Review of Quantitative Finance and Accounting, 15 (2000): 81–97

Equity Manager Selection and Performance

BRUCE COLLINS
President, FinCast LLC.
Professor of Finance, Western Connecticut State University.

FRANK FABOZZI
Editor, Journal of Portfolio Management.
Adjunct Professor of Finance, Yale University.

Abstract. This paper introduces a methodology to select money managers and examines their performance over a 12 year period. We assess performance empirically by utilizing a series of well-known statistical procedures applied to other data. The value of this study is in the uniqueness of the data and the fact that it is the first study to use the performance of a set of pension fund money managers selected on the basis of a specific selection criteria.

Key words: alpha, market timing, stock selectivity, Jensen's selectivity index, Treynor and Mazuy model, Bhattacharya and Pfleiderer model, Henriksson and Merton model, Hamilton equity yardstick, meta-analysis

1. Introduction

There have been numerous studies assessing and evaluating the performance of mutual fund managers. The general conclusion of most of these studies over the last 30 years is that managers exhibit some skill at security selection, but have little or no market timing ability.[1] However, most of the positive performance measures have been small or insignificant. Most studies have examined the performance of mutual funds because of the availability of data. There is, however, a shortage of studies that address the investment performance of pensions fund managers. It is important to pursue this line of investigation because the size of the pension assets has grown to surpass $5 trillion in the 1990s with the proportion allocated to equities estimated at about $2.6 trillion. This is about three times the value of the dollar allocation to U.S. equity investments among mutual funds. Furthermore, it is estimated that 20–25% of U.S. equity investments is indexed. The amount of indexing is particularly pronounced for the 200 largest U.S. pension plans that may have as much as 50% of equity allocation indexed. The establishment and expansion of index funds is the result of the belief that markets are efficient in the long term and any short-term inefficiencies do not deliver sufficient risk-adjusted returns to satisfy investors. This belief has been supported by a series of academic studies that to date have failed to identify consistent superior performance by active managers. Nonetheless, despite these setbacks and trends, the largest component of equity investments among pension plans

remains actively managed. Moreover, although the average returns of active managers suggest a mediocre performance record, there are some active managers that do produce superior returns.

Statements about the performance of professional money managers are incomplete without a close examination of the performance of pension fund managers. Because of the lack of data, only a few studies have examined equity pension fund managers. Brinson et al. (1986) studied 91 managers from the SEI universe of managers and found that on average they underperformed their respective benchmarks by 1.1% per annum. However, the authors did acknowledge that some individual managers outperformed. Ippolito and Turner (1987) found that superior performance in active portfolios is not high enough to offset research costs. Lakonishik et al. (1992) examined the annual returns of a sample of equity pension funds for the period 1983 to 1989 and found that on average pension fund managers seem to reduce rather than add value compared with the performance of the S&P 500 Index. However, this study made no adjustment for risk, used only the S&P benchmark, and did not distinguish between market timing and selectivity skill. Coggin et al. (CFR) (1993) studied a random sample of pension fund managers from the Frank Russell database of pension fund managers and found that, on average, the study group exhibited positive selectivity skill and negative market timing skill. Their conclusions indicated that performance attribution is sensitive to the choice of a benchmark and that managers perform better against their respective benchmarks than versus the market as a whole. Furthermore, the authors find significant variation in performance with some managers delivering substantial risk-adjusted returns. This is the first study to specifically examine the two components of investment returns for a sample of U.S. equity pension fund managers.

The purpose of this paper is twofold: first, to discuss a methodology to select money managers and second, to present some evidence with regard to the performance of the selected group of managers that are the subject of this study. The statistical results are not intended to break new ground in performance analysis, but only to compare performance versus several benchmarks. The focus of the paper is to present a set of managers chosen by a specific process and their returns over a 12-year period. The managers were all chosen as part of selection criteria designed to identify superior investment managers. The importance of this paper is that it addresses a fundamental concern of pension funds to select good managers and it evaluates the performance of such a group. Our study is the first applied to a group of pension fund managers selected on the basis of a set of selection criteria.

The remainder of the paper is organized in the following way. The next section outlines a methodology for manager selection followed by a leading investment advisory firm. The section that follows reviews the chosen models for measuring investment performance and separating the effects of security selection from market timing. Next, we describe the data and methodologies used in our analysis. We then present the results of our empirical studies and provide some analysis of the results. The next section presents a meta-analysis of our results with a discussion of our findings followed by a brief discussion of the problem of survivorship bias. The final section summarizes the paper and provides some concluding remarks.

2. Selection methodology

The selection methodology outlined in this section is based on conversations with Hamilton & Company and derived from documents provided by the firm.[2] A complete methodology to manager selection focuses on identifying superior investment management organizations by analyzing the organization's structure, knowing the personalities behind the organization, and understanding the backgrounds and philosophies of the managers. Furthermore, it is important that the manager's approach to investment management is practical and disciplined, yet flexible and adaptive to new opportunities and changing market conditions. The selected managers used in our study are involved in managing the core equity portfolios of prominent institutional investors. Consequently, the portfolios are geared toward large capitalization holdings and are diversified across style. The selected managers are ''total results'' oriented with proven track records of success. The selection process does not involve any bias with regard to specific style and has no methodological constraints. The selected managers share a common entrepreneurial attitude tempered by the prudence, wisdom and education of seasoned professionals. The selection is based on a set of descriptive characteristics in three areas (1) the organization, (2) the managers, and (3) the approach to equity investing. These characteristics are accompanied by performance results, which require superior investment performance versus an appropriate benchmark. An example of the selection criteria might include managers that are consistent with above average returns, intelligent, well educated with experienced personnel, a stable organization with low personnel turnover, a controlled growth in firm size, and an entrepreneurial orientation.

