

MORNINGSTAR MUTUAL FUND RATINGS REDUX

By *Matthew R. Morey, Ph.D.*,
and *Aron Gottesman, Ph.D.*

Abstract

We perform an extensive examination of how the new Morningstar rating system, introduced in June 2002, predicts future fund performance. Specifically, we examine all domestic equity funds that were rated by Morningstar as of June 30, 2002. We then examine the performance of these funds over the next three years, July 2002–June 2005. Using four different performance metrics, adjustments for loads, and three different methodologies for dealing with survivorship bias, we find widespread support for the notion that the new Morningstar rating system can predict future performance, at least within the first three years out-of-sample. Specifically, we find that higher rated funds, for the most part, significantly outperform lower rated funds. Moreover, the effect is relatively monotonic as even the next-to-lowest rated funds (two-star funds) significantly outperform the lowest rated funds (one-star funds). These results are quite different from those of Blake and Morey (2000) and Morey (2002b), which show that the older Morningstar rating system did not predict future performance well.

Introduction

How does the average U.S. investor pick a mutual fund? If one were to glance at mutual fund advertisements in the popular press, the likely answer would be Morningstar star ratings. These star ratings, which range from five stars (the highest) to one star (the lowest), are the only evidence of past winning performance in advertisements for many well-known

mutual funds. Indeed, in 2001–2004, fund companies such as American Century, Dreyfus, Fidelity, Franklin Resources, Northern Funds, and Strong Funds all ran advertisements that emphasized star ratings rather than their own return histories.

Evidence from academic research also indicates that investors care greatly about star ratings. A working paper from the Atlanta Federal Reserve Bank (Del Guercio and Tkac 2005) finds that the Morningstar star rating itself has a significant effect on fund flows. The paper finds that a fund's initial five-star rating produces inflows of 53 percent above the normal flow. In contrast, funds with rating downgrades experience significant outflows beyond what would normally be expected.

Given the importance of these Morningstar ratings to investors, a considerable amount of research has been conducted on whether the star ratings actually have predictive ability. These papers include Blake and Morey (2000), Morey (2002b), and Morey (2005). The general finding of these papers is that funds that received high Morningstar ratings did not subsequently outperform average-rated mutual funds. Specifically, the five-star funds (the top-rated funds) performed about the same as three-star funds (median-rated funds) after they were rated.¹ Hence, the moral of the story for investors was not to use high star ratings as signals of future strong performance.

The above-listed studies all were conducted using Morningstar ratings issued before June 2002. This is significant because in June 2002, Morningstar substan-

tially changed the methodology used to rate funds.² The changes were threefold. First, and most important, instead of rating funds within four broad areas (domestic equity, international equity, municipal bonds, and taxable bonds), there now are forty-eight different categories. For example, in the new ratings methodology, all large growth funds now are grouped into one category and the ratings are calculated based on how the fund performs relative to other large growth funds.³ Before the change, large growth fund ratings were determined relative to all domestic equity funds, a much larger group of funds that includes value and blended funds as well as funds with much smaller sizes.

This change in the methodology is quite significant because in the previous ratings methodology a fund could have received a high star rating simply because the style of the fund was in vogue. For example, in 1999 almost all growth funds were rated highly because growth funds generally outperformed other domestic equity funds during the three years prior to 1999.

Secondly, Morningstar changed its measure of risk to more accurately measure downside risk. In the methodology used before June 2002, risk was measured as the fund's average underperformance relative to a guaranteed investment, the ninety-day Treasury bill. Hence, if a fund's return exceeded this benchmark each month, the fund was deemed riskless. This methodology created a problem in that funds with highly variable returns could have been deemed "low-risk" funds as long as the returns were strong. As the Morningstar Principia Manual (2002) points out, Internet funds were a perfect example of this problem. Because they outperformed Treasury bills for many successive months, Internet funds displayed little Morningstar risk in 1999. Yet in the following years, these same "low-risk" funds had huge losses.

As a result of this problem, Morningstar changed its measure of risk so that it accounts for all variations in a fund's monthly performance, with more emphasis on downward variation. It rewards consistent performance and reduces the possibility of strong short-term performance masking inherent risk.

The third major change to the Morningstar ratings methodology is that funds with multiple share classes now are considered one fund in terms of the ratings

methodology. Before June 2002, each class of a mutual fund was considered a separate fund by Morningstar. Given the massive growth in multiple share classes over the past ten years, this change was important because Morningstar was not accurately measuring the number of funds that it was actually rating. This could have caused a problem in the rating system because the Morningstar ratings are computed by examining the performance of a fund relative to other funds. If the number of other funds is bloated due to multiple share classes, the rating that a fund receives could be influenced.

In this study we performed an extensive examination of how the new Morningstar rating system predicts future fund performance. To answer this question we examined all domestic equity mutual funds rated by Morningstar as of June 30, 2002 (the date of the first quarterly data disk to use the new ratings methodology). We next examined the out-of-sample performance of these funds for the subsequent three years (July 2002–June 2005), and we then examined the relationship between the Morningstar ratings on June 30, 2002 and the subsequent three-year performance.⁴

The remainder of the paper describes the data and the methods we used to combat survivorship bias, as well as our methodology, results and conclusions.

Data

To select funds we used the July 2002 Morningstar Principia Mutual Funds Data Disk, which contains data for funds as of June 30, 2002. From this disk we then selected all domestic equity funds that received an overall Morningstar mutual fund rating. This produced a sample of 3,886 funds. We used domestic equity funds because these styles are the most popular for domestic investors.⁵

We then narrowed the sample by eliminating replicate funds. Replicate funds are funds that are identical to another fund in our sample except that they are sold as different share classes. As stated above, the new ratings methodology calls for funds with multiple share classes to be considered as one fund in the revised Morningstar ratings. Because we didn't want to overcount the number of funds in the sample, we included only one of the fund's share classes in the sample. To choose the one share class that is included in the sam-

TABLE 1

DESCRIPTIVE STATISTICS

The sample consists of all domestic equity funds that received an overall Morningstar rating on June 30, 2002.

