Morningstar Ratings and Mutual Fund Performance

Christopher R. Blake
Graduate School of Business
Fordham University
113 West 60th Street
New York, NY 10023
Phone: (212) 636-6750
Fax: (212) 765-5573
email: blake@murray.fordham.edu

Matthew R. Morey
Department of Economics
Smith College
204 Pierce Hall
Northampton, MA 01063
Phone: (212) 769-0355
Fax: (718) 817-3518
email: morey@sophia.smith.edu

March 15, 1999
All Comments Welcome

The authors wish to acknowledge the financial support of the Economic and Pension Research Department of TIAA-CREF and thank Will Goetzmann and Charles Trzcinka for data and Stephen Brown, Edwin Elton, Martin Gruber, Richard C. Morey, Derrick Reagle, Emily Rosenbaum, H.D. Vinod, Mark Warshawsky, and seminar participants at the Securities and Exchange Commission for helpful comments and suggestions.

Morningstar Ratings and Mutual Fund Performance

Abstract

This study examines the degree to which the well-known Morningstar rating system is a predictor of out-of-sample mutual fund performance, an important issue given that high-rated funds receive the lion's share of investor cash inflow. We use a data set based on growth mutual funds that is free from survivorship bias and adjusted for load fees to examine the predictive qualities of the rating system. Moreover, we use various performance metrics over different time horizons and sample periods. We also compare the predictive qualities of the Morningstar rating system with those of a "naïve" predictor: simple historical average monthly returns. The results indicate three main findings. First, low ratings from Morningstar generally indicate relatively poor future performance. Second, for the most part, there is little statistical evidence that Morningstar's highest-rated funds outperform the next-to-highest and median-rated funds. Third, Morningstar ratings do no better than the "naïve" predictor for predicting fund performance.

JEL code: G23

I. Introduction

In recent years, there has been increasing attention paid to the persistence of mutual fund performance in the finance literature.[1] Yet, to date, there has been considerably less attention devoted to the predictive qualities of the Morningstar 5-star mutual fund rating service that many investors use as a guide in their mutual fund selections. This study attempts to fill that void by examining the ability of the Morningstar ratings to predict both unadjusted and risk-adjusted returns, using performance metrics common in the performance literature.

The question of whether Morningstar ratings predict out-of-sample performance is an important one, given that several studies in the performance literature have documented that new cash flows from investors are related to past performance ratings. (See, e.g., Sirri and Tufano (1992) and Gruber (1996).) In fact, there is evidence that high-rated funds experience far greater investor cash inflows than the investor cash outflows experienced by low-rated funds. (See, e.g., Sirri and Tufano (1992) and Goetzmann and Peles (1994).) Hence, examining performance across funds grouped by Morningstar rankings will indicate if these cash flows are justified by subsequent relative performance.

As evidence of the importance of the Morningstar 5-star rating service (where a 5-star rating is the best and a 1-star rating is the worst), consider a recent study reported in both the Boston Globe and the Wall Street Journal.[2] This study found that 97 percent of the money flowing into no-load equity funds between January and August 1995 was invested into funds which were rated as 5-star or 4-star funds by Morningstar, while funds with less than 3 stars suffered a net outflow of funds during the same period. Moreover, the heavy use of Morningstar ratings in mutual fund advertising suggests that mutual fund companies *believe* that investors care about Morningstar ratings. Indeed, in some cases, the only mention of return performance in the mutual fund advertisement is the Morningstar star rating. Finally, the importance of the Morningstar ratings has been underscored by some recent high-profile publications (e.g., Blume (1998) and Sharpe (1998)) which have investigated the underlying properties of the Morningstar rating system.

---

[1] For example, Hendricks, Patel and Zeckhauser (1993), Goetzmann and Ibbotson (1994), Malkiel (1995), Brown and Goetzmann (1995), Elton, Gruber and Blake (1996a) and Carhart (1997).

[2] Charles Jaffe, "Rating the Raters: Flaws Found in Each Service." Boston Globe, August 27th, 1995, p. 78. The same survey was also reported by Karen Damato, "Morningstar Edges Toward One-Year Ratings." Wall Street Journal, April 5th, 1996.

Despite the importance of the Morningstar ratings service, there is very little previous research on the predictive abilities of the Morningstar ratings. Moreover, the work that has been done has often suffered from methodological problems. For instance, Khorana and Nelling (1998) examine the question of persistence of the Morningstar ratings themselves. Specifically, the authors compare the Morningstar ratings from a group of funds in December 1992 to the ratings the funds received in June 1995. They find evidence of persistence, in that highly rated funds are still highly rated and low-rated funds are still low rated. However, there are a number of problems in that study. First, there is a survivorship bias problem, since the funds were selected at the end of the sample period rather than at beginning. Hence, any fund which had merged, liquidated or changed its name between the beginning and ending of the sample period was not included in the sample. Second, because Morningstar uses a 10-year risk-adjusted return as a major component of its ratings, and because there are only 2 and ½ years of data between the beginning and end of their sample, the ratings are based on overlapping data. Consequently, the findings of persistence in the ratings are endemic to the data. Finally, their study ignores a key question from an investor's viewpoint, which is whether or not the Morningstar ratings can be used to predict fund *performance*, not merely future Morningstar ratings.

In this paper we examine the question, Does the Morningstar five-star system have any predictive power for the future performance of funds? Our data and methodology are sensitive to many key issues in mutual fund research. Namely:

1) Our paper uses a mutual fund data set generated at the time the funds were actually rated by Morningstar. We then follow the out-of-sample performance of all of these funds. This methodology allows us to circumvent the well-known survivorship bias problem that is described by Brown, Goetzmann, Ibbotson and Ross (1992), Elton, Gruber and Blake (1996b) and others.

2) Unlike most previous studies of mutual fund performance and prediction, our returns adjust for front-end and deferred loads. This is important since the Morningstar rating system also adjusts for loads.

3) We compare the predictive qualities of the Morningstar ratings with those of a "naïve" alternative predictor: simple historical average monthly returns.

4) We examine different out-of-sample horizons, i.e., one-year, three-years and five-years, so that we can give both short- and long-term analyses of the predictive qualities of Morningstar ratings and the "naïve" predictor. Moreover, these time horizons are consistent with the historical returns that prospective investors are often provided with when considering a mutual fund.

4) We examine the predictive qualities of the Morningstar ratings and the "naïve" predictor at different times. Hence, we can examine how well they predict in up and down markets.

5) A number of studies, e.g. Brown (1999), Brown and Goetzmann (1997), Elton, Gruber, Das and Hlavka (1993) and Goetzmann and Ibbotson (1994), state that performance predictability may be due to the style of funds examined rather than skill. To control for this, we only examine funds with the same style (i.e., growth funds) at the time they were rated.

6) We measure out-of-sample performance using several well-known performance metrics including the Sharpe Ratio (1966), mean monthly excess returns, a modified version of Jensen's alpha (1968) and a 4-index alpha.

7) We analyze the results using parametric and non-parametric tests.

The rest of the paper is organized as follows. Section II extensively describes the data that we use in the paper and relates the method in which the funds where chosen, how Morningstar calculates their ratings, and how the returns data were collected and calculated. Section III describes the methodology of the paper, Section IV presents the Morningstar rating results, Section V presents the "naïve" predictor results, and Section VI provides the conclusion.

II. Data

To better organize the description of the data, this section is divided it into seven subsections: fund selection criteria, problem funds, Morningstar ratings, Morningstar scores, simple historical average monthly returns, out-of-sample evaluation periods, and the returns and load adjustments.

IIA. Fund Selection Criteria

We are careful to track a consistent and unbiased sample of funds over time. For selecting the mutual funds for our sample we use the beginning-of-the-year Morningstar On-Disk or Principia programs from 1992 to 1997.[3] We use the beginning-of-the-year disks as a way of simplifying the data so that we are always examining calendar years. Moreover, we start at the beginning of the year 1992 since this corresponds to the first beginning-of-the-year On-disk program.[4]

By using the actual Morningstar disks we know all the funds which were available to investors selecting funds based on Morningstar ratings at the time of the Morningstar evaluation. In this way, we circumvent any possible survivorship bias problems. Data previous to the beginning of the On-disk program are available from Morningstar on a proprietary basis, however, these data include only the surviving funds; funds that were rated at the time of the Morningstar rating and yet have merged or liquidated at some later date are not available.[5] Since the use of such data would introduce a severe survivorship bias, they are not used in our study.

From the beginning-of-the-year disks we then select funds based on three criteria. First, we select only growth equity funds as identified by Morningstar. These funds are described by Morningstar as "funds that pursue capital appreciation by investing primarily in equity securities," where "current income, if considered at all, is a secondary concern."[6] We choose only the growth investment category for three reasons. One, growth funds always have represented the largest number of funds in any one Morningstar investment objective[7] and hence were more likely to reflect the average investor's choices. Second, a number of studies, (e.g. Brown (1999), Brown and Goetzmann (1997), Goetzmann and Ibbotson, (1994)), state that performance predictability may by due to the style of funds examined rather than skill. We attempt to control for this by only examining growth funds. Third, because we are examining the out-of-sample forecasting ability, each fund listed by Morningstar had to be identified with out-of-sample returns. With mergers,

[3] These correspond to the January 1992 On-Disk, January 1993 On-Disk, January 1994 On-Disk, January 1995 On-Disk, January 1996 On-Disk, and the January 1997 Principia. In 1997 On-disk changed to Principia.

[4] The On-Disks begin in October 1991.

[5] We thank Peter Carrillo of Morningstar for this point.

[6] p. 92 of Morningstar Principia Reference Manual (1998).

[7] In domestic equity the other investment objectives that are defined by Morningstar are aggressive growth, growth-income, equity-income, and small company.

name changes and liquidations of funds, this identification process was quite onerous. Hence, other investment categories were excluded from the analysis.

Since we are examining the out-of-sample performance of the funds, we also examine if the funds retain their growth classification by Morningstar in the out-of-sample periods. We find that in every sample examined over 90 percent of the funds retain their growth classification at the end of the sample period.[8] Hence, the vast majority of funds do not change their style of management.

The second criterion was that each growth fund had to have at least 10 years of returns at the time it was ranked by Morningstar. In other words, funds rated by Morningstar in January 1993 had to have return data starting from, at the latest, January 1983. We used the 10-year cut off because of Morningstar's *base-line* rating system.[9] As stated earlier, Morningstar provides each mutual fund with a 1 to 5 star summary rating. To obtain this summary rating, Morningstar takes a weighted average of the 3-year, 5-year and *10-year* risk adjusted returns. Due to the importance of the 10 year time period in their rankings, we used this as a criterion in selecting funds.

The third and last selection criterion used was that funds had to be open at the time they were rated by Morningstar. Any fund that was closed to new investors at the time of the rating by

[8] To obtain this percentage we examine only the funds in the sample which did not merge nor were liquidated during the out-of-sample period. Table 3 shows the actual number of funds that did change their classification.

[9] We call this the base-line rating system since Morningstar uses this system if a fund has at least 10 years of returns. It should noted that Morningstar also provides summary ratings for funds with less than 10 years of monthly returns: funds with more than 5 years of returns but less than 10 years use a 40 percent weighting on the 3 year risk-adjusted return and a 60 percent weighting on the 5 year risk-adjusted return; funds with less than 5 years of return data use a 100 percent weighting on the 3 year risk-adjusted returns; funds with less than 3 years of returns are not rated. Blume (1998) provides evidence to suggest that funds with less than 10 years of returns are more likely to be rated as 5-star or 1-star since their weightings are based upon shorter samples. Hence, another reason for our use of the 10-year cut-off as a criterion for fund selection is to prevent such a bias from affecting our results.

Morningstar was excluded from our analysis. The purpose of this was to maintain a sample of funds that could actually be invested in at the time of the ratings.[10]

IIB. Problem Funds

In this paper we examine the out-of-sample forecasting ability of Morningstar's ratings and rankings. As described in the previous section, we select funds at the time the funds were rated by Morningstar. To examine the out-of-sample forecast ability we then obtain the out-of-sample monthly returns of these funds. For a majority of the funds, obtaining the out-of-sample returns is simply a matter of following the previously rated fund. However, because a minority of funds have either gone through a name change, a merger, a combination of both, or because they have liquidated, identifying out-of-sample returns for those funds is more complicated. In this section, we describe how we handle these problematic funds.

For name changes, we use the Morningstar data[11] and *The Wall Street Journal* to identify the name changes. We then simply use the new named fund's returns as the out-of-sample returns.

For the merger funds we used the Morningstar data and *The Wall Street Journal* to ascertain the month of the fund merger. However, when these two sources did not provide the necessary information, we called the individual mutual fund companies. Once the merger month is identified, we then collect the out-of-sample returns by the following procedure. First, until the fund merges, we simply use the out-of-sample returns of the fund in question. After the fund merges into its partner fund, we assume the investor randomly re-invests into one of the other growth funds in our sample. Hence, the out-of-sample returns from the merger month onwards are equally weighted monthly averages of the returns of all the other growth funds in our sample.

