Statistical Inference for Risk-Adjusted Performance Measure

Miranda Lam •

Abstract

This paper examines the statistical properties of and significance tests for a popular risk-adjusted performance measure, the M-squared measure. Test statistics along with asymptotic moments and probability distributions developed by prior studies are reviewed and their applications to the M-squared measure are discussed. This paper demonstrates through an example that calculation of the analytic test statistic and the corresponding p-value can be performed easily using a spreadsheet or mathematical software. Results of significance tests using bootstrapping are similar to those based on the analytic solutions, suggesting that the analytic solution approach is a useful method for conducting statistical inference. The results appear relatively robust even in small samples.

Keywords: Performance Evaluation, Hypothesis Testing, Mutual Funds
JEL Classification: C12, G12

• Salem State College, voice : 978-542-6308, e-mail: mlam@salemstate.edu

Statistical Inference for Risk-Adjusted Performance Measure

1 - Introduction

Measuring the performance of managed mutual funds is an important topic of interest to a wide audience, including economists, financial analysts, investors, and legislators. In 1995, the Security and Exchange Commission (SEC) received over 3,500 responses as a result of the Risk Concept Release (Investment Company Act Release No. 20974) requesting comments on improving the description of risk to investors by mutual fund companies. The enthusiasm of the responses was matched by their diversity and the SEC decided not to adopt a quantitative measure of risk or risk-adjusted performance at the current time. One important objection based on a survey by the Investment Company Institute (ICI) was that “quantitative risk measures have a strong potential to confuse or mislead investors.” (ICI 1996).

Modigliani and Modigliani (1997), motivated by the need to create an easy-to-understand risk-adjusted performance measure, propose the M-square Measure as an intuitive, rigorous, and flexible alternative to traditional measures based on the Capital Asset Pricing Model. The M-squared measure has gained steady popularity since its introduction, and strives as its goal to inform and not confuse investors. In practice, the M-squared measure relies on historic returns as input data and is subject to typical statistical noise of historic sampling.

There are two important principles for evaluating past performance: the first is comparing apples with apples, i.e. risk-adjustment, and the second is ensuring that the phenomenon is not spurious, i.e. statistical significance. Whereas the first principle, risk-adjusted performance, has received wide-spread attention, the second principle, ascertaining statistical significance is less understood and often ignored in practice. The purpose of this paper is to explore the statistical properties of and to demonstrate applying significance tests for the M-squared measure. We show that under the null hypothesis of zero performance, the M-squared measure can be transformed into a test statistic developed by Jobson and Korkie (1981). We present the asymptotic mean and standard deviation for this test statistic. We also illustrate how to perform statistical test using the test statistic for the M-square measure through an example containing seven mutual funds. Lastly, we conduct the same statistical tests using the bootstrapping method. Results from the bootstrapping method are very similar to those based on the asymptotic analytic solutions of the test statistic. For robustness check, we implement the same tests using sub-samples of varying sample sizes and find that the analytic approach is fairly robust. This is an important finding because the analytic approach can be easily applied using spreadsheet or mathematical software.

This paper is organized as follows. Section 2 presents the M-squared measure and its test statistic with asymptotic first and second moments. Section 3 presents an application of significance tests for the M-squared measure and compares results of significance tests based on the asymptotic distribution to the bootstrapping method. Selected results from sub-samples of different sample sizes are also presented. Section 4 provides a brief summary and conclusions.

2 - Statistical Properties of Risk-adjusted Performance Measures

Modigliani and Modigliani (1997) introduced the Risk-adjusted performance measure (RAP) for mutual funds, using a benchmark portfolio as the reference point for risk. The term benchmark will be used in the remainder of the paper to denote the unmanaged diversified portfolio against which the performance of the mutual funds will be evaluated. Most importantly, the risk and return of the benchmark portfolio serve as a reference for the market opportunity cost of risk. The RAP is defined as

$$RAP_i = \left(\frac{\sigma_M}{\sigma_i}\right)\mu_i + \left(1 - \frac{\sigma_M}{\sigma_i}\right) r_F \qquad (1)$$

where

σ_M = standard deviation of return on the benchmark portfolio;

σ_i = standard deviation of return on the mutual fund;

μ_i = mean return on the mutual fund;

r_F = risk-free interest rate;