VARIABLE	STARS	NUMBER	PERCENTAGE IN LEVEL
Morningstar rating	★	168	8.8%
	★★	408	21.5%
	★★★	678	35.6%
	★★★★	456	24.0%
	★★★★★	192	10.1%
Morningstar Category	Large Blend	455	23.9%
	Large Growth	349	18.3%
	Large Value	274	14.4%
	Mid-Cap Blend	88	4.6%
	Mid-Cap Growth	220	11.6%
	Mid-Cap Value	103	5.4%
	Small Blend	95	5.0%
	Small Growth	206	10.8%
	Small Value	112	5.9%
Age of Fund, July 2002	10 < Age	687	36.1%
	3 < Age < 5	461	24.3%
	5 < Age < 10	754	39.6%
Year fund disappears	July 2002–June 2003	132	6.9%
	July 2003–June 2004	131	6.9%
	July 2004–June 2005	120	6.3%
	Survived full period	1,519	79.9%
All observations	All	1,902	100.0%

ple, we used a rule of selecting the fund share class that has the earliest inception date. By eliminating these replicate funds, we attained a sample of 1,902 funds.

Table 1 presents some descriptive statistics for the sample of 1,902 funds. The table shows the number of funds that received each star rating on June 30, 2002. It also shows the number of funds in each of the nine Morningstar categories and the number of funds in different age brackets. The table also shows the number of funds that survived the entire out-of-sample period.

For each of the 1,902 funds, we then attempted to obtain the out-of-sample returns of these funds. For a large majority of the funds (approximately 80 percent) obtaining the out-of-sample returns was simply a matter of following the fund's future performance. However, for about 20 percent of the sample, the funds disappeared before June 30, 2005, as a result of liquidations or mergers.

If we simply were to have reduced our sample to include only the funds that survived the entire out-of-

sample period, we would have subjected our study to a survivorship bias. To include funds that disappeared before June 30, 2005, we used three separate and distinct survivorship bias methodologies. We list them below.

Survivorship Bias Method 1. Before the fund disappeared, we simply used the out-of-sample returns of the fund in question. After the fund disappeared, we assumed the investor randomly reinvested into one of the other surviving funds of the same Morningstar category (large growth, medium growth, small growth, large blend, medium blend, small blend, large value, medium value, small value). Hence the out-of-sample returns from the month of disappearance onward are the equally weighted average returns of all the other surviving funds in our sample with the same category.

Survivorship Bias Method 2. For this method, we included only funds with at least twelve months of out-of-sample returns, i.e., at least July 2002–June 2003. We then used the actual returns of the funds, regardless

of how long the funds survived. Thus for some funds we have only twelve monthly return observations, while for others we have thirty-six observations. While this method obviously has a survivorship bias because we are including only funds that survived at least one year out-of-sample, it has the advantage that we used the actual returns of the funds in the sample.

Survivorship Bias Method 3. Similar to method 1, before a fund disappeared we used the out-of-sample returns of the fund in question. After the fund disappeared, we found a surviving fund that closely matched the fund that had disappeared. To find the matching fund we used an approach similar to Loughran and Ritter (1997). Specifically, we used an algorithm, detailed in Appendix A, that finds a surviving fund that is matched to the fund that has just disappeared. The algorithm uses the Morningstar category, Morningstar star rating, front and deferred loads, turnover, expense ratios, and fund size to find the matching fund. Hence, in this method the out-of-sample returns from the month of disappearance onward are the returns of the matching surviving fund. The advantage of this method is that it gave us higher power than method 1 because we use an actual matching fund rather than an equally weighted average.

Load Adjusted Returns

To calculate the ratings, Morningstar uses load-adjusted returns to reflect the extra expenses that the funds charge. However, the monthly fund returns that are available as Morningstar data are not adjusted for front or deferred loads. As a result, it was important for us to adjust the out-of-sample returns for loads when investigating whether the ratings were able to predict future performance. Moreover, for investors, load-adjusted performance is what they typically are concerned with.

To adjust the returns for loads we used an approach similar to that used in Blake and Morey (2000) and Morey (2002b). This approach is described in Appendix B.

Methodology

Performance Metrics

To measure out-of-sample performance, we used four risk-adjusted performance metrics: a Sharpe ratio, a

single-index alpha, a four-index alpha, and a conditional alpha. Note that each performance metric is calculated over the entire out-of-sample period July 2002–June 2005. For each performance metric we examine both nonload-adjusted and load-adjusted versions. We now briefly explain the four performance metrics:

The Sharpe ratio is:

$$Sharpe_i = \frac{\overline{R_{it} - R_{ft}}}{\sigma_i} \quad (1)$$

where: $R_{it} - R_{ft}$ are the monthly returns, in excess of the i th thirty-day Treasury-bill rate, R_{ft} the mutual fund during the out-of-sample period, and σ_i is the standard deviation of $R_{it} - R_{ft}$.

The single-index, or Jensen, $\alpha(SI)_i$, alpha is defined as:

$$R_{it} - R_{ft} = \alpha(SI)_i + \beta_{i1} RMRF_t + \epsilon_{it} \quad (2)$$

where, $RMRF_t$ is the value weighted market return on all NYSE/AMEX/NASDAQ firms in excess of the risk-free rate.

For the four-index alpha, $\alpha(4I)_i$, the following time-series regression model is used:

$$R_{it} - R_{ft} = \alpha(4I)_i + \beta_{i1} RMRF_t + \beta_{i2} SMB_t + \beta_{i3} HML_t + \beta_{i4} UMD_t + \epsilon_{it} \quad (3)$$

where SMB_t is the difference in returns across small and big stock portfolios controlling for the same weighted average book-to-market equity in the two portfolios; HML_t is the difference in returns between high and low book-to-market equity portfolios; UMD_t is the momentum factor, the average return on two high prior return portfolios minus the average return on two low prior portfolios.⁶

To estimate the conditional alpha, $\alpha(C)_i$, we use the form specified by Ferson and Schadt (1996). We use the following regression model:

$$R_{it} - R_{ft} = \alpha(C)_i + \beta_{i1} RMRF_t + \beta_{i2} (RMRF_t * TB_{t-1}) + \beta_{i3} (RMRF_t * TS_{t-1}) + \epsilon_{it} \quad (4)$$

where TB_{t-1} is the lagged one-month Treasury bill rate and TS_{t-1} is the lagged Treasury slope measure (i.e., the ten-year Treasury yield minus the three-month Treasury bill yield). The conditional alpha is a measure that compares a fund's return with the return of a dynamic strategy that attempts to match the fund's risk exposures.