For the liquidated funds we first define when the fund was liquidated. Again, this information was obtained from Morningstar or *The Wall Street Journal*. As with the merger funds, from the month of liquidation and onwards, we assume the investor randomly re-invests in the current sample of growth funds.

---

[10] The number of closed funds in each of the years examined is as follows: January 1992: 3 closed funds; January 1993: 4 closed funds; January 1994: 4 closed funds; January 1995: 6 closed funds; January 1996: 8 closed funds; January 1997: 15 closed funds.

[11] The Morningstar On-Disk and Prinicipia disks both provide a list of funds that have recently undergone name changes, mergers and liquidiations.

IIC. Morningstar ratings

To calculate its ratings, Morningstar first classifies funds into one of four categories: Domestic Equity, Foreign Equity, Municipal Bond and Taxable Bond.[12] The ratings are then based upon an aggregation of the 3-year, 5-year and 10-year risk-adjusted return. The risk-adjusted return is calculated in the following manner. First they calculate a load-adjusted return for the fund by adjusting the returns for expenses such as 12b-1 fees, management fees and other costs automatically taken out of the fund, and then by adjusting for front-end and deferred loads.[13] Next, they calculate a "Morningstar Return" in which they take the expense- and load-adjusted excess return divided by the higher of two variables: the excess average return of the fund category (domestic stock, international stock, taxable bond, or municipal bond) or the average 90-day U.S. T-bill rate:

$$\frac{\text{(Expense and Load Adjusted Return on the Fund - T-Bill)}}{\text{Higher of (Average Category Return-T-Bill or T-Bill)}} \qquad (1)$$

Morningstar divides through by one of these two variables to prevent distortions caused by having low or negative average excess returns in the denominator of equation (1). Such a situation might occur in a protracted down market. [14]

Morningstar then calculates a "Morningstar Risk" measure. This measure is calculated differently from traditional risk measures, such as beta and standard deviation, which both see greater-than and less-than-expected returns as added volatility. Morningstar believes that for most investors their greatest fear is losing money which they define as under performing the risk-free rate of return an investor can earn from the 90 day Treasury Bill. Hence, their risk measure only

[12] Note that originally Morningstar used only three categories: Domestic Equity, Municipal Bond, and Taxable Bond. The Foreign Equity funds were placed in the domestic equity category.

[13] Blume (1998), p. 4-5, provides an excellent description of how Morningstar accounts for loads in the Morningstar Returns. The load adjustment process is the following. Assume L is the load adjustment. If there is no load of any type, then L is equal to 1. If there is a load, L is less than one, i.e., a 4 percent front-end load, would make L equal to 0.96. The load-adjusted return is then the (return of the fund)*L. Note that the front-end load is always assumed to the be the maximum possible load. The deferred load adjustment is reduced as the holding period is increased. Later in the data section of the paper we explain more about how we adjust the return data for loads.

[14] Principia Manual, p. 97.

focuses on downside risk. To calculate the Morningstar risk, they plot the monthly returns in relation to T-bill returns. They add up the amounts by which the fund trails the T-Bill return each month and then divide that total by the time horizon's total number of months. This number, the average monthly underperformance statistic, is then compared with those of other funds in the same broad investment category to assign the risk scores. The resultant Morningstar risk score expresses how risky the fund is relative to the average fund in its category.[15]

To illustrate the Morningstar risk calculation, we provide an example where we define the time horizon as 1 year. Table 1 presents hypothetical results for a mutual fund.

To calculate a fund's summary star-rating, Morningstar calculates the 3-year, 5-year and 10-year Morningstar Return and Risk. For each time horizon, the Morningstar Risk scores are then subtracted from the Morningstar Return scores. The three numbers (one for each time horizon) are then given subjective weights.[16] The 3-year number receives a 20 percent weighting, the 5-year a 30 percent weighting, and the 10-year a 50 percent weighting. With these weights, the three numbers are then added together. The resulting number is then plotted along a bell curve to determine the fund's star rating. If the fund scores in the top 10 percent of its broad investment category, it receives a rating of 5 stars; if the fund falls in the next 22.5 percent it receives 4 stars; if it falls in the middle 35 percent it receives 3 stars; if it lies in the next 22.5 percent the fund receives 2 stars, and if it is in the bottom 10 percent it receives 1 star. Morningstar, with a few minor exceptions, has used this same summary rating system throughout its history.[17]

Table 2 presents the distribution and average star ratings in our January 1992 through January 1997 samples. Several qualities about the data should be noted here. One, the number of funds in each sample grows. This is not surprising, since with each year the number of funds that meet the criteria grow. Two, there are more 5-star funds than 1-star funds and the average star rating of each sample is above 3. This skewness in the ratings of the sample indicates that growth funds and funds with 10 years or more of returns performed slightly better than other funds in the Morningstar domestic equity category. Three, the standard deviation of the ratings is about the same in each sample indicating that the distribution of the ratings does not differ much from one sample to another. Four, in every sample, the majority of the funds are load-funds although the

[15] Principia Manual, p. 98.

[16] Morey and Morey (1999) present a methodology that endogenously determines these weights.

[17] The Morningstar technical staff verified this point. See Blume (1998) p. 3 for more on this issue.

percentage falls off slightly as the sample moves closer to the present. Five, for the load-funds, most have front-end loads and relatively few have deferred loads.

IID. Morningstar Scores

Since January 1994, Morningstar has provided the 3, 5 and 10-year Morningstar Return and Risk numbers for all the mutual funds that it evaluates. This information, plus the subjective weights, (20%, 30% and 50% for the 3, 5 and 10-year horizons) allows us to calculate the resultant score and hence to numerically rank the funds evaluated here. These scores allow us to conduct non-parametric rank correlation tests.

IIE. Simple Historical Average Monthly Returns: The "Naïve" Predictor

We use each fund's average monthly return during the ten-year period prior to the out-of-sample evaluation period as an alternative predictor. We then compare the predictive qualities of this "naïve" predictor with those of the Morningstar rankings and scores.

IIF. Out-Of-Sample Evaluation Periods

Investors, when evaluating performance, are typically presented with the 1-year, 3-year, 5-year and (when possible) the 10-year past performance windows. Similarly, we use 1-, 3- and 5-year periods to examine the out-of-sample forecasting ability of Morningstar's ratings (the 10-year window is outside the bounds of our sample). This provides us with 12 subsamples for performance evaluation. Table 3 presents, for each sample period, the number of funds, the number of merger funds, the number of liquidated funds and the number of funds who changed their Morningstar objective during the out-of-sample evaluation period (i.e., from growth to some other objective).

IIG. Returns Data and Load Adjustments

For the out-of-sample returns and the in-sample returns used in the naïve predictor, the data consist of monthly returns from the Morningstar On-Disk and Prinicipia programs. These return data are adjusted to account for management, administrative, and 12b-1 fees and other costs automatically taken out of fund assets. However, unlike the Morningstar risk-adjusted ratings, the monthly return data do not adjust for sales charges such as a front-end and deferred loads.[18] Consequently, if we

[18] Principia Manual (1998), p. 107.

use the monthly return data for the out-of-sample returns, the returns on load funds are overstated. The question is, thus, how to incorporate loads into the monthly return data?

Very little attention in the mutual fund performance literature is given to the treatment of loads in return data. Hendricks, Patel and Zeckhauser (1993), Elton, Gruber and Blake (1996a), Malkiel (1995), and Carhart (1997) provide no adjustment for loads in their returns data. However loads are important, especially in this paper since the Morningstar ratings encompass load-adjusted returns. Evaluating the out-of-sample performance of load-adjusted ratings by using non-load adjusted returns is somewhat inappropriate to say the least. But the question is how to deal with loads? There is not a simple answer. For example do you use front-end loads, deferred loads, or both? When and for how long to do you apply the load? What if the mutual fund has reduced its load over time (especially the deferred load)? Do you use an average load adjustment for each month or do you use an annualized load? If you decide to use an annualized load what interest rate do you use to discount the load factor?

In light of all these difficulties, we adjust the monthly returns of each mutual fund using an approach similar to Rea and Reid (1998). The approach is the following. For both front-end and deferred loads, we consider an investor who buys and holds the load shares for a fixed number of months, i.e., 12 months (1 year), 36 months (3 years) or 60 months (5 years). For front-end loads, the investor buying the fund pays a load in a lump sum at the time the fund is purchased. To spread the front-end load across the period that the shares are held, we use Rea and Reid's assumption that the investor borrows the amount necessary to pay the load up front and then repays the loan as an annuity in equal, monthly installments during the holding period. Hence, the monthly load adjustment reflects the amount that was borrowed and the interest on the loan.

Mathematically, our front-end load adjustment process is the following:

$$f^m = \frac{f}{\sum_{j=1}^{h} (1+r)^{-j}} \qquad (2)$$

where

r = the monthly interest rate (average monthly 1-, 3-, or 5-year Treasury rate over the holding period)

f = the front-end load (expressed as a percent)

h = the number of months the fund is held

f^m = the monthly front-end load adjustment

Hence, the front-end load adjusted returns are:

$m^{fla} = m - f^m$, where

m = the monthly return of the mutual fund

m^{fla} = the monthly front-end-load-adjusted return of the mutual fund

As an example of the above adjustment, consider a one-year investment in Fidelity's Magellan fund starting in January 1992. As of January 1992, that fund had a front-end load of 3%, and the 1-year Treasury yield was 3.84%, giving a monthly average rate of 0.31%. Therefore, for the 1-year holding (out-of-sample) period, f = 3%, r = 0.0031, and h = 12, giving f^m = 0.255%. We then subtract 0.255% from each of the Magellan fund's 12 monthly returns during 1992 to obtain the load-adjusted returns.

For the deferred-load adjustment, the process is slightly different. The difference lies in the fact that the payment of the deferred load does not occur until the end of the holding period. To convert the deferred load into a monthly payment, the investor is assumed to prepay the load in equal monthly installments. The amount of the monthly prepayment reflects the deferred load less the interest earned on the prepayments.

Thus the equation for the monthly deferred-load adjustment is:

$$d^m = \frac{d}{\sum_{j=1}^{h} (1+r)^j} \qquad (3)$$

where

d = the deferred load (expressed as a percent)

d^m = the monthly deferred-load adjustment

Hence, the deferred-load-adjusted returns are:

$m^{dla} = m - d^m$, where

m = the monthly return of the mutual fund

m^{dla} = the monthly deferred-load-adjusted return of the mutual fund

As with the front-end loads, we use the average monthly 1-, 3-, or 5-year Treasury rate over the holding period for the interest rate. However, in contrast to the front-end load adjustment, we reduce the amount of the deferred load as the holding period, *h*, increases. We do this because Morningstar also reduces the deferred load as the holding period increases. Hence, for a holding period of 12 months, the full amount of the deferred load is imposed. For the 36-month holding period we apply only ½ of the original deferred load and in the 60-month holding period the deferred load completely disappears. Table 2 presents some summary data on the load structure of the funds in our samples.[19]

III. Methodology

To measure out-of-sample performance we use four performance metrics: The Sharpe (1966) ratio, mean monthly excess returns, a modified version of Jensen's (1969) alpha, and a 4-index alpha. To examine the out-of-sample predictive performance of the Morningstar ratings and the "naïve" predictor, we use three methods: Dummy variable regression analysis, pooled dummy variable regression, and the non-parametric Spearman-Rho rank correlation test.

IIIA. Out-of-Sample Performance Measurement

We use four performance metrics from the existing performance literature to measure out-of-sample performance: The Sharpe (1966) ratio, mean monthly excess returns, a modified version of Jensen's (1969) alpha, and a 4-index alpha.

Specifically the Sharpe ratio for fund *i* is:

$$Sharpe_i = \frac{\overline{}^{LA}_{i}}{\sigma_i}$$

where

$\overline{}^{LA}_{i}$ = the mean excess (net of the 30-day T-bill rate) load-adjusted monthly return for the *i*th mutual fund during the evaluation (out-of-sample) period.

σ_i = the standard deviation of the excess load-adjusted monthly returns for the *i*th mutual fund during the evaluation period.

[19] We also examined all the results with the returns not adjusted for loads. In this analysis we simply used a dummy variable to control of the difference in load and no-load funds. The results were very similar to those presented later in the paper.

The mean monthly excess returns are simply equal to $\overline{R_i}^{LA}$.

The modified Jensen and 4-index alphas are calculated using a methodology similar to that of Elton, Gruber and Blake (1996a). The following time-series regression model is used:

$$R_{it} = a_i + \sum_{k=1}^{K} \beta_{ik} I_{kt} + \varepsilon_{it} \qquad (5)$$

where

R_{it} = the excess total return (net of the 30-day T-bill return) for fund i in month t, not load adjusted

a_i = the intercept for fund i, upon which the fund's performance alphas are based

b_{ik} = the sensitivity of fund i's excess return to index k

I_{kt} = the return for index k in month t

e_{it} = the random error for fund i in month t

For the modified Jensen alphas, K = 1 and I_{1t} = the excess total return of the S&P 500 in month t. For the 4-index alphas, K = 4, I_{1t} = the excess total return of the S&P 500 in month t, I_{2t} = the excess total return of Lehman Aggregate Bond Index in month t, I_{3t} = is the difference in return between a small-cap and large-cap stock portfolio based on Prudential Bache indexes in month t, and I_{4t} = is the difference in return between a growth and value stock portfolio based on Prudential Bache indexes in month t.[20] We utilize the 4-index model because, as shown in Elton, Gruber and Blake (1996a), this model provides for better risk adjustment for mutual funds than does the single-index model and because this model is especially well suited for "growth" mutual fund performance evaluation.