The selection of a specific manager depends upon the needs of the sponsor. This may include for some public funds legal constraints and the ''Prudent Man Rule'' guidelines provided by ERISA regulations. The fund may have specific requirements associated with investment style or target specific markets such as international or small capitalization stocks. The final decision will be part quantitative and part qualitative. The process of selection will involve several steps listed below and is the same whether a style is a consideration or not.

(1) Preparation of investor objectives and guidelines.
(2) Preliminary research and screening to produce a list of candidates.
(3) Detailed analysis of prospective candidates to narrow list of candidates.
(4) Final selection.
(5) Implementation.

The preparation stage of the selection process involves establishing the specifics of the sponsor's objectives and a general set of criteria for manager selection. Examples of a sponsor's objective might include performance objectives such as outperforming an S&P 500 benchmark on a risk-adjusted basis. Performance combined with a set of risk parameters. The objectives provide the critical mass from which a preliminary list of managers is selected. This list is derived from the criteria related to three general areas:

(1) Type of investment management firm.
(2) Nature of people who make up the firm.
(3) Investment approach.
(4) Historical investment performance.
(5) Prospects for futures investment performance.

The second step involves identifying a universe of managers and creating screening criteria. The central issue in this step is to find managers that fit the specification of the plan sponsor. The universe will be narrowed by such criteria as style, performance, philosophy, and organizational structure. The result is a set of candidates to subject to further review and analysis. This may involve contacting the candidates to obtain additional data and other information about the firm including the possibility of an on sight interview.

At this phase of the evaluation and selection process, the prospect for future performance becomes an essential focus. How has the manager performed during falling markets versus rising markets? The evaluation process allows the investigator to develop an understanding of the manager's investment philosophy and how it adapts to change over time. The analysis at this stage requires the use of techniques to help understand how the manager was able to produce a performance record. The thrust of this investigation is to assess the probability of superior performance in the future. Essentially, the manager selection process is a forecasting methodology that is complex in nature. It is not a simple matter to examine returns over the last n years and conclude from them the likely outcome in the next year or the next n years. The circumstances surrounding performance are a vital part of the evaluation. It is therefore important to assess performance in terms of a benchmark and over market cycles and different risk environments. Also, during this stage of the evaluation distinguishing features are explored that might separate one manager from another. These might include reputation, client relations, fee structure, accessibility, and turnover.

After consulting again with the client, the advisor selects a set of candidates for final evaluation. The finalist would all be excellent choices just as all actors nominated for an academy award might all be deserving of the award. Nonetheless, it is necessary to rank the managers according to highest recommendation.

3. Measures of investment performance

We measure performance by first looking at Jensen's selectivity index, which is a more traditional measures of relative performance or ''alpha'' and does not take market timing in account. Second, we use performance models that consider both selectivity and market timing.

The Jensen selectivity index is derived from the risk premium version of the capital asset pricing model (CAPM) and based on the assumptions of the model formulates a return-generating model Jensen developed a regression equation shown below as equation (1) that he used to test for superior performance.[3]

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

The CAPM predicts that the intercept term should be zero in the absence of superior returns. Therefore, significant alphas can be viewed as an indication of superior performance. Alpha is the measure of security selection ability, beta measures the sensitivity of returns to market returns, and the error term is a random variable with an expected value of zero.

The weakness of the first model of performance measurement is that it fails to consider stock selectivity and market timing. Because the model assumes that the risk level of the portfolio is stationary over time it does not consider the market timing aspect of investment strategies. The fact is that managers may alter the risk characteristics of their portfolio over time in anticipation of general price movements. Thus, if managers actively practice market timing, the model considered above may produce misleading results.[4]

Several empirical methods have been developed to determine whether managers are practicing market timing. We use three models that take both stock selection and market timing into account. All add a second term to the Jensen relationship to take market timing into account. In general, for all three models, alpha measures stock selection and the second coefficient measures market timing.[5] Treynor and Mazuy (1966) were the first to empirically consider market timing performance. They argued that the simple regression model given by Jensen is not specified correctly if managers engaged in market timing. The proportion of the market portfolio held will change according to the manager's market forecasts. Consequently, the manager's portfolio returns would be a nonlinear function of market returns. They attempt to capture this effect in equation (2) by adding a quadratic independent variable to the relationship above. A positive value for the coefficient of the quadratic term would indicate positive market timing skill.

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

Bhattacharya and Pfleiderer (1983) also extended the work of Jensen (1972).[6] The relationship they derive is a refinement of the Treynor and Mazuy model and presented as equation (3).

$$R_{pt} = \eta_0 + \eta_1 R_{mt} + \eta_2 (R_{mt})^2 + \omega_t. \tag{3}$$

The coefficient of squared excess market return is also an indication of market timing skill. It is the first model that analyzes the error term to identify a manager's forecasting ability. The econometric issues are addressed in the next section.