Dummy Variable Regressions

The method we used to examine the out-of-sample predictive performance is a cross-sectional dummy variable regression analysis. Specifically, we estimated the following equation:

$$S_i = \gamma_0 + \gamma_1 D4_i + \gamma_2 D3_i + \gamma_3 D2_i + \gamma_4 D1_i + u_i \quad (5)$$

where:

S_i = out-sample performance metric for fund I ; i.e., Sharpe, $\alpha(SI)$, $\alpha(4I)$, or $\alpha(C)$.

$D4$ = 1 if the fund received four-star overall Morningstar rating as of June 30, 2002, 0 if not,

$D3$ = 1 if the fund received a three-star overall Morningstar rating as of June 30, 2002, 0 if not,

$D2$ = 1 if the fund received a two-star overall Morningstar rating as of June 30, 2002, 0 if not,

$D1$ = 1 if the fund received a one-star overall Morningstar rating as of June 30, 2002, 0 if not,

i = 1 through N , where N is the total number of funds in the sample.

In the above equation, the five-star fund group for the Morningstar sample is the reference group for the dummy variable regressions. Hence, when using the Sharpe ratio as the out-of-sample performance measure, the coefficient γ_0 represents the expected Sharpe ratio when all the dummy variables are equal to 0, and the coefficients γ_1 through γ_4 represent the differences between the dummy variables and the reference group. For example, a negative γ_1 implies the group of four-star funds performs worse than the group of five-star funds; a positive γ_1 implies the group of four-star funds outperforms the five-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 five-star funds as a reference group because they provide a ceiling with which we can compare the performance of the lower-rated group funds. If the ratings accurately predict out-of-sample performance, we should see increasingly negative (and significant) coefficients as we move from γ_1 to γ_4 .

Results

Our results are presented in tables 2–7. Tables 2–4 present the results of equation 5. More specifically, table 2 presents the results of

Using all three survivorship bias methods, and all four performance metrics, we find that the five-star funds as of June 30, 2002, significantly outperformed the other funds over the three-year out-of-sample period July 2002–June 2005.

equation 5 using survivorship bias method 1; table 3 presents the results using survivorship bias method 2; and table 4 provides the results for survivorship bias method 3.⁷ Tables 5–7 provide the tests of differences of the coefficients used in equation 5.

The results of tables 2–4 show solid support for the idea that the revised Morningstar ratings accurately predict future performance for the time period of the study. Using all three survivorship bias methods, and all four performance metrics, we find that the five-star funds as of June 30, 2002, significantly outperformed the other funds over the three-year out-of-sample period July 2002–June 2005. This can be seen by the fact that the coefficients γ_1 through γ_4 are always negative and significant. Furthermore, we find increasingly negative (and significant) coefficients as we move from γ_1 to γ_4 , indicating that five-star funds outperformed one-star funds much more often than they outperformed four- or three-star funds.⁸

Tables 5–7 provide the results of tests of differences in the coefficients of equation 5 and illustrate the ability of the ratings to predict future performance. Specifically, table 5 presents the tests of the differences of the coefficients in table 2, table 6 presents the tests of the differences of the coefficients in table 3, and table 7 presents the tests of the differences of the coefficients of table 4. The results in these tables show support for a monotonic relationship between the Morningstar ratings and future performance. For most performance metrics and survivorship bias methodologies, we found that the four-star funds had significantly higher average performance than three-star, two-star, and one-star

TABLE 2

DUMMY VARIABLE REGRESSIONS USING SURVIVORSHIP BIAS METHOD 1

The Morningstar ratings were taken from June 30, 2002. They are for all domestic equity mutual funds that received an overall Morningstar rating as of June 30, 2002. The out-of-sample performance was then measured from July 2002–June 2005. The t-statistics are in parentheses. Note that Survivorship Bias Method 1 assumes that if a fund disappears from the sample during the out-of-sample period (July 2002–June 2005), its monthly returns after it disappears are a weighted average of the surviving funds of the same Morningstar category.

OUT-OF-SAMPLE PERFORMANCE MEASURE	γ_0 (CONSTANT)	γ_1 (4-STAR FUNDS)	γ_2 (3-STAR FUNDS)	γ_3 (2-STAR FUNDS)	γ_4 (1-STAR FUNDS)	N	ADJ-R ²
Non-Load Adjusted	0.1809***	-0.0138**	-0.0249***	-0.0329***	-0.0570***	1902	0.0366
Sharpe Ratio	(35.8193)	(-2.2955)	(-4.3540)	(-5.3704)	(-7.7102)		
Non-Load Adjusted	0.0019	-0.0832***	-0.1259***	-0.1702***	-0.3136***	1902	0.0502
Jensen Alpha	(0.0829)	(-3.0401)	(-4.8406)	(-6.1125)	(-9.3318)		
Non-Load Adjusted	-0.1257***	-0.0673***	-0.0883***	-0.1148***	-0.2587***	1902	0.0354
Four-index Alpha	(-5.7566)	(-2.5845)	(-3.5704)	(-4.3344)	(-8.0898)		
Non-Load Adjusted	0.0528**	-0.0731**	-0.1211***	-0.1778***	-0.3240***	1902	0.0435
Conditional Alpha	(2.0108)	(-2.3335)	(-4.0677)	(-5.5809)	(-8.4220)		
Load Adjusted	0.1729***	-0.0159**	-0.0291***	-0.0408***	-0.0626***	1902	0.0440
Sharpe Ratio	(32.8691)	(-2.5296)	(-4.8767)	(-6.4039)	(-8.1343)		
Load Adjusted	-0.0309	-0.0921***	-0.1453***	-0.2062***	-0.3444***	1902	0.0592
Jensen Alpha	(-1.2954)	(-3.2410)	(-5.3799)	(-7.1345)	(-9.8698)		
Load Adjusted	-0.1585***	-0.0762***	-0.1077***	-0.1510***	-0.2896***	1902	0.0436
Four-index Alpha	(-7.0121)	(-2.8299)	(-4.2075)	(-5.5099)	(-8.7536)		
Load Adjusted	0.0200	-0.0819**	-0.1406***	-0.2134***	-0.3542***	1902	0.0517
Conditional Alpha	(0.7394)	(-2.5408)	(-4.5897)	(-6.5091)	(-8.9500)		

*** indicates significance at the 1-percent level ** indicates significance at the 5-percent level * indicates significance at the 10-percent level

TABLE 3

DUMMY VARIABLE REGRESSIONS USING SURVIVORSHIP BIAS METHOD 2

The Morningstar ratings were taken from June 30, 2002. They are for all domestic equity mutual funds that received an overall Morningstar rating as of June 30, 2002. The out-of-sample performance was then measured from July 2002–June 2005. The t-statistics are in parentheses. Note that Survivorship Bias Method 2 uses only funds that had at least twelve months of out-of-sample monthly returns. Hence, if a fund disappeared in December 2002 (only six months of out-of-sample data), it was excluded from the sample.