Unlike the Sharpe and mean returns measures, for the single and multi-index measures, we utilize monthly non-load-adjusted returns. We use non-load adjusted returns since we use both out-of-sample and in-sample data for these measures, and there is no clear method for adjusting the in-sample data for loads.[21]

[20] See Elton, Gruber and Blake (1996a) for a detailed description of the Prudential Bache portfolios used in the 4-index model.

[21] We also examined the results for the Sharpe ratios and mean monthly returns using *non-load-adjusted* returns, and the results were essentially the same as those with the load-adjusted returns. Hence, it is unlikely that using load-adjusted returns would affect the results for the Jensen and multi-index alphas.

Specifically, for each sub-sample, we utilize a ten-year time series of monthly non-load returns to obtain an estimate of the intercept from either the single index or 4-index model regression. For our 1-year-evaluation subsamples, we utilize a time series starting 9 years before the selection date and ending 1 year after the selection date. For our 3-year-evaluation subsamples, we utilize a time series starting 7 years before the selection date and ending 3 years after the selection date. For our 5-year-evaluation subsamples, we utilize a time series starting 5 years before the selection date and ending 5 years after the selection date.

To obtain the alphas, we add the average monthly residual during the evaluation period to the intercept. For example, to obtain a modified Jensen alpha for a fund's 1-year out-of-sample performance measure, we run the 1-index model on 10 years of monthly returns starting 9 years before the selection date and ending 1 year after the selection date to obtain an estimate of the intercept. We then add the average of the fund's residuals during the 1 year after the selection date (the evaluation period) to the estimated intercept to obtain the fund's modified Jensen alpha.

To obtain alphas for funds that merged or liquidated during the evaluation period, we proceed as follows. First, we run 2 regressions: (1) a regression using the fund's returns starting in the same month as the surviving funds and ending in the month prior to the fund's disappearance, and (2) a regression run over the entire 10-year regression period using the returns on an equally weighted portfolio formed each month from the existing funds in the sample. We then form a weighted average of: (1) the fund's estimated intercept plus the fund's average residual during the time it survived in the evaluation period and (2) the estimated intercept plus the average residual during the remaining time in the evaluation period of the equally weighted portfolio, where the fund's weight is the fraction of the evaluation period it survived and the equally weighted portfolio's weight is the remaining fraction. This provides a performance measure for an investor who buys a remaining fund in the sample at random if the original fund merges or liquidates.

IIIB. Dummy variable regression analysis

The first method we use to examine out-of-sample predictive performance is a cross-sectional dummy variable regression analysis. This approach allows us to examine the Morningstar star ranking group differences in performance predictability.

In addition, in order to make the results for the "naïve" predictor comparable to those for the Morningstar star groups, we also divide the funds into five subgroups after ranking them in descending order by their 10-year historical monthly return averages (their "naïve" predictors).

These five "naïve" predictor groups are not quintiles, since we wanted to preserve the same number funds in each "naïve" predictor group as we have in each of the five Morningstar star groups. As an example, consider our January 1993 subsamples. The same 134 funds are in each of these subsamples: 11 5-star funds, 38 4-star funds, 68 3-star funds, 16 2-star funds, and 1 1-star fund (see Table 2). Therefore, for our 1993 subsamples, "naïve" predictor group 5 has the 11 funds with the highest "naïve" predictors, group 4 has the next highest 38 funds, etc.

For the dummy variable regression analysis, we estimate the following equation for each of our 12 subsamples shown in Table 3 when using the Sharpe ratio or mean monthly excess return performance metric:

$$S_i = \gamma_0 + \gamma_1 D4_i + \gamma_2 D3_i + \gamma_3 D2_i + \gamma_4 D1_i + u_i \qquad (6)$$

where:

S_i = out-sample performance metric for fund i,
$D4$ = 1 if a 4-star fund or if in "naïve" predictor group 4, 0 if not,
$D3$ = 1 if a 3-star fund or if in "naïve" predictor group 3, 0 if not,
$D2$ = 1 if a 2-star fund or if in "naïve" predictor group 2, 0 if not,
$D1$ = 1 if a 1-star fund or if in "naïve" predictor group 1, 0 if not,
i = 1 through N, where N is the total number of funds in the subsample.

In the above equation, the 5-star fund group or the "naïve" predictor group 5 is the reference group for the dummy variable regression.[22] Hence, when using the load-adjusted Sharpe ratio as the out-of-sample performance measure, the coefficient, g_0 represents the expected load-adjusted Sharpe ratio when all the dummy variables are equal to 0, and the coefficients g_1 through g_4 represent the differences between the dummy variables and the reference group. For example, a negative g_1 implies the group of 4-star funds performs worse than the group of 5-star funds; a positive g_1 implies the group of 4-star funds outperforms the 5-star fund group. The t-statistics on the coefficients provide a test of the significance of the difference between an individual dummy group and the reference group.

We use the 5-star funds or "naïve" predictor group 5 as a reference group because they provide a ceiling from which we can compare the performance of the lower group funds. If the star

[22] It should be noted here that we also performed all of the dummy variable regressions using the 3-star funds or the "naïve" predictor group 3 as the reference group. The results did not change when using this reference group. These results are available from the authors upon request.

ratings or "naïve" predictors accurately predict out-of-sample performance we should see increasingly negative (and significant) coefficients as we move from g_1 to g_4.[23]

For the Jensen and 4-index alpha out-of-sample performance metrics, we use equation (6) plus an added (0,1) dummy variable to account for differences between no-load and load funds. (We do this, since the Jensen and 4-index alphas are based upon non-load adjusted returns.)

IIIC. Pooled Dummy Variable Regression Analysis

In order to obtain summary results from the dummy variable regression analysis we use a pooled cross-sectional regression approach for each of the 3 evaluation periods. Thus we have three different pooled regressions:

1) A "one-year" pooled regression that pools the 92-1year, 93-1year, 94-1year, 95-1year, 96-1year and 97-1year subsamples.

2) A "three-year" pooled regression that pools the 92-3year, 93-3year, 94-3year, 95-3year subsamples.

3) A "five-year" pooled regression that pools the 92-5year, 93-5year subsamples.

Note that these pooled regressions are pooling unbalanced cross-sections (e.g., the sample of funds in the 1992-1year group is similar but not equal to the sample of funds in the 1993-1year group). Moreover, the number of time series periods is very short. Indeed, the 5 year pooled regression only includes two time periods (the 1992-5year and 1993-5year samples). Hence, with the unbalanced sample pool, and because the small number of time periods invalidates the use of GLS procedures such as the Parks method,[24] we do the following. First, before pooling the samples together we test each individual sample for cross-sectional heteroscedasticity using the White (1980) test. Upon finding no evidence of problems, we then use a least squares dummy variable approach in which we run least squares using yearly dummy variables for each additional year that

[23] As a test of robustness we also examined the results with the 3-star group (the median star-group ranking) as the reference group. The results were the same as those presented below.

[24] According to Beck and Katz (1995) the Parks method for pooled regressions requires that the number of time periods be at least as large as the number of cross-sections.

pooled into the sample. The annual dummies control for the differences between years. After pooling the data we then examine and, if needed, correct the standard errors for heteroscedasticity once more using the White procedure. Thus, for example, the "one-year" pooled regression was estimated by the following equation:

$$S_{i1} = \gamma_0 + \gamma_1 D4_i + \gamma_2 D3_i + \gamma_3 D2_i + \gamma_4 D1_i + \gamma_5 D93_i + \gamma_6 D94_i + \gamma_7 D95_i + \gamma_8 D96_i + \gamma_9 D97_i + u_i \quad (7)$$

where,

S_{i1} = 1-year out-sample performance metric for fund i,
$D4$ = 1 if a 4-star fund or if in "naïve" predictor group 4, 0 if not,
$D3$ = 1 if a 3-star fund or if in "naïve" predictor group 3, 0 if not,
$D2$ = 1 if a 2-star fund or if in "naïve" predictor group 2, 0 if not,
$D1$ = 1 if a 1-star fund or if in "naïve" predictor group 1, 0 if not,
$D93$ = 1 if in the 1993 sample, 0 if not in the 1993 sample,
$D94$ = 1 if in the 1994 sample, 0 if not in the 1994 sample,
$D95$ = 1 if in the 1995 sample, 0 if not in the 1995 sample,
$D96$ = 1 if in the 1996 sample, 0 if not in the 1996 sample,
$D97$ = 1 if in the 1997 sample, 0 if not in the 1997 sample,
i = 1 through N, where N is the total number of funds in the pooled sample.

In the pooled regressions, either the 5-star or "naïve" predictor group 5 from the 1992 sample fund group is the reference group for the dummy variable regression. The representation of coefficients g_1 through g_4 is the same as in the individual regressions. The coefficients g_5 through g_9 represent the differences between the various sample groups and the 1992 sample and are used to control for differences between sample periods. A significant coefficient indicates that one sample period is significantly different from the 1992 sample.

Also, as with the individual non-pooled regressions, for the pooled Jensen and 4-index alpha regressions we also add a (0,1) dummy variable to account for differences between no-load and load funds.

IIID. Spearman-Rho Rank Correlation Test

As a final test we use the two-tailed Spearman-Rho rank correlation test to examine the rank correlations of both the Morningstar scores and the "naïve" predictors with the out-of-sample performance measures. Since Morningstar provides the data to rank the funds beginning in 1994, we only examine this test for samples that begin in 1994 or later. The Spearman-Rho has a null hypothesis of no correlation between the two rankings and is a non-parametric test.

For this test we follow the methodology of Elton, Gruber and Blake (1996a). For each fund in the sample, we examine the four different out-of-sample measures: the (load-adjusted) Sharpe ratios, the (load-adjusted) mean monthly excess returns, the Jensen alphas, and the 4-index alphas. We first sort all the funds in descending order by either their in-sample Morningstar scores or , in the case of the "naïve" predictor, by their in-sample 10-year mean monthly return. We then organize the data into deciles and compute the average for each decile. Our goal is then to examine whether the decile ranking given by either the Morningstar scores or by the "naïve" predictors corresponds to the decile rankings of the four out-of-sample performance measures. If the Morningstar system or the "naïve" predictor predicts well out-of-sample, then there should be close correlation between the in-sample rankings and the out-of-sample rankings.

IV. Morningstar Rating Results

IV.A Dummy Variable Regression Analysis

Tables 4 through 7 presents the dummy variable regression analysis for the Sharpe ratio, mean monthly excess returns, Jensen and 4-index alphas respectively. Each table first presents the one-year samples, and then presents the three-year and five-year samples. Note that all the regressions were tested for heteroscedasticity using the White (1980) test. None of regression residuals exhibited evidence of heteroscedasticity at the 5 percent level.

IV.A1 The Sharpe Ratio

The Sharpe ratio results (table 4) show several interesting findings. First, the g_0 coefficients, the constants in the dummy variable regressions, are quite different from sample to sample. The 1992 constant is close to zero and insignificant, the 1994 constant is well below zero and significant, and the 1993 and 1995-1997 constants are all positive and significant. These results indicate that the reference group (the 5-star funds) perform quite differently in different years. The up and down performance of the 5-star Sharpe ratios is consistent with the performance of the Standard and Poor's 500 (S&P500) index's mean excess monthly returns. For comparison, table 8 presents the mean monthly excess return of the S&P500 for the different sample periods.

Second, the results show that the 4-star and 3-star funds do not diverge from the 5-star funds in terms of out-sample performance. None of the 24 coefficients (g_1 and g_2 for the 12 samples) are significant—indicating that for most samples there is not a significant difference in out-of-sample performance of median-rated funds and the top-rated funds.

Third, there is some evidence that the Morningstar's ratings do seem to predict the low performing funds. The g_3 and g_4 coefficients are generally negative and significant (12 of the 24 g_3 and g_4 coefficients), indicating that the performance of 1- and 2-star funds is significantly worse than that of the 5-star funds.

Fourth, the R^2 and F-statistic values for the samples differ dramatically. For example, the 1992-1year sample has an R^2 of 0.02 while the 1997-1year sample has an R^2 of 0.22.