The RAP is, in effect, a complete portfolio containing the mutual fund and the risk-free asset and this complete portfolio has the same standard deviation as the benchmark portfolio. A popular transformation of the RAP, dubbed the M-squared measure, M_i, is expressed as the difference between the RAP and return on the benchmark portfolio (Bodie, Kane, Marcus, 2001):

$$M_i = RAP_i - \mu_M \qquad (2)$$

where RAP_i is the risk-adjusted performance defined in (1) and μ_M is the mean return on the benchmark portfolio. The Sharpe ratio (Sharpe 1966) is closely related to the M-squared measure and the RAP in that they all use standard deviation as the measure of risk. The advantage of the M-squared measure over the Sharpe ratio is that the average investor is more familiar with basis points than with ratios. Graham and Harvey (1997) introduced a performance measure (GH2) that is also based on adjusting the standard deviation of a complete portfolio of T-bill and the risky asset being evaluated to match the S&P500 (benchmark portfolio). The GH2 measure allows returns on the T-bill to have non-zero variance and covariance with other risky assets. The GH2 measure is a more generalized form that includes the M-square measure as a special case. Graham and Harvey (1997) noted that for well-diversified mutual funds where substantial leverage is not needed to match the benchmark portfolio's volatility, the assumption of a risk-free asset by the M-squared measure is not an important issue. Thus, the M-square measure remains a useful performance evaluation tool for stock mutual funds. The following section explains how statistical inference based on parametric testing can be conducted for the M-square measure. The statistical property of the GH2 measure is beyond the scope of this paper.

In practice, the true values of the mean and standard deviation of returns on the mutual fund and on the benchmark portfolio are unknown. When sample means and sample standard deviations are used in place of the true values, the results are estimators of the RAP and the M-squared measure:

$$\overline{RAP}_i = \left(\frac{s_M}{s_i}\right)\bar{r}_i + \left(1 - \frac{s_M}{s_i}\right) r_F \qquad (3)$$

and

$$\overline{M}_i = \overline{RAP}_i - \bar{r}_M \qquad (4)$$

where

$$\bar{r}_i = \frac{1}{T}\sum_{t=1}^{T} r_{it} \quad = \text{ sample mean of returns on mutual fund i;}$$

$$\bar{r}_M = \frac{1}{T}\sum_{t=1}^{T} r_{Mt} \quad = \text{ sample mean of returns on the benchmark;}$$

$$s_i = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(r_{it} - \bar{r}_i\right)^2} \quad = \text{ sample standard deviation of returns on mutual fund i;}$$

$$s_M = \sqrt{\frac{1}{T-1}\sum_{t=1}^{T}\left(r_{Mt} - \bar{r}_M\right)^2} \quad = \text{ sample standard deviation of returns on the benchmark;}$$

For significance testing, the null hypothesis for the M-squared measure is[1]

$$H_O\text{: } M_i = 0. \tag{5}$$

A prime candidate for the test statistic is the sample estimator for the M-squared measure, $\overline{M}_i$. The asymptotic statistical properties, including the first and second moments, of the estimator, $\overline{M}_i$, can be derived directly using Taylor series expansions. Expanding the terms of $\overline{RAP}_i$, equation (4) becomes

$$\overline{M}_i = \left(\frac{s_M}{s_i}\right)\left(\bar{r}_i - r_F\right) - \left(\bar{r}_M - r_F\right) \tag{6}$$

Since the risk-free rate is a constant, the statistical properties of $\overline{M}_i$ depend only on the sample means and sample standard deviations of the mutual fund and the benchmark. Under the null hypothesis that $M_i = 0$, equation (6) can be rewritten as

$$\overline{M}_i' = s_M\left(\bar{r}_i - r_F\right) - s_i\left(\bar{r}_M - r_F\right) \tag{7}$$

[1] The null hypothesis for testing the RAP is H_0: $RAP_i = \mu_M$. Since this equation is a linear transformation of (6) and (6) is a more familiar format in hypothesis testing, the discussion in the paper focuses on deriving statistical tests for (6). Derivations of the test statistic for RAP_i and empirical results for RAP_i are available from the author.

Jobson and Korkie (1981) derive a test statistic, which is identical in form to (7), to test the difference between two Sharpe ratios.[2] Following Jobson and Korkie (1981), the asymptotic distribution of $\overline{M}_i'$ is normal with mean M_i' and variance ϖ_i^2, i.e.

$$\overline{M}_i' \sim N(M_i', \varpi_i^2) \tag{8}$$

where

$$M_i' = \sigma_M (\mu_i - r_F) - \sigma_i (\mu_M - r_F) \tag{9}$$

and

$$\varpi_i^2 = \frac{1}{T}\left[2\sigma_i^2\sigma_M^2 - 2\sigma_i\sigma_M\sigma_{iM} + \frac{1}{2}\mu_i^2\sigma_M^2 + \frac{1}{2}\mu_M^2\sigma_i^2 - \frac{1}{2}\left(\frac{\mu_i\mu_M}{\sigma_i\sigma_M}\right)\left(\sigma_{iM}^2 + \sigma_i^2\sigma_M^2\right)\right] \tag{10}$$