Equation (4) is a model developed by Henriksson and Merton (1981) that also allows fund managers to assess both the selectivity and market timing activities of investors.

$$R_{pt} = \alpha_p + \beta_{1p} R_{mt} + \beta_{2p} D_t (R_{mt})^2 + \varepsilon_{pt}. \tag{4}$$

Their approach assumes managers forecast the market and have two target systematic risk levels. The manager adjusts the beta in response to a forecast choosing a low beta when they believe default-free bonds will outperform stocks and choosing a high beta when stocks are expected to outperform default-free bonds. They add a dummy variable to

the second term, which equals 0 if the market is up and equals -1 otherwise. The alpha can be used to measure stock selection and the second beta can be used to measure market timing.

We adopt the approaches mentioned above to take into account more than one method of evaluating performance consistent with CFR (1993) and Fletcher (1995). Although other approaches have been taken to separate stock selectivity from market timing as a source of performance, we believe that the three alternatives we have chosen represent our best chance at accomplishing that objective.

4. Data and methodology

The data for our study consist of quarterly total returns for the period beginning with the first quarter of 1980 and ending with the second quarter of 1994 for a set of 37 managers representing Hamilton & Company's client selection list for the period. We only use managers who have been part of the client selection list for at least five years. During this period there has been less than a handful of managers that have been removed from the list due to special circumstances.[7] In addition, of the 37 managers 26 were on the client selection list for the whole time period. The other 11 joined the list at a later date, but have a minimum of five years of history. The managers represent a broad range of asset management organizations with the characteristics described above and have been employed as domestic core equity managers by fund sponsors. They all have the discretion to engage in stock selection, market timing or both. Although the managers may have specific style orientations, due to their overall investment approach and Hamilton & Company's selection criteria, they are not limited to a particular style and the style may change over time. However, the portfolio holdings are part of the core equity holdings of pension funds and are predominantly large capitalization, which will allow us to use large capitalization benchmarks. The identities of the individual managers and fund sponsors are not included.

Hamilton & Company also provided data for the risk-free security proxy and four alternative benchmarks. In all cases, the data consist of quarterly total returns. The risk-free proxy is the 91-day Treasury bill rate. The four benchmarks include the Hamilton Equity Yardstick^{TM}, the S&P 500 capitalization-weighted total returns index, the S&P 500 equally-weighted total returns index, and the Frank Russell 3,000 capitalization-weighted total returns index. The use of alternative benchmarks allows us to examine the sensitivity of performance to the choice of a measurement standard. We include an equal-weighted benchmark to demonstrate the small-capitalization effect. An estimate of the variance of excess return on the market was estimated from observed returns for each benchmark following a procedure from Merton (1980) and offered in Lee and Rahman (1990).

The Hamilton & Company Equity Yardstick is a comparative database representative of the performance that is achieved by other managers with similar investment objectives. The Yardstick allows the researcher to compare the performance of individual equity managers to their peers in the industry. It is composed of banks, insurance companies, investment companies and investment counselors. Hamilton & Company collect the returns on a monthly basis directly from the investment managers. The Yardstick funds are

large and diversified, representing real portfolios. They do not have unusually restrictive investment provisions or unmarketable securities. The fund consists of pooled or commingled portfolios typically representing employee benefit money. The fund managers have full discretion over the securities. The Yardstick assets total over $300 billion from a nationwide selection of portfolios, which represent institutions managing a far greater amount. The funds range in size from $50 million to $34 billion. The Yardstick is constructed of real portfolios and contains an element of prudent investment not found in market indexes. The Yardstick is not affected by survivorship bias, which is a byproduct of most other compilations of funds such as client-based universes.[8] Moreover, the Yardstick removes factor bias by randomly selecting a set of portfolios from the universe of core mangers. The Yardstick equally weights the set of portfolios.

In our study, we examine the null hypothesis that the data reveal no significant total returns versus a randomly selected group represented by various benchmarks. The first statistical data include general statistics such as total return performance comparisons measured as geometric and arithmetic averages. These data summarize performance and volatility characteristics over the sample period. We then conduct a set of empirical tests estimating regression models based on the CAPM beginning with Jensen's selectivity index. This is followed by empirical tests for the three models that test simultaneously for selectivity and market timing. These models require an adjustment for heteroscedasticity. The error terms in the three will exhibit conditional heteroscedasticity due to the manager's attempt to time the market.[9] We use two alternative methods for adjusting for heteroscedasticity. To accomplish this we use heteroscedasticity-consistent standard errors proposed by White (1980), Hansen (1982), and Hsieh (1983).[10] The *t*-statistics for the significance tests are adjusted for heteroscedasticity. In the Bhattacharya and Pfleiderer model, we use a generalized least square method described in Lee and Rahman and employed in CFR (1993).

CFR (1993) responding to an observation of Coggin and Hunter (1993) make adjustments to the Treynor and Mazuy model and the Bhattacharya and Pfleiderer model to allow for negative market timing skill. In the Treynor and Mazuy model this means that managers hold less of the market portfolio than implied when the market return is high and more than implied when the market return is low. This translates into a negative value for γ in equation (2). In the Bhattacharya and Pfleiderer model, negative market timing is the result of a negative correlation between the manager's beta and the market return. This could be a consequence of the manager's inability to forecast the market. We examine the value of η in equation (3) for an indication of a negative correlation between the manager's forecast and the expected return of the market. In the Henriksson and Merton model, we examine the sign of β_{2p} in equation (4) as an indication of market timing skill.

