PART I

OVERVIEW OF PERFORMANCE EVALUATION

COPYRIGHTED MATERIAL

CHAPTER 1

EVALUATING PORTFOLIO PERFORMANCE

Jeffery V. Bailey, CFA

Thomas M. Richards, CFA

David E. Tierney

The *ex post* analysis of investment performance stands as a prominent and ubiquitous feature of modern investment management practice. Investing involves making decisions that have readily quantifiable consequences and that, at least on the surface, lend themselves to elaborate dissection and review. We broadly refer to the measurement and assessment of the outcomes of these investment management decisions as **performance evaluation.** At the institutional investor level, and to a lesser (but growing) extent on the individual investor level, a large industry has developed to satisfy the demand for performance evaluation services. Although some observers contend that performance evaluation is misguided, frequently misapplied, or simply unattainable with any reasonable degree of statistical confidence, we believe that analytic techniques representing best practices can lead to valid insights about the sources of past returns, and such insights can be useful inputs for managing an investment program.

The purpose of this chapter is to provide an overview of current performance evaluation concepts and techniques. Our focus will be on how institutional investors—both fund sponsors and investment managers—conduct performance evaluation. Individual investors tend to use variations of the performance evaluation techniques employed by institutional investors. We define fund sponsors to be owners of large pools of investable assets, such as corporate and public pension funds, endowments, and foundations. These organizations typically retain multiple investment management firms deployed across a range of asset categories. Fund sponsors have the challenge of evaluating not only the performance of the individual managers, but also the investment results within the asset categories and for their total investment programs.

Reprinted from *Managing Investment Portfolios: A Dynamic Process*, 3rd Edition (John Wiley & Sons, 2007):717–780.

In the section titled The Importance of Performance Evaluation, we distinguish between the perspectives of the fund sponsor and the investment manager. In The Three Components of Performance Evaluation, we divide the broad subject of performance evaluation into three components: **performance measurement, performance attribution**, and **performance appraisal**. Under the topic of performance measurement, we discuss several methods of calculating portfolio performance. The next section introduces the concept of performance benchmarks. Turning to performance attribution, we consider the process of analyzing the sources of returns relative to a designated benchmark both from the total fund (fund sponsor) level and from the individual portfolio (investment manager) level. This is followed by performance appraisal, which deals with assessing investment skill. The chapter ends by addressing key issues in the practice of performance evaluation.

THE IMPORTANCE OF PERFORMANCE EVALUATION

Performance evaluation is important from the perspectives of both the fund sponsor and the investment manager.

The Fund Sponsor's Perspective

A typical fund sponsor would consider its investment program incomplete without a thorough and regular evaluation of the fund's performance relative to its investment objectives. Applied in a comprehensive manner, performance evaluation is more than a simple exercise in calculating rates of return. Rather, it provides an exhaustive "quality control" check, emphasizing not only the performance of the fund and its constituent parts relative to objectives, but the sources of that relative performance as well.

Performance evaluation is part of the feedback step of the investment management process. As such, it should be an integral part of a fund's investment policy and documented in its investment policy statement. As discussed in Ambachtsheer (1986) and Ellis (1985), investment policy itself is a combination of philosophy and planning. On the one hand, it expresses the fund sponsor's attitudes toward a number of important investment management issues, such as the fund's mission, the fund sponsor's risk tolerance, the fund's investment objectives, and so on. On the other hand, investment policy is a form of long-term strategic planning. It defines the specific goals that the fund sponsor expects the fund to accomplish, and it describes how the fund sponsor foresees the realization of those goals.

Investment policy gives an investment program a sense of direction and discipline. Performance evaluation enhances the effectiveness of a fund's investment policy by acting as a feedback and control mechanism. It identifies an investment program's strengths and weaknesses and attributes the fund's investment results to various key decisions. It assists the fund sponsor in reaffirming a commitment to successful investment strategies, and it helps to focus attention on poorly performing operations. Moreover, it provides evidence to fund trustees, who ultimately bear fiduciary responsibility for the fund's viability, that the investment program is being conducted in an appropriate and effective manner.

Fund sponsors are venturing into nontraditional asset categories and hiring a larger assortment of managers exhibiting unique investment styles, with the addition of hedge fund managers representing the latest and perhaps most complex example of this trend. Some fund sponsors are taking more investment decisions into their own hands, such as tactical asset allocation and style timing. Others are taking a quite different direction, giving their managers broad discretion to make asset allocation and security selection decisions. As a consequence of these developments, alert trustee boards are demanding more information from their investment staffs. The staffs, in turn, are seeking to better understand the extent of their own contributions and those of the funds' investment managers to the funds' investment results. The increased complexity of institutional investment management has brought a correspondingly greater need for sophisticated performance evaluation from the fund sponsor's perspective.

The Investment Manager's Perspective

Investment managers have various incentives to evaluate the performance of the portfolios that they manage for their clients. Virtually all fund sponsors insist that their managers offer some type of accounting of portfolio investment results. In many cases, performance evaluation conducted by the investment manager simply takes the form of reporting investment returns, perhaps presented alongside the returns of some designated benchmark. Other clients may insist on more sophisticated analyses, which the managers may produce in-house or acquire from a third party.

Some investment managers may seriously wish to investigate the effectiveness of various elements of their investment processes and examine the relative contributions of those elements. Managing investment portfolios involves a complex set of decision-making procedures. For example, an equity manager must make decisions about which stocks to hold, when to transact in those stocks, how much to allocate to various economic sectors, and how to allocate funds between stocks and cash. Numerous analysts and portfolio managers may be involved in determining a portfolio's composition. Just as in the case of the fund sponsor, performance evaluation can serve as a feedback and control loop, helping to monitor the proficiency of various aspects of the portfolio construction process.

THE THREE COMPONENTS OF PERFORMANCE EVALUATION

In light of the subject's importance to fund sponsors and investment managers alike, we want to consider the primary questions that performance evaluation seeks to address. In discussing performance evaluation we shall use the term *account* to refer generically to one or more portfolios of securities, managed by one or more investment management organizations. Thus, at one end of the spectrum, an account might indicate a single portfolio invested by a single manager. At the other end, an account could mean a fund sponsor's total fund, which might involve numerous portfolios invested by many different managers across multiple asset categories. In between, it might include all of a fund sponsor's assets in a particular asset category or the aggregate of all of the portfolios managed by an investment manager according to a particular mandate. The basic performance evaluation concepts are the same, regardless of the account's composition.

With the definition of an account in mind, three questions naturally arise in examining the investment performance of an account:

1. What was the account's performance?
2. Why did the account produce the observed performance?
3. Is the account's performance due to luck or skill?

In somewhat simplistic terms, these questions constitute the three primary issues of performance evaluation. The first issue is addressed by performance measurement, which calculates rates of return based on investment-related changes in an account's value over specified time periods. Performance attribution deals with the second issue. It extends the results of performance measurement to investigate both the sources of the account's performance relative to a specific investment benchmark and the importance of those sources. Finally, performance appraisal tackles the third question. It attempts to draw conclusions concerning the quality (that is, the magnitude and consistency) of the account's relative performance.

PERFORMANCE MEASUREMENT

To many investors, performance measurement and performance evaluation are synonymous. However, according to our classification, performance measurement is a component of performance evaluation. Performance measurement is the relatively simple procedure of calculating returns for an account. Performance evaluation, on the other hand, encompasses the broader and much more complex task of placing those investment results in the context of the account's investment objectives.

Performance measurement is the first step in the performance evaluation process. Yet it is a critical step, because to be of value, performance evaluation requires accurate and timely rate-of-return information. Therefore, we must fully understand how to compute an account's returns before advancing to more involved performance evaluation issues.

Performance Measurement without Intraperiod External Cash Flows

The rate of return on an account is the percentage change in the account's market value over some defined period of time (the evaluation period), after accounting for all external cash flows.[1] (External cash flows refer to contributions and withdrawals made to and from an account, as opposed to internal cash flows such as dividends and interest payments.) Therefore, a rate of return measures the relative change in the account's value due solely to investment-related sources, namely capital appreciation or depreciation and income. The mere addition or subtraction of assets to or from the account by the account's owner should not affect the rate of return. Of course, in the simplest case, the account would experience no external cash flows. In that situation, the account's rate of return during evaluation period t equals the market value (MV_1) at the end of the period less the market value at the beginning of the period (MV_0), divided by the beginning market value.[2] That is,

$$r_t = \frac{MV_1 - MV_0}{MV_0} \tag{1.1}$$

Example 1.1 illustrates the use of Equation 1.1.

EXAMPLE 1.1 Rate-of-Return Calculations When There Are No External Cash Flows

Winter Asset Management manages institutional and individual accounts, including the account of the Mientkiewicz family. The Mientkiewicz account was initially valued at \$1,000,000. One month later it was worth \$1,080,000. Assuming no external cash flows and the reinvestment of all income, applying Equation 1.1, the return on the Mientkiewicz account for the month is

$$r_t = \frac{\$1{,}080{,}000 - \$1{,}000{,}000}{\$1{,}000{,}000} = 8.0\%$$

Fund sponsors occasionally (and in some cases frequently) add and subtract cash to and from their managers' accounts. These external cash flows complicate rate-of-return calculations. The rate-of-return algorithm must deal not only with the investment earnings on the initial assets in the account, but also with the earnings on any additional assets added to or subtracted from the account during the evaluation period. At the same time, the algorithm must exclude the direct impact of the external cash flows on the account's value.

An account's rate of return may still be computed in a straightforward fashion if the external cash flows occur at the beginning or the end of the measurement period when the account is valued. If a contribution is received at the start of the period, it should be added to (or, in the case of a withdrawal, subtracted from) the account's beginning value when calculating the account's rate of return for that period. The external cash flow will be invested alongside the rest of the account for the full length of the evaluation period and will have the same investment-related impact on the account's ending market value and, hence, should receive a full weighting. Thus, the account's return in the presence of an external cash flow at the beginning of the evaluation period should be calculated as

$$r_t = \frac{MV_1 - (MV_0 + CF)}{MV_0 + CF} \tag{1.2}$$

If a contribution is received at the end of the evaluation period, it should be subtracted from (or, in the case of a withdrawal, added to) the account's ending value. The external cash flow had no opportunity to affect the investment-related value of the account, and hence it should be ignored.

$$r_t = \frac{(MV_1 - CF) - MV_0}{MV_0} \tag{1.3}$$

EXAMPLE 1.2 Rate-of-Return Calculations When External Cash Flows Occur at the Beginning or End of an Evaluation Period

Returning to the example of the Mientkiewicz account, assume that the account received a $50,000 contribution at the beginning of the month. Further, the account's ending and beginning market values equal the same amounts previously stated, $1,080,000 and $1,000,000, respectively. Applying Equation 1.2, the rate of return for the month is therefore

$$r_t = \frac{\$1{,}080{,}000 - (\$1{,}000{,}000 + \$50{,}000)}{\$1{,}000{,}000 + \$50{,}000} = 2.86\%$$

If the contribution had occurred at month-end, using Equation 1.3, the account's return would be

$$r_t = \frac{(\$1{,}080{,}000 - \$50{,}000) - \$1{,}000{,}000}{\$1{,}000{,}000} = 3.00\%$$

Both returns are less than the 8 percent return reported when no external cash flows took place because we are holding the ending account value fixed at $1,080,000. In the case of the beginning-of-period contribution, the account achieves an ending value of $1,080,000 on a beginning value that is higher than in Example 1.1, so its return must be less than 8 percent. In the case of the end-of-period contribution, the return is lower than 8 percent because the ending value of $1,080,000 is assumed to reflect an end-of-period contribution that is removed in calculating the return. In both instances, a portion of the account's change in value from $1,000,000 to $1,080,000 resulted from the contribution; in Example 1.1, by contrast, the change in value resulted entirely from positive investment performance by the account.[3]

The ease and accuracy of calculating returns when external cash flows occur, if those flows take place at the beginning or end of an evaluation period, lead to an important practical recommendation: Whenever possible, a fund sponsor should make contributions to, or withdrawals from, an account at the end of an evaluation period (or equivalently, the beginning of the next evaluation period) when the account is valued. In the case of accounts that are valued on a daily basis, the issue is trivial. However, despite the increasing prevalence of daily valued accounts, many accounts are still valued on an audited basis once a month (or possibly less frequently), and the owners of those accounts should be aware of the potential for rate-of-return distortions caused by intraperiod external cash flows.

What does happen when external cash flows occur between the beginning and the end of an evaluation period? The simple comparison of the account's value relative to the account's beginning value must be abandoned in favor of more intricate methods.

Total Rate of Return

Interestingly, widely accepted solutions to the problem of measuring returns in the presence of intraperiod external cash flows are relatively recent developments. Prior to the 1960s, the issue received little attention, largely because the prevailing performance measures were unaffected by such flows. Performance was typically measured on an income-only basis, thus excluding the impact of capital appreciation. For example, current yield (income-to-price) and yield-to-maturity were commonly quoted return measures.

The emphasis on income-related return measures was due to several factors:

- **Portfolio management emphasis on fixed-income assets.** Particularly in the low-volatility interest rate environment that existed prior to the late 1970s, bond prices tended to be stable. Generally high allocations to fixed-income assets made income the primary source of investment-related wealth production for many investors.
- **Limited computing power.** Accurately accounting for external cash flows when calculating rates of return that include capital appreciation requires the use of computers. Access to the necessary computing resources was not readily available. The income-related return measures were simpler and could be performed by hand.
- **Less competitive investment environment.** Investors, as a whole, were less sophisticated and less demanding of accurate performance measures.

As portfolio allocations to equity securities increased, as computing costs declined, and as investors (particularly larger institutional investors) came to focus more intently on the performance of their portfolios, the demand grew for rate-of-return measures that correctly incorporated all aspects of an account's investment-related increase in wealth. This demand led to the adoption of total rate of return as the generally accepted measure of investment performance.

Total rate of return measures the increase in the investor's wealth due to both investment income (for example, dividends and interest) and capital gains (both realized and unrealized). The total rate of return implies that a dollar of wealth is equally meaningful to the investor whether that wealth is generated by the secure income from a 90-day Treasury bill or by the unrealized appreciation in the price of a share of common stock.

Acceptance of the total rate of return as the primary measure of investment performance was assured by a seminal study performed in 1968 by the Bank Administration Institute (BAI). The BAI study (which we refer to again shortly) was the first comprehensive research conducted on the issue of performance measurement. Among its many important contributions, the study strongly endorsed the use of the total rate of return as the only valid measure of investment performance. For our purposes, henceforth, it will be assumed that rate of return refers to the total rate of return, unless otherwise specified.

The Time-Weighted Rate of Return

We now return to considering the calculation of rates of return in the context of intraperiod external cash flows. To fully appreciate the issue at hand, we must think clearly about the meaning of "rate of return." In essence, the rate of return on an account is the investment-related growth rate in the account's value over the evaluation period. However, we can envision this growth rate being applied to a single dollar invested in the account at the start of the evaluation period or to an "average" amount of dollars invested in the account over the evaluation period. This subtle but important distinction leads to two different measures: the time-weighted and the money-weighted rates of return.

The **time-weighted rate of return** (TWR) reflects the compound rate of growth over a stated evaluation period of one unit of money initially invested in the account. Its calculation requires that the account be valued every time an external cash flow occurs. If no such flows take place, then the calculation of the TWR is trivial; it is simply the application of Equation 1.1, in which the change in the account's value is expressed relative to its beginning value. If external cash flows do occur, then the TWR requires computing a set of subperiod returns (with the number of subperiods equaling one plus the number of dates on which external cash flows occur). These subperiod returns must then be linked together in computing the TWR for the entire evaluation period.

EXAMPLE 1.3 Calculating Subperiod Rates of Return

Returning again to the Mientkiewicz account, let us assume that the account received two cash flows during month *t*: a contribution of \$30,000 on day 5 and a contribution of \$20,000 on day 16. Further, assume that we use a daily pricing system that provides us with values of the Mientkiewicz account (inclusive of the contributions) of \$1,045,000 and \$1,060,000 on days 5 and 16 of the month, respectively. We can then calculate three separate subperiod returns using the rate-of-return computation applicable to situations when external cash flows occur at the end of an evaluation period, as given by Equation 1.3:

Subperiod 1 = Days 1–5

Subperiod 2 = Days 6–16

Subperiod 3 = Days 17–30

For subperiod 1:

$$
\begin{aligned}
r_{t,1} &= [(\$1{,}045{,}000 - \$30{,}000) - \$1{,}000{,}000]/\$1{,}000{,}000 \\
&= 0.0150 \\
&= 1.50\%
\end{aligned}
$$

For subperiod 2:

$$
\begin{aligned}
r_{t,2} &= [(\$1{,}060{,}000 - \$20{,}000) - \$1{,}045{,}000]/\$1{,}045{,}000 \\
&= -0.0048 \\
&= -0.48\%
\end{aligned}
$$

For subperiod 3:

$$
\begin{aligned}
r_{t,3} &= (\$1{,}080{,}000 - \$1{,}060{,}000)/\$1{,}060{,}000 \\
&= 0.0189 \\
&= 1.89\%
\end{aligned}
$$

The subperiod returns can be combined through a process called **chain-linking**. Chain-linking involves first adding 1 to the (decimal) rate of return for each subperiod to create a set of wealth relatives. A **wealth relative** can be thought of as the ending value of one unit of money (for example, one dollar) invested at each subperiod's rate of return. Next, the wealth relatives are multiplied together to produce a cumulative wealth relative for the full period, and 1 is subtracted from the result to obtain the TWR. Note that this process of chain-linking implicitly assumes that the initially invested dollar and earnings on that dollar are reinvested (or compounded) from one subperiod to the next. The cumulative wealth relative from the chain-linking of the subperiod wealth relatives can be interpreted as the ending value of one dollar invested in the account at the beginning of the evaluation period. Subtracting 1 from this wealth relative produces the TWR for the account:

$$r_{\text{twr}} = (1 + r_{t,1}) \times (1 + r_{t,2}) \times \cdots \times (1 + r_{t,n}) - 1 \tag{1.4}$$

Note that unless the subperiods constitute a year, the time-weighted rate of return will not be expressed as an annual rate. Example 1.4 illustrates the calculation of a time-weighted rate of return.

