# Risk-Adjusted Performance of Mutual Funds

Katerina Simons

*Economist, Federal Reserve Bank of Boston.* *The author is grateful to Richard Kopcke and Peter Fortune for helpful comments and to Jay Seideman for excellent research assistance.*

The number of mutual funds has grown dramatically in recent years. The Financial Research Corporation data base, the source of data for this article, lists 7,734 distinct mutual fund portfolios. Mutual funds are now the preferred way for individual investors and many institutions to participate in the capital markets, and their popularity has increased demand for evaluations of fund performance. *Business Week, Barron's, Forbes, Money,* and many other business publications rank mutual funds according to their performance. Information services, such as Morningstar and Lipper Analytical Services, exist specifically for this purpose. There is no general agreement, however, about how best to measure and compare fund performance and on what information funds should disclose to investors.

The two major issues that need to be addressed in any performance ranking are how to choose an appropriate benchmark for comparison and how to adjust a fund's return for risk. In March 1995, the Securities and Exchange Commission (SEC) issued a Request for Comments on "Improving Descriptions of Risk by Mutual Funds and Other Investment Companies." The request generated a lot of interest, with 3,600 comment letters from investors. However, no consensus has emerged and the SEC has declined for now to mandate a specific risk measure.

Risk and performance measurement is an active area for academic research and continues to be of vital interest to investors who need to make informed decisions and to mutual fund managers whose compensation is tied to fund performance. This article describes a number of performance measures. Their common feature is that they all measure funds' returns relative to risk. However, they differ in how they define and measure risk and, consequently, in how they define risk-adjusted performance. The article also compares rankings of a large sample of funds using two popular measures. It finds a surprisingly good agreement between the two measures for both stock and bond funds during the three-year period between 1995 and 1997.

Section I of the article describes simple measures of fund return, and Section II concentrates on several measures of risk. Section III describes a number of measures of risk-adjusted performance and their agreement with each other in ranking the three-year performance of a sample of bond, domestic stock, and international stock funds. Section IV describes measures of risk and return based on modern portfolio theory. Section V suggests some additional information that fund managers could provide to help investors choose funds appropriate to their needs. In particular, investors would benefit from better estimates of future asset returns, risks, and correlations. Fund managers could help investors make more informed decisions by providing estimates of expected future asset allocations for their funds.

*Mutual funds are now the preferred way for individual investors and many institutions to participate in the capital markets, and their popularity has increased demand for evaluations of fund performance.*

## I. Simple Measures of Return

The return on a mutual fund investment includes both income (in the form of dividends or interest payments) and capital gains or losses (the increase or decrease in the value of a security). The return is calculated by taking the change in a fund's net asset value, which is the market value of securities the fund holds divided by the number of the fund's shares during a given time period, assuming the reinvestment of all income and capital-gains distributions, and dividing it by the original net asset value. The return is calculated net of management fees and other expenses charged to the fund. Thus, a fund's monthly return can be expressed as follows:

$$R_t = \frac{NAV_t + DIST_t - NAV_{t-1}}{NAV_{t-1}} \tag{1}$$

where $R_t$ is the return in month t, $NAV_t$ is the closing net asset value of the fund on the last trading day of the month, $NAV_{t-1}$ is the closing net asset value of the fund on the last day of the previous month, and $DIST_t$ is income and capital gains distributions taken during the month.

Note that because of compounding, an arithmetic average of monthly returns for a period of time is not the same as the monthly rate of return that would have produced the total cumulative return during that period. The latter is equivalent to the geometric mean of monthly returns, calculated as follows:

$$R = \sqrt[T]{\Pi(1 + R_t)} \tag{2}$$

where R is the geometric mean for the period of T months. The industry standard is to report geometric mean return, which is always smaller than the arithmetic mean return. As an illustration, the first column of Table 1 provides a year of monthly returns for a hypothetical XYZ mutual fund and shows its monthly and annualized arithmetic and geometric mean returns.

Investors are not interested in the returns of a mutual fund in isolation but in comparison to some alternative investment. To be considered, a fund should meet some minimum hurdle, such as a return on a completely safe, liquid investment available at the time. Such a return is referred to as the "risk-free rate" and is usually taken to be the rate on 90-day Treasury bills. A fund's monthly return minus the monthly risk-free rate is called the fund's monthly "excess return." Column 2 of Table 1 shows the risk-free rate as represented by 1996 monthly returns on a money market fund investing in Treasury bills. Column 4 shows monthly excess returns of XYZ Fund, derived by subtracting monthly returns on the money market fund from monthly returns on XYZ Fund. We see that XYZ Fund had an annual (geometric) mean return of 20.26 percent in excess of the risk-free rate.

Comparing a fund's return to a risk-free investment is not the only relevant comparison. Domestic equity funds are often compared to the S&P 500 index, which is the most widely used benchmark for diversified domestic equity funds. However, other benchmarks may be more appropriate for some types of funds. Assume that XYZ is a "small-cap" fund, namely, that it invests in small-capitalization stocks, or stocks of companies with a total market value of less than $1 billion. Since XYZ Fund does not have any of the stocks that constitute the S&P 500, a more appropriate benchmark would be a "small-cap" index. Thus, we will use returns on a small-cap index fund as

Table 1
*XYZ Equity Fund Monthly Returns and Summary Statistics*

| Month | XYZ Return (%) (1) | Risk-Free Rate (%) (2) | Benchmark Return (%) (3) | XYZ Excess Return (%) (4) | Benchmark Excess Return (%) (5) | XYZ Excess Return over Benchmark (%) (6) |
|---|---|---|---|---|---|---|
| 1 | −1.66 | .46 | .16 | −2.12 | −.30 | −1.82 |
| 2 | 3.37 | .41 | 3.43 | 2.96 | 3.02 | −.06 |
| 3 | 3.26 | .43 | 1.87 | 2.83 | 1.44 | 1.39 |
| 4 | 4.61 | .41 | 5.59 | 4.20 | 5.18 | −.98 |
| 5 | 4.40 | .43 | 3.93 | 3.97 | 3.51 | .47 |
| 6 | −1.45 | .42 | −3.79 | −1.87 | −4.21 | 2.34 |
| 7 | −6.23 | .44 | −8.45 | −6.67 | −8.89 | 2.22 |
| 8 | 4.82 | .44 | 5.94 | 4.38 | 5.50 | −1.12 |
| 9 | 3.86 | .43 | 3.76 | 3.43 | 3.33 | .10 |
| 10 | 1.56 | .44 | −1.45 | 1.12 | −1.89 | 3.01 |
| 11 | 4.36 | .42 | 4.36 | 3.94 | 3.94 | .00 |
| 12 | 3.51 | .44 | 2.41 | 3.07 | 1.97 | 1.10 |
| Geometric Mean (percent) | | | | | | |
| Monthly | 1.98 | | 1.40 | 1.55 | .97 | .54 |
| Annualized | 26.53 | | 18.11 | 20.26 | 12.22 | 6.72 |
| Arithmetic Mean (percent) | | | | | | |
| Monthly | 2.03 | | 1.48 | 1.60 | 1.05 | .55 |
| Annualized | 24.41 | | 17.77 | 19.25 | 12.60 | 6.64 |
| Standard Deviation (percent) | | | | | | |
| Monthly | 3.27 | | 4.06 | 3.28 | 4.06 | 1.43 |
| Annualized | 11.34 | | 14.06 | 11.36 | 14.08 | 4.97 |