5. Empirical results

5.1. General statistics and performance results

In this section, we examine the performance of the Hamilton manager selection set versus the four benchmarks. In Table 1, we report average performance measures in terms of the

Table 1. General statistics. Managers vs benchmarks: 1980–1994

Statistic	Managers	S&P 500	Yardstick	R3,000	S&P Equal
Average	19.23	16.87	15.43	16.65	19.28
Geometric	17.82	15.55	14.15	15.12	17.4
Median	19.89	21.78	18.61	17.62	19.16
SD	15.92	15.66	15.29	16.65	18.31
Min	− 20.33	− 22.53	− 20.95	− 22.99	− 23.9
Max	20.45	21.36	19.62	20.58	20.58

annualized arithmetic average, the geometric average, and the median. The results indicate that the managers fared very well versus the three capitalization-weighted benchmarks and did almost as well as the equal-weighted benchmark.[11] On a risk-adjusted basis using standard deviation to normalize returns, the managers come out ahead in all cases. The results suggest that the managers have some superior ability versus the benchmarks on a risk-adjusted basis.

5.2. Measures of selectivity and market timing skills

In this section we introduce the results of estimating the models of selectivity and market timing skills. We begin by reporting the results of the Jensen model for stock selectivity or alpha and then proceed to report the results of the models that include both selectivity and market timing. The conventional approach to assessing manager performance coming out of early studies on mutual fund performance is based on the CAPM. Jensen (1968) argued that security selection performance could be captured by the portfolio excess return compared to excess return of a randomly selected portfolio with the same systematic risk. The simple regression of excess portfolio returns and excess market returns provides a framework for this analysis where the alpha is a measure of selectivity performance. The model assumes that the portfolio manager maintains a constant level of systematic risk over the sample period and focuses on stock selection to achieve superior returns. Table 2 provides a summary of each regression model estimated. The dependent variable is the total returns from an equally-weighted portfolio of all managers in the selection list. The results of the Jensen model are included in Table 2 and the results for the individual managers are summarized on a directional and significance count basis in Table 3. From Table 2, we can clearly establish that the group exhibits superior selection skill provided the model's assumptions are realistic. The first column of Table 2 under the benchmark and selectivity headings is the alpha estimate and the second column is the corresponding *t*-statistic. The managers as a group exhibit superior selection skill versus all the benchmarks at the 10% level of significance and three of four at the 5% level. Table 3 confirms the positive overall manager performance by indicating that more managers have positive selection skills than suggested by a random sample as indicated by the *z*-scores in row 2. In addition, more have significant superior returns as indicated by the *z*-score in row 4. None have a negative alpha. These results seem to support our earlier findings of

Table 2. Investment manager performance: Regression Results. Jensen Selectivity Index, Henriksson and Merton Model, Treynor and Mazuy Model, Bhattacharya and Phleiderer Model

	S&P 500				Hamilton Yardstick			
	Selectivity		Timing		Selectivity		Timing	
Jensen	0.529	2.677	NA	NA	0.797	6.661	NA	NA
H&M	0.2726	0.8583	0.0800	1.0985	0.7285	4.0313	0.0216	0.5466
T&M	0.0045	1.8701	0.1110	0.6671	0.0079	5.4331	0.0074	0.0891
B&P	0.0045	1.8813	0.1013	0.7479	0.0079	5.3496	0.0106	0.0782
	Russell 3,000				S&P 500 Equal			
	Selectivity		Timing		Selectivity		Timing	
Jensen	0.673	5.170	NA	NA	0.359	1.536	NA	NA
H&M	0.4821	2.4418	0.0559	1.3588	0.3038	0.7957	0.0153	0.2335
T&M	0.0060	3.8925	0.0983	1.3810	0.0036	1.2674	0.0015	0.0130
B&P	0.0060	3.7083	0.1514	1.1259	0.0036	1.2499	0.0020	0.0146

Table 3. Investment manager performance: Jensen index. Directional and significance count

	S&P 500	Hamilton Yardstick	Russell 3,000	S&P Equal
$N>0$	33	37	36	31
Z-stat	4.768	6.083	5.754	4.110
$t>1.64$	12	25	22	6
Z-stat	7.656	17.462	15.199	3.130
$N<0$	0	0	0	0
Z-stat	− 1.395	− 1.395	− 1.395	− 1..395

positive and significant performance of the managers as group versus these particular benchmarks. However, we know that for many managers the beta of their portfolio often changes over time. Consequently, this is equivalent to engaging in some form of market timing either deliberately or inadvertently. As long as the beta changes over time, the Jensen model is inappropriate.

We have employed three alternative models that take into account both stock selection and market timing. These models are given by equations (2)–(4) given in Section 3. Table 2 reports the results from estimating the Henriksson and Merton model given in equation (4) for the managers as a group. The results indicate that performance significance for the managers as a group depends on the benchmark. The *t*-statistics for selectivity are significant versus the Hamilton Yardstick, indicating superior selection skill versus a randomly selected group of managers. The *t*-statistics for selectivity is also significant versus the broad-based Russell 3,000 but not against either S&P 500 index. The results for market timing indicate that only one *t*-statistic is significant at the 10% level and that is versus the Russell 3,000 index.