To test the null hypothesis, H_O: $M_i = 0$, a standardized test statistic, MT_i, is formed by substituting sample estimators for the means, variances, and covariances in (8), (9) and (10):

$$MT_i = \frac{\overline{M}_i'}{\overline{\varpi}_i}. \tag{11}$$

The test statistic, MT_i, has an asymptotic standard normal distribution. The sample estimator for $\overline{M}_i'$ is given by (7) and the sample estimator for $\overline{\varpi}_i^2$ is

$$\overline{\varpi}_i^2 = \frac{1}{T}\left[2s_i^2 s_M^2 - 2s_i s_M s_{iM} + \frac{1}{2}\bar{r}_i^2 s_M^2 + \frac{1}{2}\bar{r}_M^2 s_i^2 - \frac{1}{2}\left(\frac{\bar{r}_i\bar{r}_M}{s_i s_M}\right)\left(s_{iM}^2 + s_i^2 s_M^2\right)\right] \tag{12}$$

where $\bar{r}_i$, $\bar{r}_M$, s_i, and s_M are defined under (4) and s_{iM} is the sample covariance between returns on mutual fund i and returns on the benchmark:[3]

$$s_{iM} = \frac{1}{T-1}\sum_{t=1}^{T}\left(r_{it} - \bar{r}_i\right)\left(r_{Mt} - \bar{r}_M\right) \tag{13}$$

---

[2] Jobson and Korkie (1981) derive test statistics to compare two or more Sharpe ratios. They explore the statistical properties of several transformations of the difference between two Sharpe ratios. They present in their paper the most successful transformation, which is identical to (4)'.

[3] Note that $\overline{M}_i'$ is a biased estimator of M_i, $E(\overline{M}_i') = M_i (1 - ¼ T^{-1} + 1/32\, T^{-2})$, and the bias to $O(T^{-2})$ is $M_i (- ¼ T^{-1} + 1/32\, T^{-2})$. See Jobson and Korkie (1981) for a discussion on the behavior of the estimator for the variance, $\overline{\varpi}_i^2$, based on simulation experiments.

The next section presents an example to demonstrate the application of significance tests for the M-squared measure. Results obtained using the bootstrapping method are compared to results from the analytic test statistics presented in this section.

3 – Estimation and Significance Tests for the M-Squared Measure

Data

The mutual fund sample consists of the seven funds from the Modiglianni and Modiglianni (1990) study. The sample period is from January 1988 through April 2002, resulting in 172 monthly observations.[4] These funds represent six different categories, including Large Growth, Large Value, Large Blend, Mid-cap Growth, Small Growth, and Domestic Hybrid. Two benchmarks, the S&P 500 Index, the most popular benchmark for performance evaluation, and the Russell 3000 Index, which includes the 3000 largest U.S. companies by market capitalization, are studied. Returns on 3-month Treasury Bills are used as proxy for risk-free returns. In practice, standard deviation of returns on a 3-month Treasury Bill is not zero. However, for the purpose of evaluating performance, since the same risk-free return is subtracted from all returns, mutual funds and benchmarks, we follow conventional treatment and use return on a 3-month Treasury Bill as proxy for the risk-free rate. Returns on the mutual funds, benchmarks, and Treasury Bills are computed as:

$$r_{i,t} = (P_{i,t} – P_{i,t-1}) / P_{i,t-1} \qquad (14)$$

where $P_{i,t}$ is the closing price for month t. Monthly prices, adjusted for dividends and capital gains distribution, for mutual funds and closing values for benchmarks are obtained from www.yahoo.com. Prices for 3-month treasury bills are computed as:

$$P_t = 10000 \text{ x } (1 – \text{T-Bill Rate}_t \text{ x } 90 / 360) \qquad (15)$$

where T-Bill Rate_t is the discount rate of 3-month Treasury Bill traded in the secondary market in month t. Data on T-Bill Rates are from the Federal Reserve Board of Governors. Excess monthly return, defined as the difference between total monthly return and return on the T-bill, is used in the following empirical calculations.

---

[4] Data for the American Income Fund of America is only available beginning May 1996. Hence the sample period for this fund is from May 1996 through April 2002.

Summary statistics in Exhibit 1 show that average excess monthly returns on the mutual funds range from –0.3299% to 0.7677% and average excess returns on the S&P 500 Index is 0.4748% and on the Russell 3000 Index is 0.4718%. Sample standard deviations for the mutual funds range between 2.6196% to 7.7522%, and sample standard deviation for the S&P 500 Index is 4.0569% and for the Russell 3000 Index is 4.0672%. Returns on all mutual funds are positively correlated with returns on the benchmarks.[5] With the exception of 2 funds, higher sample standard deviations are associated with higher average monthly returns.