OUT-OF-SAMPLE PERFORMANCE MEASURE	γ_0 (CONSTANT)	γ_1 (4-STAR FUNDS)	γ_2 (3-STAR FUNDS)	γ_3 (2-STAR FUNDS)	γ_4 (1-STAR FUNDS)	N	ADJ-R ²
Non-Load Adjusted	0.1806***	-0.0150**	-0.0304***	-0.0381***	-0.0878***	1770	0.0592
Sharpe Ratio	(30.7871)	(-2.1445)	(-4.5497)	(-5.2858)	(-9.6889)		
Non-Load Adjusted	0.0025	-0.0834***	-0.1398***	-0.1742***	-0.4223***	1770	0.0701
Jensen Alpha	(0.0996)	(-2.7837)	(-4.8807)	(-5.6329)	(-10.8643)		
Non-Load Adjusted	-0.1262***	-0.0726**	-0.1075***	-0.1286***	-0.3788***	1770	0.0564
Four-index Alpha	(-5.0614)	(-2.4460)	(-3.7867)	(-4.1990)	(-9.8374)		
Non-Load Adjusted	0.0494*	-0.0707**	-0.1292***	-0.1866***	-0.4234***	1770	0.0546
Conditional Alpha	(1.6642)	(-2.0017)	(-3.8260)	(-5.1199)	(-9.2417)		
Load Adjusted	0.1726***	-0.0169**	-0.0342***	-0.0465***	-0.0931***	1770	0.0655
Sharpe Ratio	(28.5815)	(-2.3569)	(-4.9678)	(-6.2614)	(-9.9767)		
Load Adjusted	-0.0300	-0.0927***	-0.1583***	-0.2145***	-0.4570***	1770	0.0792
Jensen Alpha	(-1.1526)	(-2.9887)	(-5.3414)	(-6.7034)	(-11.3613)		
Load Adjusted	-0.1587***	-0.0819***	-0.1260***	-0.1690***	-0.4136***	1770	0.0649
Four-index Alpha	(-6.1879)	(-2.6810)	(-4.3157)	(-5.3614)	(-10.4389)		
Load Adjusted	0.0168	-0.0800**	-0.1477***	-0.2269***	-0.4581***	1770	0.0632
Conditional Alpha	(0.5535)	(-2.2075)	(-4.2658)	(-6.0712)	(-9.7500)		

*** indicates significance at the 1-percent level ** indicates significance at the 5-percent level * indicates significance at the 10-percent level

TABLE 4

DUMMY VARIABLE REGRESSIONS USING SURVIVORSHIP BIAS METHOD 3

The Morningstar ratings were taken from June 30, 2002. They are for all domestic equity mutual funds that received an overall Morningstar rating as of June 30, 2002. The out-of-sample performance was then measured from July 2002–June 2005. The t-statistics are in parentheses. Note that Survivorship Bias Method 3 assumes that if a fund disappears from the sample during the out-of-sample period (July 2002–June 2005), its monthly returns after it disappears are equal to a matching surviving fund determined by an algorithm in appendix A.

OUT-OF-SAMPLE PERFORMANCE MEASURE	γ_0 (CONSTANT)	γ_1 (4-STAR FUNDS)	γ_2 (3-STAR FUNDS)	γ_3 (2-STAR FUNDS)	γ_4 (1-STAR FUNDS)	N	ADJ-R ²
Nonload-Adjusted Sharpe Ratio	0.1820*** (34.8633)	-0.0145** (-2.3344)	-0.0252*** (-4.2520)	-0.0352*** (-5.5533)	-0.0652*** (-8.4343)	1879	0.0438
Nonload-Adjusted Jensen Alpha	0.0085 (0.3461)	-0.0869*** (-2.9537)	-0.1292*** (-4.6213)	-0.1797*** (-6.0010)	-0.3609*** (-9.8870)	1879	0.0565
Nonload-Adjusted Four-index Alpha	-0.1188*** (-5.0427)	-0.0722** (-2.5689)	-0.0898*** (-3.3616)	-0.1221*** (-4.2667)	-0.2985*** (-8.5554)	1879	0.0405
Nonload-Adjusted Conditional Alpha	0.0592** (2.1273)	-0.0754** (-2.2675)	-0.1248*** (-3.9510)	-0.1886*** (-5.5762)	-0.3829*** (-9.2840)	1879	0.0529
Load-Adjusted Sharpe Ratio	0.1742*** (31.9941)	-0.0166** (-2.5531)	-0.0293*** (-4.7440)	-0.0442*** (-6.6789)	-0.0710*** (-8.8111)	1879	0.0518
Load-Adjusted Jensen Alpha	-0.0232 (-0.9081)	0.0960*** (3.1489)	-0.1487*** (-5.1325)	-0.2203*** (-7.1010)	-0.3938*** (-10.4109)	1879	0.0659
Load-Adjusted Four-index Alpha	-0.1505*** (-6.1881)	-0.0813*** (-2.8023)	-0.1093*** (-3.9624)	-0.1627*** (-5.5081)	-0.3315*** (-9.2024)	1879	0.0490
Load-Adjusted Conditional Alpha	0.0275 (0.9607)	-0.0844** (-2.4685)	-0.1443*** (-4.4416)	-0.2292*** (-6.5892)	-0.4154*** (-9.7954)	1879	0.0617

*** indicates significance at the 1-percent level ** indicates significance at the 5-percent level * indicates significance at the 10-percent level

TABLE 5

TESTS OF DIFFERENCES IN COEFFICIENTS USED IN EQUATION 5 USING SURVIVORSHIP BIAS METHOD 1

The table reports the results of the test of the difference of the coefficients used in equation 5. A negative (positive) coefficient indicates that the lower-rated fund group performed (better) worse on average than the higher-rated fund group. Significance implies that the difference is significant at traditional levels. The t-statistics are listed in parentheses.