We also summarized the results for the individual managers on a directional and significance count basis.[12] These statistics indicate more positive superior selection than expected versus three of the four benchmarks. Furthermore, the results show that a significant number of managers have a positive and significant selection skill, but the number of managers with positive market timing results is no better than a random sample. However, more managers than expected have significant and positive market timing skills, than have negative market timing skills. These results in general seem to support previous results that managers have better stock selection skills than market timing skills. Our results are consistent with CFR (1993) and Lee and Rahman (1990), who both found some evidence of superior stock selection skill and some market timing skill. The difference here is we find superior stock selection skill not only against a broad-based index but also against a randomly selected group of managers. Our market timing results differ from previous studies in that we don't find more negative market timing than positive, but just the opposite.[13] Fletcher (1995) found negative market timing performance and positive stock selection performance.

The next set of results includes estimates for both stock selection and markets timing performance and is based on equation (2) attributed to Treynor and Mazuy (1966). Table 2 presents the results for the managers as group and indicates positive and significant stock selection performance versus three of the four benchmarks at the 5% level of significance. Under market timing the second column indicates little evidence of market timing performance. Again the results are consistent with previous studies. We also summarized the results of a directional and significance counts. Our results indicate a greater number of positive and significant stock selection performances than expected and a greater number of positive and significant market timing performance. The stock selection performance is consistent across all benchmarks with the exception of the S&P 500 equal-weighted index. There is evidence of more positive market timing performance than negative. Once again, the results in this case reinforce the results of the previous model.

In the next study, we look at estimates for both stock selection performance and market-timing performance based on Bhattacharya and Pfleiderer (1983) as expressed in equation (3). Table 2 presents the results of estimating equation (3) for the managers as a group. The results are consistent with the Treynor and Mazuy model. Stock selection performance is positively significant against three of the four benchmarks and little market timing performance is indicated for the managers as a group. These results are consistent in every way with the Treynor and Mazuy model indicating positive and significant stock selection performance, some market timing performance and more positive market timing performance than negative market timing performance.

The conclusion we reach with regard to the empirical results suggest that the Hamilton manager selection list offers superior risk-adjusted returns versus a randomly selected group of managers or the market portfolio. The individual managers are better at stock selection than market timing although the ones engaged in market timing indicate more positive market timing performance than negative. One extension of this study would involve identifying more precisely how style may change during various market scenarios and examining performance versus a broader range of benchmarks. Although we recognize the importance of benchmark selection, our purpose is only to be regarded a first look at a specialized data set. We have attempted to address the issue by including several

benchmarks. We acknowledge that no individual benchmark is perfect, but we believe it is important to conduct the analysis against as many benchmarks as possible. CFR (1993) found that the results of their analysis were consistent across models and benchmarks. Their results differ from Lemann and Modest (1987) and Grinbatt and Titman (1989) who found the opposite. However, the latter two studies did not use a model to take into account market timing while CFR did. We also include a benchmark consisting of actual managed portfolios and compiled by Hamilton & Company over time. Although our study is confined to single index benchmarks, the Hamilton Yardstick is essentially a multifactor benchmark consisting of a number of portfolios with diversified styles. The one common denominator is the equity portfolios represent core holdings of pension funds and are predominately large cap in nature. We conclude, therefore, that our choice of the S&P 500 and the Russell 3,000 is appropriate. We include an equal-weighted S&P index to see if it yields different results and the Hamilton Yardstick because it accounts for diversified styles.

6. Results from meta-analysis

In the 1990s, Fletcher (1995), Coggin and Hunter (1993), and CFR (1993) have applied meta-analysis to empirical studies of portfolio manager performance. Meta-analysis is a method of statistically integrating the results of a set of independent studies. It produces a single set of numbers that describe and summarize the results of the studies. One of the features of meta-analysis is that it deals with summary statistics and therefore does not require the use of the original data. Coggin and Hunter have applied meta-analysis to the regression estimates of stock selection and market timing performance. Any difference in performance across studies can either be due to real differences or to sampling error variation. Coggin and Hunter have shown that for the purpose of meta-analysis each of the manager's performance can be treated as an independent study and the results can be cumulated across managers. In addition, meta-analysis has advantages over alternative multivariate techniques because the formulas used in meta-analysis depend only on the distribution of the summary statistics and not on the original data. Meta-analysis applied to performance studies can test whether variations across stock selection or market timing performance are due to real differences or statistical and methodological artifactual differences resulting from sampling error. It addresses the question of whether observed differences represent differences in the underlying population or are the result of sampling error. We can also estimate the proportion of observed variation that is attributable to sampling error variation.

We apply meta-analysis to the excess returns study, the Jensen model, and the three models that separate stock selection and market timing.[14] The information necessary to perform the meta-analysis includes the regression coefficients, the standard errors, and the number of observations associated with each manager's performance. Since not all managers were part of the selection list for the whole time period, we limit our analysis to the 26 managers who have complete data sets. The formulas used in the analysis are from Coggin and Hunter and include adjustments to account for residual correlation between the managers' performances. Tables 4 and 5 present the results of the meta-analysis for the studies mentioned above. The tables include the frequency-weighted mean, a chi-square

statistic for the ratio of observed variance to the sampling error variance, and the estimate of the proportion of total observed variance accounted for by sampling error, $(1 - r)S_e^2/S_b^2$ The null hypothesis is that there is no real variation in performance. The chi-square critical value has N-1 degrees of freedom where N is the number of managers in the analysis. The critical chi-square is 37.65.