IV.A2 Mean Monthly Returns

The results pertaining to the out-of-sample mean monthly returns are very similar to those pertaining to the Sharpe ratios. There is little, if any, significant difference in the out-of-sample returns between the 5-star funds and the 4- and 3-star funds. There is however, again, limited evidence that low rated funds do significantly worse than the 5-star rated funds (8 of the 24 g_3 and g_4 are negative and significant). Lastly, there are wide swings in the values of the constants, g_0, and the R-square values

IV.A3 The Modified Jensen Alpha and 4-Index Alpha

As with the Sharpe ratio and mean monthly excess returns, the modified Jensen and 4-index alphas continue to show the same patterns: little if any significant difference between the 5-star, 4-star and 3-star rated funds (with the 1993-5year sample providing the only evidence of significance), some evidence of negative and significant differences between the low- rated funds and the 5-star funds, and wide swings in the constant and R-square values. In addition, the 1-index and 4-index models show that in most cases the 5-star funds have negative (and sometimes significant) alphas (the γ_0 coefficient).

IV. B. Pooled Dummy Variable Regression Analysis

Table 9 presents the pooled dummy variable regressions. The results illustrate the same pattern as seen in the individual regressions. The 3-star and 4-star funds are generally not significantly different from the 5-star funds; only in the 5-year pooled 4-index alpha regression are the 3-star funds significantly different from the 5-star funds (at the 5 percent level). Moreover, the 2-star and 1-star funds are generally significantly different from the 5-star funds. Indicating that the low rated funds continue to perform poorly out-of-sample.

IV.C. Spearman-Rho Rank Correlation Tests

Tables 10A-F display the Spearman-Rho Rank Correlation Tests. Each table shows the decile averages of the performance measures, the Spearman-rho rank correlations on the entire 10 deciles, and the Spearman-rho correlations for both the top-5 deciles and the bottom-5 deciles. The results show the same basic pattern found in the regression analysis: the low scores predict poor future performance and the high scores have at best only mixed ability to predict future performance. In examining the rank correlation coefficients on all 10 deciles, several of the performance measures are relatively well correlated with the in-sample Morningstar Scores. Indeed in 4 of the 6 samples for the Sharpe and 3 of 6 samples for the Jensen we cannot reject the null hypothesis of no correlation in the rankings at the 95 percent confidence levels. However, upon examination of the correlation coefficients of the top 5-deciles and bottom 5-deciles, we see that overall rank correlation results are largely based on the ability of the low-scores to predict poor future performance. In most cases, the correlation coefficients for the bottom 5-decile is much larger and closer than 1 than the top-5 decile. In fact, generally, the rank correlation coefficients for the top-5 deciles are actually negative, indicating that the high scores do not accurately predict future performance.

V. "Naïve" Predictor Results

The results so far indicate that Morningstar ratings do not predict superior fund performance but do have some predictive power for poor-performing funds. Could an investor do as well by simply choosing funds based on the "naïve" predictor of a fund's average monthly return over the past ten years prior to investment? We examine that question in this section.[25] (The reader should note that tables 4 through 7 and 9 through 10F from the earlier sections are directly comparable to tables 11 through 16F in this section, the only difference being that the former set of tables uses Morningstar star rankings or scores as performance predictors, whereas the latter set for this section uses the "naïve" predictor of historical average monthly returns.)

V.A Dummy Variable Regression Analysis

Tables 11 through 14 presents the dummy variable regression analysis for the Sharpe ratio, mean monthly excess returns, Jensen and 4-index alphas respectively, using the "naïve" predictor. For both the individual and pooled dummy variable regressions, recall that we form five groups of

[25] We thank Stephen Brown for suggesting the examination in this section.

funds based on their "naïve" predictors, where each group contains the same number of funds as those in the corresponding Morningstar star group and where group 5 contains the funds with the highest "naïve" predictors, group 4 the next highest, etc. Each table first presents the one-year samples, and then presents the three-year and five-year samples. As in tables 4 through 7, all the regressions were tested for heteroscedasticity using the White (1980) test, and none of regression residuals exhibited evidence of heteroscedasticity at the 5 percent level.

V.A1 The Sharpe Ratio

Table 11 shows the dummy variable regression results for the load-adjusted Sharpe ratio. Comparing table 11 with table 4, we see that the signs and significance levels are quite similar. In every subsample, the reference group (group 5) intercepts have the same signs as those for the 5-star reference group in table 4 (except for the 1-year 1992 sample, which is not siginificant at the 10% level in either case), and the significance levels are the same in every case. Only five of the 24 coefficients for groups 4 and 3 are significant across subsamples using the "naïve" predictor; none are using the Morningstar star groups. Among the top three groups, neither predictor is able to differentiate future performance based on Sharpe ratios. Looking at the coefficients for the lowest two groups, we see that there are 8 significant and negative coefficients for the "naïve" predictor groups compared to 12 for the Morningstar star groups, indicating that Morningstar groups may do a slightly better job at predicting poor-performing funds based on Sharpe ratios.

V.A2 Mean Monthly Returns, The Modified Jensen Alpha and 4-Index Alpha

Table 12 shows the dummy variable regression using the out-of-sample mean monthly returns. As with the Sharpe ratio we see that in general, the naïve predictor does a reasonable job of predicting low-performing funds and yet has little predictive power in picking the highest performing funds. Indeed there are several instances (the 1992-1year and 1997-1year samples) were the 4 and 3 group funds performing significantly better than than the group 5 funds. Tables 13 and 14 provide the results using the Jensen and 4-index alpha measures. Again, they show the same general patterns as seen in the Sharpe and mean returns.

V. B. Pooled Dummy Variable Regression Analysis and Spearman-Rho Rank Correlation Tests

The pooled dummy variable regression analysis and Spearman-Rho Rank Correlation tests are shown in Tables 15 and 16A-F. The pooled dummy variables illustrate much the same

pattern as seen in the Morningstar Star pooled regression analysis (Table 9): The group 4 and group 3 funds do perform significantly worse than the group 5 funds, and yet the group 1 and 2 funds do perform significantly worse than the top rated funds.

The Spearman-Rho tests show, in general, lower rank correlation test than when using the Morningstar Scores. Upon examination of the 10 decile-rank correlation coefficients, none of the Sharpe measures are significant, and two are negative (although insignificant). This is in comparison to 4 significantly positive measures in the Morningstar Score analysis. However, the rank correlation coefficients for the bottom 5 deciles tend to be more positive than the top-five deciles, indicating again that the naïve predictor is able to predict the low performing funds and not the high performing funds.

VI. Conclusions

This paper has investigated the degree to which the well-known Morningstar 5-star rating system is a predictor of out-of-sample mutual fund performance. This is an important issue because several past studies have shown that highly ranked funds attract the greatest investor cash inflow. We have used a data set based on growth mutual funds that is free from survivorship bias, adjusted for load fees, and which allows us to examine the predictive qualities of the rating system over different time horizons, periods, and with different out-of-sample performance metrics. We have also compared these results with those from using a "naïve" predictor of ten-year historical monthly average returns. Moreover, we examined the data with parametric and non-parametric methods.

The results indicate several interesting findings. First, Morningstar is able to "predict" low-performing funds. Generally speaking, funds with less than 3 stars generally have much worse performance than other groups of funds. Second, there is only weak statistical evidence that the 5-star (highest rated) funds out-perform the 4- and 3-star funds (next-to-highest and median-rated funds). These results suggest that investors should be very cautious about associating a highly rated fund with having superior future performance. Also, given that previous studies have shown that high-rated funds attract the bulk of investor cash inflows, our results suggest that those cash inflows are not necessarily justified by subsequent superior performance.

These results are broadly consistent with much of the mutual fund performance persistence literature. Many authors have shown that, while it is relatively easy to predict poor performance, it is much more difficult to predict superior performance.

Our results also show that using average historical monthly return as a "naïve" performance predictor does just as well as using Morningstar ratings to predict performance for future "winners" and that perhaps Morningstar ratings are slightly better for identifying future "turkeys."

Finally, it should be noted that these results do not refute the Morningstar rating system. Indeed, in almost all their publications, Morningstar states that the star ratings are not predictors of future performance, but rather "achievement" marks. However, the fact is that that many investors and mutual funds do see the ratings as a sign of future performance. Studies show that high Morningstar ratings are strongly related to large capital inflows and are well-used in marketing mutual funds to the public. This research has provided an answer to an important question that investors should ask: Do the star ratings actually predict out-of-sample performance?

References

Beck, Nathaniel and Jonathan N. Katz, 1995, "What to do (and not to do) with Time-Series Cross-Section Data." *American Political Science Review*, 89, no.3, September, 634-647.

Blume, Marshall, 1998, "An Anatomy of Morningstar Ratings." *Financial Analysts Journal.*

Brown, Stephen J. and William Goetzmann, 1997, "Mutual Fund Styles", *Journal of Financial Economics*, 43, 373-399.

Brown, Stephen J., 1999, "Mutual Fund Styles" Conference Proceedings from Computational Finance Conference, New York University, January 6-8.

Stephen J. Brown, William M. Goetzmann, Roger G. Ibbotson and Stephen Ross, 1992, "Survivorship Bias in Performance Studies." *Review of Financial Studies*, 5, 553-580.

Carhart, Mark, M, 1997, "On Persistence in Mutual Fund Performance." *Journal of Finance.*

Damato, Karen, 1996, "Morningstar Edges Toward One-Year Ratings." Wall Street Journal (April 5):C1.

Elton, Edwin J., Martin J. Gruber and Christopher R. Blake, 1996a, "The Persistence of Risk-Adjusted Mutual Fund Performance." *Journal of Business*, 69, 2, 133-157.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 1996b, "Survivorship Bias and Mutual Fund Performance." *Review of Financial Studies*, 9, 1097-1120.

Elton, Edwin J., Martin J. Gruber, Sanjiv Das and Matthew Hlavka, 1993, "Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios." *Review of Financial Studies*, 6, 1-22.

Goetzmann, William N. and Roger G. Ibbotson, 1994, "Do Winners Repeat?" *Journal of Portfolio Management*, Winter, 9-18.

Goetzmann, William N. and N. Peles, 1994, "Cognitive Dissonance and Mutual Fund Investors." Working Paper, Columbia University.

Gruber, Martin, J., 1996, "Another Puzzle: The Growth in Actively Managed Mutual Funds." *Journal of Finance*, 51, 783-810.

Hendricks, Darryll, Jayendu Patel and Richard Zeckhauser, 1993, "Hot Hands in Mutual Funds: Short-run Persistence of relative performance, 1974-1988." *Journal of Finance*, 48, 93-130.

Khorana, Ajay and Edward Nelling, 1998, "The Determinants and Predictive Ability of Mutual Fund Ratings." *The Journal of Investing*, November, 61-66.

Jaffe, Charles1995, "Rating the Raters: Flaws Found in Each Service." Boston Globe, August 27th, 1995, p. 78.

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

Malkiel, Burton G., 1995, "Returns from Investing in Equity Mutual Funds 1971 to 1991." *Journal of Finance*, 51, 783-810.

Morey, Matthew R. and Richard C. Morey, 1999, "A Mutual Performance Appraisals: A Multi-Horizon Perspective with Endogenous Benchmarking." *Omega: The International Journal of Management Science* (forthcoming).

Morningstar Principia Manual (1998), Morningstar Publications, Chicago, Il.

Rea, John D. and Brian K. Reid, 1998, "Trends In The Ownership Cost of Equity Mutual Funds." *Investment Company Institute Perspective*, vol. 4, no. 3 (November), 1-15.

Sharpe, William, 1966, "Mutual Fund Performance." *Journal of Business*, vol.39, no.1, (January), 119-138.

Sharpe, William, 1998, "Morningstar Performance Measures", *Financial Analysts Journal*, July/August.

Sirri, Erik R. and Peter Tufano, 1992, "The Demand for Mutual Fund Services by Individual Investors", Working Paper, Harvard Business School.

White, Halbert, 1980, "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity", *Econometrica*, 48, 817-838.