a benchmark. A comparison with this benchmark would show whether or not investing in XYZ Fund would have been better than investing in small cap stocks through the index fund.

Column 3 of Table 1 shows monthly returns on a small-cap index fund from a large mutual fund family specializing in index funds. Column 6 shows the difference between XYZ monthly returns and the monthly returns on the small-cap index fund. This difference shows how well the manager of XYZ Fund was able to pick stocks in the small-cap category. In our example, XYZ Fund was able to beat its benchmark by 6.72 percent in 1996.

## II. *Measures of Risk*

Investors are interested not only in funds' returns but also in risks taken to achieve those returns. We can think of risk as the uncertainty of the expected return, and uncertainty is generally equated with variability. Investors demand and receive higher returns with increased variability, suggesting that variability and risk are related.

### *Standard Deviation*

The basic measure of variability is the standard deviation, also known as the volatility. For a mutual fund, the standard deviation is used to measure the variability of monthly returns, as follows:

$$STD = \sqrt{1/T * \Sigma(R_t - AR)^2} \tag{3}$$

where STD is the monthly standard deviation, AR is the average monthly return, and T is the number of months in the period for which the standard deviation is being calculated. The monthly standard deviation can be annualized by multiplying it by the square root of 12.

For mutual funds, we are most often interested in the standard deviation of excess returns over the risk-free rate. To continue with our example, XYZ Fund had a monthly standard deviation of excess returns equal to 3.27 percent, or an annualized standard deviation of 11.34 percent. Mutual fund companies are sometimes interested in how well their fund managers are able to track the returns on some benchmark index related to the fund's announced purpose.

This can be measured as the standard deviation of the difference in returns between the fund and the appropriate benchmark index. The latter is sometimes referred to as "tracking error." In our example, XYZ Fund had a monthly tracking error of 1.43 percent and an annualized tracking error of 4.97 percent.

### *Downside Risk*

Standard deviation is sometimes criticized as being an inadequate measure of risk because investors do not dislike variability per se. Rather, they dislike losses but are quite happy to receive unexpected gains. One way to meet this objection is to calculate a measure of downside variability, which takes account of losses but not of gains. For example, we could calculate a measure of average monthly underperformance as follows: 1) Count the number of months when the fund lost money or underperformed Treasury bills, that is, when excess returns were negative. 2) Sum these negative excess returns. 3) Divide the sum by the total number of months in the measurement period. If we count negative excess returns for XYZ Fund in Table 1, we see it had negative excess returns in three out of 12 months and their sum was 10.66 percent. Thus, its downside risk, measured as average monthly underperformance, was 0.89 percent, compared to its monthly standard deviation of 3.27 percent.

*Investors do not dislike variability per se. Rather, they dislike losses but are quite happy to receive unexpected gains. Downside risk may be a better reflection of investors' attitudes toward risk.*

While downside risk may be a better reflection of investors' attitudes towards risk, empirical evidence suggests that the distinction between downside risk and the standard deviation is not as important as it seems because the two measures are highly correlated. Sharpe (1997) analyzed monthly standard deviations of excess returns and average monthly underperformance in a sample of 1,286 diversified equity funds in the three-year period between 1994 and 1996. He found a close relationship between these two measures, with a correlation coefficient of 0.932. Such a close correlation is not surprising, since monthly stock returns generally follow a symmetrical bell-shaped distribution. Therefore, stocks with larger downside deviations will also have larger standard deviations.

A more relevant measure is the ability to predict downside risk on the basis of both standard deviation and expected returns. Using the same sample of funds, Sharpe found that a regression of average underperformance on the standard deviation and expected return yields an R-squared of 0.999, which means that using only expected returns and standard deviations of these funds, one can explain 99.9 percent of the variation in average underperformance.

The average underperformance does not appear to yield much new information over and above the standard deviation. It is noted here chiefly because it is used by Morningstar, Inc. in its popular ratings of mutual funds, Morningstar ratings, which are discussed in the next section.

### *Value at Risk*

In recent years, Value at Risk has gained prominence as a risk measure. Value at Risk, also known as VAR, originated on derivatives trading desks at major banks and from there spread to currency and bond trading. Its popularity was much enhanced by the 1993 study by the Group of Thirty, *Derivatives: Practices and Principles,* which strongly recommended VAR analysis for derivatives trading. Essentially, it answers the question, "How much can the value of a portfolio decline with a given probability in a given time period?" The period used in measuring VAR for a bank's trading desk ranges from one day to two weeks, while the probability level is usually set in the range of 1 to 5 percent. Therefore, if we choose a period of one week and a probability level of 1 percent, a portfolio with a VAR of 5 percent might lose 5 percent or more of its value no more than 1 percent of the time. VAR is not a measure of maximum loss; instead, for given odds, it reports how great the range of losses is likely to be.

We will use the example of XYZ Fund returns to illustrate the simplest version of VAR calculation. Suppose that an investor put $1,000 into XYZ Fund and wishes to know the VAR for this investment for the next month. We can easily answer this question if we make certain assumptions about the statistical distribution of the fund's returns.

The most common assumption is that returns follow a normal distribution. One of the properties of the normal distribution is that 95 percent of all observations occur within 1.96 standard deviations from the mean. This means that the probability that an observation will fall 1.96 standard deviations *below* the mean is only 2.5 percent. For the purposes of calculating VAR we are interested only in losses, not gains, so this is the relevant probability. Recall that XYZ Fund had an (arithmetic) average monthly return of 2.03 percent and a standard deviation of 3.27 percent. Thus, its monthly VAR at the 2.5 percent probability level is $2.03\% - 1.96 * 3.27 = -4.38\%$, or $43.80 for a $1,000 investment, meaning that the probability of losing more than this is 2.5 percent.