An important concern to investors is whether the higher average return represents appropriate compensation for the higher risk. Currently, the SEC requires mutual funds to present risk/return data in a bar chart format. The following except from the SEC's Final Release 33-7513 explains the rationale for adopting the Bar Chart in the presentation standard for mutual funds.

> "The proposed risk/return summary would require a fund's profile to include a bar chart showing the fund's annual returns for each of the last 10 calendar years and a table comparing the fund's average annual returns for the last 1-, 5-, and 10-fiscal years to those of a broad-based securities market index. The bar chart reflects the Commission's determination that investors need improved disclosure about the risks of investing in a fund. The bar chart is intended to illustrate graphically the variability of a fund's returns (e.g., whether a fund's annual returns for a 10-year period have varied significantly from year to year or were relatively even over the period)." (SEC 1998)

The SEC mandated bar chart applies only to historic returns on the mutual fund and does not require the bar chart to contain returns on the benchmark. Figure 1 includes returns on both the mutual fund and the

[5] The two benchmarks, the S&P 500 Index and the Russell 3000 Index, are highly correlated with a correlation coefficient of 0.9894. Empirical results for performance and statistical tests are very similar using both benchmarks. In the interest of brevity, only results using the S&P500 Index are reported in the paper. Results using the Russell 3000 Index are available from the author.

benchmark. Exhibit 2 contains such a bar chart using annual returns for 10 years from the Aim Constellation fund. While a bar chart shows the variability of annual returns from year to year, it does not address the question of whether the reward, in the form of average returns, is adequate compensation for the risk. This question is particularly important for investors who desire investments with different risk level than the benchmark. A bar chart, currently required by the SEC, does not convey explicitly the direct trade-off between risk and return.

Risk-Adjusted Performance

Exhibit 3 shows the estimated values for several risk-adjusted performance measures, including the Sharpe ratio, the Risk-Adjusted Performance and the M-squared Measure for the sample mutual funds against the S&P 500 Index as the benchmark. The estimated Sharpe ratio for the S&P 500 Index during the sample period is 0.1170 and its average excess return is 0.4748%. The Risk-adjusted Performance (RAP) measure of a mutual fund represents an investment strategy combining the mutual fund and a 3-month Treasury Bill to achieve the same standard deviation as the S&P 500 Index. A mutual fund with a higher Sharpe ratio or a higher RAP relative to those of the benchmark is considered to have superior performance. Note that the RAP and the M-squared measure are consistent with the Sharpe ratio, i.e. a fund demonstrating superior performance using the RAP or the M-squared measure will also demonstrate superior performance using the Sharpe ratio as the evaluation criteria. The M-squared measure is the difference between a mutual fund's RAP and the average excess return on the benchmark. The M-squared measure gives the absolute advantage, in basis points, of the investment strategy of holding a portfolio of the mutual fund and a 3-month Treasury Bill over investing in the benchmark alone. It is a risk-adjusted representation of performance because both strategies have the same standard deviation. In the sample, 4 funds have negative M-squared measures and 3 funds have positive M-squared measures and the values range from – 0.9570% to 0.2130%. Since these values are computed using sample statistics, significance tests are needed to ascertain whether the findings represent true measure of performance or the results are spurious due to noise in the sample. The next section discusses a simple approach to determine statistical significance for the estimated M-squared measure.

Statistical Test for Risk-adjusted Performance

Since the M-squared measure is computed using historic data, it is a sample estimator and proper interpretation of performance must include statistical inference. Equation (11) presented a test statistic, MT_i, for testing the null hypothesis of zero risk-adjusted performance. In this section, two methods are used to compute the p-value for MT_i. The first method evaluates the asymptotic mean and variance of MT_i using sample estimators as outlined in equations (7) and (12). We refer to this method as the analytical method and present its results in Panel A of Exhibit 4. The p-values range from 0.0290 to 0.9124. Out of the seven funds in the example, only the performance of the Magellan Fund is statistically significant. Performance measures for the remaining six funds are statistically indistinguishable from zero.