Table 1: Understanding Morningstar Risk

Month	Fund Return(%)	T-Bill Return	Underperformance
1	2.0	0.5	NA
2	-1.5	0.5	2.0
3	3.2	0.5	NA
4	1.2	0.4	NA
5	-4.0	0.6	4.6
6	2.1	0.5	NA
7	0.7	0.5	NA
8	2.3	0.5	NA
9	-1.7	0.5	2.2
10	2.4	0.4	NA
11	1.2	0.6	NA
12	-3.1	0.5	3.4
Total Underperformance			**13.2**

$$\frac{\text{Total Underperformance}}{\text{Total Number of Months}} = \frac{13.2}{12} = 1.10$$ is the average monthly underperfomance

$$\frac{\text{Average Monthly Underperformance}}{\text{Average monthly Underperformance of investment category}} = \text{1-year Morningstar Risk}$$

Table 2: The Distribution, Average Star Ratings and Load Information for the Various Samples

January 1992
Total funds: 135
5-star funds: 11
4-star funds: 41
3-star funds: 69
2-star funds: 13
1-star funds: 1
Avg. Star Rating: 3.35
Std. Dev. of Star Rating: 0.80
Load-Funds:83 (61.48% of the sample)
Front-Load Funds: 77
Avg. Annual Front-Load (of the 77): 5.46
Deferred-Load Funds: 6
Avg. Annual Def. Load (of the 6): 4.33
No-Load Funds: 52

January 1993
Total funds: 134
5-star funds: 11
4-star funds: 38
3-star funds: 68
2-star funds: 16
1-star funds: 1
Avg. Star Rating: 3.31
Std. Dev. of Star Rating: 0.82
Load Funds: 81 (60.44% of the sample)
Front-Load Funds: 74
Avg. Annual Front-Load (of the 74): 5.46
Deferred-Load Funds: 7
Avg. Annual Def. Load (of the 7): 4.43
No-Load Funds: 53

January 1994
Total funds: 141
5-star funds: 12
4-star funds: 28
3-star funds: 73
2-star funds: 26
1-star funds: 2
Avg. Star Rating: 3.16
Std. Dev. of Star Rating: 0.87
Load Funds: 79 (56.03% of the sample)
Front-Load Funds: 71
Avg. Annual Front-Load (of the 71): 5.19
Deferred-Load Funds: 8
Avg. Annual Def. Load (of the 8): 4.50
No-Load Funds: 62

January 1995
Total funds: 158
5-star funds: 7
4-star funds: 42
3-star funds: 67
2-star funds: 35
1-star funds: 7
Avg. Star Rating: 3.04
Std. Dev. of Star Rating: 0.92
Load Funds: 85 (53.80% of the sample)
Front-Load Funds: 75
Avg. Annual Front-Load (of the 75): 5.06
Deferred-Load Funds: 10
Avg. Annual Def. Load (of the 10): 4.00
No-Load Funds: 52

January 1996
Total funds: 170
5-star funds: 12
4-star funds: 49
3-star funds: 70
2-star funds: 35
1-star funds: 4
Avg. Star Rating: 3.18
Std. Dev. of Star Rating: 0.92
Load Funds: 92 (54.11% of the sample)
Front-Load Funds: 81
Avg. Annual Front-Load (of the 81): 5.14
Deferred-Load Funds: 11
Avg. Annual Def. Load (of the 11): 3.91
No-Load Funds: 78

January 1997
Total funds: 184
5-star funds: 11
4-star funds: 50
3-star funds: 65
2-star funds: 54
1-star funds: 4
Avg. Star Rating: 3.06
Std. Dev. of Star Rating: 0.95
Load Funds: 99 (53.80% of the sample)
Front-Load Funds: 84
Avg. Annual Front-Load (of the 84): 5.11
Deferred-Load Funds: 15
Avg. Annual Def. Load (of the 15): 3.87
No-Load Funds: 86

Table 3: Summary of number of funds in each out-of-sample period

Sample Name	Date of Morningstar Rating	Out of Sample Period (date)	Total number of funds in the sample	Number of merger funds	Number of liquidations	Number of funds* that change their Morningstar objective by the end of the out-of-sample period
92-1year	Beginning of year 1992	1-year (1992)	135	6	0	3
92-3year	Beginning of year 1992	3-year (1992-1994)	135	13	0	7
92-5year	Beginning of year 1992	5-year (1992-1996)	135	16	2	13
93-1year	Beginning of year 1993	1-year (1993)	134	4	0	3
93-3year	Beginning of year 1993	3-year (1993-1995)	134	9	0	8
93-5year	Beginning of year 1993	5-year (1993-1997)	134	14	2	12
94-1year	Beginning of year 1994	1-year (1994)	141	4	0	4
94-3year	Beginning of year 1994	3-year (1994-1996)	141	8	3	8
95-1year	Beginning of year 1995	1-year (1995)	158	2	1	7
95-3year	Beginning of year 1995	3-year (1995-1997)	158	10	4	9
96-1year	Beginning of year 1996	1-year (1996)	170	6	3	5
97-1year	Beginning of year 1997	1-year (1997)	184	4	0	3

* funds that merged or were liquidated by the end of the sample period were not counted.

Table 4: Dummy Variable Regressions Using Morningstar Stars

Out-of-Sample Performance Measure: The Load-Adjusted Sharpe Ratio

Sample	γ_0 (constant)	γ_1 (4-star)	γ_2 (3-star)	γ_3 (2-star)	γ_4 (1-star)	R^2	F-stat
One-Year							
1992-1year	0.03	-0.02	-0.02	0.05	0.02	0.02	0.48
	(0.38)	(0.22)	(0.24)	(0.58)	(0.07)		
1993-1year	0.25**	-0.08	-0.09	-0.20**	-0.09	0.03	1.11
	(3.29)	(0.93)	(1.14)	(2.02)	(0.35)		
1994-1year	-0.20**	-0.01	-0.02	-0.12*	-0.36**	0.08	3.01**
	(3.70)	(0.15)	(0.34)	(1.76)	(2.50)		
1995-1year	0.75**	0.06	-0.03	-0.11	-0.45**	0.11	4.87**
	(6.56)	(0.52)	(0.23)	(0.84)	(2.76)		
1996-1year	0.25**	0.05	0.01	-0.03	-0.16*	0.06	2.64**
	(5.50)	(0.99)	(0.21)	(0.62)	(1.77)		
1997-1year	0.33**	0.01	-0.03	-0.09*	-0.49**	0.22	12.35**
	(7.42)	(0.11)	(0.68)	(1.91)	(5.68)		
Three-year							
1992-3year	0.01	0.03	0.02	0.01	0.03	0.01	0.08
	(0.30)	(0.54)	(0.37)	(0.20)	(0.18)		
1993-3year	0.26**	-0.05	-0.06	-0.15**	-0.10	0.07	2.50**
	(6.38)	(0.96)	(1.39)	(2.86)	(0.70)		
1994-3year	0.26**	0.01	-0.03	-0.09**	-0.38**	0.15	5.96**
	(7.41)	(0.01)	(0.86)	(2.01)	(4.09)		
1995-3year	0.42**	0.01	-0.03	-0.05	-0.23**	0.11	4.61**
	(7.88)	(0.18)	(0.53)	(0.88)	(3.02)		
Five-year							
1992-5year	0.23**	-0.01	-0.02	-0.04	-0.12	0.01	0.41
	(7.01)	(0.37)	(0.57)	(0.84)	(1.05)		
1993-5year	0.31**	-0.05	-0.06	-0.12**	-0.14	0.06	1.98*
	(8.89)	(1.17)	(1.60)	(2.59)	(1.16)		

T-statistics are in parenthesis. * indicates significance at the 10 percent level.
**indicates significance at the 5 percent level.

Table 5: Dummy Variable Regressions using Morningstar Stars

Out-of-Sample Performance Measure: The Load-Adjusted Excess Mean Monthly Returns.

Sample	γ_0 (constant)	γ_1 (4-star)	γ_2 (3-star)	γ_3 (2-star)	γ_4 (1-star)	R^2	F-stat
One-Year							
1992-1year	0.08	-0.07	-0.07	0.17	0.07	0.04	1.00
	(0.46)	(0.38)	(0.36)	(0.73)	(1.08)		
1993-1year	0.54**	-0.09	-0.15	-0.35	-0.12	0.02	0.70
	(2.92)	(0.45)	(0.75)	(1.49)	(0.19)		
1994-1year	-0.61**	0.01	-0.11	-0.39**	-1.26**	0.11	4.31**
	(3.89)	(0.05)	(0.67)	(2.05)	(3.06)		
1995-1year	1.82**	-0.10	-0.28	-0.34	-1.32**	0.16	7.26**
	(8.31)	(0.43)	(1.21)	(1.42)	(4.29)		
1996-1year	0.83	0.12	0.03	-0.12	-0.35	0.04	1.75
	(5.80)	(0.77)	(0.19)	(0.70)	(1.22)		
1997-1year	1.41**	-0.01	-0.08	-0.27	-2.03**	0.19	10.72**
	(7.57)	(0.03)	(0.37)	(1.30)	(5.60)		
Three-year							
1992-3year	0.02	0.06	0.05	0.08	0.04	0.01	0.10
	(0.21)	(0.46)	(0.45)	(0.52)	(0.42)		
1993-3year	0.67**	-0.06	-0.12	-0.33**	-0.17	0.05	1.86
	(6.01)	(0.48)	(1.00)	(2.29)	(0.45)		
1994-3year	0.72**	0.05	-0.02	-0.14	-0.99**	0.12	4.66**
	(6.99)	(0.39)	(0.18)	(1.16)	(3.65)		
1995-3year	1.44**	0.01	-0.12	-0.12	-0.73**	0.11	4.60**
	(8.97)	(0.01)	(0.70)	(-0.71)	(3.21)		
Five-year							
1992-5year	0.62**	0.01	-0.01	0.02	-0.02	0.01	0.03
	(7.20)	(0.12)	(0.05)	(0.16)	(0.07)		
1993-5year	0.95**	-0.08	-0.12	-0.26*	-0.34	0.04	1.23
	(9.25)	(0.71)	(1.11)	(1.94)	(0.96)		

T-statistics are in parenthesis. * indicates significance at the 10 percent level.
**indicates significance at the 5 percent level.

Table 6: Dummy Variable Regressions using Morningstar Stars

Out of Sample Performance Measure: The Non-load adjusted Jensen index model

Sample	$\gamma_{0\ (constant)}$	$\gamma_{1\ (4\text{-star})}$	$\gamma_{2\ (3\text{-star})}$	$\gamma_{3\ (2\text{-star})}$	$\gamma_{4\ (1\text{-star})}$	Load	R^2	F-stat
One-Year								
1992-1year	-0.10	0.09	0.22	0.01	-0.20	-0.11	0.04	1.00
	(0.60)	(0.51)	(1.32)	(0.08)	(0.37)	(1.21)		
1993-1year	0.25	-0.17	-0.15	-0.37	-0.19	0.04	0.02	0.58
	(1.37)	(0.87)	(0.83)	(1.63)	(0.32)	(0.43)		
1994-1year	-0.22	-0.03	-0.05	-0.30*	-1.28**	-0.05	0.13	4.01**
	(1.52)	(0.18)	(0.30)	(1.82)	(3.55)			
1995-1year	-0.16	-0.01	-0.14	-0.31	-1.12**	-0.09	0.17	6.23**
	(0.75)	(0.02)	(0.65)	(1.40)	(3.89)	(0.98)		
1996-1year	-0.34**	0.26**	0.16	0.01	0.05	-0.16**	0.10	3.54**
	(2.74)	(1.98)	(1.25)	(0.07)	(0.22)	(2.44)		
1997-1year	-0.55**	0.17	0.18	-0.04	-0.92**	-0.10	0.09	3.35**
	(2.87)	(0.83)	(0.89)	(0.20)	(2.53)	(1.00)		
Three-year								
1992-3year	-0.08	0.04	0.08	0.05	-0.08	-0.08	0.01	0.35
	(0.71)	(0.35)	(0.66)	(0.35)	(0.20)	(1.12)		
1993-3year	-0.10	-0.10	-0.17	-0.42**	-0.21	0.09	0.08	2.33**
	(0.84)	(0.76)	(1.47)	(2.91)	(0.54)	(1.43)		
1994-3year	-0.24**	0.03	-0.05	-0.21*	-0.90**	-0.07	0.13	4.10**
	(2.29)	(0.24)	(0.46)	(1.70)	(3.39)	(1.15)		
1995-3year	-0.34**	0.12	0.04	-0.06	-0.47**	-0.08	0.12	4.00**
	(2.30)	(0.78)	(0.27)	(0.37)	(2.36)	(1.31)		
Five-year								
1992-5year	-0.11	-0.02	-0.03	-0.08	-0.11	-0.05	0.01	0.34
	(1.11)	(0.20)	(0.32)	(0.70)	(0.36)	(1.01)		
1993-5year	-0.13	-0.08	-0.18*	-0.32**	-0.18	0.05	0.07	2.00*
	(1.34)	(0.76)	(1.83)	(2.66)	(0.56)	(0.89)		

T-statistics are in parenthesis. * indicates significance at the 10 percent level.
**indicates significance at the 5 percent level.