OUT-OF-SAMPLE PERFORMANCE MEASURE	γ_1 (4-STAR FUNDS) VERSUS γ_2 (3-STAR FUNDS)	γ_1 (4-STAR FUNDS) VERSUS γ_3 (2-STAR FUNDS)	γ_1 (4-STAR FUNDS) VERSUS γ_4 (1-STAR FUNDS)	γ_2 (3-STAR FUNDS) VERSUS γ_3 (2-STAR FUNDS)	γ_2 (3-STAR FUNDS) VERSUS γ_4 (1-STAR FUNDS)	γ_3 (2-STAR FUNDS) VERSUS γ_4 (1-STAR FUNDS)
Nonload-Adjusted Sharpe Ratio	-0.0111*** (-2.6165)	-0.0191*** (-3.9991)	-0.0432*** (-6.8370)	-0.0080* (-1.8205)	-0.0321*** (-5.3212)	-0.0241*** (-3.7583)
Nonload-Adjusted Jensen Alpha	-0.0427** (-2.2156)	-0.0870*** (-4.0121)	-0.2304*** (-8.0254)	-0.0443** (-2.2220)	-0.1877*** (-6.8474)	-0.1434*** (-4.9187)
Nonload-Adjusted Four-index Alpha	-0.0210 (-1.1482)	-0.0475** (-2.3036)	-0.1914*** (-7.0059)	-0.0265 (-1.3956)	-0.1703*** (-6.5299)	-0.1439*** (-5.1850)
Nonload-Adjusted Conditional Alpha	-0.0480** (-2.1761)	-0.1047*** (-4.2213)	-0.2509*** (-7.6341)	-0.0568** (-2.4879)	-0.2029*** (-6.4654)	-0.1461*** (-4.3779)
Load-Adjusted Sharpe Ratio	-0.0132*** (-2.9895)	-0.0250*** (-5.0309)	-0.0468*** (-7.1104)	-0.0118*** (-2.5820)	-0.0336*** (-5.3453)	-0.0218*** (-3.2605)
Load-Adjusted Jensen Alpha	-0.0532*** (-2.6581)	-0.1141*** (-5.0709)	-0.2523*** (-8.4637)	-0.0610*** (-2.9459)	-0.1991*** (-6.9954)	-0.1382*** (-4.5630)
Load-Adjusted Four-index Alpha	-0.0315* (-1.6595)	-0.0748*** (-3.5035)	-0.2134*** (-7.5489)	-0.0433** (-2.2064)	-0.1819*** (-6.7392)	-0.1386*** (-4.8276)
Load-Adjusted Conditional Alpha	-0.0587*** (-2.5861)	-0.1315*** (-5.1517)	-0.2724*** (-8.0545)	-0.0729*** (-3.1033)	-0.2137*** (-6.6175)	-0.1408*** (-4.1001)

*** indicates significance at the 1-percent level ** indicates significance at the 5-percent level * indicates significance at the 10-percent level

TABLE 6

TESTS OF DIFFERENCES IN COEFFICIENTS USED IN EQUATION 5 USING SURVIVORSHIP BIAS METHOD 2

The table reports the results of the test of the difference of the coefficients used in equation 5. A negative (positive) coefficient indicates that the lower-rated fund group performed (better) worse on average than the higher rated fund group. Significance implies that the difference is significant at traditional levels. The t-statistics are listed in parentheses.

OUT-OF-SAMPLE PERFORMANCE MEASURE	$\gamma 1$ (4-STAR FUNDS) VERSUS	$\gamma 1$ (4-STAR FUNDS) VERSUS	$\gamma 1$ (4-STAR FUNDS) VERSUS	$\gamma 2$ (3-STAR FUNDS) VERSUS	$\gamma 2$ (3-STAR FUNDS) VERSUS	$\gamma 3$ (2-STAR FUNDS) VERSUS
	$\gamma 2$ (3-STAR FUNDS)	$\gamma 3$ (2-STAR FUNDS)	$\gamma 4$ (1-STAR FUNDS)	$\gamma 3$ (2-STAR FUNDS)	$\gamma 4$ (1-STAR FUNDS)	$\gamma 4$ (1-STAR FUNDS)
Nonload-Adjusted Sharpe Ratio	-0.0154*** (-3.1093)	-0.0231*** (-4.0936)	-0.0728*** (-9.2435)	-0.0077 (-1.4669)	-0.0574*** (-7.5482)	-0.0497*** (-6.1528)
Nonload-Adjusted Jensen Alpha	-0.0564*** (-2.6541)	-0.0908*** (-3.7459)	-0.3389*** (-10.0287)	-0.0344 (-1.5222)	-0.2825*** (-8.6578)	-0.2481*** (-7.1618)
Nonload-Adjusted Four-index Alpha	-0.0349* (-1.6556)	-0.0560** (-2.3337)	-0.3062*** (-9.1470)	-0.0212 (-0.9463)	-0.2714*** (-8.3949)	-0.2502*** (-7.2895)
Nonload-Adjusted Conditional Alpha	-0.0585** (-2.3351)	-0.1159*** (-4.0584)	-0.3527*** (-8.8557)	-0.0574** (-2.1576)	-0.2942*** (-7.6507)	-0.2368*** (-5.7991)
Load-Adjusted Sharpe Ratio	-0.0172*** (-3.3734)	-0.0295*** (-5.0758)	-0.0761*** (-9.3862)	-0.0123** (-2.2724)	-0.0589*** (-7.5239)	-0.0466*** (-5.6049)
Load-Adjusted Jensen Alpha	-0.0657*** (-2.9862)	-0.1218*** (-4.8587)	-0.3643*** (-10.4187)	-0.0562** (-2.4038)	-0.2986*** (-8.8454)	-0.2425*** (-6.7638)
Load-Adjusted Four-index Alpha	-0.0441** (-2.0373)	-0.0871*** (-3.5265)	-0.3317*** (-9.6305)	-0.0430* (-1.8671)	-0.2876*** (-8.6471)	-0.2446*** (-6.9269)
Load-Adjusted Conditional Alpha	-0.0677*** (-2.6377)	-0.1470*** (-5.0178)	-0.3781*** (-9.2579)	-0.0792*** (-2.9025)	-0.3104*** (-7.8702)	-0.2312*** (-5.5203)

*** indicates significance at the 1-percent level ** indicates significance at the 5-percent level * indicates significance at the 10-percent level

TABLE 7

TESTS OF DIFFERENCES IN COEFFICIENTS USED IN EQUATION 5 USING SURVIVORSHIP BIAS METHOD 3

The table reports the results of the test of the difference of the coefficients used in equation 5. A negative (positive) coefficient indicates that the lower rated fund group performed (better) worse on average than the higher rated fund group. Significance implies that the difference is significant at traditional levels. The t-statistics are listed in parentheses.