The second method to compute p-value for the test statistic employs a bootstrapping procedure and does not require explicit evaluation of the mean and variance of the test statistic, MT_i. The bootstrap procedure generates 1000 repetitions of random sampling, with replacement, from the original sample of 172 monthly observations. In each repetition, 172 data points are selected randomly for each mutual fund and the benchmark. Sample mean and standard deviation, computed in each iteration, are used to calculate the test statistic MT_i. The bootstrapping method produces 1000 estimates of MT_i and the p-value for the mean value of these 1000 estimates is presented in Panel B of Exhibit 4. Results from the bootstrapping procedures are virtually identical to those obtained using the asymptotic moments. This is good news to practitioners and investors because computing the test statistics, MT_i, using the asymptotic moments does not require sophisticated software and since MT_i has a standard normal distribution, its p-value is readily available.[6]

[6] For example, the Sharpe ratios, RAPs, and the M-squared measures in Table 2 and the asymptotic moments of the test statistics, MTi, and the p-values in Panel A of Table 3 are computed using MS Excel®. The bootstrapping procedure, which produces the results in Panel B of Table 3, is conducted using SAS® and requires considerably more work.

Exhibit 5 presents the M-squared measure and their p-values computed using the analytical approach and the bootstrap procedure. Similar to findings for the test statistic, MT_i, estimates for the M-squared measures using both methods are very similar. P-values, in parenthesis, are indicated next to the M-squared measures to give a complete picture of both economic and statistical significance of performance. Calculation of the p-values for the M-squared measures using the analytical approach is identical to the calculation of p-value for the test statistic MT_i because under the null hypothesis, MT_i is a linear transformation of the M-squared measure. The bootstrap estimation for the M-squared p-value also employs 1000 repetitions of random sampling, with replacement, from the original sample of 172 monthly observations. In each repetition, 172 data points are selected randomly for each mutual fund and the benchmark. Sample mean and standard deviation are computed in each iteration and are used to calculate the M-squared measure. The bootstrapping method produces 1000 estimates of M-squared measure for each mutual and the p-value for the mean value of these 1000 estimates is presented in Panel B of Exhibit 5.

Out of the sample of 7 funds, only one fund, Fidelity Magellan, has significant positive performance with a M-squared measure of 0.2130%. Since Magellan has a higher standard deviation, 4.3590%, than the S&P 500 Index, 4.0569%, the M-squared measure represents the gain of a strategy that invests 93% in the Magellan fund and 7% in a 3-month T-Bill. The M-squared measures of the other 6 funds are not statistically significant even though the estimated values can be quite large. This example demonstrates the importance of including significance test when presenting risk-adjusted performance information.

Sample Size and Robustness

The analytical approach is based on the asymptotic statistical properties of the test statistic and it is important to examine robustness, especially in smaller samples.[7] We create five sub-samples from the original sample of 172 monthly observations by random selection

[7] We thank participants at the 2002 FMA conference for suggesting the examination of robustness in smaller samples.

stratified across years. The samples sizes are 24, 36, 48, 60, and 120 monthly observations. For each sample size, we repeated the statistical tests described in the previous section. [8] Results in Exhibit 6 show that inference based on p-values computed using the analytical approach is similar to those based on bootstrap estimation. Panels A and B present the results for the M-squared measure and Panels C and D present the results for the test statistic, MT_i. For example, in Panels A and B, the analytical approach generated a p-value of 0.1703 while the bootstrap estimation generated a p-value of 0.1477 for the sub-sample with 36 observations. Both methods will lead to not rejecting the null hypothesis and conclude that the M-squared measure is not statistically different from zero. In fact, the two methods lead to the same conclusion for four of the five sub-samples examined, suggesting that the analytical approach is quite robust even in small samples.

4 – Summary and Conclusions

This paper illustrates that the two important principles of performance evaluation, risk-adjustment and significance testing, can be easily implemented. We present a test statistic developed by Jobson and Korkie (1981) and demonstrate that this test statistic can be used to perform significance test for the M-squared measure. Two approaches are used to compute the p-value for the test statistic. The first approach uses the analytical solutions to the asymptotic moments of the test statistics with values from sample estimators. The second approach uses bootstrapping. P-values of the M-squared measures for the sample funds obtained using both approaches are very similar and the results are relatively robust even in small samples. This is good news for the investors and financial advisors because the analytical approach can be applied easily without sophisticated software or mathematical programming.

[8] In the interest of brevity, only results for the Magellan Fund are reported for the sub-sample robustness check.

References

Bodie, Kane, and Marcus, 2001, *Essentials of Investments*, 4th Edition, New York: McGraw Hill, pp. 607-612.

Graham, J. and Harvey C., 1997, Grading the Performance of Market Timing Newsletters, with John Graham, *Financial Analysts Journal* 53, 6, 1997, pp. 54-66.

Gibbons, M., Ross, S., and Shanken, J., 1989, A Test of the Efficiency of a Given Portfolio, *Econometrica*, September 1989, pp. 1121-1152.