VAR is often said to have an advantage over other risk measures in that it is more forward-looking. For example, in a recent article in *Risk* Magazine, Glauber (1998) describes the advantages of using VAR in this way: "A common analogy is that without VAR, management has to drive forward by looking out of the rear window. All the information available is about past performance. By using VAR management can use the latest tools to keep their eyes firmly focused in front."

While it can be described as forward-looking, VAR still relies on historical volatilities. However, the strength of VAR models is that they allow us to construct a measure of risk for the portfolio not from its *own* past volatility but from the volatilities of *risk factors* affecting the portfolio as it is constituted today. Risk factors are any factors that can affect the value of a given portfolio. They include stock indexes, interest rates, exchange rates, and commodity prices. A measure based on risk factors rather than on the portfolio's own volatility is especially important for funds that range far and wide in their choice of investments, use futures and options, and abruptly change their commitments to various asset classes. (This description applies to many hedge funds, though not perhaps to many of the regular mutual funds available to retail investors.)

Clearly, if the present composition of the fund's portfolio is significantly different than it was during the past year, then historical measures would not predict its future performance very accurately. However, as long as we know the fund's current composition and can assume that it will stay the same during the period for which we want to know the VAR, we can use a model based on the historical data about the risk factors to make statistical inferences about the probability distribution of the fund's future returns. In fact, for certain portfolios it is necessary to have a model based on risk factors even if one does not trade the portfolio at all. This is particularly true for portfolios consisting of bonds and/or options and futures, because such portfolios "age," that is, their characteristics change from the passage of time alone. In particular, as bonds approach maturity, their value approaches face value and their volatility diminishes and disappears altogether at maturity, when the bond can be redeemed at face value. Options, on the other hand, tend to lose value as they approach expiration, all other things being equal. This is one of the reasons why VAR analysis is used more frequently in derivatives and fixed-income investment and is less widespread for equities.

*VAR answers the question, "How much can the value of a portfolio decline with a given probability in a given time period?"*

Nevertheless, VAR models can provide useful information for equities also. For example, the manager of XYZ Fund can consider all the stocks currently in the portfolio to be separate risk factors. As long as the manager has the data on past returns for each stock, he can estimate their volatilities and correlations. This will enable the manager to calculate the VAR of the portfolio as it exists at the moment, not as it has been in the past. Risk managers at mutual fund companies may also be interested in the value at risk as it applies to underperforming the fund's chosen benchmark. This measure, known as "relative" or "tracking" VAR, can be thought of as the VAR of a portfolio consisting of long positions in all the stocks the fund currently owns and a short position in the fund's benchmark. While VAR provides a view of risk based on low-probability losses, for symmetrical bell-shaped distributions such as those typically followed by stock returns, VAR is highly correlated with volatility as measured by the standard deviation. In fact, for normally distributed returns, value at risk is directly proportional to standard deviation.

## III. Risk-Adjusted Performance

Two risk measures discussed in the previous section, namely the standard deviation and the downside risk, have been used to adjust mutual fund returns to obtain measures of risk-adjusted performance. This section describes two measures of risk-adjusted performance based on the standard deviation, namely, the Sharpe ratio and the Modigliani measure, and Morningstar ratings, which are based on downside risk.

### *Sharpe Ratio*

The most commonly used measure of risk-adjusted performance is the Sharpe ratio (Sharpe 1966), which measures the fund's excess return per unit of its risk. The Sharpe ratio can be expressed as follows:

$$\textit{Sharpe ratio} = \frac{\textit{fund's average excess return}}{\textit{standard deviation of fund's excess return}}. \quad (4)$$

Column 4 of Table 1 shows that the (arithmetic) monthly mean excess return of XYZ Fund is 1.60 percent, while the monthly standard deviation of its excess return is 3.28 percent.[1] Thus, the fund's monthly Sharpe ratio is 1.60%/3.28% = .49. The annualized Sharpe ratio is computed as the ratio of annualized mean excess return to its annualized standard deviation, or, equivalently, as the monthly Sharpe ratio times the square root of 12. Thus, XYZ's annualized Sharpe ratio is 19.25%/11.36% = 1.69.

The Sharpe ratio is based on the trade-off between risk and return. A high Sharpe ratio means that the fund delivers a lot of return for its level of volatility. The Sharpe ratio allows a direct comparison of the risk-adjusted performance of any two mutual funds, regardless of their volatilities and their correlations with a benchmark.

It is important to keep in mind that the relevance of a risk-adjusted measure such as the Sharpe ratio for choosing a mutual fund depends critically on investors' ability to do two things: 1) combine an investment in a mutual fund with an investment in the riskless asset, and 2) leverage the investment by, for example, borrowing money to invest in the mutual fund. (For the result to hold exactly, the investor must be able to borrow and lend at the same risk-free rate.) This is because the combination of investing in any given mutual fund and in a riskless asset allows one to lower the risk of the combined investment at the price of the corresponding reduction in expected return. Alternatively, leveraging one's investment in the fund allows one to increase expected return at the price of the corresponding increase in risk. Thus, any level of risk can be achieved with the given fund, and so the investor can achieve the best combination of risk and return by investing in the fund with the highest Sharpe ratio, regardless of the investor's own degree of risk tolerance.[2]

As an example, consider an investor who has $1,000 to invest and is choosing whether to invest in Fund X or Fund Y (but not both). Fund X has an expected excess return of 12 percent and a standard deviation of 9 percent. Fund Y has an expected excess return of 6 percent and a standard deviation of 4 percent. Fund X has a Sharpe ratio of 1.33 while Fund Y has a Sharpe ratio of 1.5. Because Fund Y has a higher Sharpe ratio it is a better choice, even for investors who wish to earn an expected excess return of 12 percent. Instead of investing in Fund X, those investors can borrow another $1,000 and invest the resulting $2,000 in Fund Y. (See point Y′ on Figure 1.) This leveraged investment provides twice the risk and twice the expected return of unleveraged investment in Fund Y, namely expected excess return of 12 percent and a standard deviation of 8 percent, better than the 9 percent standard deviation the investor could get by investing in Fund X. Figure 1 shows these risk/return combinations. The slopes of the lines drawn from the origin through the points representing risk and return of Funds X and Y are equal to the funds' Sharpe ratios. Clearly, all funds that lie along a higher line are better investments than the funds on a lower line, so that a fund with a higher Sharpe ratio is preferable to a fund with a lower one.