EXAMPLE 1.4 Calculating the TWR

Continuing with the Mientkiewicz account, its TWR is

$$\begin{aligned} r_{\text{twr}} &= (1 + 0.0150) \times (1 + -0.0048) \times (1 + 0.0189) - 1 \\ &= 0.0292 \\ &= 2.92\% \end{aligned}$$

The TWR derives its name from the fact that each subperiod return within the full evaluation period receives a weight proportional to the length of the subperiod relative to the length of the full evaluation period. That relationship becomes apparent if each subperiod return is expressed as the cumulative return over smaller time units. In the Mientkiewicz account example, the return in the first subperiod is 0.015 over five days. On a daily compounded basis that return is $0.0030[=(1 + 0.015)^{1/5} - 1]$. Performing the same calculation for the other two subperiods yields the following:

$$\begin{aligned} r_{\text{twr}} &= (1 + 0.0030)^5 \times (1 + -0.0004)^{11} \times (1 + 0.0013)^{14} - 1 \\ &= 0.0292 = 2.92\% \text{ (allowing for rounding)} \end{aligned}$$

From this expression for the TWR, we can see that the subperiods 1, 2, and 3 receive compounding "weights" of 5/30, 11/30, and 14/30, respectively.

The Money-Weighted Rate of Return

The **money-weighted rate of return** (MWR) measures the compound growth rate in the value of all funds invested in the account over the evaluation period. In the corporate finance literature, the MWR goes by the name **internal rate of return,** or IRR. Of importance for performance measurement, the MWR is the growth rate that will link the ending value of the account to its beginning value plus all intermediate cash flows. With MV_1 and MV_0 the values of the account at the end and beginning of the evaluation period, respectively, in equation form the MWR is the growth rate R that solves

$$MV_1 = MV_0(1 + R)^m + CF_1(1 + R)^{m-L(1)} + \cdots + CF_n(1 + R)^{m-L(n)} \tag{1.5}$$

where

m = number of time units in the evaluation period (for example, the number of days in the month)

CF_i = the ith cash flow

$L(i)$ = number of time units by which the ith cash flow is separated from the beginning of the evaluation period

Note that R is expressed as the return per unit of time composing the evaluation period. For example, in the case of monthly performance measurement, where the constituent time unit is one day, R would be the daily MWR of the account. Extending this thought, $[(1 + R)^m - 1]$ can be seen to be the account's MWR for the entire evaluation period, as $(1 + R)^m = (1 + r_{mwr})$. Therefore, in the case of no external cash flows, with some algebraic manipulation Equation 1.4 reduces to Equation 1.1, the simple expression for rate of return:

$$\begin{aligned} MV_1 &= MV_0(1 + R)^m + 0 \\ (1 + R)^m &= MV_1/MV_0 \\ (1 + r_{mwr}) &= MV_1/MV_0 \\ r_{mwr} &= (MV_1 - MV_0)/MV_0 \\ &= r_t \end{aligned}$$

EXAMPLE 1.5 Calculating the MWR

Consider the Mientkiewicz account again. Its MWR is found by solving the following equation for R:

$$\$1{,}080{,}000 = \$1{,}000{,}000(1 + R)^{30} + \$30{,}000(1 + R)^{30-5} + \$20{,}000(1 + R)^{30-16}$$

There exists no closed-form solution for R. That is, Equation 1.4 cannot be manipulated to isolate R on the left-hand side. Consequently, R must be solved iteratively through a trial-and-error process. In our example, we begin with an initial guess of $R = 0.001$. The right-hand side of the equation becomes \$1,081,480. Thus our initial guess is too high and must be lowered. Next, try a value $R = 0.0007$. In this case the right-hand side now equals \$1,071,941. Therefore, our second guess is too low.

We can continue this process. Eventually, we will arrive at the correct value for R, which for the Mientkiewicz account is 0.0009536. Remember that this value is the Mientkiewicz account's daily rate of return during the month. Expressed on a monthly basis, the MWR is 0.0290 $[=(1 + 0.0009536)^{30} - 1]$, or 2.90%.

As one might expect, a computer is best suited to perform the repetitious task of calculating the MWR. Spreadsheet software to perform these computations is readily available.

TWR versus MWR

The MWR represents the average growth rate of all money invested in an account, while the TWR represents the growth of a single unit of money invested in the account. Consequently, the MWR is sensitive to the size and timing of external cash flows to and from the account, while the TWR is unaffected by these flows. Under "normal" conditions, these two return measures will produce similar results. In the example of the Mientkiewicz account, the MWR was 2.90 percent for the month and the TWR was 2.92 percent.

However, when external cash flows occur that are large relative to the account's value and the account's performance is fluctuating significantly during the measurement period, then the MWR and the TWR can differ materially.

EXAMPLE 1.6 When TWR and MWR Differ

Consider the Charlton account, worth \$800,000 at the beginning of the month. On day 10 it is valued at \$1.8 million after receiving a \$1 million contribution. At the end of the month, the account is worth \$3 million. As a result, the Charlton account's MWR is 87.5 percent, while its TWR is only 66.7 percent.
For subperiod 1:

$$r_{t,1} = [(\$1,800,000 - \$1,000,000) - \$800,000]/\$800,000$$
$$= 0.0 \text{ or } 0\%$$

For subperiod 2:

$$r_{t,2} = (\$3,000,000 - \$1,800,000)/\$1,800,000$$
$$= 0.6667 \text{ or } 66.7\%$$

Then

$$r_{\text{twr}} = (1 + 0) \times (1 + 0.667) - 1$$
$$= 0.667 \text{ or } 66.7\%$$

For MWR, we need to solve:

$$\$3,000,000 = \$800,000(1 + R)^{30} + \$1,000,000(1 + R)^{30-10}$$

By trial and error, R comes out to be 0.020896. Expressed on a monthly basis, MWR is 0.859709 or 86.0%$[=(1 + 0.020896)^{30} - 1]$.

If funds are contributed to an account prior to a period of strong performance (as in the case of the Charlton account in Example 1.6), then the MWR will be positively affected compared to the TWR, as a relatively large sum is invested at a high growth rate. That is, in the case of the Charlton account, a contribution was made just prior to a subperiod in which a dollar invested in the account earned 66.7 percent. In the prior subperiod, the account earned 0.0 percent. Thus, on average, the account had more dollars invested earning

66.7 percent than it had dollars earning 0.0 percent, resulting in an MWR greater than the TWR. Conversely, if funds are withdrawn from the account prior to the strong performance, then the MWR will be adversely affected relative to the TWR. (The opposite conclusions hold if the external cash flow occurred prior to a period of weak performance.)

As noted, the TWR is unaffected by external cash flow activity. Valuing the account at the time of each external cash flow effectively removes the impact of those flows on the TWR. Consequently, the TWR accurately reflects how an investor would have fared over the evaluation period if he or she had placed funds in the account at the beginning of the period.

In most situations, an investment manager has little or no control over the size and timing of external cash flows into or out of his or her accounts. Therefore, practitioners generally prefer a rate-of-return measure that is not sensitive to cash flows if they want to evaluate how a manager's investment actions have affected an account's value. This consideration led the authors of the Bank Administration Institute study to recommend that the TWR be adopted as the appropriate measure of account performance. That recommendation has received almost universal acceptance since the study's publication. (Note that the Global Investment Performance Standards [GIPS®] generally require a TWR methodology.)

However, one can readily conceive of situations in which the MWR may prove useful in evaluating the returns achieved by an investment manager. The most obvious examples are those situations in which the investment manager maintains control over the timing and amount of cash flows into the account. Managers of various types of private equity investments typically operate under arrangements that permit them to call capital from their investors at the managers' discretion and ultimately to determine when the original capital, and any earnings on that capital, will be returned to investors. In these "opportunistic" situations, it is generally agreed that the MWR is the more appropriate measure of account returns.[4]

The Linked Internal Rate of Return

Despite its useful characteristics, the TWR does have an important disadvantage: It requires account valuations on every date that an external cash flow takes place. Thus, calculation of the TWR typically necessitates the ability to price a portfolio of securities on a daily basis. Although daily pricing services are becoming more common, marking an account to market daily is administratively more expensive and cumbersome, and potentially more error-prone, than traditional monthly accounting procedures. For these reasons, use of pure TWR is not yet standard practice, with the prominent exception of the mutual fund industry.[5] The MWR, on the other hand, despite its sensitivity to the size and timing of external cash flows, requires only that an account be valued at the beginning and end of the evaluation period and that the amounts and dates of any external cash flows be recorded.

The complementary advantages and disadvantages of the TWR and the MWR led the authors of the BAI study to make an important recommendation: The TWR should be approximated by calculating the MWR over reasonably frequent time intervals and then chain-linking those returns over the entire evaluation period. This process is referred to as the Linked Internal Rate of Return (LIRR) method and originally was developed by Peter Dietz (1966). The BAI study estimated that if the LIRR method were applied to an account experiencing "normal" cash flow activity, then using monthly valuations and daily dating of external cash flows, the calculated rate of return on average would fall within 4 basis points per year of the true TWR.[6] Given the inaccuracies inherent in the pricing of even the most liquid portfolios, this slight difference appears immaterial.

EXAMPLE 1.7 An Example of LIRR

Suppose, in a given month, the Mientkiewicz account's MWR is calculated each week. These MWRs are 0.021 in week 1, 0.0016 in week 2, −0.014 in week 3, and 0.018 in week 4. The LIRR is obtained by linking these rates:

$$\begin{aligned} R_{\text{LIRR}} &= (1 + 0.021) \times (1 + 0.0016) \times (1 + -0.014) \times (1 + 0.018) - 1 \\ &= 0.0265 \text{ or } 2.65\% \end{aligned}$$

The BAI study concluded that only under unusual circumstances would the LIRR fail to provide an acceptable representation of the TWR. Specifically, the LIRR would fail if both large external cash flows (generally over 10 percent of the account's value) and volatile swings in subperiod performance occurred during the evaluation period. With an evaluation period as short as one month, the chances of such a joint event occurring for an account are low. Nevertheless, if it should happen, the BAI study recommended valuing the account on the date of the intramonth cash flow.

Annualized Return

For comparison purposes, rates of return are typically reported on an annualized basis. As defined here, the annualized return represents the compound average annual return earned by the account over the evaluation period. The calculation is also known as the compound growth rate or geometric mean return. An annualized return is computed by employing the same chain-linking method used to calculate linked internal rates of return, except that the product of the linking is raised to the reciprocal of the number of years covering the evaluation period (or equivalently, taking the appropriate root of the linked product, where the root is the number of years in the measurement period).

EXAMPLE 1.8 Annualized Return

If in years 1, 2, and 3 of a three-year evaluation period an account earned 2.0 percent, 9.5 percent, and −4.7 percent, respectively, then the annualized return for the evaluation period would be:

$$\begin{aligned} r_a &= [(1 + 0.02) \times (1 + 0.095) \times (1 - 0.047)]^{1/3} - 1 \\ &= 0.021 \text{ or } 2.1\% \end{aligned}$$

If 12 quarterly returns had been available for the account instead of three yearly returns, then those quarterly returns would have been similarly linked and the cube root of the product would have been calculated to produce the account's annualized return over the three-year period.

In general, with measurement periods shorter than a full year, it is inadvisable to calculate annualized returns. Essentially, the person calculating returns is extrapolating the account's returns over a sample period to the full year. Particularly for equity accounts, returns can fluctuate significantly during the remaining time in the evaluation period, making the annualized return a potentially unrealistic estimate of the account's actual return over the full year.

Data Quality Issues

The performance measurement process is only as accurate as the inputs to the process. Often performance report users fail to distinguish between rates of return of high and low reliability. In the case of accounts invested in liquid and transparently priced securities and experiencing little external cash flow activity, the reported rates of return are likely to be highly reliable performance indicators. They will accurately reflect the experience of an investor who entered such an account at the beginning of the evaluation period and liquidated his or her investment at the end of the period. Conversely, for accounts invested in illiquid and infrequently priced assets, the underlying valuations may be suspect, thereby invalidating the reported rates of return. For example, due to the inaccuracy inherent in estimation techniques, quarterly valuations of venture capital funds typically have limited economic content. An investor may not be able to enter or leave the account at a value anywhere near the reported valuations. As a result, monthly or even annual performance measurement of such funds should be viewed with caution.

Various services exist that collect data on recent market transactions for a wide range of fixed-income and equity securities. Particularly for many thinly traded fixed-income securities, a current market price may not always be available. In that case, estimated prices may be derived based on dealer-quoted prices for securities with similar attributes (for example, a security with a similar credit rating, maturity, and economic sector). This approach is referred to as **matrix pricing.** For highly illiquid securities, reasonable estimates of market prices may be difficult or impossible to obtain. Investment managers may carry these securities at cost or the price of the last trade in those securities. It is outside the scope of this discussion to address in detail the subject of account valuation. Suffice it to say that *caveat emptor*—"let the buyer beware"—should be the motto of any user of performance measurement reports who deals with securities other than liquid stocks and bonds.

In addition to obtaining accurate account valuations and external cash flow recognition, reliable performance measurement requires appropriate data collection procedures. For example, account valuations should be reported on a trade-date, fully accrued basis. That is, the stated value of the account should reflect the impact of any unsettled trades and any income owed by or to the account but not yet paid. Such a valuation process correctly represents the best available statement of the account's position at a point in time. Other methods, such as settlement date accounting and the exclusion of accrued income, incorrectly reflect the account's market value.

BENCHMARKS

Performance evaluation cannot be conducted in a vacuum. By its nature, performance evaluation is a relative concept. Absolute performance numbers mean little. Even so-called "absolute return" managers should provide some sense of how alternative uses of their clients' money would have performed if exposed to similar risks. Consider how one interprets a 7 percent return on a well-diversified common stock portfolio during a given month. If you knew that the broad stock market had declined 15 percent during the month, you might be impressed. Conversely, if the market had advanced 25 percent, you might be disappointed. If we are to conduct meaningful performance evaluation, then we must develop an appropriate benchmark against which an account's performance can be compared.

Concept of a Benchmark

The *Merriam-Webster Dictionary* defines a benchmark as a "standard or point of reference in measuring or judging quality, value, etc." Applying this general definition to investment management, a benchmark is a collection of securities or risk factors and associated weights that represents the persistent and prominent investment characteristics of an asset category or manager's investment process. At the asset category level, we can think of a benchmark as the collection of securities that the fund sponsor would own if the fund sponsor were required to place all of its investments within the asset category in a single, passively managed portfolio. (In other words, the benchmark is the fund sponsor's preferred index fund for the asset category.) At the manager level, we can think of a benchmark as a passive representation of the manager's investment style, incorporating the salient investment features (such as significant exposures to particular sources of systematic risk) that consistently appear in the manager's portfolios. More metaphorically, a manager's benchmark encompasses the manager's "area" of expertise. Just as an angler has a favorite fishing hole, an investment manager also has distinct preferences for certain types of securities and risk exposures. The opportunity set that represents the manager's area of expertise may be broad or narrow, reflecting the resources and investment processes that the manager brings to bear on the portfolio selection problem.

A little algebra succinctly conveys these concepts. Begin with the simple identity of an investment manager's portfolio; that is, any portfolio is equal to itself:[7]

$$P = P$$

Now, consider an appropriately selected benchmark B. If we add and subtract B from the right-hand side of this identity, effectively adding a zero to the relationship, the result is

$$P = B + (P - B)$$

Additionally, if we define the manager's active investment judgments as being the difference between the manager's portfolio P and the benchmark B so that $A = (P - B)$, then the equation just given becomes

$$P = B + A \tag{1.6}$$

Thus, the managed portfolio P can be viewed as a combination of (1) the benchmark B and (2) active management decisions A composed of a set of over- and underweighted positions in securities relative to the benchmark. We can extend this relationship by introducing a market index M. Adding and subtracting M from the right-hand side of Equation 1.6 gives

$$P = M + (B - M) + A$$

The difference between the manager's benchmark portfolio and the market index ($B - M$) can be defined as the manager's investment style S. If we let $S = (B - M)$, then the equation just given becomes

$$P = M + S + A \tag{1.7}$$

Equation 1.7 states that a portfolio has three components: market, style, and active management.

There are several interesting applications of Equation 1.7. First, note that if the portfolio is a broad market index fund, then $S = (B - M) = 0$ (that is, no style biases) and $A = (P - B) = 0$ (that is, no active management). Consequently, Equation 1.7 reduces to $P = M$; the portfolio is equivalent to the market index.