OUT-OF-SAMPLE PERFORMANCE MEASURE	$\gamma 1$ (4-STAR FUNDS) VERSUS	$\gamma 1$ (4-STAR FUNDS) VERSUS	$\gamma 1$ (4-STAR FUNDS) VERSUS	$\gamma 2$ (3-STAR FUNDS) VERSUS	$\gamma 2$ (3-STAR FUNDS) VERSUS	$\gamma 3$ (2-STAR FUNDS) VERSUS
	$\gamma 2$ (3-STAR FUNDS)	$\gamma 3$ (2-STAR FUNDS)	$\gamma 4$ (1-STAR FUNDS)	$\gamma 3$ (2-STAR FUNDS)	$\gamma 4$ (1-STAR FUNDS)	$\gamma 4$ (1-STAR FUNDS)
Nonload-Adjusted Sharpe Ratio	-0.0106** (-2.4175)	-0.0207*** (-4.1758)	-0.0507*** (-7.6322)	-0.0100** (-2.2007)	-0.0400*** (-6.3017)	-0.0300*** (-4.4476)
Nonload-Adjusted Jensen Alpha	-0.0424** (-2.0379)	-0.0928*** (-3.9699)	-0.2740*** (-8.7428)	-0.0504** (-2.3435)	-0.2316*** (-7.7256)	-0.1812*** (-5.6922)
Nonload-Adjusted Four-index Alpha	-0.0176 (-0.8874)	-0.0499** (-2.2326)	-0.2262*** (-7.5531)	-0.0322 (-1.5678)	-0.2086*** (-7.2795)	-0.1764*** (-5.7965)
Nonload-Adjusted Conditional Alpha	-0.0495** (-2.1071)	-0.1133*** (-4.2893)	-0.3075*** (-8.6846)	-0.0638*** (-2.6236)	-0.2580*** (-7.6168)	-0.1942*** (-5.4004)
Load-Adjusted Sharpe Ratio	-0.0127*** (-2.7700)	-0.0276*** (-5.3424)	-0.0544*** (-7.8658)	-0.0149*** (-3.1273)	-0.0417*** (-6.3015)	-0.0269*** (-3.8211)
Load-Adjusted Jensen Alpha	-0.0528** (-2.4495)	-0.1243*** (-5.1333)	-0.2978*** (-9.1698)	-0.0716*** (-3.2096)	-0.2450*** (-7.8867)	-0.1735*** (-5.2584)
Load-Adjusted Four-index Alpha	-0.0280 (-1.3656)	-0.0814** (-3.5291)	-0.2501*** (-8.0876)	-0.0534** (-2.5140)	-0.2221*** (-7.5068)	-0.1687*** (-5.3708)
Load-Adjusted Conditional Alpha	-0.0600** (-2.4828)	-0.1448*** (-5.3339)	-0.3310*** (-9.0915)	-0.0849*** (-3.3953)	-0.2711*** (-7.7817)	-0.1862*** (-5.0340)

*** indicates significance at the 1-percent level ** indicates significance at the 5-percent level * indicates significance at the 10-percent level

funds. Indeed, the only time the four-star fund group did not show significantly higher average performance against a lower-rated fund group was in a comparison of four-star and three-star fund performance using the nonload-adjusted alpha (this only is when using survivorship bias methods 1 and 3). Similar to the results for the four-star funds, we also found that three-star funds generally had significantly higher performance than two- and one-star funds, and that two-star funds always had significantly higher performance than one-star funds. The only cases where this did not hold were when comparing three-star fund performance to two-star fund performance with the nonload-adjusted Sharpe, Jensen, and four-index alphas for the survivorship bias method 2 (see table 6).⁹

Conclusions

We performed an extensive examination of how the new Morningstar rating system, introduced in June 2002, predicts future fund performance. We investigated the predictive power of the Morningstar rating system using all domestic equity funds that were rated by Morningstar

as of June 30, 2002. We then examined the performance of all these funds over the next three years out-of-sample, July 2002–June 2005. In addition, we used nonload- and load-adjusted returns, four different performance metrics, and three different methodologies for dealing with survivorship bias.

We found significant supporting evidence that the new Morningstar rating system predicted future performance over the time periods we studied. Specifically, we found that higher-rated funds, for the most part, significantly outperformed lower-rated funds. Moreover, the effect was relatively monotonic: even the next-to-lowest rated funds (two-star funds) significantly outperformed the lowest-rated funds (one-star funds). These results are quite different from those of Blake and Morey (2000) and Morey (2002b), which indicated that the previous Morningstar rating system did not predict future performance.

What is the reason for the great disparity in the results of this study and the previous studies? One obvious explanation is that the focus on style-adjusted ratings rather than broad-category ratings likely is a large source of the ability to predict future performance.

TABLE 8

EXPENSE RATIOS AND TURNOVER ORGANIZED BY STAR GROUPS

Sample 1: All diversified domestic equity funds (funds with styles of aggressive growth, equity-income, growth, growth-income, and small company) from March 31, 2002 (rating determined by the previous Morningstar ratings methodology). All data as of March 31, 2002.