Investment Company Act Release No. 20974 (Mar. 29, 1995) [60 FR 17172], Risk Concept Release.

Investment Company Institute, 1996, *Shareholder Assessment of Risk Disclosure Methods*, Washington DC: Investment Company Institute.

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

Jobson, J., and Korkie B., 1981, Performance Hypothesis Testing with the Sharpe and Treynor Measures, *The Journal of Finance*, September 1981, pp. 889-908.

Jobson, J., and Korkie B., 1989, A Performance Interpretation of Multivariate Tests of Asset Set Intersection, Spanning, and Mean-variance Efficiency, *Journal of Financial and Quantitative Analysis* 24/2, June 1989, pp. 185-204.

Judge, G., Hill, R., Griffiths, W., Lutkepohl H, and Lee, T., 1988, *Introduction to the Theory and Practice of Econometrics*, New York: John Wiley & Sons.

Kandel, S. and Stambaugh R., 1989, A Mean-variance Framework for Tests of Asset Pricing Models, *The Review of Financial Studies,* 2/2 1989, pp. 125-156.

Kennedy, P., 1998, *A Guid to Econometrics* 4th Edition, Cambridge, MA: the MIT Press.

Modigliani, F., and L. Modigliani, 1997, Risk-Adjusted Performance, *The Journal of Portfolio Management*, Winter 1997, pp. 45-54.

Sharpe, W., 1966, Mutual Fund Perfomance, *Journal of Business* A Supplement, January 1966, pp. 119-138.

Treynor, J., 1965, How to Rate Management of Investment Funds, *Harvard Business Review* 43, January-February 1965, pp. 63-75.

Exhibit 1 **Summary Statistics**

The sample period is from January 1988 through April 2002. Monthly NAVs, dividends and capital gain distributions for the mutual funds and closing values for the S&P 500 index are obtained from finance.yahoo.com. Mutual Fund categories, expense ratio and total assets as of December 2001 are obtained from www.morningstar.com. Monthly return is computed as $r_{i,t} = (NAV_{i,t} + D_{i,t} - NAV_{i,t-1}) / NAV_{i,t-1}$; excess return is the difference between monthly return and return on a 3-month Treasury Bill, $R_{i,t} = r_{i,t} - r_{f,t}$.

Assets	Average Monthly Excess Return (%)	Standard Deviation (%)	Mutual Fund Category	Expense Ratio	Total (Mils)
3-month U.S. Treasury Bills	0.4388	0.1481	N/A N/A	N/A	
S&P 500 Index	0.4748	4.0569	N/A N/A	N/A	
Russell 3000 Index	0.4718	4.0672	N/A N/A	N/A	
Aim Constellation A	0.7677	6.2920	Large Growth	1.14	9,000
American Century Vista Investors	0.6339	7.7522	Mid-cap Growth	1.00	1,207
T. Rowe Price New Horizons	0.5537	6.5409	Small Growth	0.88	4,622
Fidelity Magellan	0.7390	4.3590	Large Blend	0.88	71,893
Vanguard Windsor	0.4192	4.6253	Large Value	0.41	14,592
Fidelity Puritan	0.3608	2.6196	Domestic Hybrid	0.63	20,363
American Income Fund of America A	-0.3299	2.7910	Domestic Hybrid	0.62	20,949
Correlation between the S&P 500 and the Russell 3000 Index	0.9894				

Correlation Coefficient Matrix	S&P500	CSGTX	TWCVX	PRNHX	FMAGX	VWNDX	FPURX	AMECX
S&P 500 Index	1.0000							
Aim Constellation A (CSGTX)	0.8259	1.0000						
American Century Vista Investors (TWCVX)	0.6146	0.8490	1.0000					
T. Rowe Price New Horizons (PRNHX)	0.7144	0.9138	0.8486	1.0000				
Fidelity Magellan (FMAGX)	0.9526	0.8853	0.7031	0.7758	1.0000			
Vanguard Windsor (VWNDX)	0.7807	0.6452	0.4347	0.5941	0.7705	1.0000		
Fidelity Puritan (FPURX)	0.8570	0.6767	0.4744	0.5787	0.8420	0.8103	1.0000	
American Income Fund of America (AMECX)	0.5644	0.4222	0.1492	0.4400	0.5349	0.8362	0.7028	1.0000