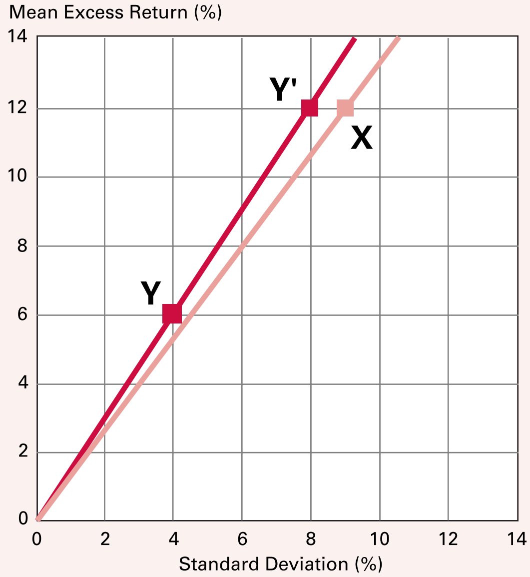
Despite its near universal acceptance among academics and institutional investors, the Sharpe ratio is not well known among the general public and financial advisors. A recent newspaper column, comment-

[1] Academic literature generally uses the arithmetic mean in the calculation of the Sharpe ratio because of its better statistical properties. For example, the Sharpe ratio based on the arithmetic mean times the square root of the number of observations can be interpreted as a T-statistic for the hypothesis that the fund's excess return is significantly different from zero.

[2] This is exactly the conclusion of the Capital Asset Pricing Model (CAPM), which holds that a portfolio of assets exists (known as the market portfolio) that provides the highest return per unit of risk and is appropriate for all investors. The CAPM is discussed in more detail in the next section.

Figure 1

*Two Funds with Different Sharpe Ratios*




| | Sharpe Ratio | Excess Return | Standard Deviation |
|---|---|---|---|
| Fund X | 1.33 | 12% | 9% |
| Fund Y | 1.5 | 6% | 4% |

ing on the contents of the CFP (Certified Financial Planner) examination, had this to say about the Sharpe ratio:

> But I do not know of a single financial planner—and I asked dozens of them after taking the test—who has ever had a client come in and ask for the calculation of Sharpe Measure of Performance on a mutual fund. In fact, none of the planners I queried could actually calculate the Sharpe Index without the formula in front of them. (The Sharpe is so esoteric that most mainstream financial dictionaries ignore it, most planners can't adequately explain it, and I am not even going to attempt it here.) Yet the Sharpe Index is on the CFP exam (Jaffe 1998).

### *Modigliani Measure*

The view that the Sharpe ratio may be too difficult for the average investor to understand is shared by Modigliani and Modigliani (1997), who propose a somewhat different measure of risk-adjusted performance. Their measure expresses a fund's performance relative to the market in percentage terms and they believe that the average investor would find the measure easier to understand. The Modigliani measure can be expressed as follows:

$$\textit{Modigliani measure} = \frac{\textit{fund's average excess return}}{\textit{standard deviation of fund's excess return}} \times \textit{standard deviation of index excess return}. \qquad (5)$$

Modigliani and Modigliani propose to use the standard deviation of a broad-based market index, such as the S&P 500, as the benchmark for risk comparison, but presumably other benchmarks could be used. In essence, for a fund with any given risk and return, the Modigliani measure is equivalent to the return the fund would have achieved *if it had the same risk as the market index.* Thus, the fund with the highest Modigliani measure, like the fund with the highest Sharpe ratio, would have the highest return for any level of risk. Since their measure is expressed in percentage points, Modigliani and Modigliani believe that it can be more easily understood by average investors.

To continue with our example of XYZ Fund, its annualized mean (arithmetic) excess return is 19.25 percent and its annualized standard deviation is 11.36 percent. If the standard deviation of the excess return on the S&P 500 market index is 15 percent, XYZ's Modigliani measure is 19.25%/11.36% × 15% = 25.42%. This 25.42 percent return can be interpreted as follows: An investor who is willing to accept the higher standard deviation of the S&P500 can improve his return by investing in XYZ and leveraging that investment to achieve the standard deviation of 15 percent. This would result in the return of 25.42 percent, which is the fund's Modigliani measure. Performance measures for the XYZ Fund are summarized in Table 2.

As the preceding example makes clear, the Modigliani measure has the same limitation as the Sharpe ratio in that it is of limited practical use to investors who are unable to use leverage in their

Table 2

*Risk-Adjusted Performance Measures for XYZ Fund*

| | Alpha | Sharpe Ratio | Modigliani Measure (%) |
|---|---|---|---|
| Monthly | .803 | .49 | 1.98 |
| Annualized | 9.63 | 1.69 | 25.42 |

mutual fund investing. As the Modigliani measure is very new, it remains to be seen if it will meet with more understanding and acceptance than the Sharpe ratio.

### *Morningstar Ratings*

Morningstar, Incorporated, calculates its own measures of risk-adjusted performance that form the basis of its popular star ratings.[3] Star ratings are well known among individual investors. One study found that 90 percent of new money invested in equity funds in 1995 flowed to funds rated 4 or 5 stars by Morningstar (Damato 1996).

For the purpose of its star ratings, Morningstar divides all mutual funds into four asset classes—domestic stock funds, international stock funds, taxable bond funds, and municipal bond funds. First, Morningstar calculates an excess return measure for each fund by adjusting for sales loads and subtracting the 90-day Treasury bill rate. These load-adjusted excess returns are then divided by the average excess return for the fund's asset class.[4] This can be summarized as follows:

*Morningstar return*

$$= \frac{\textit{load-adjusted fund excess return}}{\textit{average excess return for asset class}}. \quad (6)$$

Second, Morningstar calculates a measure of downside risk by counting the number of months in which the fund's excess return was negative, summing up all the negative excess returns and dividing the sum by the total number of months in the measurement period. The same calculation of average monthly underperformance is then done for the fund's asset class as a whole. Their ratio constitutes Morningstar risk:

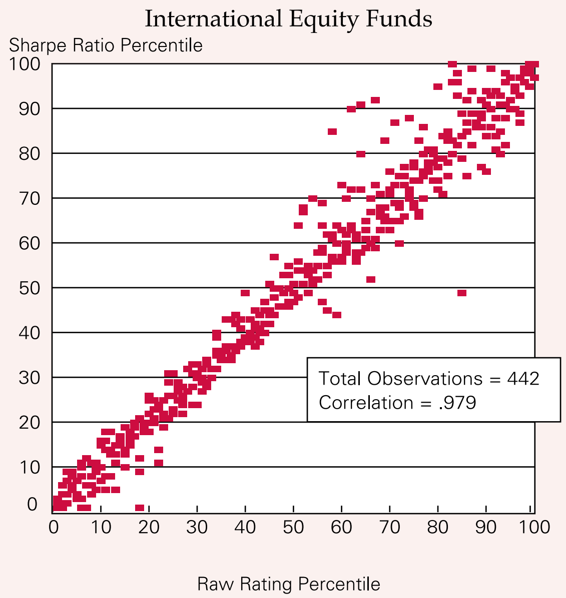
*Morningstar risk*

$$= \frac{\textit{fund's average underperformance}}{\textit{average underperformance of its asset class}}. \quad (7)$$