STAR RATING GROUP	AVERAGE EXPENSE RATIO (%)	AVERAGE TURNOVER RATIO (%)
Five-star funds	1.398	103.843
Four-star funds	1.362	91.622
Three-star funds	1.286	94.525
Two-star funds	1.465	105.579
One-star funds	1.808	158.461

Sample 2: All diversified domestic equity funds (funds with Morningstar categories of large blend, large growth, large value, mid-cap blend, mid-cap growth, mid-cap value, small blend, small growth, and small value) from June 30, 2002 (rating determined by the revised Morningstar ratings methodology). All data as of June 30, 2002

STAR RATING GROUP	AVERAGE EXPENSE RATIO (%)	AVERAGE TURNOVER RATIO (%)
Five-star funds	1.191	113.804
Four-star funds	1.086	92.715
Three-star funds	1.141	94.149
Two-star funds	1.260	112.774
One-star funds	1.877	187.037

TABLE 9

CORRELATION OF THE TWELVE-MONTH PERCENTILE RANKING AND MORNINGSTAR RATINGS

The table presents the correlations of the twelve-month percentile ranking and the overall Morningstar rating over five separate data sets (June 30, 2001, September 30, 2001, December 31, 2001, March 31, 2002, and June 30, 2002). The first four data sets use the previous Morningstar ratings methodology; the June 30, 2002, data set uses the revised Morningstar ratings methodology. The twelve-month percentile ranking provides the percentile ranks of funds in the same Morningstar category. Funds are ranked by their past twelve-month performance where performance is determined using the same methodology that Morningstar uses to determine performance for its overall Morningstar ratings. For the percentile ranking, 1 is the best possible ranking and 100 is the worst possible ranking. Only funds that are listed in the following categories were considered (large blend, medium blend, small blend, large growth, medium growth, small growth, large value, medium value, and small value).

SAMPLE	CORRELATION OF TWELVE-MONTH PERCENTILE RANK AND OVERALL MORNINGSTAR RATING
June 30, 2001 (previous Morningstar rating system)	-0.403
September 30, 2001 (previous Morningstar rating system)	-0.419
December 31, 2001 (previous Morningstar rating system)	-0.416
March 31, 2002 (previous Morningstar rating system)	-0.368
June 30, 2002 (revised Morningstar rating system)	-0.595

Another explanation may be fund expenses. Carhart (1997) found that fund expenses are inversely related to future fund performance. That is, funds with low expenses seem to perform better in the future than those with high expenses. Thus, an explanation for our results could be that before the change in the ratings methodology, five-star funds had expenses similar to the other funds and hence did not perform better in the future than other, lower-rated funds. Yet, after the change in the methodology, five-star funds may have significantly lower expenses than other funds and thus the high star rating can predict strong future performance.

While compelling, this explanation is not supported by the data. As can be seen in table 8, *before and after* the ratings change (i.e., using old ratings methodology data from March 31, 2002, and new ratings methodology data from June 30, 2002) five-star funds had slightly higher expense ratios and turnover ratios than four- and three-star funds.¹⁰

One last explanation may be the “hot hands phenomenon,” first described by Hendricks, Patel, and Zeckhauser (1988). They found short-term persistence in fund performance. Given that our out-of-sample period is only three-years, we may be capturing this effect. Indeed, if the methodology of the rating system was flawed before June 2002, it is possible that the

hot hands effect would not have shown up in performance prediction results from the old Morningstar ratings methodology. Now, with the improvements in the ratings methodology, the hot hands phenomenon may surface.

Acknowledgments

The authors thank Michele Gambera, chief economist at Morningstar, for her comments, and the Investment Management Consultants Association for financial support.

APPENDIX A: Algorithm for Finding Matching Funds Used in Survivorship Bias Method 3.

Using the Morningstar data disks, we created twelve samples of funds from which to choose the closest surviving fund. These are July 2002 (data as of June 30, 2002); October 2002; January 2003; April 2003, July 2003, October 2003, January 2004, April 2004, July 2004, October 2004, January 2005, and April 2005. When a fund disappeared, we chose the closest surviving fund from the sample associated with the time when the fund disappeared. For example, if a fund disappeared in April 2004, the closest surviving fund was chosen from a sample of funds from the April 2004 Morningstar data disk (data as of March 31, 2004). Using

this sampling methodology, we then used the following algorithm to identify the closest surviving fund.

1. To choose the matching fund, we first sampled only funds from the quarter before the fund disappeared that had the same Morningstar category and Morningstar rating as the fund that disappeared. The Morningstar category and Morningstar rating are defined as of the beginning of the sample, i.e., June 30, 2002.
2. We then identified all funds that survived the previous step for which the expense ratio was within 25 percent and 200 percent of the fund that eventually disappeared. Then,
 - a. If only one fund was identified that satisfies this filter, we eliminated all other funds and went to the final step.
 - b. If no funds were identified, we didn't eliminate any other funds, and went to the next step.
 - c. If more than one fund survived this filter, we eliminated all other funds and went to the next step.
3. We then identified all funds that survived the previous step for which turnover was within 25 percent and 200 percent of the fund that eventually disappeared. Then,
 - a. If only one fund was identified that satisfies this filter, we eliminated all other funds and went to the final step.
 - b. If no funds were identified, we didn't eliminate any other funds, and went to the next step.
 - c. If more than one fund survived this filter, we eliminated all other funds and went to the next step.
4. We then identified all funds that survived the previous step for which asset size was within 25 percent and 200 percent of the fund that eventually disappeared. Then,
 - a. If only one fund was identified that satisfies this filter, we eliminated all other funds and went to the final step.
 - b. If no funds were identified, we didn't eliminate any other funds, and went to the next step.
 - c. If more than one fund survived this filter, we eliminated all other funds and went to the next step.
5. We then identified the fund(s) with load that is (are) closest to the load of the fund that eventually disappeared. Then,
 - a. If only one fund was identified that satisfies this filter, we eliminated all other funds and went to the final step.
 - b. If more than one fund survived this filter, we eliminated all other funds and went to the next step.
6. We then identified all funds that survived the previous step for which expense ratio was within 75 percent and 125 percent of the fund that eventually disappeared. Then,
 - a. If only one fund was identified that satisfies this filter, we eliminated all other funds and went to the final step.
 - b. If no funds were identified, we didn't eliminate any other funds, and went to the next step.
 - c. If more than one fund survived this filter, we eliminated all other funds and went to the next step.
7. We then identified all funds that survived the previous step for which turnover was within 75 percent and 125 percent of the fund that eventually disappeared. Then,
 - a. If only one fund was identified that satisfies this filter, we eliminated all other funds and went to the final step.
 - b. If no funds were identified, we didn't eliminate any other funds, and went to the next step.
 - c. If more than one fund survived this filter, we eliminated all other funds and went to the next step.
8. We then identified all funds that survived the previous step for which asset size was within 75 percent and 125 percent of the fund that eventually disappeared. Then,
 - a. If only one fund was identified that satisfies this filter, we eliminated all other funds and went to the final step.
 - b. If no funds were identified, we didn't eliminate any other funds, and went to the next step.
 - c. If more than one fund survived this filter, we eliminated all other funds and went to the next step.