Second, if we define the benchmark as the market index [that is, $S = (B - M) = 0$, or no style], then Equation 1.7 reduces to Equation 1.6 and substituting M for B gives

$$P = M + A$$

Because many managers and fund sponsors have been willing to define a manager's benchmark as a broad market index (for example, the S&P 500 in the case of U.S. common stock managers), both parties are implicitly stating that they believe that the manager has no distinct investment style. However, most practitioners would agree that the vast majority of managers pursue specific investment styles. Specialization has become the hallmark of our postindustrial society, and it should not be surprising that, with respect to a subject as complex as portfolio management, many managers have chosen to focus their skills on certain segments of that subject.

EXAMPLE 1.9 Returns Due to Style and Active Management

Suppose the Mientkiewicz account earns a total return of 3.6 percent in a given month, during which the portfolio benchmark has a return of 3.8 percent and the market index has a return of 2.8 percent. Then the return due to the portfolio manager's style is

$$S = B - M = 3.8\% - 2.8\% = 1\%$$

and the return due to active management is

$$A = P - B = 3.6\% - 3.8\% = -0.2\%$$

Properties of a Valid Benchmark

Although in practice an acceptable benchmark is simply one that both the manager and the fund sponsor agree fairly represents the manager's investment process, to function effectively in performance evaluation, a benchmark should possess certain basic properties. A valid benchmark is

- **Unambiguous.** The identities and weights of securities or factor exposures constituting the benchmark are clearly defined.
- **Investable.** It is possible to forgo active management and simply hold the benchmark.
- **Measurable.** The benchmark's return is readily calculable on a reasonably frequent basis.
- **Appropriate.** The benchmark is consistent with the manager's investment style or area of expertise.
- **Reflective of current investment opinions.** The manager has current investment knowledge (be it positive, negative, or neutral) of the securities or factor exposures within the benchmark.
- **Specified in advance.** The benchmark is specified prior to the start of an evaluation period and known to all interested parties.
- **Owned.** The investment manager should be aware of and accept accountability for the constituents and performance of the benchmark. It is encouraged that the benchmark be embedded in and integral to the investment process and procedures of the investment manager.

The failure of a benchmark to possess these properties compromises its utility as an effective investment management tool. A benchmark represents an equivalent risk opportunity cost to the fund sponsor. The properties listed merely formalize intuitive notions of what constitutes a fair and relevant performance comparison. It is interesting to observe that a number of commonly used benchmarks fail to satisfy these properties.

Types of Benchmarks

At the investment manager level, a benchmark forms the basis of a covenant between the manager and the fund sponsor. It reflects the investment style that the fund sponsor expects the manager to pursue, and it becomes the basis for evaluating the success of the manager's investment management efforts. Many different benchmarks may satisfy the criteria for an acceptable benchmark and, if agreed upon by both parties, could be implemented. In general, there are seven primary types of benchmarks in use.

1. *Absolute*. An absolute return can be a return objective. Examples include an actuarial rate-of-return assumption or a minimum return target that the fund strives to exceed. Unfortunately, absolute return objectives are not investable alternatives and do not satisfy the benchmark validity criteria.[8]
2. *Manager universes.* Consultants and fund sponsors frequently use the median manager or fund from a broad universe of managers or funds as a performance evaluation benchmark. As discussed in more detail later, a median manager benchmark fails all the tests of benchmark validity except for being measurable.
3. *Broad market indexes.* Many managers and fund sponsors use **broad market indexes** as benchmarks. Prominent examples of broad market indexes used by U.S. investors include the S&P 500, Wilshire 5000, and Russell 3000 indexes for U.S. common stocks; the Lehman Aggregate and the Citigroup Broad Investment-Grade (U.S. BIG) Bond Indexes for U.S. investment-grade debt; and the Morgan Stanley Capital International (MSCI)

Europe, Australasia, and Far East (EAFE) Index for non-U.S. developed-market common stocks. Market indexes are well recognized, easy to understand, and widely available, and satisfy several properties of valid benchmarks. They are unambiguous, generally investable, and measurable, and they may be specified in advance. In certain situations, market indexes are perfectly acceptable as benchmarks, particularly as benchmarks for asset category performance or for "core" type investment approaches in which the manager selects from a universe of securities similar in composition to the benchmark. However, in other circumstances, the manager's style may deviate considerably from the style reflected in a market index. For example, assigning a micro-capitalization U.S. growth stock manager an S&P 500 benchmark clearly violates the appropriateness criterion.

4. *Style indexes.* Broad market indexes have been increasingly partitioned to create **investment style indexes** that represent specific portions of an asset category: for example, subgroups within the U.S. common stock asset category. Four popular U.S. common stock style indexes are (1) large-capitalization growth, (2) large-capitalization value, (3) small-capitalization growth, and (4) small-capitalization value. (Mid-capitalization growth and value common stock indexes are also available.) The Frank Russell Company, Standard & Poor's, and Morgan Stanley Capital International produce the most widely used U.S. common stock style indexes. International common stock style indexes are more recent developments.

   Fixed-income style indexes are produced in a similar manner. In many ways, investment-grade bonds are a more convenient asset category for developing style indexes because the broad market indexes are easily segregated into various types of securities. For example, broad bond market indexes, such as the Lehman Aggregate for U.S. debt, can be broken up into their constituent parts, such as the Lehman Government/Credit Index, the Lehman Mortgage Index, and so on. The Lehman Aggregate can also be decomposed along the lines of maturity (or duration) and quality.

   Similar to broad market indexes, investment style indexes are often well known, easy to understand, and widely available. However, their ability to pass tests of benchmark validity can be problematic. Some style indexes contain weightings in certain securities and economic sectors that are much larger than what many managers consider prudent. Further, the definition of investment style implied in the benchmark may be ambiguous or inconsistent with the investment process of the manager being evaluated. Differing definitions of investment style at times can produce rather extreme return differentials. In 1999, the S&P Large Value Index had a return of 12.72 percent, and the Russell Large Value Index had a return of 7.35 percent. These large return differences among indexes presumably designed to represent the results of the same investment style are disconcerting. Users of style indexes should closely examine how the indexes are constructed and assess their applicability to specific managers.

5. *Factor model based.* Factor models provide a means of relating one or more systematic sources of return to the returns on an account.[9] As a result, a specified set of factor exposures could potentially be used as a **factor model–based benchmark.** The simplest form of factor model is a one-factor model, such as the familiar **market model.** In that relationship, the return on a security, or a portfolio of securities, is expressed as a linear function of the return on a broad market index, established over a suitably long period (for example, 60 months):

$$R_p = a_p + \beta_p R_I + \varepsilon_p \tag{1.8}$$

where R_p represents the periodic return on an account and R_I represents the periodic return on the market index.[10] The market index is used as a proxy for the underlying systematic return factor (or factors). The term ε_p is the residual, or nonsystematic, element of the relationship. The term β_p measures the sensitivity of the returns on the account to the returns on the market index; it is typically estimated by regressing the account's returns on those of the market index. The sensitivity term is called the beta of the account. Finally, the intercept a_p is the "zero factor" term, representing the expected value of R_p if the factor value was zero.

EXAMPLE 1.10 Returns from a Market Model

Consider an account with a zero-factor value of 2.0 percent and a beta of 1.5. Applying Equation 1.8, a return of 8 percent for the market index generates an expected return on the account of 14% (= 2.0% + 1.5 × 8%).

Some managers hold accounts that persistently display a beta greater than 1.0, while other managers hold accounts with betas persistently less than 1.0. Out of these patterns arises the concept of a benchmark with a "normal beta" consistent with these observed tendencies. For example, suppose that an analysis of past account returns, combined with discussions with the manager, suggests a normal beta of 1.2. This normal beta becomes the basis for the benchmark that specifies the level of return that the account would be expected to generate in the absence of any value added by active management on the part of the manager.

Incorporating multiple sources of systematic risk can enhance the richness of the factor model approach. That is, Equation 1.8 can be extended to include more than one factor. For example, a company's size, industry, growth characteristics, financial strength, and other factors may have a systematic impact on a portfolio's performance. Generalizing Equation 1.8 produces

$$R_p = a_p + b_1 F_1 + b_2 F_2 + \cdots + b_K F_K + \varepsilon_p$$

where $F_1, F_2, \ldots F_K$ represent the values of factors 1 through K, respectively. Numerous commercially available multifactor risk models have been produced. Rosenberg and Marathe (1975) pioneered the development of these models, and their work was extended to create performance evaluation benchmarks. The concept of a "normal beta" in a multifactor context leads to the concept of a normal portfolio. A **normal portfolio** is a portfolio with exposures to sources of systematic risk that are typical for a manager, using the manager's past portfolios as a guide.

Benchmarks based on factor exposures can be useful in performance evaluation. Because they capture the systematic sources of return that affect an account's performance, they help managers and fund sponsors better understand a manager's investment style. However, they are not always intuitive to the fund sponsor and particularly to the investment managers (who rarely think in terms of factor exposures when designing investment strategies), are not always easy to obtain, and are potentially expensive to use.

In addition, they are ambiguous. We can build multiple benchmarks with the same factor exposures, but each benchmark can earn different returns. For example, we can construct two different portfolios, each with a beta of 1.2 ("normal beta"), but the portfolios can have materially different returns. Also, because the composition of a factor-based benchmark is not specified with respect to the constituent securities and their weights, we cannot verify all the validity criteria (the benchmark may not be investable, for example).

6. *Returns based.* Sharpe (1988, 1992) introduced the concept of **returns-based benchmarks.** These benchmarks are constructed using (1) the series of a manager's account returns (ideally, monthly returns going back in time as long as the investment process has been in place) and (2) the series of returns on several investment style indexes over the same period. These return series are then submitted to an allocation algorithm that solves for the combination of investment style indexes that most closely track the account's returns.[11]

   For example, assume that we have 10 years of monthly returns of a U.S. equity mutual fund. Also, assume that we have the monthly returns of four U.S. equity style indexes—(1) large-cap growth, (2) large-cap value, (3) small-cap growth, and (4) small-cap value—over the same time period. If we submit these return series to a properly constructed allocation algorithm, we can solve for a particular set of allocation weights for the four style indexes that will track most closely the return series of the manager's actual portfolio. The returns-based benchmark is represented by these allocation weights.

   Returns-based benchmarks are generally easy to use and are intuitively appealing. They satisfy most benchmark validity criteria, including those of being unambiguous, measurable, investable, and specified in advance. Returns-based benchmarks are particularly useful in situations where the only information available is account returns. One disadvantage of returns-based benchmarks is that, like the style indexes that underlie the benchmarks, they may hold positions in securities and economic sectors that a manager might find unacceptable. Further, they require many months of observation to establish a statistically reliable pattern of style exposures. In the case of managers who rotate among style exposures, such a pattern may be impossible to discern.

7. *Custom security based.* An investment manager will typically follow an investment philosophy that causes the manager to focus its research activities on certain types of securities. The manager will select those securities that represent the most attractive investment opportunities that the research process has identified. As the financial and investment characteristics of securities will change over time, a manager's research universe will similarly evolve.

   A **custom security-based benchmark** is simply a manager's research universe weighted in a particular fashion. Most managers do not use a security weighting scheme that is exactly an equal weighting across all securities or one that exactly assigns weights according to market capitalization. Consequently, a custom benchmark reflecting a particular manager's unique weighting approach can be more suitable than a published index for a fair and accurate appraisal of that manager's performance.

   The overwhelming advantage of a custom security-based benchmark is that it meets all of the required benchmark properties and satisfies all of the benchmark validity criteria, making it arguably the most appropriate benchmark for performance evaluation purposes. In addition, it is a valuable tool for managers to monitor and control their investment processes and for fund sponsors to effectively allocate or budget risk across teams of investment managers. One major disadvantage is that custom security-based benchmarks are expensive to construct and maintain. In addition, as they are not composed of published indexes, the perception of a lack of transparency can be of concern.

Building Custom Security-Based Benchmarks

A valid custom security-based benchmark is the product of discussions between the client or the client's consultant and the manager and of a detailed analysis of the manager's past security holdings. The construction of such a benchmark involves the following steps:

1. Identify prominent aspects of the manager's investment process.
2. Select securities consistent with that investment process.
3. Devise a weighting scheme for the benchmark securities, including a cash position.
4. Review the preliminary benchmark and make modifications.
5. Rebalance the benchmark portfolio on a predetermined schedule.

For the purpose of custom benchmark construction, an analysis of the manager's past portfolios will identify prominent aspects of the manager's investment process. The selection of benchmark portfolio securities requires both a broad universe of potential candidates and a set of screening criteria consistent with the manager's investment process. Weighting schemes may include aspects of equal weighting and capitalization weighting, depending on the manager's investment process and client restrictions. Following these steps, a preliminary benchmark portfolio is selected. At this point, the benchmark's composition is reviewed and final modifications are made. Ultimately, keeping the benchmark portfolio current with the manager's investment process necessitates rebalancing the portfolio at regularly scheduled intervals.

These steps, though simple in appearance, constitute a complex task. A proper benchmark must make a fine distinction between the manager's "normal" or policy investment decisions and the manager's active investment judgments. Considerable resources are required, including a comprehensive security database, an efficient computer screening capability, a flexible security weighting system, and a means of maintaining the integrity of the benchmark over time.

Critique of Manager Universes as Benchmarks

Fund sponsors have a natural interest in knowing how their investment results compare to those achieved by similar institutions and how the returns earned by the managers they have selected compare to those earned by managers they might have engaged. To facilitate peer group comparisons, some consulting firms and custodial banks have developed databases or "universes" of account returns ranked in descending order. Fund sponsors often use the median account in a particular peer group as a return benchmark. For instance, the investment policy statement of a public fund might specify that the fund's objective is to perform in the top half of a certain universe of public funds, and the guidelines for a domestic large-cap equity account might state that the manager's results are expected to exceed those of the median account in a certain universe consisting of portfolios with large-cap value mandates or characteristics.

With the exception of being measurable, the median account in a typical commercially available universe does not have the properties of a valid benchmark described above. One of the most significant deficiencies is that, although the universe can be named, the median account cannot be *specified in advance*. Universe compilers can only establish the median account on an *ex post* basis, after the returns earned by all accounts have been calculated and ranked. Prior to the start of an evaluation period, neither the manager nor the fund sponsor has any knowledge of who the median manager will be at period end. In addition, different accounts will fall at the median from one evaluation period to another. For these reasons, the benchmark is not *investable* and cannot serve as a passive alternative to holding the account that is under analysis. Even after the evaluation period concludes, the identity of the median manager typically remains unknown, preventing the benchmark from satisfying the *unambiguous* property. The ambiguity of the median manager benchmark makes it impossible to verify its *appropriateness* by examining whether the investment style it represents adequately corresponds to the account being evaluated. The fund sponsor who chooses to employ universes for peer group comparisons can only rely on the compiler's representations that accounts have been rigorously screened against well-defined criteria for inclusion, the integrity of the input data is scrupulously monitored, and a uniform return calculation methodology has been used for all accounts in all periods.

One other disadvantage merits attention. Because fund sponsors terminate underperforming managers, universes are unavoidably subject to "survivor bias." Consider the hypothetical universe represented in Table 1.1, where a shaded cell indicates that a particular account existed for a given year and an X indicates that a rate of return can be calculated for the referenced evaluation period.

In this example, there were six accounts in the universe at the end of year 1, and there were six at the end of year 7. They were not all the same accounts, however; in fact, only two have survived for the full period to achieve seven-year returns. The other four in the year 1 cohort were no longer present because the sponsors reallocated funds or possibly because the managers' performance was unsatisfactory. In any event, it is likely that the two survivors were among the best-performing in the group of accounts that existed in year 1; sponsors are naturally reluctant to dismiss strong performing managers. Because the survivors' returns were presumably high, the actual median seven-year return for this universe will be higher than the median of a hypothetical return distribution from which no accounts were removed.

Why are these deficiencies of the median manager benchmark of concern? From the perspective of performance evaluation, the question becomes, "To what is the manager expected

TABLE 1.1 Survivor Bias in a Manager Universe

								Annualized Returns At End of Year 7			
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	1 Year	3 Years	5 Years	7 Years
Manager 1											
Manager 2								X	X	X	X
Manager 3								X			
Manager 4								X	X		
Manager 5											
Manager 6											
Manager 7								X	X	X	X
Manager 8								X	X	X	
Manager 9								X			
Manager 10											
Observations	6	6	7	7	7	5	6	6	4	3	2

to add value?" Without a valid reference point, superior performance remains an elusive notion. Placing above the median of a universe of investment managers or funds may be a reasonable investment *objective*, but the performance of a particular manager or fund is not a suitable performance benchmark that can be used to assess investment skill.[12]

Tests of Benchmark Quality

In many organizations, benchmarks have become an important element of the investment management process. Moreover, benchmark use has expanded beyond performance evaluation. Benchmarks are now an integral part of risk management, at both the investment manager and fund sponsor levels. Most forms of risk budgeting use benchmarks to estimate the risks to which a fund sponsor's investment program is exposed at the asset category and investment manager levels.

Given the important uses of benchmarks, it is in the interests of all parties involved (fund sponsors, consultants, and managers) to identify good benchmarks and to improve or replace poor benchmarks. Good benchmarks increase the proficiency of performance evaluation, highlighting the contributions of skillful managers. Poor benchmarks obscure manager skills. Good benchmarks enhance the capability to manage investment risk. Poor benchmarks promote inefficient manager allocations and ineffective risk management. They also increase the likelihood of unpleasant surprises, which can lead to counterproductive actions and unnecessary expense on the part of the fund sponsor.