Exhibit 2

Bar Chart for Aim Constellation

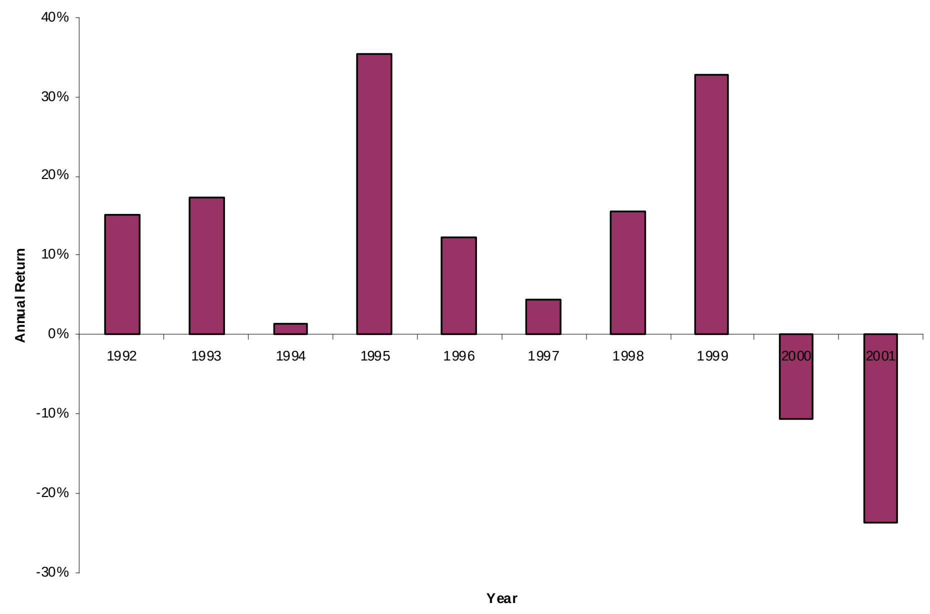

Exhibit 3 **Risk-adjusted Performance**

The sample period is from January 1988 through April 2002. Monthly NAVs, dividends and capital gain distributions for the mutual funds and closing values for the S&P 500 index are obtained from finance.yahoo.com. Monthly return is computed as $r_{i,t} = (NAV_{i,t} + D_{i,t} - NAV_{i,t-1}) / NAV_{i,t-1}$; excess return is the difference between monthly return and return on a 3-month Treasury Bill, $R_{i,t} = r_{i,t} - r_{f,t}$; sample mean of monthly excess return is $\overline{R}_i = T^{-1} \Sigma R_{it}$; sample standard deviation of monthly excess return is $s_i = ((T-1)^{-1}\Sigma(R_{it} - \overline{R}_i)^2)^{½}$; sample Sharpe ratio is $\overline{SH}_i = \overline{R}_i / s_i$; sample Risk-Adjusted Performance in excess return is $\overline{RAP}_i = (s_M/s_i)\,\overline{R}_i$. M-squared measures computed using sample statistics is $\overline{M}_i = (s_M/s_i)\,\overline{R}_i - \overline{R}_M$.

	Sharpe Ratios	RAP (in %)	M-square measures (in %)
Benchmark: S&P 500 Index	*0.1170*	*0.4748*	
Mutual Funds			
Aim Constellation A	0.1220	0.4950	0.0202
American Century Vista Investors	0.0818	0.3318	-0.1431
T. Rowe Price New Horizons	0.0846	0.3434	-0.1314
Fidelity Magellan	0.1695	0.6878	0.2130
Vanguard Windsor	0.0906	0.3677	-0.1072
Fidelity Puritan	0.1377	0.5587	0.0839
American Income Fund of America A[1]	-0.1182	-0.5865	-0.9570

[1]Sample period for the American Income Fund of America is from May 1996 through April 2002. The Sharpe ratio for the S&P 500 Index during the same period was 0.0747 and the RAP (in %) for the S&P 500 Index in the same period was 0.3705.

Exhibit 4

Results for the Test Statistic, MT_i, Calculated Using the Analytical Approach and the Bootstrap Approach

The sample period is from January 1988 through April 2002. Monthly NAVs, dividends and capital gain distributions for the mutual funds and closing values for the S&P 500 index are obtained from finance.yahoo.com. Monthly return is computed as $r_{i,t} = (NAV_{i,t} + D_{i,t} - NAV_{i,t-1}) / NAV_{i,t-1}$; excess return is the difference between monthly return and return on a 3-month Treasury Bill, $R_{i,t} = r_{i,t} - r_{f,t}$. All values are in 10^{-3} except for p-value.