Third, Morningstar calculates its raw rating by subtracting the Morningstar risk score from the Morningstar return score. Finally, all funds are ranked by their raw rating within their asset class and assigned their

Figure 2

*Correlation of Percentiles for Sharpe Ratios and Morningstar Raw Ratings*




stars as follows: top 10 percent—5 stars; next 22.5 percent—4 stars; middle 35 percent—3 stars; next 22.5 percent—2 stars; and bottom 10 percent—1 star. Stars are calculated for three-, five-, and 10-year periods and then combined into an overall rating. Funds with a track record of less than three years are not rated.

In addition to its star ratings, Morningstar also calculates category ratings for each fund. The main difference between stars and category ratings is that category ratings are not based on four asset classes but on more narrowly defined categories, with each fund assigned to one (and only one) category among 44 altogether: 20 domestic stock categories, 9 international stock categories, 10 taxable bond categories, and 5 municipal bond categories. In addition, category ratings are not adjusted for sales load and are calculated only for a three-year period.

### *Relationships among Performance Measures*

An important question in comparing performance measures is whether or not they would lead to a similar ranking of mutual funds. The first thing to note is that, as long as one uses the same benchmark,

[3] The following description is based on Harrel (1998).
[4] When average excess return for the asset class is negative, the T-bill rate is substituted in the denominator.

Figure 2 continued

Domestic Equity Funds

Total Observations = 1,737
Correlation = .947

Sharpe Ratio Percentile

Figure 2 continued

Taxable Bond Funds

[redacted]

Sharpe ratios and Morningstar star ratings had a correlation coefficient of 0.955.[5]

## IV. Modern Portfolio Theory

The measures of risk-adjusted performance discussed above are subject to the same limitation as the risk measures on which they are based, namely, that they describe each fund in isolation and not in terms of its contribution to the investor's existing portfolio. For example, the Sharpe ratio can be used by an investor to choose one fund in combination with either borrowing or investing in the risk-free asset, depending on the investor's degree of risk tolerance. However, because the Sharpe ratio does not take into account correlations between fund returns, this would not be the best way for an investor to choose several mutual funds or to add a fund to an existing portfolio. Recall that in the example in the previous section, an investor had to choose between Fund X and Fund Y. Fund Y was the better choice in combination with an investment in a risk-free asset than Fund X because Fund Y had a higher Sharpe ratio. However, as long as the returns on X and Y were not perfectly correlated, the investor could do even better with a combination of Funds X and Y and the riskless asset.

It is easy enough to find the efficient portfolio of funds (the one with the lowest risk for a given level of expected return) when one has a choice of a few funds, but it is not so easy to do with a choice of thousands of funds. One way around this problem is provided by modern portfolio theory, first developed by Markowitz (1952). It introduces the concept of the "market portfolio," that is, the portfolio consisting of every security traded in the market held in proportion to its current market value. Moreover, modern portfolio theory divides the risk of each security (or each portfolio of securities such as a mutual fund) into two parts: systematic and unsystematic. Systematic risk (or market risk) is the risk associated with the correlation between the return on the security and the return on the market portfolio. Unsystematic risk (also known as specific risk) is the "leftover" risk, which is associated with the variability of returns of that security alone. The distinction between the two components of risk is important because they behave differently as one increases the number of securities in the portfolio. The unsystematic component of risk can be diversified away because it gets "averaged out" as the number of securities gets larger, and so it can be ignored in a well-diversified portfolio. Systematic risk, on the other hand, cannot be diversified away and investors expect to be compensated for bearing it.

The distinction between systematic and unsystematic risk is the foundation of the Capital Asset Pricing Model (CAPM) developed by Sharpe (1964) and Lintner (1965). The CAPM states that the expected return on a given security or portfolio is determined by three factors: the sensitivity of its return to that of the market portfolio (known as beta), the return on the market portfolio itself and the risk-free rate. (See the Box for a more detailed discussion of the CAPM.)

### *Empirical Estimates of Beta*

The beta can be estimated empirically from a time series of the historical returns on a given investment and the historical returns on the market portfolio. Five years of monthly returns (60 months) are commonly used to estimate beta. The return on the market portfolio is traditionally represented by the return on the S&P 500, though a value-weighted index of all securities in the market may be preferable, given the definition of the market portfolio.

The most common way to estimate beta is a linear regression of the excess return of the given portfolio on the excess return of the market portfolio, where beta is the slope of the regression line:

$$R_p - R_f = \alpha + \beta(R_m - R_f) + \epsilon_p \tag{8}$$

Alpha is the intercept of that regression and can be interpreted as the "extra" return for the fund's level of systematic risk, or the "value added" by the fund's manager. This interpretation of alpha as a measure of performance adjusted for systematic risk was first suggested by Jensen (1968). However, it is important to be careful in the way one interprets this measure in the CAPM framework. In theory, any alpha other than zero is inconsistent with the CAPM because, if the market portfolio is efficient, then the expected return on every security or portfolio of securities is completely determined by its relationship to the market portfolio, as measured by beta. Thus, it is logically inconsistent to apply the CAPM to measure a mutual fund's return over and above the return required to compensate investors for the fund's systematic risk,

[5] Sharpe finds that correlations between Sharpe ratios and Morningstar measures tend to be high when the average fund performs well and the funds have returns and risks tightly clustered around the average fund. Conversely, the correlations are lower when the average fund does poorly and the funds display more variability around the average.

since according to the CAPM it is impossible to earn such extra return. On the other hand, if investors have portfolios that are markedly different from the market portfolio, then a fund's alpha and beta found with reference to the market portfolio may not be relevant for them.