Bailey (1992b) presents a heuristic set of benchmark quality criteria designed to distinguish good benchmarks from poor benchmarks. These criteria are based on the fundamental properties of valid benchmarks discussed previously and on a logical extension of the purposes for which benchmarks are used. Although none of the criteria alone provides a definitive indicator of benchmark quality, taken together they provide a means for evaluating alternative benchmarks.

- **Systematic biases.** Over time, there should be minimal systematic biases or risks in the benchmark relative to the account. One way to measure this criterion is to calculate the historical beta of the account relative to the benchmark; on average, it should be close to 1.0.[13]

  Potential systematic bias can also be identified through a set of correlation statistics. Consider the correlation between $A = (P - B)$ and $S = (B - M)$. The contention is that a manager's ability to identify attractive and unattractive investment opportunities should be uncorrelated with whether the manager's style is in or out of favor relative to the overall market. Accordingly, a good benchmark will display a correlation between A and S that is not statistically different from zero.

  Similarly, let us define the difference between the account and the market index as $E = (P - M)$. When a manager's style (S) is in favor (out of favor) relative to the market, we expect both the benchmark and the account to outperform (underperform) the market. Therefore, a good benchmark will have a statistically significant positive correlation coefficient between S and E.
- **Tracking error.** We define tracking error as the volatility of A or $(P - B)$. A good benchmark should reduce the "noise" in the performance evaluation process. Thus, the volatility (standard deviation) of an account's returns relative to a good benchmark should be less than the volatility of the account's returns versus a market index or other alternative benchmarks. Such a result indicates that the benchmark is capturing important aspects of the manager's investment style.

- **Risk characteristics.** An account's exposure to systematic sources of risk should be similar to those of the benchmark over time.[14] The objective of a good benchmark is to reflect but not to replicate the manager's investment process. Because an active manager is constantly making bets against the benchmark, a good benchmark will exhibit risk exposures at times greater than those of the managed portfolio and at times smaller. Nevertheless, if the account's risk characteristics are always greater or always smaller than those of the benchmark, a systematic bias exists.
- **Coverage.** Benchmark coverage is defined as the proportion of a portfolio's market value that is contained in the benchmark. For example, at a point in time, all of the securities and their respective weights that are contained in the account and the benchmark can be examined. The market value of the jointly held securities as a percentage of the total market value of the portfolio is termed the *coverage ratio.* High coverage indicates a strong correspondence between the manager's universe of potential securities and the benchmark. Low coverage indicates that the benchmark has little relationship, on a security level, with the opportunity set generated by the manager's investment process.
- **Turnover.** Benchmark turnover is the proportion of the benchmark's market value allocated to purchases during a periodic rebalancing of the benchmark. Because the benchmark should be an investable alternative to holding the manager's actual portfolio, the benchmark turnover should not be so excessive as to preclude the successful implementation of a passively managed portfolio.
- **Positive active positions.** An active position is an account's allocation to a security minus the corresponding weight of the same security in the benchmark. For example, assume an account has a 3 percent weighting in General Electric (GE). If the benchmark has a 2 percent weighting in GE, then the active position is 1 percent (3% − 2%). Thus, the manager will receive positive credit if GE performs well. Actively managed accounts whose investment mandates permit only long positions contain primarily securities that a manager considers to be attractive. When a good custom security-based benchmark has been built, the manager should be expected to hold largely positive active positions for actively managed long-only accounts.[15] Note that when an account is benchmarked to a published index containing securities for which a long-only manager has no investment opinion and which the manager does not own, negative active positions will arise. A high proportion of negative active positions is indicative of a benchmark that is poorly representative of the manager's investment approach.

Hedge Funds and Hedge Fund Benchmarks

Hedge funds have become increasingly popular among institutional and high-net-worth investors in recent years. Although the term *hedge fund* covers a wide range of investment strategies, there are some common threads that link these strategies. In general, **hedge funds** attempt to expose investors to a particular investment opportunity while minimizing (or hedging) other investment risks that could impact the outcome. In most cases, hedging involves both long and short investment positions.

The term *hedge fund* is believed to have originated as a description of an investment strategy developed by Alfred Winslow Jones.[16] The basic strategy involved shorting stocks that managers considered to be overvalued and using the proceeds to invest in stocks that were deemed to be undervalued. In addition, an incentive fee was established, and Jones committed his own capital to assure investors that his interests were aligned with their interests.

In essence, the Jones strategy is the same as the standard long-only strategy in that, relative to the benchmark, a long-only manager will overweight undervalued securities and underweight overvalued securities. The difference is that the long-only manager is limited to a minimum investment of zero in any security. As a result, the maximum "negative bet" that a long-only manager can place on a security that is rated as overvalued is not to hold it (a weight of zero). For example, because approximately 450 companies in the S&P 500 have weights less than 0.5 percent, a long-only manager with an S&P 500 benchmark and a negative opinion on any of these stocks would be limited to, at most, a −0.5 percent active position. By removing the zero weight constraint (that is, allowing shorting), a manager can further exploit overvalued stocks.

There are, however, performance measurement issues as well as numerous administrative and compliance issues that are created when there are short positions in an account. Recall that earlier in the chapter (Equation 1.1), we stated that an account's rate of return is equal to its market value (MV_1) at the end of a period less its market value at the beginning of the period (MV_0), divided by the beginning market value:

$$r_t = \frac{MV_1 - MV_0}{MV_0}$$

In theory, the net assets of a long-short portfolio could be zero; the value of the portfolio's long positions equal the value of the portfolio's short positions. In this case, the beginning market value, MV_0, would be zero and the account's rate of return would be either positive infinity or negative infinity. In the real world of long-short investing, an account will typically have a positive net asset value due to various margin and administrative requirements. However, as the net asset value gets smaller and approaches zero, the account's return will become nonsensically extreme (large positive or large negative).

To address this problem, we need to revise our performance measurement methodology. One approach would be to think in terms of performance impact, which is discussed in more detail later in the chapter. That is,

$$r_v = r_p - r_B \tag{1.9}$$

where

r_v = value-added return
r_p = portfolio return
r_B = benchmark return

Here, the term r_v is the value-added return on a long-short portfolio where the active weights sum to zero, which is the same situation as a zero-net asset hedge fund. Although the active weights sum to zero, a return can be determined by summing the performance impacts of the n individual security positions (both long and short).

$$\sum_{i=1}^{n} w_{vi} = \left(\sum_{i=1}^{n} w_{pi} - \sum_{i=1}^{n} w_{Bi} \right) = 0; \text{ and}$$

$$r_v = \sum_{i=1}^{n} [w_{vi} \times r_i] = \sum_{i=1}^{n} [(w_{pi} - w_{Bi}) \times r_i] = \sum_{i=1}^{n} (w_{pi} \times r_i) - \sum_{i=1}^{n} (w_{Bi} \times r_i) = r_p - r_B$$

where w_{vi}, w_{pi}, and w_{Bi} are, respectively, the *active* weight of security i in the portfolio, the weight of security i in the portfolio, and the weight of security i in the benchmark. A return could be calculated for the period during which the individual security positions were maintained. Once an individual security position changed, the return period would end and a new return period would start.[17]

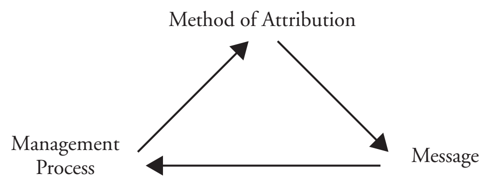
The application of benchmarks to long-only portfolios has reached a mature status. Issues regarding the quality of various benchmark designs and the concerns of overly constraining active management strategies by somehow tying performance too closely to benchmarks remain contentious issues. (For example, see Bernstein 2003.) Nevertheless, it is the rare fund sponsor or investment manager who does not make reference to account performance relative to some benchmark. The advent of hedge funds, however, added a new dynamic to the discussion of the use and design of benchmarks. Some practitioners eschew the use of benchmarks entirely for hedge fund managers, contending that the "absolute return" mandate associated with hedge funds implies that relative performance comparisons are meaningless.

The discussion of hedge fund benchmarks is confounded by the vagueness of the definition of hedge funds. A wide variety of active investment strategies fall under the category of hedge funds. The implications of that diversity for benchmark design are considerable. Underlying all long-only benchmark designs are references to the opportunity set available to the manager. Some hedge fund managers have very clearly definable investment universes composed of highly liquid, daily priced securities. For example, many long-short equity managers also manage long-only portfolios. The universe of securities from which they select on the short side often closely resembles the universe of securities from which they select on the long side. Given information regarding the historical returns and holdings of a long-short equity manager's long and short portfolios, we could use either returns-based or security-based benchmark building approaches to construct separate long and short benchmarks for the manager. These benchmarks could be combined in appropriate proportions to create a valid benchmark. Other hedge fund managers, such as macro hedge fund managers, take rapidly changing long-short leveraged positions in an array of asset categories ranging from equities to commodities, which present significant benchmark building challenges.

The ambiguity of hedge fund manager opportunity sets has led to the widespread use of the Sharpe ratio to evaluate hedge fund manager performance. As discussed later in this chapter, the traditional Sharpe ratio is a measure of excess returns (over a risk-free return) relative to the volatility of returns; notably, it can be calculated without reference to the manager's underlying investment universe. Typically, a hedge fund's Sharpe ratio is compared to that of a universe of other hedge funds that have investment mandates assumed to resemble those of the hedge fund under evaluation. Unfortunately, this approach is exposed to the same benchmark validity criticisms leveled against standard manager universe comparisons. Further, the standard deviation as a measure of risk (the denominator of the Sharpe ratio) is questionable when an investment strategy incorporates a high degree of optionality (skewness), as is the case for the strategies of many hedge funds.

PERFORMANCE ATTRIBUTION

We now move to the second phase of performance evaluation, performance attribution. Fama (1972) proposed the first approach to analyzing the sources of an account's returns.



Practitioners use various forms of performance attribution, but the basic concept remains the same: a comparison of an account's performance with that of a designated benchmark and the identification and quantification of sources of differential returns. Further, a unifying mathematical relationship underlies all performance attribution approaches: Impact equals weight times return. We will return to that relationship shortly.

Performance attribution provides an informed look at the past. It identifies the sources of different-from-benchmark returns (**differential returns**) and their impacts on an account's performance. Presuming that one of the objectives of performance attribution is to gain insights helpful for improving the portfolio management process, that process should dictate the method of attribution. The result will be information or a message that will directly relate to the inputs that have gone into the portfolio management process.

When performance attribution is conducted in this manner, the message will either (1) *reinforce* the effectiveness of the management process or (2) cause a *rethinking* of that process.

Effective performance attribution requires an appropriate analytical framework for decomposing an account's returns relative to those of the benchmark. There is no single correct approach. The appropriate framework will depend on the context of the analysis. In particular, the appropriate framework should reflect the decision-making processes of the organizations involved.

We will consider two basic forms of performance attribution from the standpoints of the fund sponsor and the investment manager. Each form seeks to explain the sources of differential returns. We refer to the performance attribution conducted on the fund sponsor level as **macro attribution.** Performance attribution carried out on the investment manager level we call **micro attribution.** The distinction relates to the specific decision variables involved, as opposed to which organization is actually conducting the performance attribution. While it is unlikely that an investment manager would be in a position to carry out macro attribution, one can easily envision situations in which a fund sponsor may wish to conduct both macro and micro attribution.

Impact Equals Weight Times Return

A manager can have a positive impact on an account's return relative to a benchmark through two basic avenues: (1) selecting superior (or avoiding inferior) performing assets and (2) owning the superior (inferior) performing assets in greater (lesser) proportions than are held in the benchmark. This simple concept underlies all types of performance attribution. The assets themselves may be divided or combined into all sorts of categories, be they economic sectors, financial factors, or investment strategies. In the end, however, the fundamental rule prevails that impact equals (active) weight times return.

The nature of this concept is illustrated through Example 1.11.

EXAMPLE 1.11 An Analogy to the Expression for Revenue

Consider a business that sells widgets. Its total revenue is determined by the formula

$$\text{Revenue} = \text{Price} \times \text{Quantity sold}$$

This year, revenue has risen. The company wants to know why. Based on the above formula, the increase in revenues can be attributed to changes in the unit prices or quantity sold or both (perhaps offsetting to a degree). Figure 1.1 displays the situation in which both price and quantity sold have risen. The old revenue was equal to $P_1 \times Q_1$. The new revenue is equal to $P_2 \times Q_2$. The difference in revenues is a bit more complicated, however. It is due in part to an increase in price [$(P_2 - P_1) \times Q_1$; Area 1], in part to an increase in quantity sold [$(Q_2 - Q_1) \times P_1$; Area 2], and in part to the interaction of both variables [$(P_2 - P_1) \times (Q_2 - Q_1)$; Area 3]. Making the connection to performance attribution, the change in quantity is roughly analogous to a difference in weights between securities held in the account and the benchmark, while the change in price represents the difference in returns between securities held in the account and the benchmark.

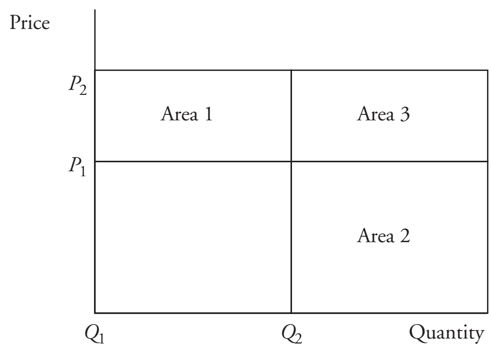


FIGURE 1.1 A Price–Quantity Analogy

Macro Attribution Overview

Let us assume for the moment that for a fund sponsor the term account refers to a total fund consisting of investments in various asset categories (for example, domestic stocks, international stocks, domestic fixed income, and so on) and that the investments are managed by various investment managers. For ease of exposition, we will call this particular account the "Fund." The fund sponsor controls a number of variables that have an impact on the performance of the Fund. For example, the fund sponsor determines the broad allocation of assets to stocks, bonds, and other types of securities. Further, because the fund sponsor retains multiple investment managers to invest the assets of the Fund, decisions must be made regarding allocations across the various investment styles offered by the managers and allocations to the individual managers themselves.

Macro attribution can be carried out solely in a rate-of-return metric. That is, the results of the analysis can be presented in terms of the effects of decision-making variables on the differential return. Most forms of macro attribution follow that approach. The analysis can be enriched by considering the impacts of the decision-making variables on the differential returns in monetary terms. Consider that it is one thing to report that a fund sponsor's active managers added, say, 0.30 percent to the Fund's performance last month. It is quite another thing to state that the 30 basis points of positive active management added US$5 million to the value of the Fund. Performance attribution expressed in a value metric (as opposed to a return metric) can make the subject more accessible not only to investment professionals, but particularly to persons not regularly exposed to the subtle issues of performance attribution. We will present examples of both approaches.

Macro Attribution Inputs

Three sets of inputs constitute the foundation of the macro attribution approach:

1. Policy allocations
2. Benchmark portfolio returns
3. Fund returns, valuations, and external cash flows

With these inputs in hand we can decompose the Fund's performance from a macro perspective.

In the following, we illustrate each concept with data for a hypothetical fund sponsor, the Michigan Endowment for the Performing Arts (MEPA). We use the data for MEPA in the subsequent section to illustrate a macro performance attribution analysis.

Policy Allocations

As part of any effective investment program, fund sponsors should determine normal weightings (that is, policy allocations) to the asset categories within the Fund and to individual managers within the asset categories. By "normal" we mean a neutral position that the fund sponsor would hold in order to satisfy long-term investment objectives and constraints. Policy allocations are a function of the fund sponsor's risk tolerance, the fund sponsor's long-term expectations regarding the investment risks and rewards offered by various asset categories and money managers, and the liabilities that the Fund is eventually expected to satisfy.

Table 1.2 displays the policy allocations of MEPA. It has divided the Fund's assets between two asset categories, with 75 percent assigned to domestic equities and 25 percent assigned to domestic fixed income. Within each asset category, MEPA has retained two active managers. It has allocated 65 percent of the domestic equities to Equity Manager 1 and the remaining 35 percent to Equity Manager 2. Similarly, the fund sponsor has assigned 55 percent of the domestic fixed income to Fixed-Income Manager 1 and 45 percent to Fixed-Income Manager 2.

Benchmark Portfolio Returns

We defined benchmarks earlier. Table 1.3 presents the benchmarks that MEPA has selected for its two asset categories and the managers within those asset categories. The fund sponsor uses broad market indexes as the benchmarks for asset categories, while it uses more narrowly focused indexes to represent the managers' investment styles.[18]

TABLE 1.2 Michigan Endowment for the Performing Arts Investment Policy Allocations

Asset Category	Policy Allocations
Domestic Equities	75.0%
Equity Manager 1	65.0
Equity Manager 2	35.0
Domestic Fixed Income	25.0
Fixed-Income Manager 1	55.0
Fixed-Income Manager 2	45.0
Total Fund	100.0%

TABLE 1.3 Michigan Endowment for the Performing Arts Benchmark Assignments

Asset Category	Benchmark
Domestic Equities	S&P 500
Equity Manager 1	Large-Cap Growth Index
Equity Manager 2	Large-Cap Value Index
Domestic Fixed Income	Lehman Govt/Credit Index
Fixed-Income Manager 1	Lehman Int Govt/Credit Index
Fixed-Income Manager 2	Lehman Treasury Index

Returns, Valuations, and External Cash Flows

Macro attribution in a return-only metric requires fund returns. These returns must be computed at the level of the individual manager to allow an analysis of the fund sponsor's decisions regarding manager selection. If macro attribution is extended to include a value-metric approach, then account valuation and external cash flow data are needed not only to calculate accurate rates of return, but also to correctly compute the value impacts of the fund sponsor's investment policy decision making.