Table 7: Dummy Variable Regressions using Morningstar Stars

Out-of-Sample Performance Measure: The non-load adjusted 4-index model

Sample	γ_0 (constant)	γ_1 (4-star)	γ_2 (3-star)	γ_3 (2-star)	γ_4 (1-star)	Load	R^2	F-stat
One-Year								
1992-1year	-0.06	0.08	0.11	-0.17	-0.15	-0.03	0.05	1.25
	(0.49)	(0.63)	(0.88)	(1.02)	(0.37)	(0.43)		
1993-1year	0.20	-0.19	-0.13	-0.30	-0.19	0.07	0.02	0.63
	(1.24)	(1.10)	(0.79)	(1.49)	(0.35)	(0.80)		
1994-1year	-0.10	-0.05	-0.08	-0.32**	-1.17**	-0.06	0.12	3.84**
	(0.79)	(0.31)	(0.56)	(2.04)	(3.39)	(0.77)		
1995-1year	0.02	0.03	-0.13	-0.17	-1.03**	-0.05	0.17	6.18**
	(0.11)	(0.13)	(0.70)	(0.86)	(4.02)	(0.65)		
1996-1year	0.01	0.20	0.08	0.01	0.13	-0.14**	0.06	2.19*
	(0.04)	(1.45)	(0.59)	(0.08)	(0.54)	(2.13)		
1997-1year	-0.06	0.04	0.15	0.08	-0.88**	-0.04	0.09	3.37**
	(0.41)	(0.26)	(0.92)	(0.51)	(3.04)	(0.63)		
Three-year								
1992-3year	-0.07	0.04	0.01	0.05	-0.08	-0.03	0.01	0.31
	(0.74)	(0.44)	(0.12)	(0.45)	(0.26)	(0.50)		
1993-3year	-0.04	-0.09	-0.14	-0.36**	-0.19	0.11*	0.08	2.14*
	(0.38)	(0.74)	(1.30)	(2.70)	(0.52)	(1.72)		
1994-3year	-0.05	-0.01	-0.06	-0.19*	-0.91**	-0.05	0.13	4.00**
	(0.47)	(0.10)	(0.59)	(1.67)	(3.70)	(0.96)		
1995-3year	-0.06	0.09	-0.02	-0.04	-0.44**	-0.03	0.11	3.89**
	(0.46)	(0.68)	(0.17)	(0.29)	(2.70)	(0.68)		
Five-year								
1992-5year	-0.01	-0.01	-0.07	-0.10	-0.09	-0.01	0.02	0.51
	(0.09)	(0.13)	(0.83)	(1.05)	(0.36)	(0.21)		
1993-5year	0.01	-0.07	-0.16*	-0.27**	-0.16	0.06	0.07	2.05*
	(0.10)	(0.85)	(1.92)	(2.67)	(0.60)	(1.28)		

T-statistics are in parenthesis. * indicates significance at the 10 percent level.
**indicates significance at the 5 percent level.

Table 8: Mean Monthly Returns of Excess Standard and Poor 500 Index Returns

Sample	Mean Monthly Return
1992	0.35
1993	0.57
1994	-0.17
1995	2.24
1996	1.37
1997	2.10
1992-1994	0.25
1993-1995	0.88
1994-1996	1.15
1995-1997	1.90
1992-1996	0.87
1993-1997	1.22

Excess Returns calculated by subtracting the one-month T-Bill rate from the Monthly return.

Table 9: Pooled Dummy variable regressions: Using Morningstar Stars.

Out of Sample Performance Measure

Sharpe Ratios (load-adjusted returns)

						Coefficients for Dummy Variables							
Sample	g_0	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	**N**	**R^2**	**F-stat**
1-year ++	0.04	0.01	-0.03	-0.09**	-0.34**	0.14**	-0.25**	0.72**	0.25**	0.29**	922	0.65	185.83**
pooled	(1.46)	(0.12)	(1.12)	(3.13)	(3.54)	(4.75)	(9.87)	(23.12)	(10.76)	(12.59)			
3-year ++	0.05**	-0.01	-0.03	-0.07**	-0.23**	0.17**	0.20**	0.38**	NA	NA	568	0.50	80.39**
pooled	(2.61)	(0.18)	(1.43)	(3.04)	(2.78)	(10.29)	(12.88)	(22.62)					
5-year	0.25**	-0.03	-0.04	-0.08**	-0.13	0.04**	NA	NA	NA	NA	269	0.06	3.42**
pooled	(10.05)	(1.11)	(1.56)	(2.51)	(1.57)	(3.04)							

Mean Monthly Returns (load-adjusted returns)

						Coefficients for Dummy Variables							
Sample	g_0	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	**N**	**R^2**	**F-stat**
1-year ++	0.13*	-0.02	-0.10	-0.23**	-1.04**	0.35**	-0.77**	1.55**	0.83**	1.26**	922	0.62	186.48**
pooled	(1.66)	(0.20)	(1.40)	(2.83)	(3.82)	(4.75)	(11.27)	(22.17)	(12.89)	(1.26)			
3-year ++	0.11**	0.01	-0.05	-0.12**	-0.62**	0.48**	0.62**	1.29**	NA	NA	568	0.60	120.90**
pooled	(2.06)	(0.24)	(0.92)	(1.96)	(2.78)	(10.72)	(14.13)	(27.20)					
5-year	0.67**	-0.04	-0.06	-0.13	-0.18	-0.21**	NA	NA	NA	NA	269	0.11	6.67**
pooled	(9.78)	(0.04)	(0.87)	(1.45)	(0.78)	5.56							

Jensen Alpha (non-load adjusted returns)

						Coefficients for Dummy Variables								
Sample	g_0	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	**Load**	**N**	**R^2**	**F-stat**
1-year ++	-0.01	0.07	0.05	-0.16*	-0.74**	0.14**	0.30**	-0.31**	-0.22**	-0.45**	-0.08**	922	0.19	21.25**
pooled	(0.08)	(0.84)	(0.63)	(1.85)	(3.84)	(2.05)	(5.00)	(4.94)	(3.92)	(6.98)	(2.08)			
3-year	-0.01	0.01	-0.03	-0.16**	-0.53**	-0.13**	-0.26**	-0.26**	NA	NA	-0.04	568	0.15	12.52**
pooled	(0.21)	(0.21)	(0.52)	(2.39)	(4.29)	(3.01)	(6.02)	(6.13)			(1.26)			
5-year	-0.08	-0.05	-0.11	-0.20**	-0.12*	-0.08**	NA	NA	NA	NA	0.01	269	0.05	2.23**
pooled	(1.22)	(0.74)	(1.64)	(2.15)	(1.93)	(2.31)					(0.15)			

4-index Alpha (non-load adjusted returns)

						Coefficients for Dummy Variables								
Sample	g_0	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	**Load**	**N**	**R^2**	**F-stat**
1-year ++	0.02	0.02	0.01	-0.11	-0.68**	0.11*	-0.24**	-0.09*	0.06	0.02	-0.04	922	0.10	10.54**
pooled	(0.26)	(0.37)	(0.10)	(1.64)	(3.56)	(1.90)	(4.68)	(1.70)	(1.25)	(0.49)	(1.27)			
3-year ++	-0.02	0.01	-0.06	-0.15**	-0.50**	-0.04	-0.07*	0.01	NA	NA	-0.01	568	0.07	5.29**
pooled	(0.42)	(0.04)	(1.28)	(2.69)	(3.02)	(1.17)	(1.82)	(0.39)			(0.24)			
5-year ++	0.01	-0.04	-0.11**	-0.19**	-0.11**	-0.03	NA	NA	NA	NA	0.02	269	0.04	1.98*
pooled	(0.37)	(1.00)	(2.69)	(2.26)	(2.99)	(0.94)					(0.69)			

T-statistics are in parenthesis

*indicates significance at the 10 percent level. NA indicates there was no sample to pool.

** indicates significance at the 5 percent level.

++ Indicates that we use the White Heteroskedastic Consistent Standard Errors.

Table 10a: Average Performance Values by Decile and Spearman-Rho Rank Correlation Tests using Morningstar Scores

Sample: 94-1 year		Out-of-Sample-Performance Measures				
Decile		Morningstar Score	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top	1	0.83	-0.60	-0.19	-0.25	-0.16
	2	0.47	-0.82	-0.26	-0.31	-0.21
	3	0.28	-0.64	-0.21	-0.24	-0.15
	4	0.20	-0.75	-0.21	-0.32	-0.27
	5	0.11	-0.59	-0.19	-0.21	-0.13
	6	0.04	-0.82	-0.30	-0.43	-0.31
	7	0.00	-0.86	-0.27	-0.40	-0.33
	8	-0.13	-0.75	-0.28	-0.31	-0.29
	9	-0.25	-0.90	-0.29	-0.49	-0.43
	10	-0.55	-0.89	-0.32	-0.52	-0.38
Rank Correlation of Morningstar Score to Out-of-Sample Performance:			.636	.588	.745	.794
Two-tailed p-value:			(.048)	(.074)	(.013)	(.006)
Rank Correlations of Top-5 Deciles:			-.300	-.100	-.300	-.300
Two-tailed p-value:			(.624)	(.873)	(.624)	(.624)
Rank Correlations of Bottom-5 Deciles:			-.100	.600	.600	.600
Two-tailed p-value:			(.873)	(.285)	(.285)	(.285)

Table 10b: Average Performance Values by Decile and Spearman-Rho Rank Correlation Tests using Morningstar Scores

Sample: 95-1 year		Out-of-Sample-Performance Measures				
Decile		Morningstar Score	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top	1	1.16	1.54	0.67	-0.46	-0.18
	2	0.60	1.56	0.71	-0.13	0.12
	3	0.32	1.96	1.00	0.05	0.20
	4	0.12	1.59	0.74	-0.42	-0.20
	5	-0.06	1.52	0.71	-0.38	-0.15
	6	-0.19	1.68	0.82	-0.15	0.06
	7	-0.34	1.63	0.69	-0.49	-0.23
	8	-0.55	1.37	0.64	-0.41	-0.18
	9	-0.80	1.45	0.61	-0.64	-0.28
	10	-1.48	1.03	0.52	-0.89	-0.59
Rank Correlation of Morningstar Score to Out-of-Sample Performance:			.539	.624	.612	.648
Two tailed p-value:			(.108)	(.054)	(.060)	(.043)
Rank Correlations of Top-5 Deciles:			.100	-.300	-.200	-.100
Two-tailed p-value:			(.873)	(.624)	(.747)	(.873)
Rank Correlations of Bottom-5 Deciles:			.900	1.000	.900	.900
Two-tailed p-value:			(.037)	(.000)	(.037)	(.037)

Table 10C: Average Performance Values by Decile and Spearman-Rho Rank Correlation Tests using Morningstar Scores

Sample: 96-1 year		**Out-of-Sample Performance Measures**				
Decile		Morningstar Score	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top	1	0.88	0.85	0.26	-0.37	-0.06
	2	0.51	0.96	0.29	-0.13	0.18
	3	0.35	0.97	0.31	-0.17	0.10
	4	0.21	0.93	0.27	-0.09	0.17
	5	0.10	1.08	0.33	-0.07	0.18
	6	0.01	0.84	0.26	-0.28	0.03
	7	-0.10	0.72	0.21	-0.40	-0.15
	8	-0.20	0.66	0.19	-0.39	-0.10
	9	-0.37	0.72	0.20	-0.47	-0.05
	10	-0.78	0.72	0.21	-0.37	-0.05
Rank Correlation of Morningstar Score to Out-of-Sample Performance:			.733	.709	.515	.455
Two-tailed p-value:			(.016)	(.022)	(.128)	(.187)
Rank Correlations of Top-5 Deciles:			-.700	-.700	-.900	-.400
Two-tailed p-value:			(.188)	(.188)	(.037)	(.505)
Rank Correlations of Bottom-5 Deciles:			.500	.300	.300	.100
Two-tailed p-value:			(.391)	(.624)	(.624)	(.873)

Table 10D: Average Performance Value by Decile and Spearman-Rho Rank Correlation Tests using Morningstar Scores

Sample: 97-1 year		Out-of-Sample Performance Measures			
Decile	Morningstar Score	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top 1	0.86	1.43	0.34	-0.46	-0.04
2	0.54	1.47	0.36	-0.28	0.10
3	0.40	1.39	0.34	-0.38	-0.06
4	0.25	1.34	0.30	-0.51	-0.09
5	0.11	1.46	0.32	-0.38	0.12
6	0.00	1.34	0.31	-0.40	0.10
7	-0.14	1.28	0.29	-0.46	0.00
8	-0.31	1.21	0.25	-0.54	0.02
9	-0.44	1.23	0.26	-0.70	0.02
10	-0.81	0.53	0.10	-1.08	-0.35
Rank Correlation of Morningstar Score to Out-of-Sample Performance:		.891	.939	.648	.067
Two-tailed p-value:		(..001)	(.000)	(.029)	(.855)
Rank Correlations of Top-5 Deciles:		.200	.800	.000	-.100
Two-tailed p-value:		(.747)	(.104)	(1.000)	(.873)
Rank Correlations of Bottom-5 Deciles:		.900	.900	1.000	.600
Two-tailed p-value:		(.047)	(.047)	(0.00)	(.285)

Table 10E: Average Performance Values by Decile and Spearman-Rho Rank Correlation Tests using Morningstar Scores

Sample: 94-3 year		Out-of-Sample Performance Measures				
Decile		Morningstar Score	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top	1	0.83	0.76	0.26	-0.35	-0.14
	2	0.47	0.74	0.24	-0.26	-0.07
	3	0.28	0.75	0.24	-0.33	-0.08
	4	0.20	0.78	0.25	-0.33	-0.14
	5	0.11	0.74	0.26	-0.22	-0.07
	6	0.04	0.43	0.14	-0.51	-0.30
	7	0.00	0.60	0.19	-0.47	-0.28
	8	-0.13	0.63	0.22	-0.33	-0.15
	9	-0.25	0.59	0.19	-0.32	-0.11
	10	-0.55	0.56	0.18	-0.38	-0.20
Rank Correlation of Morningstar Score to Out-of-Sample Performance:			.648	.661	.164	.479
Two-tailed p-value:			(.043)	(.038)	(.651)	(.162)
Rank Correlations of Top-5 Deciles:			.000	.100	-.700	-.300
Two-tailed p-value:			(1.000)	(.873)	(.188)	(.624)
Rank Correlations of Bottom-5 Deciles:			.600	.600	-.700	-.700
Two-tailed p-value:			(.285)	(.285)	(.188)	(.188)