Thus far, we seem to have come to a paradoxical conclusion: We can measure the risk and the risk-adjusted return of a mutual fund on an individual basis by using measures such as the standard deviation and the Sharpe ratio, but this measure does not take account of the effects of diversification. Alternatively, we can use the empirical form of the CAPM to derive the fund's alpha and beta with respect to the market portfolio. However, the CAPM implies that all investors hold the market portfolio, in which case there is no point in analyzing mutual funds, since they would all be inferior to the market portfolio. Also, if the market portfolio is not efficient for all investors, then they would hold different portfolios and alpha and beta may be no more relevant than a simple Sharpe ratio. In theory, an investor can construct an individual efficient portfolio out of mutual funds, subject to the investor's tax status, expectations, holdings of illiquid assets, and so on. But with 5,000 funds to choose from, an investor would have to consider 12.5 million correlations between them to find the efficient portfolio. An even more basic problem for many investors is simply to know what assets they hold in their portfolios at any given time. For an investor with half a dozen funds each holding hundreds of securities, it is not a trivial problem to know what the portfolio consists of, let alone how efficient it is or whether it resembles the market portfolio, however defined.

### *Asset-Class Factor Models*

It is generally agreed that a large part of the differences in investors' portfolio returns can be explained by the allocation of the portfolio among key asset classes. Thus, it is not crucial to consider each individual security separately for inclusion in the portfolio. Instead, one can use an asset-class factor model to evaluate the performance of portfolio managers and construct a portfolio of mutual funds. To do so, one must first define the "key" asset classes and measure how sensitive mutual fund returns are to the variations in asset-class returns.

For example, an asset-class model might include the following: Treasury bills, intermediate-term government bonds, long-term government bonds, corporate bonds, mortgage-based bonds, non-U.S. bonds, U.S. stocks, European stocks, and Asia/Pacific stocks. What makes a "good" asset-class model? According to Sharpe (1992), while not strictly necessary, it is desirable for asset classes to be mutually exclusive and exhaustive and to have returns that "differ." This means that no security should be in more than one asset class, as many securities as possible should be included in a given asset class, and asset returns on different classes of assets should have low correlations and, if this is not possible, different standard deviations. Generally, a good model of asset classes is the one that can explain a large portion of the variance of returns on the assets. If two models can explain this variance equally well, the one with fewer asset classes is preferable because fewer classes are more likely to represent stable economic relationships. An additional practical consideration is that widely available, reliable indexes that could be used as benchmarks should represent the returns on each class of assets accurately.

*Generally, a good model of asset classes is the one that can explain a large portion of the variance of returns on the assets.*

### *Constructing a Benchmark*

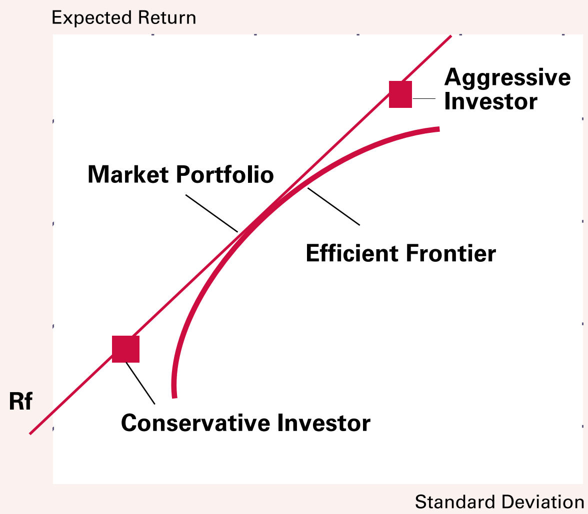
If the asset classes span the market portfolio, the investor still has the problem of comparing the returns on his mutual funds to the return on the whole collection of asset classes. It would be convenient if the investment objectives of every fund neatly corresponded to one asset class. In this case, the index representing the asset class in question would be an appropriate benchmark for measuring the fund's performance. For example, the Russell 1000 index could be used as a benchmark for a fund invested in large-capitalization U.S. stocks, while the MSCI EAFE (Europe/Australia/Far East) Index could be used to benchmark an international stock fund. However, this one-to-one correspondence rarely happens. Many funds invest in a number of asset classes and finding an appropriate benchmark consisting of a "blend" of appropriate indexes is not a straightforward exercise. Some funds shift their asset allocation through time,

## The Capital Asset Pricing Model

The Capital Asset Pricing Model rests on a number of simplifying assumptions. All investors are assumed to be risk averse and to have identical preferences about risk and return. Investors are assumed to care *only* about risk and return, so that their utility function admits only the mean and the variance of the distribution of returns. In addition, the model assumes that all investors have identical expectations about the future risks and returns of all securities, have the same tax rates, and are able to borrow and lend at the risk-free rate without limits on the amounts borrowed or lent, and that no risky assets are excluded from the investment portfolio. Finally, the model assumes that there are no transaction costs and no costs of research.

Figure B-1

*Security Market Line*




To see the implications of this more clearly, we can plot the risks against the expected returns of a number of possible portfolios, as shown in Figure B-1. Among all possible portfolios there will be those where no other combination of (risky) assets would produce a better return for the same level of risk, or equivalently, lower risk for the same return. Such portfolios are known as mean-variance efficient. If we plot a line through them, the result will be the "efficient frontier," as shown in Figure B-1. If no borrowing or lending was allowed, all investors would hold one of these efficient portfolios, depending on their risk tolerance. However, if borrowing and lending are possible, investors can do even better than being somewhere on the efficient frontier. We can see this clearly if we draw a tangent from the efficient frontier starting at the risk-free rate. If investors hold a combination of risky securities that is the same as the one where the

which further complicates the issue. Consider, for example, a balanced fund that is invested 50 percent in U.S. common stocks and 50 percent in U.S. long-term government bonds. Suppose also that during the past five years the fund's stock allocation ranged from 30 percent to 70 percent depending on the manager's view of the market. We could try to construct a benchmark that would mimic the fund's shifts in asset allocation through time. However, such a benchmark would be of questionable value to an investor, even if it were possible to know a fund's asset allocation at any given moment. To be useful, a benchmark for the fund's performance should be a viable investment strategy that can be followed by an investor at a low cost and it should not depend upon the benefit of hindsight. For example, a strategy consisting of investing in a mix of index funds and holding this mix for five years would meet these requirements.

A fixed benchmark can be constructed using either a historical or a hypothetical approach. To use a historical benchmark we would estimate the fund's *average* asset allocation throughout the last five years and compare a return on this asset mix to the fund's own return. A hypothetical approach would be to use the fund's current asset allocation (in this case half stocks and half bonds) and compare the performance on this mix over the last five years to the fund's actual performance. Note that neither of these approaches would require an investor to trade in and out of asset classes.