For the month of June 20XX, Table 1.4 shows the beginning and ending values, external cash flows, and the actual and benchmark returns for MEPA's total fund, asset categories, and investment managers.

With the inputs for our hypothetical fund sponsor in hand, we turn to an example of a macro performance attribution analysis in the next section.

Conducting a Macro Attribution Analysis

One can envision a number of different variables of interest when evaluating the fund sponsor's decision-making process. Below, we present six levels or components of investment

TABLE 1.4 Michigan Endowment for the Performing Arts Account Valuations, Cash Flows, and Returns: June 20XX

Asset Category	Beginning Value	Ending Value	Net Cash Flows	Actual Return	Benchmark Return
Domestic Equities	$143,295,254	$148,747,228	$(1,050,000)	4.55%	4.04%
Equity Mgr 1	93,045,008	99,512,122	1,950,000	4.76	4.61
Equity Mgr 2	50,250,246	49,235,106	(3,000,000)	4.13	4.31
Domestic Fixed Income	43,124,151	46,069,371	2,000,000	2.16	2.56
Fixed-Income Mgr 1	24,900,250	25,298,754	0	1.60	1.99
Fixed-Income Mgr 2	18,223,900	20,770,617	2,000,000	2.91	2.55
Total Fund	$186,419,405	$194,816,599	$950,000	3.99%	3.94%

policy decision making into which the Fund's performance might be analyzed. We do not imply that these are the only correct variables—they are simply logical extensions of a typical fund sponsor's decision-making process.

Specifically, those levels (which we later refer to as *investment strategies* for reasons to become apparent shortly) are:

1. Net contributions
2. Risk-free asset
3. Asset categories
4. Benchmarks
5. Investment managers
6. Allocation effects

Macro attribution analysis starts with the Fund's beginning-of-period and end-of-period values. Simply put, the question under consideration is: How much did each of the decision-making levels contribute, in either a return or a value metric, to the Fund's change in value over an evaluation period? Macro attribution takes an incremental approach to answering this question. Each decision-making level in the hierarchy is treated as an investment strategy, and its investment results are compared to the cumulative results of the previous levels. That is, each decision-making level represents an unambiguous, appropriate, and specified-in-advance investment alternative: in other words, a valid benchmark. The fund sponsor has the option to place all of the Fund's assets in any of the investment strategies. The strategies are ordered in terms of increasing volatility and complexity. Presumably, the fund sponsor will move to a more aggressive strategy only if it expects to earn positive incremental returns. Macro attribution calculates the incremental contribution that the choice to move to the next strategy produces.

In the previous section, we gave the inputs necessary to conduct a macro performance attribution analysis for a hypothetical fund sponsor, MEPA, for the month of June 20XX. We apply the macro attribution framework just outlined to MEPA in the following discussion. Table 1.5 summarizes the results.

TABLE 1.5 Michigan Endowment for the Performing Arts Monthly Performance Attribution: June 20XX

Decision-Making Level (Investment Alternative)	Fund Value	Incremental Return Contribution	Incremental Value Contribution
Beginning value	$186,419,405	—	—
Net contributions	187,369,405	0.00%	$950,000
Risk-free asset	187,944,879	0.31	575,474
Asset category	194,217,537	3.36	6,272,658
Benchmarks	194,720,526	0.27	502,989
Investment managers	194,746,106	0.01	25,580
Allocation effects	194,816,599	0.04	70,494
Total Fund	194,816,599	3.99	8,397,194

We now examine each of the six levels in turn.

Net Contributions

Table 1.5 indicates that the starting point of the analysis is the Fund's beginning market value. In our example, at the beginning of June 20XX, the market value of the Fund was $186,419,405. During a given month, the Fund may experience contributions and/or withdrawals. The Net Contributions investment strategy specifies that the net inflows are invested at a zero rate of return and, therefore, the Fund's value changes simply by the total amount of these flows. During June 20XX, net contributions to the Fund were a positive $950,000. Adding this amount to the Fund's beginning value produces a value of $187,369,405 for the Fund under the Net Contributions investment strategy. Although no fund sponsor would deliberately follow this investment strategy, it provides a useful baseline to begin the analysis.

Risk-Free Asset

One highly conservative (but certainly reasonable) investment strategy open to a fund sponsor is to invest all of the Fund's assets in a risk-free asset, such as 90-day Treasury bills.[19] Assuming that the Fund's beginning value and its net external cash inflows (accounting for the dates on which those flows occur) are invested at the risk-free rate, the Fund's value will increase by an additional amount over the value achieved under the Net Contributions investment strategy with its zero rate of return. The Risk-Free Asset investment strategy, using a risk-free rate during June 20XX of 0.31 percent, produces an incremental increase in value of $575,474 (=$187,944,879 − 187,369,405) over the results of the Net Contributions investment strategy, for a total fund value of $187,944,879.[20]

Asset Category

Most fund sponsors view the Risk-Free Asset investment strategy as too risk-averse and therefore overly expensive. Instead, they choose to invest in risky assets, based on the widely held belief that, over the long run, the market rewards investors who bear systematic risk. The

Asset Category investment strategy assumes that the Fund's beginning value and external cash flows are invested passively in a combination of the designated asset category benchmarks, with the specific allocation to each benchmark based on the fund sponsor's policy allocations to those asset categories.

In essence, this approach is a pure index fund approach. The Fund's value under this investment strategy will exceed or fall below the value achieved under the Risk-Free Asset investment strategy depending on whether the capital markets fulfill the expectation that risk-taking investors are rewarded. From a return-metric perspective, the incremental return contribution is

$$r_{AC} = \sum_{i=1}^{A} w_i \times (r_{Ci} - r_f) \tag{1.10}$$

where r_{AC} is the incremental return contribution of the Asset Category investment strategy, r_{Ci} is the return on the ith asset category, r_f is the risk-free return, w_i is the policy weight assigned to asset category i, and A is the number of asset categories. From a value-metric perspective, the incremental contribution of the Asset Category investment strategy is found by investing each asset category's policy proportion of the Fund's beginning value and all net external cash inflows at the differential rate between the asset category's benchmark rate of return and the risk-free rate, and then summing across all asset categories.

In the Fund's case, investing 75 percent of the Fund's beginning value and net external cash inflows in the S&P 500 and 25 percent in the Lehman Brothers Government/Credit Bond Index (for a combined return of 3.67 percent in the month, or 3.36 percent above the risk-free rate) increases the Fund's market value by \$6,272,658 (= \$194,217,537 − \$187,944,879) over the value produced by the Risk-Free Asset investment strategy. As a result, the Fund's value totals \$194,217,537 under the Asset Category investment strategy.

It would be entirely appropriate for a fund sponsor to stop at the Asset Category investment strategy. In fact, an efficient markets proponent might view this all-passive approach as the most appropriate course of action. Nevertheless, fund sponsors typically choose to allocate their funds within an asset category among a number of active managers, most of whom pursue distinctly different investment styles. Importantly for macro attribution, when fund sponsors hire active managers, they are actually exposing their assets to two additional sources of investment returns (and risks): investment style and active management skill.

An investment manager's performance versus the broad markets is dominated by the manager's investment style. With respect to U.S. common stocks, for example, active managers cannot realistically hope to consistently add more than 2–3 percentage points (if that much) annually to their investment styles, as represented by appropriate benchmarks. Conversely, the difference in performance between investment styles can easily range from 15 to 30 percentage points per year.

Benchmarks

The macro attribution analysis can be designed to separate the impact of the managers' investment styles (as represented by the managers' benchmarks) on the Fund's value from the effect of the managers' active management decisions. In this case, the next level of analysis assumes that the Fund's beginning value and net external cash inflows are passively invested in the aggregate of the managers' respective benchmarks. An aggregate manager benchmark return is calculated as a weighted average of the individual managers' benchmark returns.

The weights used to compute the aggregate manager benchmark return are based on the fund sponsor's policy allocations to the managers. From a return-metric perspective,

$$r_{IS} = \sum_{i=1}^{A}\sum_{j=1}^{M} w_i \times w_{ij} \times (r_{Bij} - r_{Ci}) \tag{1.11}$$

where r_{IS} is the incremental return contribution of the Benchmarks strategy, r_{Bij} is the return for the jth manager's benchmark in asset category i, r_{Ci} is the return on the ith asset category, w_i is the policy weight assigned to the ith asset category, w_{ij} is the policy weight assigned to the jth manager in asset category i, and A and M are the number of asset categories and managers, respectively.[21] (Note that summed across all managers and asset categories, $w_i \times w_{ij} \times r_{Bij}$ is the aggregate manager benchmark return.) From a value-metric perspective, the incremental contribution of the Benchmarks strategy is calculated by multiplying each manager's policy proportion of the total fund's beginning value and net external cash inflows by the difference between the manager's benchmark return and the return of the manager's asset category, and then summing across all managers.

In the case of the Fund, the aggregate manager benchmark return was 3.94 percent in June 20XX. Investing the Fund's beginning value and net external cash inflows at this aggregate manager benchmark return produces an incremental gain of \$502,989 (= \$194,720,526 − \$194,217,537) over the Fund's value achieved under the Asset Category investment strategy. As a result, under the Investment Style investment strategy, the Fund's value grows to \$194,720,526.

Paralleling the Asset Category investment strategy, the Benchmarks strategy is essentially a passively managed investment in the benchmarks of the Fund's managers. The difference in performance between the aggregate of the managers' benchmarks and the aggregate of the asset category benchmarks is termed *misfit return* or, less formally, *style bias*. In June 20XX, the Fund's misfit return was (3.94% − 3.67%), or a positive 0.27 percent. Although the expected value of misfit return is zero, it can be highly variable over time. That variability can be particularly large for a fund sponsor who has retained investment manager teams within the fund's various asset categories that display sizeable style biases relative to their respective asset category benchmarks. Some fund sponsors employ special risk-control strategies to keep this misfit risk within acceptable tolerances.

Investment Managers

In the next level of analysis, to discern the impact of the managers' active management decisions on the change in the Fund's value, macro attribution analysis calculates the value of the Fund as if its beginning value and net external cash flows were invested in the aggregate of the managers' actual portfolios. Again, the weights assigned to the managers' returns to derive the aggregate manager return will come from the policy allocations set by the fund sponsor. A relationship similar to Equation 1.11 describes the return-metric contribution of the Investment Managers strategy:

$$r_{IM} = \sum_{i=1}^{A}\sum_{j=1}^{M} w_i \times w_{ij} \times (r_{Aij} - r_{Bij}) \tag{1.12}$$

where r_{Aij} represents the actual return on the jth manager's portfolio within asset category i and the other variables are as defined previously.

The difference in the Fund's value under the Investment Managers strategy relative to the Benchmarks strategy will depend on whether the managers, in aggregate, exceeded the return on the aggregate benchmark. In the case of the Fund, the aggregate actual return of the managers (calculated using policy weights) was 3.95 percent, as opposed to 3.94 percent return on the aggregate manager investment style benchmark. This modestly positive excess return translates into an incremental increase in the fund's value of $25,580 (= $194,746,106 – $194,720,526) over the value produced under the Benchmarks strategy, for a total value of $194,746,106 under the Investment Managers investment strategy.

It should be emphasized that macro attribution calculates the value added by the Fund's managers based on the assumption that the fund sponsor has invested in each of the managers according to the managers' policy allocations. Of course, the actual allocation to the managers will likely differ from the policy allocations. However, if we wish to correctly isolate the contributions of the various levels of fund sponsor decision making, we must distinguish between those aspects of the Fund's investment results over which the fund sponsor does and does not have control. That is, the fund sponsor sets the allocation of assets to the Fund's managers but has no influence over their investment performance. Conversely, the managers have control over their investment performance, but they do not generally determine the amount of assets placed under their management.

In examining the value added by the Fund's managers, we should assume they were funded at their respective policy allocations and ask the question, "What would the managers have contributed to the Fund's performance if the fund sponsor consistently maintained the stated policy allocations?" However, in examining the contribution of the fund sponsor, it makes sense to calculate the impact of the differences between the managers' actual and policy allocations on the Fund's performance and thus ask the question, "How did the fund sponsor's decisions to deviate from investment manager policy allocations affect the Fund's performance relative to a strategy of consistently maintaining the stated policy allocations?" The analysis performed at the Investment Managers level attempts to answer the former question. The analysis done at the Allocation Effects level begins to answer the latter question.

Allocation Effects

The final macro attribution component is Allocation Effects. In a sense, the Allocation Effects incremental contribution is a reconciling factor—by definition, it is the difference between the Fund's ending value and the value calculated at the Investment Managers level. If the fund sponsor had invested in all of the managers and asset categories precisely at the established policy allocations, then the Allocation Effects investment strategy's contribution would be zero. However, most fund sponsors deviate at least slightly from their policy allocations, thereby producing an allocation effect. The Fund's actual ending value was $194,816,599, which represents a $70,494 increase (= $194,816,599 − $194,746,106) over the value achieved through the Investment Managers investment strategy. By implication, then, MEPA's actual weightings of the asset categories and managers versus the policy weightings contributed positively to the Fund's value in the month of June 20XX.

Micro Attribution Overview

As implied by its name, micro attribution focuses on a much narrower subject than does macro attribution. Instead of examining the performance of a total fund, micro attribution concerns itself with the investment results of individual portfolios relative to designated benchmarks. Thus, let us define the term *account* to mean a specific portfolio invested by a specific investment manager which we will refer to as the "Portfolio." The Portfolio can be formed of various types of securities. Our illustrations will initially be based on a portfolio of U.S. common stocks. We shall address fixed-income attribution later in this section.

Over a given evaluation period, the Portfolio will produce a return that is different from the return on the benchmark. This difference is typically referred to as the manager's value-added or active return. As shown earlier in Equation 1.9, a manager's value added can be expressed as

$$r_v = r_p - r_B$$

Because the return on any portfolio is the weighted sum of the returns on the securities composing the portfolio, Equation 1.9 can be rewritten as

$$r_v = \sum_{i=1}^{n} w_{pi} r_i - \sum_{i=1}^{n} w_{Bi} r_i \tag{1.13}$$

where w_{pi} and w_{Bi} are the proportions of the Portfolio and benchmark, respectively, invested in security i, r_i is the return on security i, and n is the number of securities.[22]

Rearranging the last equation demonstrates that the manager's value added is equal to the difference in weights of the Portfolio and benchmark invested in a security times the return on that security, summed across all n securities in the Portfolio and benchmark:

$$r_v = \sum_{i=1}^{n} [(w_{pi} - w_{Bi}) \times r_i]$$

With further manipulation,[23] it can be shown that

$$r_v = \sum_{i=1}^{n} [(w_{pi} - w_{Bi}) \times (r_i - r_B)] \tag{1.14}$$

where r_B is the return on the Portfolio's benchmark.

Equation 1.14 offers the simplest form of micro performance attribution: a security-by-security attribution analysis. In this analysis, the manager's value added can be seen to come from two sources: the weights assigned to securities in the Portfolio relative to their weights in the benchmark and the returns on the securities relative to the overall return on the benchmark.

There are four cases of relative-to-benchmark weights and returns for security i to consider. Table 1.6 gives those cases and their associated performance impacts versus the benchmark.

TABLE 1.6 Relative-to-Benchmark Weights and Returns

	$w_{pi} - w_{Bi}$	$r_i - r_B$	Performance Impact versus Benchmark
1.	Positive	Positive	Positive
2.	Negative	Positive	Negative
3.	Positive	Negative	Negative
4.	Negative	Negative	Positive

A manager can add value by overweighting (underweighting) securities that perform well (poorly) relative to the benchmark. Conversely, the manager can detract value by overweighting (underweighting) securities that perform poorly (well) relative to the benchmark.

Security-by-security micro attribution generally is unwieldy and typically provides little in the way of useful insights. The large number of securities in a well-diversified portfolio makes the impact of any individual security on portfolio returns largely uninteresting. A more productive form of micro attribution involves allocating the value-added return to various sources of systematic returns.

Underlying most micro attributions is a factor model of returns. A factor model assumes that the return on a security (or portfolio of securities) is sensitive to the changes in various factors. These factors represent common elements with which security returns are correlated. Factors can be defined in a number of ways: They might be sector or industry membership variables; they might be financial variables, such as balance sheet or income statement items; or they might be macroeconomic variables, such as changes in interest rates, inflation, or economic growth.

The market model, introduced previously, relates a security's or portfolio's return to movements of a broad market index, with the exposure to that index represented by the beta of the security. Recall that Equation 1.8 provides one expression of the market model:

$$R_p = a_p + \beta_p R_I + \varepsilon_p$$

Example 1.12 illustrates the calculation of value added (active return) relative to a one-factor model.

EXAMPLE 1.12 Active Return Relative to a One-Factor Model

Assume that the Portfolio has a zero-factor value of 1.0 percent and a beta of 1.2 at the beginning of the evaluation period. During the period, the excess return on the market index was 7 percent. The market model, expressed in Equation 1.8, states that the Portfolio should return 9.4 percent (= 1.0% + 1.2 × 7%). Further, assume that the Portfolio was assigned a custom benchmark with its own market model parameters, a zero-factor value of 2.0 percent and a beta of 0.8, and which thus has an expected return of 7.6 percent (= 2.0% + 0.8 × 7%). If the Portfolio's actual return was 10.9 percent, then the differential return of 3.3 percent could be attributed in part to the Portfolio's differential expected returns. That is, the Portfolio held a zero factor of 1.0 versus the 2.0 of the benchmark, while the Portfolio had a beta of 1.2 versus the benchmark's beta of 0.8. The incremental expected return of the Portfolio versus the benchmark was 1.8 percent [= (1.0% − 2.0%) + (1.2 − 0.8) × 7%]. The remaining 1.5 percent of differential return would be attributed to the investment skill of the manager.