Table 10F: Average Performance Values by Decile and Spearman-Rho Rank Correlation Tests using Morningstar Scores

Sample: 95-3 year		Out-of-Sample Performance Measures				
Decile		Morningstar Score	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top	1	1.16	1.37	0.42	-0.34	-0.05
	2	0.60	1.27	0.39	-0.22	-0.03
	3	0.32	1.54	0.47	-0.18	0.09
	4	0.12	1.40	0.43	-0.30	-0.12
	5	-0.06	1.26	0.39	-0.38	-0.13
	6	-0.19	1.47	0.44	-0.23	-0.02
	7	-0.34	1.36	0.38	-0.48	-0.12
	8	-0.55	1.19	0.33	-0.44	-0.09
	9	-0.80	1.35	0.34	-0.55	-0.15
	10	-1.48	1.08	0.33	-0.57	-0.27
Rank Correlation of Morningstar Score to Out-of-Sample Performance:			.527	.745	.806	.648
Two-tailed p-value:			(.117)	(.013)	(.005)	(.043)
Rank Correlations of Top-5 Deciles:			.200	.200	.300	.600
Two-tailed p-value:			(.747)	(.747)	(.624)	(.285)
Rank Correlations of Bottom-5 Deciles:			.900	.900	.900	.900
Two-tailed p-value:			(.037)	(.037)	(.037)	(.037)

Table 11: Dummy Variable Regressions Using "Naïve" Predictor Groups

Out-of-Sample Performance Measure: The Load-Adjusted Sharpe Ratio

Sample	$\gamma_{0\ (constant)}$	$\gamma_{1(group\ 4)}$	$\gamma_{2\ (group\ 3)}$	$\gamma_{3\ (group\ 2)}$	$\gamma_{4\ (group\ 1)}$	R^2	F-stat
One-Year							
1992-1year	-0.09	0.13*	0.12*	0.09	-0.02	0.03	0.86
	(1.34)	(1.66)	(1.66)	(1.00)	(0.07)		
1993-1year	0.28**	-0.15*	-0.12	-0.19*	-0.12	0.03	0.99
	(3.69)	(1.71)	(1.45)	(1.87)	(.47)		
1994-1year	-0.24**	0.01	0.01	-0.02	-0.33**	0.04	1.56
	(4.27)	(0.10)	(0.14)	(0.36)	(2.21)		
1995-1year	0.69**	0.06	0.07	-0.04	-0.31*	0.07	2.86**
	(5.89)	(0.48)	(0.55)	(0.29)	(1.87)		
1996-1year	0.20**	0.05	0.07	0.07	-0.12	0.04	1.92
	(4.37)	(1.01)	(1.52)	(1.37)	(1.28)		
1997-1year	0.20**	0.09*	0.10*	0.10*	-0.15	0.07	3.14**
	(4.05)	(1.68)	(1.90)	(1.89)	(1.58)		
Three-year							
1992-3year	0.02	0.02	0.01	0.01	-0.14	0.01	0.40
	(0.38)	(0.49)	(0.32)	(0.19)	(1.00)		
1993-3year	0.24**	-0.02	-0.03	-0.15**	-0.07	0.08	2.97**
	(5.79)	(0.49)	(0.58)	(2.72)	(0.52)		
1994-3year	0.24**	-0.03	0.01	-0.03	-0.36**	0.12	4.51**
	(6.70)	(0.72)	(0.03)	(0.85)	(3.79)		
1995-3year	0.35**	0.04	0.06	0.05	-0.15*	0.08	3.51**
	(6.44)	(0.75)	(0.99)	(0.92)	(1.95)		
Five-year							
1992-5year	0.19**	0.02	0.03	-0.01	-0.13	0.03	0.90
	(5.83)	(0.62)	(0.87)	(0.14)	(1.12)		
1993-5year	0.28**	-0.01	-0.02	-0.11**	-0.11	0.07	2.42*
	(8.06)	(0.23)	(0.62)	(2.35)	(0.91)		

T-statistics are in parenthesis. * indicates significance at the 10 percent level.
**indicates significance at the 5 percent level.

Table 12: Dummy Variable Regressions Using "Naïve" Predictor Groups

Out-of-Sample Performance Measure: The Load-Adjusted Excess Mean Monthly Returns.

Sample	γ_{0} (constant)	γ_{1}(group4)	γ_{2} (group 3)	γ_{3} (group 2)	γ_{4} (group 1)	R^2	F-stat
One-Year							
1992-1year	-0.26	0.34*	0.34*	0.33	-0.38	0.04	1.31
	(1.54)	(1.79)	(1.87)	(1.45)	(.66)		
1993-1year	0.76**	-0.38*	-0.36*	-0.64**	-0.34	0.06	1.90
	(4.22)	(1.86)	(1.86)	(2.75)	(0.55)		
1994-1year	-0.76**	-0.02	0.05	0.03	-1.11**	0.06	2.14*
	(4.74)	(0.12)	(0.27)	(0.13)	(2.61)		
1995-1year	1.69**	-0.04	-0.04	-0.35	-1.06**	0.14	6.35**
	(7.65)	(0.16)	(0.18)	(1.46)	(3.39)		
1996-1year	0.70**	0.17	0.20	0.13	-0.29	0.03	1.22
	(4.84)	(1.03)	(1.27)	(0.77)	(-1.00)		
1997-1year	0.94**	0.46**	0.36*	0.30	-0.66*	0.07	3.46**
	(4.67)	(2.09)	(1.68)	(1.37)	(1.70)		
Three-year							
1992-3year	0.03	0.09	0.04	0.04	-0.50	0.02	0.77
	(0.27)	(0.72)	(0.33)	(0.25)	(1.33)		
1993-3year	0.71**	-0.11	-0.11	-0.55**	-0.21	0.16	6.08**
	(6.71)	(0.89)	(1.00)	(4.02)	(0.57)		
1994-3year	0.75**	-0.06	-0.02	-0.19	-1.02**	0.12	4.79**
	(7.27)	(0.45)	(0.17)	(1.51)	(3.76)		
1995-3year	1.26**	0.16	0.14	0.01	-0.65**	0.14	6.35**
	(7.98)	(0.93)	(0.82)	(0.06)	(2.92)		
Five-year							
1992-5year	0.62**	0.02	0.01	-0.06	-0.40	0.02	0.72
	(7.29)	(0.19)	(0.11)	(0.56)	(1.34)		
1993-5year	0.96**	-0.06	-0.10	-0.49**	-0.35	0.16	6.29**
	(10.09)	(0.56)	(1.00)	(3.94)	(1.06)		

T-statistics are in parenthesis. * indicates significance at the 10 percent level.
**indicates significance at the 5 percent level.

Table 13: Dummy Variable Regressions Using "Naïve" Predictor Groups

Out of Sample Performance Measure: The Non-Load-Adjusted Jensen Alpha

Sample	$\gamma_{0\ (constant)}$	$\gamma_{1(group\ 4)}$	$\gamma_{2\ (group\ 3)}$	$\gamma_{3\ (group\ 2)}$	$\gamma_{4\ (group\ 1)}$	Load	R^2	F-stat
One-Year								
1992-1year	-0.10	0.13	0.12	0.21	-0.15	-0.07	0.02	.42
	(0.59)	(0.72)	(0.70)	(0.96)	(0.27)	(0.77)		
1993-1year	0.43**	-0.39**	-0.34*	-0.49**	-0.36	0.03	0.04	1.09
	(2.41)	(2.03)	(1.86)	(2.21)	(0.60)	(0.29)		
1994-1year	-0.27*	-0.00	-0.02	-0.05	-1.22**	-0.07	0.09	2.59**
	(1.81)	(0.01)	(0.11)	(0.30)	(3.29)	(0.91)		
1995-1year	-.42*	0.07	0.23	0.07	-0.36	-0.14	0.06	2.09*
	(1.87)	(0.28)	(1.02)	(0.27)	(1.17)	(1.53)		
1996-1year	-0.35**	0.16	0.19	0.23	0.01	-0.18**	0.06	2.26*
	(2.75)	(1.18)	(1.44)	(1.61)	(0.03)	(2.77)		
1997-1year	-1.03**	0.51**	0.58**	0.74**	0.20	-0.11	0.08	3.21**
	(5.26)	(2.47)	(2.84)	(3.55)	(0.54)	(1.17)		
Three-year								
1992-3year	-.05	0.05	0.01	-0.01	-0.52	-0.06	0.03	0.80
	(0.39)	(0.41)	(0.02)	(0.07)	(1.44)	(.89)		
1993-3year	-0.14	-0.07	-0.05	-0.40**	-0.12	0.06	0.10	2.74**
	(1.25)	(0.60)	(0.46)	(2.82)	(0.32)	(0.86)		
1994-3year	-0.32**	-0.03	0.06	-0.01	-0.82**	-0.08	0.10	3.01**
	(2.88)	(0.27)	(0.56)	(0.08)	(3.02)	(1.36)		
1995-3year	-0.50**	0.12	0.23	0.28*	-0.05	-0.11*	0.07	2.33**
	(3.30)	(0.76)	(1.49)	(1.76)	(0.25)	(1.75)		
Five-year								
1992-5year	-0.20**	0.08	0.06	0.05	-0.25	-0.04	0.02	0.55
	(1.99)	(.78)	(0.62)	(0.37)	(0.81)	(0.81)		
1993-5year	-0.24**	-0.00	-0.01	-0.15	-0.03	0.01	0.03	0.66
	(2.44)	(.04)	(0.05)	(1.26)	(0.10)	(0.27)		

T-statistics are in parenthesis. * indicates significance at the 10 percent level.
**indicates significance at the 5 percent level.

Table 14: Dummy Variable Regressions Using "Naïve" Predictor Groups

Out-of-Sample Performance Measure: The Non-Load-Adjusted 4-Index Alpha

Sample	γ_0 (constant)	γ_1 (group 4)	γ_2 (group 3)	γ_3 (group 2)	γ_4 (group 1)	Load	R^2	F-stat
One-Year								
1992-1year	0.03	0.01	-0.07	0.06	-0.45	-0.02	0.02	0.64
	(0.22)	(.10)	(.55)	(.36)	(1.07)	(.32)		
1993-1year	0.43**	-0.46**	-0.35**	-0.58**	-0.41	0.07	0.08	2.16*
	(2.72)	(2.69)	(2.19)	(3.00)	(0.79)	(0.82)		
1994-1year	-0.13	-0.02	-0.09	-0.14	-1.14**	-0.09	0.09	2.54**
	(.88)	(.10)	(.62)	(.85)	(3.22)	(1.15)		
1995-1year	-0.02	-0.08	0.02	-0.13	-0.56*	-0.09	0.06	2.01*
	(0.11)	(0.37)	(0.08)	(0.62)	(2.06)	(1.17)		
1996-1year	0.07	0.11	0.03	0.01	-0.15	-0.16**	0.05	1.68
	(0.54)	(0.81)	(0.23)	(0.07)	(0.63)	(2.47)		
1997-1year	-0.12	0.22	0.08	0.12	-0.45	-0.03	0.04	1.66
	(0.73)	(1.31)	(0.46)	(0.72)	(1.53)	(0.39)		
Three-year								
1992-3year	0.02	-0.02	-0.09	-0.10	-0.62**	-0.03	0.05	1.41
	(0.16)	(0.19)	(1.03)	(0.90)	(2.12)	(0.56)		
1993-3year	-0.04	-0.11	-0.09	-0.45**	-0.16	0.07	0.13	3.97**
	(0.35)	(0.95)	(0.80)	(3.50)	(0.47)	(1.29)		
1994-3year	-0.08	-0.03	0.01	-0.11	-0.86**	-0.07	0.11	3.40**
	(0.83)	(0.24)	(0.05)	(0.93)	(3.49)	(1.23)		
1995-3year	-0.18	0.14	0.15	0.13	-0.21	-0.06	0.06	2.09*
	(1.41)	(1.05)	(1.19)	(.97)	(1.24)	(1.11)		
Five-year								
1992-5year	-0.03	0.01	-0.03	-0.06	-0.32	-0.01	0.02	0.61
	(0.36)	(0.15)	(0.43)	(0.57)	(1.28)	(0.31)		
1993-5year	-0.01	-0.09	-0.08	-0.33**	-0.12	0.03	0.11	3.12**
	(0.06)	(1.07)	(0.99)	(3.32)	(0.46)	(0.75)		

T-statistics are in parenthesis. * indicates significance at the 10 percent level.
**indicates significance at the 5 percent level.