6.1. Meta-analysis of stock selectivity, and market timing

The results of a meta-analysis applied to the excess returns study and the Jensen model are found in Table 4 and consider only selectivity, while the results in Table 5 include market timing. The excess return study, which is the least robust of the studies, indicates positive mean quarterly returns and significant departures from the average. The calculated chi-square statistics all exceed the critical chi-square. Also, about half of total variance is explained by sampling error. The Jensen study shows positive mean performance attributable to selectivity. However, none of the chi-square statistics are significant indicating that in all cases, the observed differences in performance based on the Jensen model can be attributed sampling error variation. The high proportions of sampling variation to total variance further evidence this. These results differ from the excess returns and suggest that after accounting for market risk, there is no significant population variance.

Table 5 includes the results relating to selectivity and it indicates that the annualized mean selectivity values are positive for all models. For example, the annualized values for Henriksson and Merton model are 1.69% for the S&P 500 benchmark, 3.44% for the Hamilton Yardstick, 2.42% for the Russell 3,000, and 1.74% for the equal-weighted S&P 500. The chi-square values indicate, however, no real variation around the mean value across managers for the Treynor and Mazuy model or the Bhattacharya and Pfleiderer model. The Henriksson and Merton model shows some variation at the 5% level of significance, but none at the 1% level. The proportion of total variation explained by sampling error is high for all models particularly for the Jensen model, the Henriksson and Merton model and the Bhattacharya and Pfleiderer model. One observation is that the results vary for the S&P 500 equal-weighted index, which indicates that all variation among managers can be explained by sampling error.[15] The results suggest that on the basis of stock selection, the Hamilton & Company's selected managers outperform the broad market and a typical active manager represented by the Hamilton Yardstick. In addition, the lack of significant variation among the managers further suggests that the selection group is a reasonably homogeneous group of managers from a performance viewpoint who, on average, exhibit superior stock selection capability versus the average manager. This is consistent with the fact that the managers are selected on common criteria according the methodology outlined in a previous section.

Our results from market timing show negative mean values across models and benchmarks.[16] The one exception is some of the results from the Henriksson and Merton model. These results are consistent with CFR (1993) and Fletcher (1995) and a host of other studies that focused on mutual fund performance. Those studies found negative market timing performance on average. We do, however, find considerable discrepancy

Table 4. Investment manager performance: excess returns. Meta-analysis

	S&P 500	Hamilton Yardstick	Russell 3,000	S&P Equal
b	0.00496	0.00819	0.00544	− 0.00040
χ^2	51.295	56.837	53.757	41.143
$(1-r)S_e^2/S_b^2$	0.507	0.457	0.484	0.632
	Jensen Index			
	S&P 500	Hamilton Yardstick	Russell 3,000	S&P Equal
b	0.005106	0.007772	0.006519	0.003311
χ^2	30.569	34.938	32.827	22.580
$(1-r)S_e^2/S_b^2$	0.851	0.744	0.792	1.151

Table 5. Investment manager performance: Meta-analysis. H&M model

	S&P 500	Hamilton Yardstick	Russell 3,000	S&P Equal
	Selectivity			
b	0.00422	0.00859	0.00604	0.00436
χ^2	31.828	33.437	31.525	19.732
$(1-r)S_e^2/S_b^2$	0.817	0.778	0.825	1.318
	Market Timing			
b	0.02747	− 0.02558	0.01424	− 0.02876
χ^2	32.840	40.273	34.942	19.815
$(1-r)S_e^2/S_b^2$	0.792	0.646	0.744	1.312
	T&M Model Selectivity			
b	0.00511	0.00855	0.00659	0.00402
χ^2	41.266	46.81266	44.260	27.800
$(1-r)S_e^2/S_b^2$	0.630	0.555405	0.587	0.935
	Market Timing			
b	− 0.00123	− 0.12078	− 0.00964	− 0.07662
χ^2	45.861	49.354	44.439	27.416
$(1-r)S_e^2/S_b^2$	0.567	0.527	0.585	0.948
	B&P Model Selectivity			
b	0.005115	0.008550	0.006594	0.004022
χ^2	32.762	34.039	32.601	21.828
$(1-r)S_e^2/S_b^2$	0.793591	0.763828	0.797510	1.191139
	Market Timing			
b	− 0.01188	− 0.06816	− 0.02035	− 0.07200
χ^2	62.482	75.550	66.967	48.147
$(1-r)S_e^2/S_b^2$	0.41612	0.34414	0.38825	0.54001

across models and somewhat less across benchmarks with regard to explaining the source of variation. If there were no real variation about the observed mean, then the observed mean value would be the true value for each of the managers. In our analysis, we find real variation in the excess returns model, the Bhattacharya and Pfleiderer model and the Treynor and Mazuy model. There is no consistent real variation for selectivity but some for market timing. There is also a lack of real variation across all benchmarks and models for the majority of the cases. The last row of each table suggests a large percentage of total variation can be attributable to sample variation. This is somewhat different from the results in CFR (1993), Fletcher (1995), and Coggin and Hunter (1993) who found evidence of real variation in every set of selectivity and market timing values. This result supports our earlier suggestion that the performance of this set of managers is more homogeneous because they were selected on the basis of the same selection criteria.[17]