Often, we do not know the fund's asset allocation. At present, the mutual fund prospectus describes the fund's investment goals, as well as any restrictions on the fund's portfolio composition, such as the ability to use derivatives. However, the description of the fund's goals often is not specific enough to enable line is tangent to the efficient frontier, they can achieve any desired trade-off of risk and return that is possible along that line. This line is known as the security market line and it is also shown in Figure B-1. Borrowing and lending make it possible to separate investors' preferences about risk and return from the opportunities available in the capital market. Thus, each investor would hold the same portfolio of risky assets (the market portfolio) and only the mix of the market portfolio and the risk-free asset would vary.

If the model were literally true and all investors held the same mix of risky assets, talking about measuring risk or performance of mutual funds would be pointless. In fact, only one mutual fund would exist, the universal index fund consisting of the market portfolio; any fund consisting of a different combination of risky assets would be inferior. Recall that this is the same line of reasoning we used in describing why the Sharpe ratio (or the Modigliani measure) is the relevant measure of risk-adjusted performance. The Sharpe ratio measures the amount of expected excess return per unit of risk. If investors can borrow and lend, they can invest in the portfolio with the highest Sharpe ratio and mix it with the risk-free asset in different proportions. Thus, it follows that if the CAPM holds, then the market portfolio is, in fact, the portfolio with the highest Sharpe ratio.

Two problems arise in applying the CAPM to real-world investment. One problem concerns the definition of the "market portfolio," and the second, the definition of the "efficient portfolio." Roll (1977) pointed out that the CAPM can never be definitely tested because, as a practical matter, it is impossible to define the "market portfolio" with any degree of precision. Should foreign assets be included? How about commodities? Real estate? Antiques? Art? Some of these assets are traded so infrequently that it would be quite difficult to construct a reliable series of monthly returns. Finally, some assets, such as the present value of the investor's labor income, cannot be traded at all, yet they constitute an important part of the investor's overall "portfolio." Generally, as the definition of the "market" becomes broader, the estimate of its monthly returns becomes less reliable.

The second problem is that no one truly "efficient" portfolio exists that would be appropriate for all investors. Because research is costly, not all investors have access to the same information, nor do they have the same opinions and beliefs. As long as investors have differing expectations about the future risks and returns of various investments, they will not agree on the same "efficient" portfolio but rather choose securities that have the best prospects according to their own judgment. In this case, instead of being efficient in some absolute sense, the market portfolio balances the divergent assessments of all investors (Lintner 1965).

the investor to assign a specific mix of asset classes to the fund with any degree of precision. In addition to the disclosures found in the prospectus, the SEC requires management to disclose the list of securities owned by the fund every six months. In principle, by studying this list of securities the investor can determine to which class each belongs and, thus, determine the fund's asset allocation. This approach has two problems, however. First, each fund typically holds hundreds of securities and categorizing each one is a difficult and time-consuming task. Second, the investor would be looking at the fund composition six months ago, which may not necessarily represent its composition now or in the future. Reporting holdings on a more timely basis would not be an acceptable solution, however. For competitive reasons, many funds regard their current holdings as proprietary information. Disclosing them can hurt the fund's performance by enabling other market participants to trade against them, especially if some positions are large or illiquid.

Thus, it is often necessary to construct a benchmark for the fund's asset allocation without knowing the fund's actual holdings. If the fund does not shift its asset allocation through time, it is possible to analyze its historical performance with respect to a number of previously defined asset classes. For example, one can regress the fund's monthly returns on the monthly returns to the indexes chosen to represent asset classes for the set amount of time, say 60 months, as shown in Equation 9:

$$R_p - R_f = b_1(R_1 - R_f) + b_2(R_2 - R_f) + \ldots + b_n(R_n - R_f) + e_p \quad (9)$$

where $R_p$ is the expected return on the fund, $R_f$ is the return on the risk-free asset, $R_1$ through $R_n$ are the returns on asset classes 1 through n, and $b_1$ through $b_n$ are the corresponding investments of $R_p$ to these asset classes. Finally, $e_p$ is the residual, or non-factor, return on the fund. It can be seen as the "value added" (or subtracted, as the case may be) by the fund manager relative to the return the investor could get by investing in a benchmark consisting of index funds representing the same asset classes.

The resulting slope coefficients would then represent the sensitivities of the fund's returns to the returns of the corresponding asset classes. However, the results of such regressions are often difficult to interpret, because the coefficients do not sum to one and often some coefficients turn out to be negative. Mutual funds normally do not take short positions in asset classes and their investments in various assets should sum to 100 percent. Thus, to be meaningful, coefficients should be constrained to be positive or zero and to sum to one. The presence of inequality constraints ($0 \leq b_i \leq 100\%$) necessitates the use of quadratic programming for the estimation of the fund's exposure to the asset classes. This method, introduced by Sharpe, has become known as "style analysis." It involves finding the set of asset class exposures ($b_i$s) that minimize the variance of the fund's residual return VAR($e_p$) and are consistent with the above constraint. Note that style analysis represents a form of historical approach, which estimates the fund's average exposure to asset classes during the period analyzed.

*A large part of the differences in investors' portfolio returns can be explained by the allocation of the portfolio among key asset classes.*

Yet another approach to estimating risk and performance of mutual funds was recommended by the Financial Economists Roundtable in its "Statement on Risk Disclosure by Mutual Funds" issued in September 1996. This approach is future oriented because it calls for disclosure by fund managers of asset allocations they plan to have in their funds for a specified future period. Specifically, the Roundtable recommended that funds use narrowly defined asset classes for this disclosure, rather than broad ones like the S&P 500. The statement listed 14 possible asset classes that are represented well by available indexes and suggested that funds specify not just one index, but a portfolio of indexes, whenever appropriate. The Roundtable also recommended that funds report historical comparisons of their returns with the returns that could have been obtained by investing in an index fund or a portfolio of index funds corresponding to their *previously announced* index or blend of indexes. This information would be sufficient to evaluate whether or not a given fund fits the investor's portfolio in terms of asset allocation and, if it does, whether the investor would be better off investing in the fund or in an index fund representing the same asset class.

Another performance measure that is derived from comparing a fund to its benchmark is the "information ratio," defined as follows:

$$\textit{Information ratio} = \frac{\textit{fund return} - \textit{benchmark return}}{\textit{standard deviation}\ (\textit{fund return} - \textit{benchmark return})}\,. \tag{10}$$

This is another version of the Sharpe ratio, where instead of dividing the fund's return in excess of the risk-free rate by its standard deviation, we divide the fund's return in excess of the return on the benchmark index by *its* standard deviation. The information ratio can be thought of as a more general measure of which the "regular" Sharpe ratio is a special case with the return on Treasury bills used as the benchmark for all funds. It should be noted that a ranking of funds based on the information ratio will generally differ from the one based on the regular, or excess return, Sharpe ratio, and its relevance to an investor's decision-making is not obvious.