Table 15: Pooled Dummy Variable Regressions: Using "Naïve" Predictor Groups

Out of Sample Performance Measure

Sharpe Ratios (load-adjusted returns)

						Coefficients for Dummy Variables							
Sample	g_0	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	**N**	**R^2**	**F-stat**
1-year	-0.01	0.03	0.04	0.01	-0.22**	0.14**	-0.26**	0.71**	0.24**	0.28**	922	0.64	176.94**
pooled	(.39)	(1.12)	(1.61)	(.26)	(2.46)	(4.76)	(10.38)	(22.89)	(10.79)	(12.18)			
3-year ++	0.03	-0.01	0.01	-0.03	-0.21**	0.17**	0.20**	0.37**	NA	NA	568	0.49	78.01**
pooled	(1.42)	(.04)	(0.32)	(1.00)	(2.67)	(10.31)	(12.66)	(22.78)					
5-year	0.22**	0.01	0.01	-0.06	-0.12*	0.04**	NA	NA	NA	NA	269	0.07	3.85**
pooled	(8.65)	(0.24)	(0.14)	(1.86)	(1.43)	(3.08)							

Mean Monthly Returns (load-adjusted returns)

						Coefficients for Dummy Variables							
Sample	g_0	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	**N**	**R^2**	**F-stat**
1-year	-0.03	0.11	0.10	-0.02	-0.74**	0.35**	-0.78**	1.53**	.82**	1.25**	922	0.64	179.72**
pooled	(0.38)	(1.41)	(1.42)	(.24)	(3.21)	(4.84)	(11.75)	(22.79)	(13.34)	(18.09)			
3-year ++	0.10	0.01	-0.01	-0.18**	-0.74**	0.48**	0.62**	1.30**	NA	NA	568	0.62	128.55**
pooled	(1.56)	(0.17)	(0.06)	(2.31)	(3.41)	(11.05)	(14.52)	(28.68)					
5-year ++	0.68**	-0.02	-0.05	-0.29**	-0.37**	0.22**	NA	NA	NA	NA	269	0.17	10.80**
pooled	(11.30)	(0.33)	(0.77)	(2.65)	(4.49)	(5.92)							

Jensen Alpha (non-load adjusted returns)

						Coefficients for Dummy Variables								
Sample	g_0	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	**Load**	**N**	**R^2**	**F-stat**
1-year ++	-0.07	0.07	0.13	0.14	-0.33	0.13*	-0.33**	-0.36**	-0.25**	-0.50**	-0.09**	922	0.15	16.43**
pooled	(.73)	(0.86)	(1.54)	(1.54)	(1.61)	(1.95)	(5.47)	(5.55)	(4.51)	(7.68)	(2.59)			
3-year ++	-0.06	0.01	0.05	-0.01	-0.34*	-0.14**	-0.28**	-0.28**	NA	NA	-0.06*	568	0.12	9.67**
pooled	(0.79)	(0.05)	(0.73)	(0.16)	(1.85)	(3.11)	(6.59)	(6.47)			(1.73)			
5-year ++	-0.18**	0.04	-0.03	-0.06	-0.14	--0.08*	NA	NA	NA	NA	0.02	269	0.03	1.48
pooled	(2.29)	(0.53)	(0.43)	(0.48)	(-1.01)	(2.31)					(0.41)			

4-index Alpha (non-load adjusted returns)

						Coefficients for Dummy Variables								
Sample	g_0	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8	g_9	**Load**	**N**	**R^2**	**F-stat**
1-year ++	0.08	-0.03	-0.06	-0.10	-0.54**	0.10*	-0.25**	-0.10*	0.05	0.01	-0.06*	922	0.06	4.24**
pooled	(1.13)	(0.51)	(1.09)	(1.52)	(2.94)	(1.89)	(4.83)	(1.94)	(1.01)	(0.16)	(1.79)			
3-year ++	-0.03	-0.02	-0.02	-0.12*	-0.45**	-0.05	-0.07**	0.01	NA	NA	-0.02	568	0.05	7.96**
pooled	(0.47)	(0.29)	(0.33)	(1.78)	(2.61)	(1.22)	(2.00)	(0.28)			(0.87)			
5-year ++	-0.01	-0.04	-0.05	-0.20**	-0.22*	-0.03	NA	NA	NA	NA	0.01	269	0.05	2.17**
pooled	(0.14)	(0.72)	(1.23)	(2.13)	(1.88)	(0.92)					(0.25)			

T-statistics are in parenthesis

*indicates significance at the 10 percent level. NA indicates there was no sample to pool.

** indicates significance at the 5 percent level.

++ Indicates that we use the White Heteroskedastic Consistent Standard Errors.

Table 16a: Average Performance Values by Decile and Spearman-Rho Rank Correlation Tests Using "Naïve" Predictor

Sample: 94-1 year		**Out-of-Sample-Performance Measures**			
Decile	**In-Sample Mean Monthly Return (1984.01-1993.12)**	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top 1	**1.45**	-0.79	-0.24	-0.36	-0.22
2	**1.28**	-0.64	-0.19	-0.23	-0.13
3	**1.21**	-0.91	-0.27	-0.39	-0.26
4	**1.18**	-0.73	-0.23	-0.40	-0.33
5	**1.15**	-0.72	-0.22	-0.23	-0.17
6	**1.11**	-0.65	-0.21	-0.27	-0.20
7	**1.08**	-0.50	-0.16	-0.11	-0.08
8	**1.00**	-0.97	-0.30	-0.58	-0.50
9	**0.92**	-0.71	-0.24	-0.32	-0.27
10	**0.74**	-0.92	-0.34	-0.58	-0.51
Rank Correlation of "Naïve" Predictor to Out-of-Sample Performance:		.152	.382	.285	.455
Two-tailed p-value:		(.676)	(.276)	(.425)	(.187)
Rank Correlations of Top-5 Deciles:		-.200	-.200	.200	.200
Two-tailed p-value:		(.747)	(.747)	(.747)	(.747)
Rank Correlations of Bottom-5 Deciles:		.600	.800	.800	.800
Two-tailed p-value:		(.285)	(.104)	(.104)	(.104)

Table 16b: Average Performance Values by Decile and Spearman-Rho Rank Correlation Tests Using "Naïve" Predictor

Sample: 95-1 year		Out-of-Sample-Performance Measures				
Decile		**In Sample Mean Monthly Return (1985.01-1994.12)**	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top	1	**1.48**	1.69	0.67	-0.44	-0.09
	2	**1.30**	1.72	0.78	-0.40	-0.13
	3	**1.23**	1.58	0.79	-0.44	-0.14
	4	**1.19**	1.61	0.76	-0.37	-0.11
	5	**1.14**	1.64	0.81	-0.40	-0.18
	6	**1.10**	1.82	0.74	-0.11	0.08
	7	**1.05**	1.47	0.72	-0.34	-0.15
	8	**0.98**	1.66	0.75	-0.16	0.07
	9	**0.93**	1.21	0.61	-0.43	-0.23
	10	**0.71**	0.90	0.48	-0.79	-0.55
Rank Correlation of "Naïve" Predictor to Out-of-Sample Performance:			.600	.539	-.103	.442
Two tailed p-value:			(.067)	(.108)	(.777)	(.200)
Rank Correlations of Top-5 Deciles:			.500	-.700	-.600	.700
Two-tailed p-value:			(.391)	(.188)	(.285)	(.188)
Rank Correlations of Bottom-5 Deciles:			.900	.700	.900	.900
Two-tailed p-value:			(.037)	(.188)	(.037)	(.037)

Table 16C: Average Performance Values by Decile and Spearman-Rho Rank Correlation Tests Using "Naïve" Predictor

Sample: 96-1 year		**Out-of-Sample Performance Measures**				
Decile		**In Sample Mean Monthly Return (1986.01-1995.12)**	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top	1	**1.51**	0.82	0.23	-0.35	0.13
	2	**1.31**	0.85	0.24	-0.32	0.03
	3	**1.26**	0.81	0.25	-0.34	0.02
	4	**1.21**	0.89	0.24	-0.25	0.09
	5	**1.16**	1.01	0.30	-0.21	0.11
	6	**1.13**	0.75	0.22	-0.37	-0.06
	7	**1.09**	0.76	0.25	-0.32	-0.09
	8	**1.03**	1.04	0.33	-0.06	0.09
	9	**0.97**	0.92	0.28	-0.17	0.12
	10	**0.77**	0.61	0.20	-0.35	-0.20
Rank Correlation of "Naïve" Predictor to Out-of-Sample Performance:			.067	-.176	-.176	.382
Two-tailed p-value:			(.855)	(.627)	(.627)	(.276)
Rank Correlations of Top-5 Deciles:			-.700	-.900	-.900	.100
Two-tailed p-value:			(.188)	(.037)	(.037)	(.873)
Rank Correlations of Bottom-5 Deciles:			.100	.100	-.300	.100
Two-tailed p-value:			(.873)	(.873)	(.624)	(.873)

Table 16D: Average Performance Value by Decile and Spearman-Rho Rank Correlation Tests Using "Naïve" Predictor

Sample: 97-1 year		**Out-of-Sample Performance Measures**			
Decile	**In Sample Mean Monthly Return (1984.01-1993.12)**	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top 1	**1.55**	0.98	0.21	-1.01	-0.17
2	**1.37**	1.39	0.28	-0.62	0.15
3	**1.29**	1.55	0.34	-0.37	0.20
4	**1.25**	1.29	0.27	-0.60	0.05
5	**1.22**	1.48	0.34	-0.36	-0.02
6	**1.17**	1.21	0.28	-0.51	-0.07
7	**1.13**	1.23	0.28	-0.61	-0.14
8	**1.07**	1.17	0.29	-0.47	-0.16
9	**1.01**	1.24	0.31	-0.33	0.06
10	**0.79**	1.01	0.23	-0.43	-0.13
Rank Correlation of "Naïve" Predictor to Out-of-Sample Performance:		.309	-.176	-.564	.200
Two-tailed p-value:		(.385)	(.627)	(.090)	(.580)
Rank Correlations of Top-5 Deciles:		-.500	-.700	-.900	-.100
Two-tailed p-value:		(.391)	(.188)	(.037)	(.873)
Rank Correlations of Bottom-5 Deciles:		.300	.100	-.800	-.100
Two-tailed p-value:		(.624)	(.873)	(.104)	(.873)

Table 16E: Average Performance Values by Decile and Spearman-Rho Rank Correlation Tests Using "Naïve" Predictor

Sample: 94-3 year		**Out-of-Sample Performance Measures**				
Decile		**In Sample Mean Monthly Return (1984.01-1993.12)**	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top	1	**1.45**	0.70	0.23	-0.42	-0.15
	2	**1.28**	0.74	0.23	-0.37	-0.13
	3	**1.21**	0.67	0.21	-0.38	-0.15
	4	**1.18**	0.75	0.25	-0.35	-0.13
	5	**1.15**	0.71	0.21	-0.36	-0.11
	6	**1.11**	0.76	0.26	-0.24	-0.10
	7	**1.08**	0.88	0.31	-0.13	-0.01
	8	**1.00**	0.61	0.21	-0.38	-0.18
	9	**0.92**	0.60	0.19	-0.35	-0.17
	10	**0.74**	0.35	0.15	-0.52	-0.40
Rank Correlation of "Naïve" Predictor to Out-of-Sample Performance:			.406	.358	-.115	.321
Two-tailed p-value:			(.244)	(.310)	(.751)	(.365)
Rank Correlations of Top-5 Deciles:			-.300	.100	-.800	-.900
Two-tailed p-value:			(.624)	(.873)	(.104)	(.037)
Rank Correlations of Bottom-5 Deciles:			.900	.900	.800	.800
Two-tailed p-value:			(.037)	(.037)	(.104)	(.104)

Table 16F: Average Performance Values by Decile and Spearman-Rho Rank Correlation Tests Using "Naïve" Predictor

Sample: 95-3 year		**Out-of-Sample Performance Measures**				
Decile		**In Sample Mean Monthly Return (1985.01-1994.12)**	Mean-monthly return (load-adjusted)	Sharpe Ratio (load adjusted)	Jensen alpha (non-load adjusted)	4-index alpha (non-load adjusted)
Top	1	**1.48**	1.37	0.37	-0.52	-0.10
	2	**1.30**	1.45	0.41	-0.40	-0.01
	3	**1.23**	1.40	0.40	-0.40	-0.12
	4	**1.19**	1.50	0.43	-0.30	0.02
	5	**1.14**	1.48	0.43	-0.34	-0.07
	6	**1.10**	1.34	0.37	-0.38	-0.06
	7	**1.05**	1.27	0.41	-0.35	-0.18
	8	**0.98**	1.39	0.41	-0.24	0.01
	9	**0.93**	1.18	0.40	-0.27	-0.08
	10	**0.71**	0.89	0.29	-0.49	-0.32
Rank Correlation of "Naïve" Predictor to Out-of-Sample Performance:			.661	.212	-.442	.248
Two-tailed p-value:			(.038)	(.556)	(.200)	(.489)
Rank Correlations of Top-5 Deciles:			-.300	.100	-.800	-.900
Two-tailed p-value:			(.624)	(.873)	(.104)	(.037)
Rank Correlations of Bottom-5 Deciles:			.700	.400	.100	.500
Two-tailed p-value:			(.188)	(.505)	(.873)	(.391)