Sector Weighting/Stock Selection Micro Attribution

Many investment managers employ analysts to research securities and portfolio managers to then build portfolios based on that research. With this investment process, managers are interested in an attribution analysis that will disaggregate the performance effects of the analysts' recommendations and the portfolio managers' decisions to over- and underweight economic sectors and industries.

We can define the returns on the Portfolio and its benchmark to be the weighted sums of their respective economic sector returns. Therefore, just as Equation 1.13 expressed the manager's value-added return as the difference between the weighted average return on the securities in the Portfolio and the benchmark, the manager's value-added return can similarly be expressed as the difference between the weighted average return on the economic sectors in the Portfolio and the benchmark:

$$r_v = \sum_{j=1}^{S} w_{pj} r_{pj} - \sum_{j=1}^{S} w_{Bj} r_{Bj} \tag{1.15}$$

w_{pj} = Portfolio weight of sector j
w_{Bj} = benchmark weight of sector j
r_{pj} = Portfolio return of sector j
r_{Bj} = benchmark return of sector j
S = number of sectors

Continuing with the example of one of MEPA's investment managers, Table 1.7 shows the results of a micro attribution analysis based on partitioning a manager's value added into a part due to skill in sector selection and a part due to skill in security selection. In this example, the return on the Portfolio for a selected one-month period was 1.12 percent. During that same month the benchmark return was 0.69 percent, generating a value added return of 0.43 percent.

Note that this is a holdings-based or "buy-and-hold" attribution. Each sector's contribution to the total allocation and selection effects depends upon the beginning portfolio and benchmark weights in that sector and the constituent securities' returns due to price appreciation and dividend income. The buy-and-hold approach, which disregards the impact of transactions during the evaluation period, has an important practical advantage: Only the holdings and their returns need be input to the attribution system. There is, however, a disadvantage: The account's buy-and-hold return will not equal its time-weighted total return. For that reason, the attribution analysis shown above includes a reconciling item captioned "Trading and Other." In the example shown in Table 1.7, "Trading and Other" is the negative 14 basis point (−0.14 percent) difference between the account's Buy/Hold return of 1.26 percent and the actual portfolio return of 1.12 percent. The imputed "trading and other" factor reflects the net impact of cash flows and security purchases and sales during the evaluation period. In actively managed accounts with high turnover, the "trading and other" factor can be significant. Where this is a concern, transaction-based attribution analysis can be employed.[24]

The value-added return can be segmented into the impact of assigning the assets of the portfolio to various economic sectors and the impact of selecting securities within those economic sectors. Equation 1.15 can be rearranged to form the following relationship:[25]

$$r_v = \underbrace{\sum_{j=1}^{S} (w_{pj} - w_{Bj})(r_{Bj} - r_B)}_{\textit{Pure Sector Allocation}} + \underbrace{\sum_{j=1}^{S} (w_{pj} - w_{Bj})(r_{pj} - r_{Bj})}_{\textit{Allocation/Selection Interaction}} + \underbrace{\sum_{j=1}^{S} w_{Bj}(r_{pj} - r_{Bj})}_{\textit{Within-Sector Selection}} \tag{1.16}$$

where S is the number of sectors and r_B is the return on the Portfolio's benchmark.

In Equation 1.16, the **Pure Sector Allocation return** equals the difference between the allocation (weight) of the Portfolio to a given sector and the Portfolio's benchmark weight for that sector, times the difference between the sector benchmark's return and the overall Portfolio's benchmark return, summed across all sectors. The pure sector allocation return

TABLE 1.7 Results of a Micro Attribution Analysis

Economic Sectors	Portfolio Weight (%)	Sector Benchmark Weight (%)	Portfolio Return (%)	Sector Benchmark Return (%)	Performance Attribution: Pure Sector Allocation	Allocation/ Selection Interaction	Within-Sector Selection	Total Value Added
Basic materials	5.97	5.54	−0.79	−0.67	−0.01	0.00	−0.01	−0.01
Capital goods	7.82	7.99	−3.60	−3.95	0.01	0.00	0.03	0.04
Consumer durables	2.90	2.38	0.46	−0.21	0.00	0.00	0.02	0.01
Consumer nondurables	31.78	34.75	1.92	1.97	−0.04	0.00	−0.02	−0.05
Energy	7.15	6.01	0.37	0.14	−0.01	0.00	0.01	0.01
Financial	22.47	20.91	2.92	2.05	0.02	0.01	0.18	0.22
Technology	12.14	16.02	2.00	−0.30	0.04	−0.09	0.37	0.32
Utilities	8.64	6.40	0.46	−0.37	−0.02	0.02	0.05	0.05
Cash and equivalent	1.13	0.00	0.14		−0.01	0.00	0.00	−0.01
Buy/Hold + Cash	100.00	100.00	1.26	0.69	−0.02	−0.05	0.64	0.57
Trading and Other			−0.14					−0.14
Total Portfolio			1.12	0.69				0.43

assumes that within each sector the manager held the same securities as the benchmark and in the same proportions. Thus, the impact on relative performance is attributed only to the sector-weighting decisions of the manager.

EXAMPLE 1.13 The Pure Sector Allocation Return for Consumer Nondurables

Table 1.7 indicates that at the beginning of the month the Portfolio had a 31.78 percent weight in consumer nondurables, while the benchmark had a 34.75 percent weight. Because the return of the benchmark consumer nondurables sector was 1.97 percent and the return of the overall benchmark was 0.69 percent, the performance impact due to the consumer nondurables sector allocation is −0.04 percent [= (31.78% − 34.75%) × (1.97% − 0.69%)]. That is, the decision to underweight a sector that performed better than the overall benchmark resulted in a negative contribution to the performance of the Portfolio relative to the overall benchmark. The Pure Sector Allocation return is typically the responsibility of the portfolio managers who determine the Portfolio's relative allocations to economic sectors and industries.

The **Within-Sector Selection return** equals the difference between the return on the Portfolio's holdings in a given sector and the return on the corresponding sector benchmark, times the weight of the benchmark in that sector, summed across all sectors. The Within-Sector Selection return implicitly assumes that the manager weights each sector in the Portfolio in the same proportion as in the overall benchmark, although *within the sector* the manager may hold securities in different-from-benchmark weights. Thus, the impact on relative performance is now attributed only to the security selection decisions of the manager.

EXAMPLE 1.14 The Within-Sector Allocation Return for Technology

Table 1.7 shows that the return of the portfolio's technology sector was 2.00 percent, while the return of the benchmark's technology sector was −0.30 percent. Consequently, the performance impact of security selection within the technology sector was +0.37 percent {= 16.02% × [2.00% − (−0.30%)]}, where 16.02 percent is the weight of the benchmark's holdings in the technology sector. During the month, the Portfolio held technology stocks that in total performed better than the aggregate performance of the technology stocks contained in the sector benchmark, thereby contributing positively to the Portfolio's performance relative to the overall benchmark. The Within-Sector Selection impact is often the responsibility of the security analysts. Among the securities that they research, they are expected to identify significantly misvalued securities and recommend appropriate action.

The **Allocation/Selection Interaction return** is a more difficult concept because it involves the joint effect of the portfolio managers' and security analysts' decisions to assign weights to both sectors and individual securities. The Allocation/Selection Interaction equals the difference between the weight of the Portfolio in a given sector and the Portfolio's benchmark for that sector, times the difference between the Portfolio's and the benchmark's returns in that sector, summed across all sectors.

EXAMPLE 1.15 The Allocation/Selection Interaction Return for Technology

Again referring to Table 1.7, we can see that the Portfolio's relative underweight in the Technology sector of −3.88 percent (= 12.14% − 16.02%) and the Portfolio's positive relative performance in the Technology sector of 2.30 percent [= 2.00% − (−0.30%)] produced an Allocation/Selection Interaction effect of −0.09 percent during the month.

A decision to increase the allocation to a particular security adds not only to the weight in that security but also to the weight of the sector to which the security belongs, unless there is an offsetting adjustment to securities within that sector. Unless the portfolio manager is careful to make offsetting adjustments, security selection decisions can inadvertently drive sector-weighting decisions. In general, the Allocation/Selection Interaction impact will be relatively small if the benchmark is appropriate—that is, one that is devoid of any material systematic biases. Because the Allocation/Selection Interaction impact is often the source of some confusion and is usually the result of security selection decisions, some practitioners consolidate the Allocation/Selection Interaction impact with the Within-Sector Selection impact.

Fundamental Factor Model Micro Attribution

As we have noted, some type of factor model underlies virtually all forms of performance attribution. Economic sectors and industries represent only one potential source of common factor returns. Numerous practitioners and academics (for example, see Sharpe, 1982, and Fama and French, 1992) have investigated other common factor return sources. For example, with respect to common stocks, a company's size, its industry, its growth characteristics, its financial strength, and other factors seem to have an impact on account performance. Often these factors are referred to as fundamental factors. They may be combined with economic sector factors to produce multifactor models that can be used to conduct micro attribution.

As with any form of performance attribution, the exposures of the Portfolio and the benchmark to the factors of the fundamental factor model must be determined at the beginning of the evaluation period. The benchmark could be the risk exposures of a style or custom index, or it could be a set of normal factor exposures that were typical of the manager's portfolio over time. Finally, the performance of each of the factors must be determined.

EXAMPLE 1.16 Fundamental Factor Model Micro Attribution

Table 1.8 provides an example of a fundamental factor model micro attribution analysis where a U.S. growth stock manager invests the Portfolio. The performance attribution example covers a one-month period, and during that time the Portfolio generated a 6.02 percent rate of return, while the normal portfolio and the market index produced returns of 5.85 percent and 6.09 percent, respectively. During this particular month, growth stocks performed less well than the market index, largely explaining why the normal portfolio (representing the manager's investment style) underperformed the return on the market index by −0.24 percent. The performance difference between the Portfolio (6.02 percent) and the normal portfolio (5.85 percent) is a measure of the portfolio manager's investment skill (0.17 percent) or value added.

TABLE 1.8 Micro Attribution Using a Fundamental Factor Model

	Portfolio Exposure	Normal Exposure	Active Exposure	Active Impact	Return
Market Return					6.09%
Normal Portfolio Return					5.85
Cash Timing	2.36	0.00	2.36	−0.13	
Beta Timing	1.02	1.00	0.02	0.04	
Total Market Timing					−0.09
Growth	1.12	0.85	0.27	−0.15	
Size	−0.26	0.35	−0.61	−0.35	
Leverage	−0.33	−0.60	0.27	0.11	
Yield	−0.03	−0.12	−0.09	−0.22	
Total Fundamental Risk Factors				−0.61	
Basic Industry	14.10	15.00	−0.90	0.04	
Consumer	35.61	30.00	5.61	−0.07	
Energy	8.36	5.00	3.36	0.05	
Financials	22.16	20.00	2.16	−0.02	
Technology	17.42	25.00	−7.58	0.16	
Utilities	2.35	5.00	−2.65	−0.01	
Total Economic Sectors					0.15
Specific (unexplained)					0.72
Actual Portfolio Return					6.02%

The micro attribution analysis shown in Table 1.8 attributes the manager's investment skill or value added to four primary sources: (1) market timing, (2) exposures to fundamental factors, (3) exposures to economic sectors, and (4) a specific or unexplained return component. The market-timing component is made up of two performance impacts; one is due to the Portfolio's cash position, and the other relates to the Portfolio's beta. In the example, the combination of these two effects had a negative impact of −0.09 percent. The second primary performance attribute involves the exposures to the fundamental factors. The Portfolio's fundamental factor exposures are contrasted with "normal" fundamental factor exposures, represented by the manager's benchmark.[26] The Portfolio's actual factor exposures versus its "normal" exposures resulted in a negative return impact of −0.61 percent. Similarly, the Portfolio's economic sector allocations are contrasted with the Portfolio's "normal" allocations to produce performance attribution impacts. In this case, the active sector weights had a positive impact of 0.15 percent. Finally, the fundamental factor model was unable to explain a portion of the Portfolio's return; in this case, the Portfolio had a specific or unexplained return of +0.72 percent.[27] This specific return that cannot be explained by the factor model is attributed to the investment manager.

Fixed-Income Attribution

The sector weighting/stock selection approach to micro attribution is applicable to fixed-income as well as equity accounts. We mentioned in our remarks on fixed-income style indexes earlier in the chapter that broad fixed-income market indexes may be segregated into their constituent market segments. Accordingly, the sector weighting/stock selection equity attribution analysis can also be adapted for use with fixed-income accounts by substituting market segments such as government bonds, agency and investment-grade corporate credit bonds, high-yield bonds, and mortgage-backed securities, among others, for the economic sectors such as energy, financial, or utilities.

Nonetheless, bonds are unlike stocks, and an approach that merely isolates allocation and selection effects among bond market sectors will be of limited value in analyzing the sources of fixed-income account returns. Useful attribution analysis captures the return impact of the manager's investment decisions, and fixed-income managers weigh variables that differ in important ways from the factors considered by equity portfolio managers. In the interests of mathematical brevity, we will limit our discussion of fixed-income micro performance attribution to a conceptual overview.[28]

Major determinants of fixed-income results are changes in the general level of interest rates, represented by the government (default-free) yield curve, and changes in sector, credit quality, and individual security differentials, or nominal spreads, to the yield curve. As a general rule, fixed-income security prices move in the opposite direction of interest rates: If interest rates fall, bond prices rise, and vice versa. In consequence, fixed-income portfolios tend to have higher rates of return in periods of falling interest rates and, conversely, lower rates of return in periods of rising interest rates. Consider the example displayed in Figure 1.2, where the U.S. Treasury spot rate yield curve shifted upward across all maturities during the nine-month period ending June 30, 2004, and where the return for the Lehman Brothers U.S. Government Index for the nine-month period was −0.56 percent. Comparing the yield curves for September 30, 2004, and June 30, 2004, we see that in the third quarter of 2004 the change in the U.S. Treasury yield curve was more complex: Short-term rates rose, while the yields on government securities with terms to maturity longer than two years

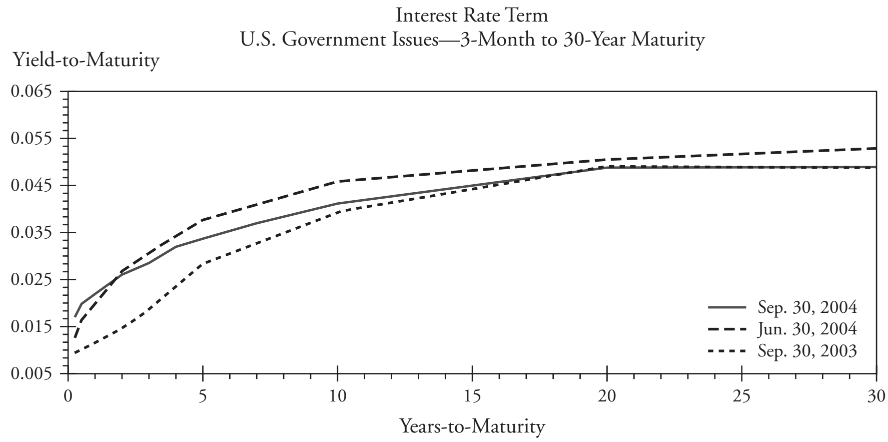


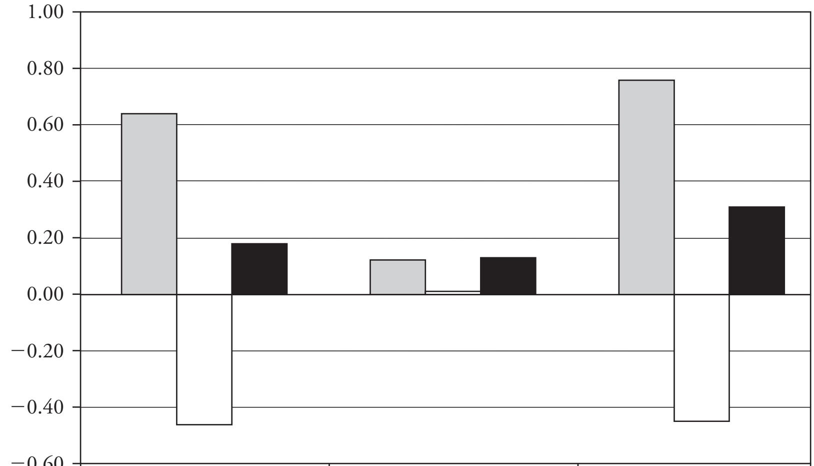
FIGURE 1.2 Interest Rate Term Structure U.S. Government Issue—3-Month to 30-Year Maturity

fell. Reflecting the decline in intermediate and long-term yields, the return on the Lehman Brothers U.S. Government Index for the three-month period was 3.11 percent.

