggin, T. D., and J. E. Hunter, 1993, A meta-analysis of mutual fund performance, *Review of Quantitative Finance and Accounting* 3, 189–201.

Connor, G., and R. Korajczyk, 1991, The attributes, behavior, and performance of U.S. mutual funds, *Review of Quantitative Finance and Accounting* 1, 5–26.

Cumby, R. E., and J. D. Glen, 1990, Evaluating the performance of international mutual funds, *Journal of Finance* 45, 497–521.

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

Fama, E. F., 1972, Components of investment performance, *Journal of Finance* 27, 551–567.

Grant, D., 1977, Portfolio performance and the cost of timing decisions, *Journal of Finance* 32, 837–846.

Grinblatt, M., and S. Titman, 1988, The evaluation of mutual fund performance: An analysis of monthly returns, Working paper, University of California, Los Angeles.

———, 1989a, Mutual fund performance: An analysis of quarterly portfolio holdings, *Journal of Business* 62, 393–416.

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

Hansen, L., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–1054.

Hedges, L. V., and I. Olkin, 1985, *Statistical Methods for Meta-Analysis* (Academic Press, Orlando, Fla.)

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

———and R. C. Merton, 1981, On the market timing and investment performance II: Statistical procedures for evaluating forecasting skills, *Journal of Business* 54, 513–533.

Hsieh, D., 1983, A heteroscedasticity-consistent covariance matrix estimator for time series regressions, *Journal of Econometrics* 22, 281–290.

Hunter, J. E., T. D. Coggin, and S. Rahman, 1992, The correlation between sampling errors in the estimation of security selection and market timing skill, Working paper, Michigan State University, East Lansing.

Hunter, J. E., and F. L. Schmidt, 1990, *Methods of Meta-Analysis* (Sage Publications, Newbury Park, Calif.).

Ippolito, R. A., 1989, Efficiency with costly information: A study of mutual fund performance, 1965–84, *Quarterly Journal of Economics* 104, 1–23.

———, and J. A. Turner, 1987, Turnover, fees and pension plan performance, *Financial Analysts Journal* 43, 16–26.

Jagannathan, R., and R. A. Korajczyk, 1986, Assessing the market timing performance of managed portfolios, *Journal of Business* 59, 217–235.

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

———, 1969, Risk, the pricing of capital assets, and the evaluation of investment portfolios, *Journal of Business* 42, 167–247.

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

Kon, S. J., 1983, The market-timing performance of mutual fund managers, *Journal of Business* 56, 323–347.

———, and F. C. Jen, 1978, Estimation of time-varying systematic risk and performance for mutual fund portfolios: An application of switching regression, *Journal of Finance* 33, 457–475.

———, 1979, The investment performance of mutual funds: An empirical investigation of timing, selectivity, and market efficiency, *Journal of Business* 52, 263–289.

Lakonishok, J., A. Shleifer, and R. W. Vishny, 1992, The structure and performance of the money management industry, in *Brookings Papers on Economic Activity: Macroeconomics*, Brookings Institution, Washington, D.C., pp. 339–391.

Lee, C. F., and S. Rahman, 1990, Market timing, selectivity, and mutual fund performance: An empirical investigation, *Journal of Business* 63, 261–278.

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

———, 1988, Empirical issues in the measurement of market timing ability, Working paper, University of California, Berkeley.

Tierney, D. E., and K. Winston, 1991, Using generic benchmarks to present manager styles, *Journal of Portfolio Management* 17, 33–36.

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

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