7. Survivorship bias

Survivorship bias can be a real problem in empirical research and has been the subject of many studies. Brown et al. (1992) discuss the salient points of this issue in the context of performance measurement. In the case of our study, survivorship bias is not a large problem because managers when dropped are done so as the result of a discretionary decision. Thus, survivorship bias is only a problem to the extent that some managers in the general universe disappeared and there is some small probability that they may have been chosen. But for the most part we do not see it as a problem except in the context of data mining. Nonetheless, the data used in our study is carefully maintained, which minimizes the problem. Perhaps, a more robust study would compare the performance of those managers chosen with those who were not. However, the very nature of the study compares those chosen with a representative group who were not chosen. The Hamilton Yardstick represents this group not chosen.

8. Summary and conclusions

One purpose of this paper was to introduce selection criteria and to examine the performance of a group of pension fund managers selected on the basis of the criteria. To accomplish this we used the well known Jensen index and we also employed three models that separate performance into stock selection and market timing. Our general conclusion is that the selected manager group exhibit superior returns on the basis of actual returns and risk-adjusted returns. However, the strength of the performance is often lost versus the S&P 500 equal-weighted index, which suggests the managers do not focus on small capitalization stocks. This is not unexpected since the portfolios are geared toward large cap issues. The managers as a group demonstrate superior stock selection skills, but not market timing skills. This result is consistent with previous studies except our results hold true versus a random sample of managers represented by the Hamilton Yardstick. This group has even better stock selection skills than the average manager does. In addition, some of the managers have market timing skills. A meta-analysis was performed to

examine the effect of sampling error. Unlike CFR (1993) and Fletcher (1995) we did not find much real variation around mean values.

We recognize the benchmark controversy, but believe it is a limited problem due to the nature of the portfolios. The Hamilton Yardstick also provides an additional alternative measure that accounts for the differing styles. In addition, we may want to compare managers who have been on the list for the entire period versus the more recent selections.

Acknowledgments

The authors would like to thank Mr Jim Hamilton, President of Hamilton & Company, for providing the information on Hamilton & Company's selection criteria. We are also grateful to the staff for providing the data used in the analysis. The authors assume responsibility for the content of the paper and any error is our own.

Notes

1. A review of these studies is not presented here. We refer to the relevant studies in our discussion throughout the paper. For a comprehensive review of mutual fund performance, see Lee and Rahman (1994). In a recent study, Fletcher (1995) evaluated the selectivity and market timing performance of a sample of UK unit trusts using two alternative methodologies. He found that selectivity was on average positive, but market timing performance was negative. He also identified a trade-off between selectivity and timing
2. Hamilton & Company is an investment advisory company focusing on investment planning, organizational supervision, manager selection and performance attribution.
3. The mathematics of the model are found in Jensen (1968, 1969).
4. Grant (1977) explained how market timing activities will influence the estimates of single factor models that ignore market timing. He demonstrated that the estimates would be biased downward and result in Type I error.
5. For the specific model specifications consult the original articles for the three multi-factor models. See Treynor and Mazuy (1966), Bhattcharya and Pfleiderer (1983), and Henriksson and Merton (1981).
6. They correct an error made by Jensen (1972) and show that a simple regression technique can be used to measure market timing and selectivity. Bhattacharya and Pfleiderer (1983) assume that managers adjust their forecasts to minimize the variance of the forecast error. The exact specification of their model is available in Bhattacharya and Pfleiderer (1983) and also presented in Lee and Rahman (1990) and Coggin et al. (1993).
7. Noted reasons for removal include organizations that fragmented or disbanded and thus cannot be tracked. Additionally, one organization was removed when abrupt changes within the organization severely altered the firm's character. The scarcity of new managers reflects the difficulty of producing a consistent and superior product. In addition, the managers on the list are not style-oriented.
8. A more detailed analysis is available from Hamilton & Company.
9. Breen, Jagannathan, and Ofer (1986) have shown using simulation techniques that correcting for heteroskedasticity can significantly affect the conclusions of the test using parametric models such as the Henriksson and Merton model. They also contend that this condition is equally true for the Treynor and Mazuy model and the Bhattacharya and Pfleiderer model.
10. Breen et al. (1986) show that this technique is particularly useful when the form of heteroskedasticity is unknown. We follow the work of Breen et al. (1986), Lehman and Modest (1987), Coggin et al. (1993) and Fletcher (1995).
11. The differences reflect in part the difference in performance between small capitalization and large capitalization stocks. Over the sample period, the performance was similar in the averages, but small stocks were much more volatile.