For fixed-income securities that are subject not only to default-free yield-curve movements but also to credit risk, spread changes represent an additional source of interest rate exposure. Companies operating within the same industry face the same business environment, and the prices of the securities they issue have a general tendency to move in the same direction in response to environmental changes. All airlines, for example, are affected by changes in business and leisure travel patterns and the cost of fuel, among other economic factors. In the corporate bond market, such commonalities are reflected in sector spreads, which widen when investors require higher yields in compensation for higher perceived business risk. In addition, rating agencies evaluate the creditworthiness of corporate bond issues, and credit quality spreads vary with changes in the required yields for fixed-income securities of a given rating. Figure 1.3 shows the combined market-based yield effect of the spot rate yield-curve and nominal spread changes for an investor holding AA-rated 10-year industrial bonds. For example, for the nine-month period ending June 30, 2004, increases in the 10-year spot rate and the 10-year AA spread of 0.64 percent and 0.12 percent, respectively, combined to result in a total change of 0.76 percent in the yield of AA-rated 10-year industrial bonds.

Table 1.9 shows the total returns of the Lehman U.S. Government and the Lehman AA Industrials Indexes for the same evaluation periods. The AA Industrials Index modestly underperformed the Government Index in the nine-month period ended June 30, 2004, when the yield curve rose and the nominal spread widened, and significantly outperformed in the subsequent quarter, when the yield curve fell and the nominal spread was essentially unchanged. In addition, of course, the spreads of individual 10-year AA-rated industrial bonds may vary from the average reflected by the sector index, and those differences, too, will be reflected in the actual performance of a specific portfolio.

The impact of interest rate and spread movements on the investment performance of a given portfolio depends upon the nature of the market changes and the interest-sensitive characteristics of the portfolio. We have already seen two types of yield-curve changes: An upward (although nonparallel) shift in the nine-month period ended June 30, 2004, and a twist in the third quarter of 2004 when short-term rates rose and long-term rates fell. Additionally, in both cases, the slope of the yield curve changed. An indicator of the slope



	10-Yr Yield	10-Yr AA Industrial Spread	Total
9 Mos. Ended 30 June 2004	0.64	0.12	0.76
3 Mos. Ended 30 Sept 2004	−0.46	0.01	−0.45
12 Mos. Ended 30 Sept 2004	0.18	0.13	0.31

FIGURE 1.3 Yield Curve and Nominal Spread Changes

TABLE 1.9 Total Returns Data

	Total Returns	
	Lehman U.S. Government Index	Lehman AA Industrials Index
9 months ended June 30, 2004	−0.56%	−0.58%
3 months ended September 30, 2004	3.11%	3.71%
12 months ended September 30, 2004	2.52%	3.11%

is the difference between the 2-year and the 10-year yield-curve rates. The difference was 2.48 percent on September 30, 2003, 1.90 percent on June 30, 2004, and 1.52 percent on September 30, 2004. Thus, over this time frame, the U.S. government spot rate yield curve flattened from one measurement point to the next.

The external interest rate environment is not under the control of the manager; the manager can dictate only the composition of the Portfolio. Subject to the constraints established by the investment mandate and the pertinent policies or guidelines, the manager can adjust the Portfolio's interest-sensitive characteristics in anticipation of forecasted yield-curve and spread changes. Different fixed-income instruments and portfolios will respond diversely to yield-curve movements like those shown above. For example, the resulting adjustment in the valuation of a mortgage-backed portfolio will not be the same as the valuation change of a government bond portfolio. Even portfolios made up of the same types of fixed-income securities (for instance, traditional investment-grade corporate bonds) will have different outcomes, depending upon factors including the maturity, coupon, and option features of their constituent holdings. The manager will modify the Portfolio's interest rate risk profile so as to benefit from expected advantageous movements or to attenuate the return impact of expected adverse changes.

In addition to such interest rate management, other management factors contributing to total portfolio return are the allocation of assets among market segments, economic sectors, and quality grades, and the selection of specific securities within those categories. Trading activity during the evaluation period will also have an impact.

These sources of return are displayed in Figure 1.4.[29] The forward interest rates referred to in this exhibit can be calculated from the points along the spot rate government yield curve at the beginning and the end of the performance evaluation period.

The total return of a fixed-income portfolio can be attributed to the external interest rate effect, on one hand, and the management effect, on the other. The return due to the external interest rate environment is estimated from a term structure analysis of a universe of Treasury securities and can be further separated into the return from the implied forward rates (the expected return) and the difference between the actual realized return and the market implied return from the forward rates (the unexpected return). The overall external interest rate effect represents the performance of a passive, default-free bond portfolio.

The management effect is calculated by a series of repricings and provides information about how the management process affects the portfolio returns. The management effect can be decomposed into four components:

1. **Interest rate management effect.** Indicates how well the manager predicts interest rate changes. To calculate this return, each security in the portfolio is priced as if it were a default-free security. The interest rate management contribution is calculated by subtracting the return of the entire Treasury universe from the aggregate return of these repriced securities. The interest rate management effect can be further broken down into returns due to duration, convexity, and yield-curve shape change, as shown in Table 1.10.
2. **Sector/quality effect.** Measures the manager's ability to select the "right" issuing sector and quality group. The sector/quality return is estimated by repricing each security in the portfolio using the average yield premium in its respective category. A gross return can be then calculated based on this price. The return from the sector/quality effect is calculated by subtracting the external effect and the interest rate management effect from this gross return.
3. **Security selection effect.** Measures how the return of a specific security within its sector relates to the average performance of the sector. The security selection effect for each security is the total return of a security minus all the other components. The portfolio

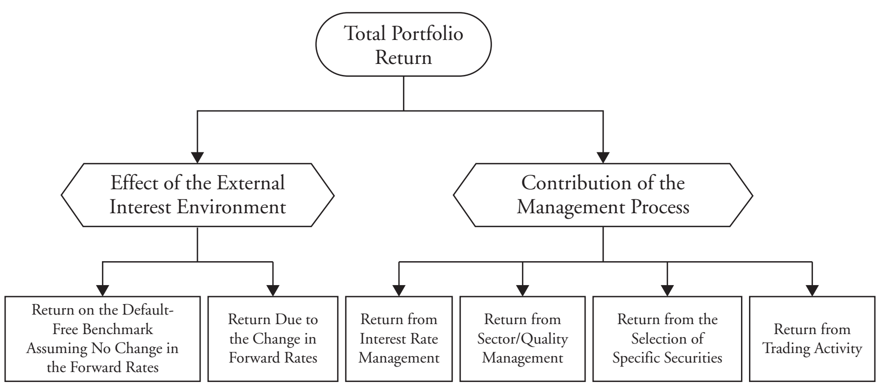


FIGURE 1.4 Sources of the Total Return of a Fixed-Income Portfolio

security selection effect is the market-value weighted average of all the individual security selection effects.

4. **Trading activity.** Captures the effect of sales and purchases of bonds over a given period and is the total portfolio return minus all the other components.

Quantifying the absolute return contributions due to the management effect by means of serial portfolio repricings is data- and computation-intensive, and conducting value-added performance attribution relative to a fixed-income benchmark is still more challenging. Fixed-income investment management organizations often use commercially developed performance measurement and attribution systems. The vendor-provided systems available vary substantially in methodology and level of analytical sophistication, and selecting a system is not a trivial exercise, but most models attempt to isolate and measure the impact of environmental and management factors like those discussed here.

The output of a representative fixed-income attribution system can be demonstrated through a brief illustration. Let us consider the case of the investment officer of the Windsor Foundation, whose consultant has analyzed the performance of two of the foundation's external fixed-income managers, Broughton Asset Management and Matthews Advisors. The consultant has prepared an attribution analysis, shown in Table 1.10, for a particular evaluation period.

TABLE 1.10 Performance Attribution Analysis for Two Fixed-Income Managers for the Windsor Foundation Year Ending December 31, 20XX

	Evaluation Period Returns (%)		
	Broughton Asset Management	Matthews Advisors	Bond Portfolio Benchmark
I. Interest Rate Effect			
1. Expected	0.44	0.44	0.44
2. Unexpected	0.55	0.55	0.55
Subtotal	0.99	0.99	0.99
II. Interest Rate Management Effect			
3. Duration	0.15	−0.13	0.00
4. Convexity	−0.03	−0.06	0.00
5. Yield-Curve Shape Change	0.04	0.13	0.00
Subtotal (options adjusted)	0.16	−0.06	0.00
III. Other Management Effects			
6. Sector/Quality	−0.09	1.15	0.00
7. Bond Selectivity	0.12	−0.08	0.00
8. Transaction Costs	0.00	0.00	0.00
Subtotal	0.03	1.07	0.00
IV. Trading Activity Return	0.10	0.08	0.00
V. Total Return (sum of I, II, III, and IV)	1.28	2.08	0.99

The consultant also included in the analysis the following summary of the investment management strategies of the two firms:

- Broughton Asset Management states that its investment strategy relies on active interest rate management decisions to outperform the benchmark index. Broughton also seeks to identify individual issues that are mispriced.
- Matthews Advisors states that its investment strategy is to enhance portfolio returns by identifying undervalued sectors while maintaining a neutral interest rate exposure relative to the benchmark index. Matthews believes it is not possible to enhance returns through individual bond selection on a consistent basis.

Does the consultant's attribution analysis validate the two firms' self-descriptions of their investment strategies?

In fact, the foundation officer and the consultant can *preliminarily* conclude on the basis of the single year under review that approximately one-half of the incremental return due to Broughton's management process can be attributed to relying on active interest rate management decisions. The total performance contribution for the interest rate management effect—the primary indicator of effective active interest rate management decisions in this analysis—was 16 basis points out of a total of 29 basis points due to the manager's active management process. In addition, the performance contribution for bond selectivity—here, the most direct measure of success in security selection—was 12 basis points. Therefore, nearly all of Broughton's positive performance of 29 basis points (1.28 percent versus 0.99 percent) was a result of its stated strategies of interest rate management (16 basis points) and security selection (12 basis points).

Interestingly, a substantial portion of Matthews' performance results are attributable to the firm's success in identifying undervalued sectors. The positive performance contribution for sector and quality was 1.15 percent, representing a large proportion of Matthews' return relative to the benchmark and indicating success over the evaluation period.

Fixed-income performance attribution is receiving increasing attention from plan sponsors and consultants, but it remains primarily the province of investment managers who have access to the requisite capital market data services as well as the scale of operations to justify the expense and the expertise needed to interpret the results in depth.

PERFORMANCE APPRAISAL

The final phase of the performance evaluation process is performance appraisal. The two preceding phases supplied information indicating how the account performed and quantifying the sources of that performance relative to a designated benchmark. Ultimately, however, fund sponsors are concerned with whether the manager of the account has displayed investment skill and whether the manager is likely to sustain that skill. The goal of performance appraisal is to provide quantitative evidence that the fund sponsor can use to make decisions about whether to retain or modify portions of its investment program.

That said, perhaps no issue elicits more frustration on the part of fund sponsors than the subject of appraising manager investment skill. The problem stems from the inherent uncertainty surrounding the outcome of active management decisions. Even the most talented managers can underperform their benchmarks during any given quarter, year, or even multiyear period due to poor luck. Conversely, ineffective managers at times may make correct decisions and outperform their benchmarks simply by good fortune. We will return to this concept later.

What do we mean by the term *investment skill?* We define **investment skill** as the ability to outperform an appropriate benchmark consistently over time. As discussed previously, a manager's returns in excess of his or her benchmark are commonly referred to as the manager's value-added return or active return. Because no manager is omniscient, every manager's value-added returns, regardless of the manager's skill, will be positive in some periods and negative in others. Nevertheless, a skillful manager should produce a larger value-added return more frequently than his or her less talented peers.

We emphasize that a skillful manager may produce a small value-added return very frequently or a larger value-added return less frequently. It is the magnitude of the value-added returns relative to the variability of value-added returns that determines a manager's skill.

When evaluating managers, many fund sponsors focus solely on the level of value-added returns produced while ignoring value-added return volatility. As a consequence, superior managers may be terminated (or not hired) and inferior managers may be retained (or hired) on the basis of statistically questionable performance data.

Risk-Adjusted Performance Appraisal Measures

Risk-adjusted performance appraisal methods can mitigate the natural fixation on rates of return. There are a number of appraisal measures that explicitly take the volatility of returns into account. A widely accepted principle of investment management theory and practice is that investors are risk averse and therefore require additional expected return to compensate for increased risk. Thus, it is not surprising that measures of performance appraisal compare returns generated by an account manager with the account's corresponding risk. Two types of risk are typically applied to deflate *ex post* returns: the account's market (or systematic) risk, as measured by its beta, and the account's total risk, as measured by its standard deviation.

Three risk-adjusted performance appraisal measures have become widely used: ***ex post*** alpha (also known as **Jensen's alpha**), the **Treynor ratio** (also known as **reward-to-volatility** or excess return to nondiversifiable risk), and the **Sharpe ratio** (also known as **reward-to-variability**). Another measure, $\mathbf{M^2}$, has also received some acceptance. A thorough discussion of these measures can be found in standard investment texts such as Sharpe, Alexander, and Bailey (1999), but we present a summary here. We consider these measures in their *ex post* (after the fact) form used to appraise a past record of performance.

Ex post Alpha

The *ex post* alpha (also known as the *ex post* Jensen's alpha—see Jensen 1968, 1969) uses the *ex post* Security Market Line (SML) to form a benchmark for performance appraisal purposes. Recall that the capital asset pricing model (CAPM) developed by Sharpe (1966), Lintner (1965), and Mossin (1966), from which the *ex post* SML is derived, assumes that on an *ex ante* (before the fact) basis, expected account returns are a linear function of the risk-free return plus a risk premium that is based on the expected excess return on the market portfolio over the risk-free return, scaled by the amount of systematic risk (beta) assumed by the account. That is, over a single period, the *ex ante* CAPM (SML) is

$$E(R_A) = r_f + \beta_A[E(R_M) - r_f] \tag{1.17}$$

where

$E(R_A)$ = the expected return on the account, given its beta
r_f = the risk-free rate of return (known constant for the evaluation period)
$E(R_M)$ = the expected return on the market portfolio
β_A = the account's beta or sensitivity to returns on the market portfolio, equal to the ratio of covariance to variance as $\text{Cov}(R_A, R_M)/\text{Var}(R_M)$

With data on the actual returns of (1) the account, (2) a proxy for the market portfolio (a market index), and (3) the risk-free rate, we can produce an *ex post* version of the CAPM relationship. Rearranging Equation 1.17, a simple linear regression can estimate the parameters of the following relationship:

$$R_{At} - r_{ft} = \alpha_A + \beta_A(R_{Mt} - r_{ft}) + \varepsilon_t \tag{1.18}$$

where for period t, R_{At} is the return on the account, r_{ft} is the risk-free return, and R_{Mt} is the return on the market proxy (market index).[30] The term α_A is the intercept of the regression, β_A is the beta of the account relative to the market index, and ε is the random error term of the regression equation. The estimate of the intercept term α_A is the *ex post* alpha. We can interpret *ex post* alpha as the differential return of the account compared to the return required to compensate for the systematic risk assumed by the account during the evaluation period. The level of the manager's demonstrated skill is indicated by the sign and value of the *ex post* alpha. Left unsaid is whether the fund sponsor prefers a manager with a large (positive) but highly variable alpha to one that produces a smaller (positive) but less variable alpha.

Treynor Measure

The Treynor measure (see Treynor 1965) is closely related to the *ex post* alpha. Like the *ex post* alpha, the Treynor measure relates an account's excess returns to the systematic risk assumed by the account. As a result, it too uses the *ex post* SML to form a benchmark, but in a somewhat different manner than the *ex post* alpha. The calculation of the Treynor ratio is

$$T_A = \frac{\bar{R}_A - \bar{r}_f}{\hat{\beta}_A} \tag{1.19}$$

$\bar{R}_A$ and $\bar{r}_f$ are the average values of each variable over the evaluation period. The Treynor ratio has a relatively simple visual interpretation, given that the beta of the risk-free asset is zero. The Treynor ratio is simply the slope of a line, graphed in the space of mean *ex post* returns and beta, which connects the average risk-free return to the point representing the average return and beta of the account. When viewed alongside the *ex post* SML, the account's benchmark effectively becomes the slope of the *ex post* SML. Thus, a skillful manager will produce returns that result in a slope greater than the slope of the *ex post* SML.

Both the *ex post* alpha and the Treynor measure will always give the same assessment of the existence of investment skill. This correspondence is evident from the fact that any account with a positive *ex post* alpha must plot above the *ex post* SML. Therefore, the slope of a line connecting the risk-free rate to this account must be greater than the slope of the *ex post* SML, the indication of skill under the Treynor ratio.

Sharpe Ratio

Both the *ex post* alpha and Treynor ratio compare excess returns on an account relative to the account's systematic risk. In contrast, the Sharpe ratio (see Sharpe 1966) compares excess returns to the total risk of the account, where total risk is measured by the account's standard deviation of returns. The *ex post* Sharpe ratio is traditionally given by

$$S_A = \frac{\bar{R}_A - \bar{r}_f}{\hat{\sigma}_A} \tag{1.20}$$

The benchmark in the case of the Sharpe ratio is based on the *ex post* capital market line (CML). The *ex post* CML is plotted in the space of returns and standard deviation of returns and connects the risk-free return and the point representing the mean return on the market index and its estimated standard deviation during the evaluation period. As with the Treynor ratio, a skillful manager will produce returns that place the account above the CML, and hence the slope of the line connecting the risk-free rate and the account will lie above the *ex post* CML. Such a manager is producing more average return relative to the risk-free rate per unit of volatility than is a passive investment in the market index.