## V. Summary and Conclusion

Portfolio theory teaches us that investment choices are made on the basis of expected risks and returns. These expectations are often formed on the basis of a historical record of monthly returns, measured for a period of time. For mutual funds, common measures include average excess return (total monthly return less the monthly return on Treasury bills) and its standard deviation, tracked for a sufficient length of time, such as three or five years. A fund's risk and return can be combined into one measure of risk-adjusted performance by dividing the average excess return by the standard deviation. The resulting measure, known as the Sharpe ratio, can help the investor to identify the most "efficient" fund, namely the one with the highest return per unit of risk. However, a universal measure such as the Sharpe ratio is useful as a guide to investment decisions only in a limited set of circumstances. In particular, the measure is useful to investors who are putting all their money into one diversified fund and are able to use leverage or invest in the risk-free asset.

Much more common is the situation where an investor constructs a portfolio of funds or adds a fund to an existing portfolio. In this case, the fund's marginal contribution to the portfolio's risk and return is more important than its individual characteristics. To construct an efficient portfolio, an investor must take account of the correlations among the investments being considered. The Capital Asset Pricing Model implies that under certain assumptions, the efficient portfolio is the same for all investors and in the aggregate constitutes the market portfolio. Taken literally, this implies that all investors should invest in a universal index fund. Because it is efficient, the universal index fund would, by definition, have the highest Sharpe ratio of all mutual funds.

Among the reasons why this does not happen is the fact that both "efficient" portfolio and "market" portfolio are difficult to define in practice. In particular, because of their different tax treatments, assessments of future asset returns, and endowments of non-tradable assets, investors cannot all have the same efficient portfolio. For example, an owner of a private business who has a substantial part of his wealth tied up in the business will have very different needs for diversification than someone who does not. Similarly, the "market portfolio" is an abstraction. In theory, it should consist of a value-weighted index of all assets, but many assets are illiquid or non-tradable and their prices are not known with any certainty.

Despite these caveats, the main insight of the CAPM remains sound: For the aggregate supply of all securities in the market to equal the aggregate demand for these securities, their expected returns must compensate investors for systematic risk. These returns tell an investor how much he can expect to be rewarded for bearing the systematic risk of a given security or fund. This approach has led to the use of risk-adjusted excess return (alpha) as a measure of performance. The excess return implies that the manager of that fund has delivered a return over and above that which is required to compensate investors for bearing market risk, where the market is represented as a broad-based index such as the S&P 500.

A different version of alpha is measured against a specific benchmark for the fund, rather than against the market as a whole. Similarly, a benchmark-related version of the Sharpe ratio, known as the information ratio, is based on excess returns over the benchmark rather than the risk-free rate. An investor can then choose one or more funds with the highest alphas or information ratios in their categories. However, if the categories themselves and their shares in the investor's portfolio are chosen arbitrarily, the resulting portfolio can be highly inefficient. This is because the excess return measured by the benchmark alpha is related not to the market risk of the fund, but to its "risk" relative to the benchmark. This, however, tells us nothing about the expected return and risk of the benchmark itself. Similarly, the information ratio uses the "tracking error," or the standard deviation of the difference between the fund return and the benchmark return, which is of questionable relevance as a measure of risk for most investors.

*To be useful, a benchmark for a fund's performance should be a viable investment strategy that can be followed by an investor at a low cost and it should not depend on the benefit of hindsight.*

An approach known as asset allocation divides all securities into several asset classes and tries to construct an efficient portfolio based on expected returns, risks, and correlations of indexes representing these asset classes. In this context, an "efficient" portfolio is simply a portfolio invested in the benchmark indexes in such a way that no other combination of these indexes would result in a portfolio with a higher return for a given level of risk. It should be emphasized, however, that this is not a fully efficient portfolio because information about correlations among individual securities within an index and across the indexes is lost in the transition from individual securities to the benchmarks that represent them.

It is quite likely that a more efficient portfolio can be constructed directly from funds that are not the best performers in their categories because they offset one another's risks better. However, the logistical problems of constructing a correlation matrix among thousands (or even hundreds) of possible funds to consider makes it an unrealistic exercise in most cases, at least for individual investors. Thus, the two-step process of choosing an asset allocation based on the information about benchmark indexes and then choosing funds in each category may be the best realistically attainable approach. To use this approach to portfolio selection effectively, investors would benefit from estimates of future asset returns, risks, and correlations, as well as from fund management's disclosure of future asset exposures and appropriate benchmarks.

## *References*


Damato, Karen. 1996. "Morningstar Edges toward One-Year Ratings." *The Wall Street Journal*, April 6, p. C1.

Financial Economists Roundtable. 1996. "Statement on Risk Disclosure by Mutual Funds." September. http://www-sharpe.stanford.edu/fer/htm.

Glauber, Robert. 1998. "Relative Values." *Risk*, vol. 11, no. 1 (January), pp. 39–40.

Group of Thirty. 1993. *Derivatives: Practices and Principles*. Washington, DC.

Harrel, David. 1998. "The Star Rating." http://text.morningstar.net/cgiin/GetNews.exe?NewsStory=MS/Investing101/StarRating/star.html

Jaffe, Charles A. 1998. "Don't Be Duped by Alphabet Soup." *The Boston Globe*, March 9, 1998, pp. A14, A16.

Jensen, Michael C. 1968. "The Performance of Mutual Funds in the Period 1945–1964." *Journal of Finance*, vol. 23, (May) pp. 389–416.

Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *The Review of Economics and Statistics*, vol. 47 (February), pp. 13–37.

Markowitz, Harry. 1952. "Portfolio Selection." *Journal of Finance*, vol. 7 (March), pp. 77–91.

Modigliani, Franco and Leah Modigliani. 1997. "Risk-Adjusted Performance." *Journal of Portfolio Management*, vol. 23 (2) (Winter), pp. 45–54.

Roll, Richard. 1977. "A Critique of the Asset Pricing Theory's Tests." *Journal of Financial Economics*, vol. 4 (March), pp. 129–76.

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk." *Journal of Finance*, vol. 19 (September), pp. 425–42.

———. 1966. "Mutual Fund Performance." *Journal of Business, Supplement on Security Prices*, vol. 39 (January), pp. 119–38.

———. 1992. "Asset Allocation: Management Style & Performance Measurement. *Journal of Portfolio Management*, vol. 18 (2), Winter 1992, pp. 7–19.

———. 1997. "Morningstar Performance Measures." http://www-sharpe.stanford.edu/stars0/htm.