	Panel A: Analytical Approach Using Sample Moments[1]				*Panel B: Bootstrap Estimation*[2]			
Mutual Funds	Sample Mean	Bias (to order T^{-2})	Standard Error	p-value	Sample Mean	Bias	Standard Error	p-value
Aim Constellation A	0.0127	0.0000	0.1153	0.9124	0.0127	0.0001	0.1141	0.9107
American Century Vista Investors	-0.1109	0.0002	0.2113	0.6003	-0.1109	-0.0009	0.2037	0.5833
T. Rowe Price New Horizons	-0.0860	0.0001	0.1535	0.5763	-0.0860	0.0010	0.1517	0.5756
Fidelity Magellan	0.0928*	-0.0001	0.0422	0.0290	0.0928*	-0.0006	0.0406	0.0232
Vanguard Windsor	-0.0496	0.0001	0.0952	0.6032	-0.0496	0.0006	0.0979	0.6168
Fidelity Puritan	0.0220	0.0000	0.0437	0.6158	0.0220	0.0000	0.0456	0.6291
American Income Fund of America A[3]	-0.2671	0.0009	0.1539	0.0870	-0.2671	0.0003	0.1553	0.0861

[1]The asymptotic mean and variance of MT^i are given in equations (9) and (10) respectively and the estimators for the asymptotic moments using samples moments are given in equations (7) and (12).
[2]Bootstrap estimation is performed based on 1000 repetitions of random sampling with replacement from the original sample of 174 monthly observations.
[3]Sample period for the American Income Fund of America is from May 1996 through April 2002.
* Statistically Significant at 5%.

Exhibit 5 **M-squared Measures and Statistical Significance Tests**

The sample period is from January 1988 through April 2002. Monthly NAVs, dividends and capital gain distributions for the mutual funds and closing values for the S&P 500 index are obtained from finance.yahoo.com. Monthly return is computed as $r_{i,t} = (NAV_{i,t} + D_{i,t} - NAV_{i,t-1}) / NAV_{i,t-1}$; excess return is the difference between monthly return and return on a 3-month Treasury Bill, $R_{i,t} = r_{i,t} - r_{f,t}$.

	Panel A: Analytical Approach Using Sample Moments		*Panel B: Bootstrap Estimation*	
Mutual Funds	M-square Measure (%)	p-value	M-square Measure (%)	p-value
Aim Constellation A	0.0202	0.9124	0.0202	0.9107
American Century Vista Investors	-0.1431	0.6003	-0.1492	0.5833
T. Rowe Price New Horizons	-0.1314	0.5763	-0.1321	0.5756
Fidelity Magellan	0.2129*	0.0290	0.2124*	0.0232
Vanguard Windsor	-0.1072	0.6032	-0.1038	0.6168
Fidelity Puritan	0.0839	0.6158	0.0878	0.6291
American Income Fund of America A[1]	-0.9570	0.0870	-0.9486	0.0861

[1]Sample period for the American Income Fund of America is from May 1996 through April 2002.

* Statistically Significant at 5%.

Exhibit 6 **Robustness in Small Samples**
(Fidelity Magellan Fund)

The sample period is from January 1988 through April 2002. Sub-samples are created by random selection stratified across years. Monthly NAVs, dividends and capital gain distributions for the mutual funds and closing values for the S&P 500 index are obtained from finance.yahoo.com. Monthly return is computed as $r_{i,t} = (NAV_{i,t} + D_{i,t} - NAV_{i,t-1}) / NAV_{i,t-1}$; excess return is the difference between monthly return and return on a 3-month Treasury Bill, $R_{i,t} = r_{i,t} - r_{f,t}$.

	Panel A: Analytical Approach		*Panel B: Bootstrap Estimation*	
Sample Size	M-square Measure (%)	p-value	M-square Measure (%)	p-value
172	0.2129	0.0290	0.2124	0.0216
120	0.3028	0.0075	0.3026	0.0045
60	0.2474	0.1128	0.2438	0.0797
48	0.1991	0.2664	0.1961	0.2172
36	0.2529	0.1703	0.2512	0.1477
24	0.3132	0.1473	0.3223	0.1265

	Panel C: Analytical Approach				*Panel D: Bootstrap Estimation*			
Sample Size	Mean of MT_i	Sample order T^{-2})	Bias (to Error	Standard p-value	Mean of MT_i	Sample Bias	Standard Error	p-value
172	0.0928	-0.0001	0.0422	0.0290	0.0928	-0.0006	0.0406	0.0232
120	0.1299	-0.0003	0.0478	0.0075	0.1299	-0.0006	0.0466	0.0057
60	0.1030	-0.0004	0.0640	0.1128	0.1030	-0.0018	0.0589	0.0860
48	0.0842	-0.0004	0.0748	0.2664	0.0842	-0.0012	0.0686	0.2270
36	0.1012	-0.0007	0.0723	0.1703	0.1012	-0.0027	0.0682	0.1493
24	0.1555	-0.0032	0.0997	0.1473	0.1555	-0.0101	0.0953	0.1270