12. The results of the directional and significance count for the three models that include market timing are not included here but are available from the authors.
13. Some previous studies cited in Coggin et al. (1993) include Kon (1983), Chang and Lewellen (1984), Henriksson (1984), Lehmann and Modest (1988), Cumby and Glen (1990) and Connor and Korajczyk (1991). This group found more evidence of negative market timing performance for mutual funds managers than positive market timing performance.
14. As reported earlier, the results of the excess returns study are not included in the body of the paper, but we have summarized the results in a meta-analysis.
15. Proportions greater than one result when the population variance is negative. This is possible when sampling error variance exceeds observed variance and when the correlation between residuals is low.
16. As pointed out in Coggin and Hunter (1993) and Coggin et al. (1993), the weakness of both the Treynor and Mazuy model and the Bhattacharya and Pfleiderer model is the failure to acknowledge negative market timing. We make the same adjustments as Coggin and Hunter (1993) to allow for negative market timing by recognizing the sign of the appropriate regression coefficient as an indicator of market timing skill. Thus, a negative coefficient on the quadratic independent variable would indicate negative market timing skill.
17. We also found that the results of examining the 80% probability interval for selectivity and market timing reveal a relatively narrow band that we have suggested is the result of the selection criteria.

References

Bello, Z. and V. Janjigian, ''A Reexamination of the Market-timing and Security-selection Performance Of Mutual Funds.'' *Financial Analysts Journal* 53, 24–31, (1997).

Bhattacharya, S. and P. Pfleiderer, ''A Note on Performance Evaluation.'' *Technical Report* Stanford University 1983.

Breen, W.R., R. Jagannathan, and R. Ofer, ''Correcting for Heteroscedasticity in Tests for Market Timing Ability.'' *Journal of Business* 59, 585–598, (1983).

Brinson, G.P, L.R. Hood, and G.L. Beebower, ''Determinants of Portfolio Performance.'' *Financial Analysts Journal* 42, 39–44, (1986).

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

Coggin, T.D. and J.E. Hunter, ''A Meta-analysis of Mutual Fund Performance.'' *Review of Quantitative Finance and Accounting* 3, 189–201, (1993).

Coggin, T.D., F.J. Fabozzi, and S. Rahman, ''The Investment Performance of US Equity Pension Fund Managers: An Empirical Investigation.'' *Journal of Finance* 48, 1039–1055, (1993).

Fama, E.F., ''Components of Investment Performance.'' *Journal of Finance* 27, 551–567, (1972).

Fletcher, J., ''An Examination of the Selectivity and Market Timing Performance of UK Unit Trusts.'' *Journal of Business Finance & Accounting* 22(1), 143–156, (1995).

Grinblatt, M. and S. Titman, ''Mutual Fund Performance: An Analysis of Quarterly Portfolio Holdings.'' *Journal of Business* 62, 393–416 (1989a).

Hansen, L., ''Large Sample Properties of Generalized Method of Moments Estimators.'' *Econometrica* 50, 1029–1054, (1982).

Henriksson, R.D., ''Market Timing and Mutual Fund Performance: An Empirical Investigation.'' *Journal of Business* 57, 73–96, (1984).

Henrikksson, R.D. and R.C., Merton, ''On the Market Timing and Investment Performance II: Statistical Procedures for Evaluating Forecasting Skills.'' *Journal of Business* 54, 513–533, (1981).

Hsieh, D., ''A Heteroscedasticity-Consistent Covariance Matrix Estimator for Time Series Regressions.'' *Journal of Econometrics* 22, 281–290, (1983).

Ippolito, R.A. and J.A. Turner, ''Turnover, Fees and Pension Plan Performance.'' *Financial Analysts Journal* 43, 16–26, (1987).

Jensen, J.C., ''The Performance of Mutual Funds in the Period 1945–1964.'' *Journal of Finance* 23, 389–416, (1968).

Jensen, J.C. Risk, ''The Pricing of Capital Assets, and the Evaluation of Investment Portfolios.'' *Journal of Business* 42, 167–247, (1969).

Jobson, J.D. and B.M. Korkie, ''Performance Hypothesis Testing with the Sharpe and Treynor Measures.'' *Journal of Finance* 36, 889–908, (1981).

Kao, G.W. and L.T. Cheng, ''International Mutual Fund Selectivity and Market Timing During Up and Market Conditions.'' *Financial Review* 33, 127–145, (1998).

Kon, S.J., ''The Market-timing Performance of Mutual Fund Managers.'' *Journal of Business* 56, 323–347, (1983).

Korkie, B., ''External vs Internal Performance Evaluation.'' *Journal of Portfolio Management* Spring, 36–42, (1993).

Lakonishok, J., A. Shleifer, and R.W. Vishny, ''The Structure and Performance of the Money Management Industry.'' In *Brookings Paper on Economic Activity Microeconomics* Brookings Institution, Washington, D.C. 339–391, 1992.

Lee, C.F. and S. Rahman, ''Market Timing, Selectivity, and Mutual Fund Performance: An Empirical Investigation.'' *Journal of Business* 62, 261–278, (1990).

Lee, C.F. and S. Rahman, ''Review, Integration, and Critique of Mutual Fund Performance Studies During 1965–1991.'' *Advances in Financial Planning and Forecasting* 5, 103–128, (1994).

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

Merton, R.C., ''On Market Timing and Investment Performance I: An Equilibrium Theory of Value of Market Forecasts.'' *Journal of Business* 54, 363–406, (1981).

Sharpe, W.F., ''Mutual Fund Performance.'' *Journal of Business* January, 119–138, (1966).

Treynor, J.L. and K. Mazuy, ''Can Mutual Funds Outguess the Market?'' *Harvard Business Review* 44, 131–136, (1966).

White, H., ''A Heteroscedasticity-Consistent Estimator and a Direct Test for Heteroscedasticity.'' *Econometrica* 48, 817–838, (1980).