9. Of the funds that survived up to this point, we then randomly chose one and went to the final step.
10. Final step: We extracted the return of the identified surviving fund.

**APPENDIX B:
Load-adjusted Returns**

For front loads, we considered an investor who buys and holds the load shares for our holding period of thirty-six months (three years). As with most front loads, we assumed that the investor buying the fund pays a load in a lump sum at the time the fund is purchased. To spread the front load across the period that the shares are held, we used the 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-load-adjustment process is the following:

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

where

r = the monthly interest rate (the monthly periodic interest rate of three-year Treasury yield in July 2002 which was equal to a monthly periodic rate was .2508 percent)

f = the front load (expressed as a percent)

h = the number of months the fund is held

f^m = the monthly front-load adjustment

Hence, the front-load-adjusted (for front loads) returns are:

$$R_{it}^{LA} = R_{it} - f^m, \text{ where}$$

R_{it} = the nonload-adjusted monthly return of fund i in month t , where t goes from 1 to 36.

R_{it}^{LA} = the monthly front-load-adjusted return of fund i in month t

For deferred-load adjustment, the process is different 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 have prepaid the load in equal monthly installments. The amount of the monthly pre-

payment reflects the deferred load less the interest earned on the prepayments. The equation for the deferred-load adjustment is

$$d^m = \frac{d}{\sum_{j=1}^h (1+r)^j} \tag{2a}$$

where d is the deferred load (expressed as a percent) and d^m is the monthly deferred load adjustment.

Hence, the deferred-load-adjusted returns are:

$$R_{it}^{DLA} = R_{it} - d^m$$

where,

R_{it} = the nonload-adjusted monthly return of fund i in month t , where t goes from 1 to 36.

R_{it}^{DLA} = the monthly front-load-adjusted return of fund i in month t

Because the holding period of our sample is three years, we reduce the amount of the deferred load to one-half of the reported deferred load to reflect that fact that deferred loads are reduced the longer the investors holds the fund.

Endnotes

1. Note that Blake and Morey (2000) and Morey (2002b) also found that the old Morningstar methodology did have some success in predicting poor performance. Specifically, one- and two-star ratings (the lowest ratings) did predict poor future fund performance.

2. For details on the Morningstar methodology before the changes, see Blume (1998), Sharpe (1998), Morey (2002a), or Morey (2002b). For information on the new Morningstar methodology, see the Morningstar Principia manual (2006) and *The Morningstar Ratings Methodology* (2003).

3. ¹ Funds within each style are rated using a fixed distribution of star ratings (similar to the old ratings methodology). Specifically, the top 10 percent of the fund group receive five stars, the next 22.5 percent receive four stars, the next 35 percent receive three stars, the next 22.5 percent receive two stars, and the bottom 10 percent receive one star.

4. Kinnel (2005) conducted an in-house study of the predictive ability of the new Morningstar ratings methodology and found that the Morningstar ratings predict future performance relatively well. However, the methodology used focuses on a much smaller sample of funds than this study and also contains a survivorship bias.

5. The domestic equity category includes all funds listed as aggressive growth, equity-income, growth, growth-income, and small company.

6. The data for the four-index alpha were obtained from Kenneth French's data library, available on the World Wide Web at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

7. Note that because survivorship bias method 2 requires twelve months of out-of-sample data, 132 funds were dropped from the sample. For survivorship bias method 3, there were twenty-three funds that did not have turnover, size, or net assets data on the last quarter that they survived the sample. That is, the funds were listed on the Morningstar disk right before they disappeared but did not have the data we needed to perform the matching algorithm. Hence, these funds were dropped from this sample.

8. It should be noted that the four-index alphas for the reference group (five-star funds) are consistently negative in all the regressions results (tables 2–4). This suggests that even the top-rated funds were performing poorly assuming the four-factor model models risk appropriately.

9. ² We found the same general results when using Wald tests to examine the differences in the coefficient values.

10. The results for the June 30, 2002, sample broken down by Morningstar category, i.e., large blend, large growth, large value, etc., also showed that five-star funds generally had slightly higher average expense and turnover ratios compared with four- and three-star funds.

References

- Blake, Christopher, and Matthew Morey. 2000. Morningstar Rating and Mutual Fund Performance. *Journal of Financial and Quantitative Analysis* 35, no 3: 451–83.
- Blume, Marshall. 1998. An Anatomy of Morningstar Ratings. *Financial Analysts Journal* 54, no. 2 (March/April): 19–27.
- Carhart, Mark. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance* 52, no.1: 57–82.
- Del Guercio, Diane, and Paula A. Tkac. 2005. The Effect of Morningstar Ratings on Mutual Fund Flows. Working paper, Atlanta Federal Reserve Bank.
- Ferson, Wayne, and Rudi Schadt. 1996. Measuring Fund Strategy and Performance in Changing Economic Conditions. *Journal of Finance* 51, no. 2: 425–462.
- Hendricks, D., J. Patel, and R. Zeckhauser. 1993. Hot Hands in Mutual Funds: Short-Run Relative Persistence of Relative Performance 1974–1988. *Journal of Finance* 48, no. 1: 93–130.
- Loughran, Tim, and Jay R. Ritter. 1997. The Operating Performance of Firms Conducting Seasoned Equity Offerings. *Journal of Finance* 52, no. 5: 1,823–1,850.
- Kinnel, Russel. 2005. Rating the Star Rating. *Morningstar Fund Investor* 14, no. 4 (December).
- Morningstar Principia 2002, 2006, manual cites missing.
- Morningstar, Inc. 2003. *The Morningstar Star Rating Methodology*. (October 1, 2003.) Available on the World Wide Web at http://corporate.morningstar.com/US/documents/MethodologyDocuments/MethodologyPapers/MorningstarFundRating_Methodology.pdf.
- Morey, Matthew, R. 2002a. Mutual Fund Age and Morningstar Ratings. *Financial Analysts Journal* 58, no. 2 (March/April): 56–63.
- Morey, Matthew. 2002b. Rating the Raters: An Investigation into Mutual Fund Rating Services. *Journal of Investment Consulting* 5, no. 2 (November/December): 30–50.
- Morey, Matthew. 2005. The Kiss of Death? A 5-Star Morningstar Rating. *Journal of Investment Management* 3, no. 2: 41–52.
- Sharpe, William. 1998. Morningstar Risk-Adjusted Ratings. *Financial Analysts Journal* 54, no. 4 (July/August): 21–33.