M^2

Like the Sharpe ratio, M^2 (see Modigliani and Modigliani 1997) uses standard deviation as the measure of risk and is based on the *ex post* CML. M^2 is the mean incremental return over a market index of a hypothetical portfolio formed by combining the account with borrowing or lending at the risk-free rate so as to match the standard deviation of the market index. M^2 measures what the account would have returned if it had taken on the same total risk as the market index. To produce that benchmark, M^2 scales up or down the excess return of the account over the risk-free rate by a factor equal to the ratio of the market index's standard deviation to the account's standard deviation.

$$M_A^2 = \bar{r}_f + \left(\frac{\bar{R}_A - \bar{r}_f}{\hat{\sigma}_A} \right) \hat{\sigma}_M \quad (1.21)$$

Visually, we can consider a line from the average risk-free rate to the point representing the average return and standard deviation of the account. Extending (or retracing) this line to a point corresponding to the standard deviation of the market index allows us to compare the return on the account to that of the market index at the same level of risk. A skillful manager will generate an M^2 value that exceeds the return on the market index.

M^2 will evaluate the skill of a manager exactly as does the Sharpe ratio. Further, as we discussed, the Jensen's alpha and the Treynor ratio will produce the same conclusions regarding the existence of manager skill. However, it is possible for the Sharpe ratio and M^2 to identify a manager as not skillful, although the *ex post* alpha and the Treynor ratio come to the opposite conclusion. This outcome is most likely to occur in instances where the manager takes on a large amount of nonsystematic risk in the account relative to the account's systematic risk. In that case, one can see by comparing Equations 1.19 and 1.20 that while the numerator remains the same, increased nonsystematic risk will lower the Sharpe ratio but leave the Treynor ratio unaffected. As the market index, by definition, has no nonsystematic risk, the account's performance will look weaker relative to the market index under the Sharpe ratio than under the Treynor ratio (and Jensen's alpha).

Information Ratio

The Sharpe ratio can be used to incorporate both risk-adjusted returns and a benchmark appropriate for the manager of the account under evaluation. In its traditional form, the numerator of the Sharpe ratio is expressed as the returns on the account in excess of the risk-free rate. Similarly, the denominator is expressed as the standard deviation of the difference in returns between the account and the risk-free return. However, by definition, in a single-period context the risk-free rate has no variability, and hence the denominator can be stated as the variability in the account's returns.

Because the Sharpe ratio is based on a differential return, it represents the results of a self-financing strategy. A certain dollar amount can be viewed as being invested in the account, with this long position funded by short-selling the risk-free asset; that is, borrowing at the risk-free rate is assumed to fund the investment in the account. In order to provide a relevant context for performance appraisal using the traditional form, we must identify an appropriate benchmark and compute the Sharpe ratio for that benchmark as well as the account. A higher Sharpe ratio for the account than for the benchmark indicates superior performance.

There is no reason, however, for insisting on appraising performance in the context of borrowing at the risk-free rate to fund the investment in the account. Instead, the Sharpe ratio can be generalized to directly incorporate a benchmark appropriate to the account manager's particular investment style. Equation 1.20 can be rewritten to show the long position in the account is funded by a short position in the benchmark:

$$\text{IR}_A = \frac{\bar{R}_A - \bar{R}_B}{\hat{\sigma}_{A-B}} \tag{1.22}$$

where $\hat{\sigma}_{A-B}$ is the standard deviation of the difference between the returns on the account and the returns on the benchmark. The Sharpe ratio in this form is commonly referred to as the **information ratio,** defined as the excess return of the account over the benchmark relative to the variability of that excess return. The numerator is often referred to as the **active return** on the account, and the denominator is referred to as the account's **active risk**. Thus, from this perspective, the information ratio measures the reward earned by the account manager per incremental unit of risk created by deviating from the benchmark's holdings.

Criticisms of Risk-Adjusted Performance Appraisal Methods

A number of criticisms of risk-adjusted performance measures have surfaced over the years, and we will return to some of those arguments later in the discussion. Perhaps the most prominent criticisms have involved the reliance of the *ex post* alpha and the Treynor ratio on the validity of the CAPM. The CAPM has come under attack for a variety of reasons, most notably the appropriateness of its underlying assumptions and the single-index nature of the model. If assets are valued according to some other equilibrium pricing model, then beta-based performance measures may give inaccurate appraisals.

Critics (for example, Roll 1978) have also pointed to problems raised by the use of surrogates (such as the S&P 500) for the true market portfolio. Roll showed that slight changes in the market portfolio surrogate can yield significantly different performance appraisal answers.

Even those appraisal methods not tied to the CAPM face implementation problems. For example, the use of a market index or custom benchmark in the appraisal of investment performance is open to criticism in that it is difficult in most cases for the account manager to replicate precisely the benchmark's return over time (see French and Henderson 1985). Transaction costs associated with initially creating and then later rebalancing the benchmark, as well as the costs of reinvesting income flows, mean that the benchmark's reported returns overstate the performance that a passive investor in the benchmark could earn.

Stability of the parameters and the estimation error involved in the risk-adjusted appraisal measures is also an issue. Even if the assumptions underlying the appraisal measures hold true, the *ex post* calculations are merely estimates of the true parameters of the actual risk–return relationships. If the estimates are recalculated over another period, they may well show conclusions that conflict with the earlier estimates, even if those relationships are stable over time. Further, that stability cannot be taken for granted; the aggressiveness of the account manager may change rapidly over time in ways that cannot be captured by the estimation procedures.

Quality Control Charts

Conveying the essence of performance appraisal to decision makers is a difficult task. A vast quantity of data needs to be synthesized into a relatively few graphs and tables if information overload is to be avoided. Yet this summary process should not come at the expense of sound data analysis. In particular, it should not preclude a consideration of the statistical and economic significance of the performance results. One effective means of presenting performance appraisal data is through the use of **quality control charts**.

Figure 1.5 presents an example of a quality control chart. It illustrates the performance of an actively managed account versus a selected benchmark. The straight horizontal line emanating from the vertical axis at zero represents the performance of the benchmark. The jagged line is the portfolio's cumulative annualized performance relative to the benchmark (that is, the manager's value-added return). The funnel-shaped lines surrounding the horizontal lines form a confidence band, a statistical concept about which we will have more to say shortly. The confidence band offers a means to evaluate the statistical significance of the account's performance relative to the benchmark.

Underlying the quality control chart's construction are three assumptions concerning the likely distribution of the manager's value-added returns. The primary assumption (and one that we will subsequently test) is referred to as the null hypothesis. The null hypothesis of the quality control chart is that the manager has no investment skill; thus, the expected value-added return is zero. With respect to Figure 1.5, we expect that the manager's value-added return line will coincide with the benchmark line.

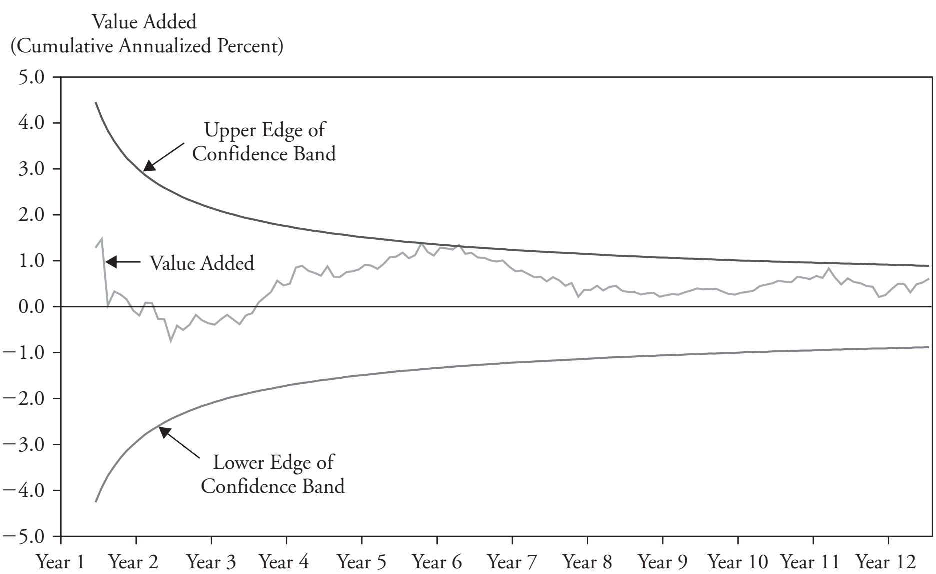


FIGURE 1.5 Quality Control Chart: Cumulative Annualized Value Added Illustrating Manager Performance within Expectations

Of course, at the end of an evaluation period it is highly unlikely that the account's return will precisely equal that of the benchmark. The account's actual return will be either above or below the benchmark's return. The null hypothesis, however, suggests that those *ex post* differences have no directional biases and are entirely due to random chance.

Our second assumption states that the manager's value-added returns are independent from period to period and normally distributed around the expected value of zero. The third assumption is that the manager's investment process does not change from period to period. Among other things, this third assumption implies that the variability of the manager's value-added returns remains constant over time.

Now consider the manager whose investment results are shown in Figure 1.5. Employing the three assumptions described above, we can completely describe the expected distribution of the manager's value-added returns, as illustrated in Figure 1.6. Corresponding to our second assumption of normally distributed value-added returns, the shape of the distribution is the familiar bell-shaped curve. Under our first assumption of no skill (the null hypothesis), the center (or mean) of the distribution is located at 0 percent. Finally, given our third assumption that the manager does not alter his or her investment process over time, we can use the manager's past performance to estimate the dispersion of the value-added return distribution. That dispersion is measured by the standard deviation of the value-added returns, which in this case is an annualized 4.1 percent. We therefore expect that two-thirds of the time, the manager's annual value-added return results will be within ±4.1 percentage points of the zero mean.

Given this information, we can compute a confidence band associated with the expected distribution of the manager's value-added returns. Based on our three assumptions, the **confidence band** indicates the range in which we anticipate that the manager's value-added returns will fall a specified percentage of the time.

In our example, suppose that we wished to determine a confidence band designed to capture 80 percent of the manager's value-added return results. Based on the properties of a normal distribution, we know that 1.28 standard deviations around the mean will capture *ex ante* 80 percent of the possible outcomes associated with a normally distributed random variable. With a 4.1 percent annual standard deviation of value-added returns, the 80 percent confidence band in our example therefore covers a range from approximately −5.2 percent to approximately +5.2 percent around the manager's expected value-added return of zero.

This range, however, corresponds to only one time period: one year from the start of the analysis. To create the confidence band at other points in time, we must transform the

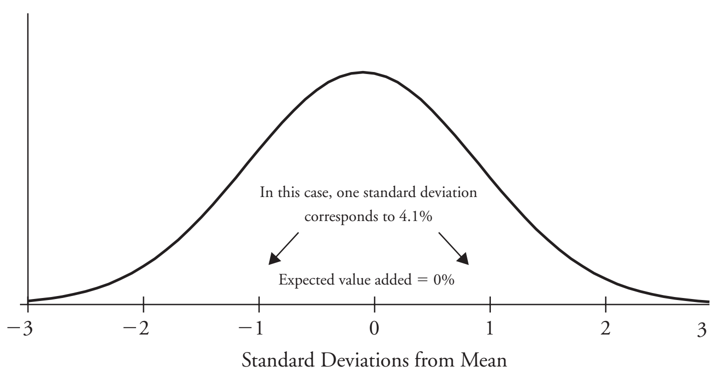


FIGURE 1.6 Expected Distribution of the Manager's Value Added

standard deviation of the manager's value-added returns to address annualized cumulative value-added returns. This transformation produces the funnel-shaped lines shown in Figure 1.5.

The standard deviation of annualized cumulative value-added returns decreases at a rate equal to the square root of time. As a result, the standard deviation of annualized cumulative value-added returns at two years is $1/\sqrt{2}$ of the one-year value, at three years it is $1/\sqrt{3}$ of the one-year value, and so on. Because the width of the confidence band depends on the standard deviation of value-added returns, as time passes, the confidence band will narrow, converging on the benchmark line.

Intuitively, that convergence means that as we collect more observations on the manager's value-added returns, the cumulative annualized results should lie closer to our expected value of zero. That is, as time passes, it becomes increasingly likely that the manager's random positive- and negative-value-added returns will offset one another. Therefore, the chances that the manager will produce a "large" cumulative annualized value-added return, on either side of the mean, declines over time.

Interpreting the Quality Control Chart

Statistical inference by its nature can be a baffling exercise in double negatives. For example, we do not *accept* the null hypothesis. Rather, lacking evidence to the contrary, we *fail to reject* it. Nevertheless, the equivocal nature of this type of analysis is well suited to the world of investments, where luck often masquerades as skill and skill is frequently overwhelmed by random events.

For example, do the data presented in Figure 1.5 tell us anything about the manager's investment skill? The answer in this case is inconclusive. Over the full period of analysis, the manager has outperformed the benchmark by about 1.0 percent per year. Based on this outcome, we might be tempted to certify the manager as being truly skillful. Before leaping to that conclusion, however, recall that our null hypothesis is that the manager has no skill. What we are really asking is, "Do the manager's performance results warrant rejecting the null hypothesis?" Remember that we assume the manager's value-added returns are normally distributed with a constant annual standard deviation of 4.1 percent. Given those assumptions, under the zero-value-added return null hypothesis, there exists a strong possibility that the manager could possess no skill and yet produce the results shown in Figure 1.5.

The quality control chart analysis provides a likely range of value-added return results for a manager who possesses no skill and who displays a specified level of value-added return variability. For a manager whose investment results are within that range (confidence band), we have no strong statistical evidence to indicate that our initial assumption of no skill is incorrect. Thus we are left with the rather unsatisfying statement, "We cannot reject the null hypothesis that the manager has no skill."

It may be true that the manager in Figure 1.5 has skill and that the 1 percent value-added return was no fluke. Unfortunately, over the limited time that we have to observe the manager, and given the variability of the manager's value-added returns, we cannot classify the manager as unambiguously skillful. Even if the manager could actually produce a 1 percent value-added return over the long run, his or her talents are obscured by the variability of his or her short-run results. That performance "noise" makes it difficult to distinguish his or her talents from those of an unskillful manager.

Now let us consider another manager who generates the value-added return series shown in Figure 1.7. The confidence interval is again designed to capture 80 percent of the potential

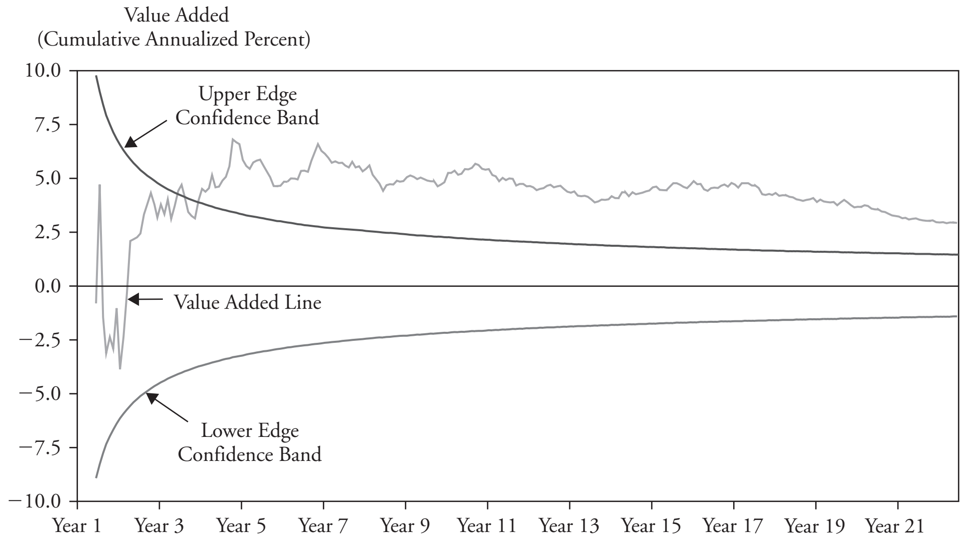


FIGURE 1.7 Quality Control Chart: Cumulative Annualized Value Added Illustrating Manager Performance Significantly Greater than Benchmark

value-added return outcomes for a zero-value-added return manager with a specified level of value-added return variability. In this case, the manager has breached the confidence band on the upside, outperforming the benchmark by about 5 percent per year over the evaluation period. How should we interpret this situation? One view is that the manager has no skill and was simply lucky. After all, there is a 2-in-10 chance that a zero-value added return manager might produce results that lie somewhere outside the confidence band (actually, a 1-in-10 chance of lying above and a 1-in-10 chance of lying below the confidence band).

On the other hand, we could reject the null hypothesis. That is, there is only a 20 percent chance that a zero-value-added return manager would produce results that lie outside the confidence band. Therefore, the occurrence of such an event might indicate that our initial assumption that the manager has no skill is incorrect. Note that our statement would then be, "We reject the null hypothesis that the manager's expected value-added return is zero." By implication, then, we accept a strategy hypothesis that the manager's expected value-added return is not zero.[31]

The quality control chart analysis is similar on the downside. That is, suppose that the manager produces a cumulative negative-value-added return yet lies above the lower edge of the confidence band. In that situation, we should not reject the null hypothesis that the manager's expected value-added return is zero. The manager might be a negative-value-added return investor (that is, be unable to earn back his or her management fees and trading costs). On the other hand, the manager might be skillful and simply be having a poor run of investment luck. In such a case, the relatively small negative-value-added return compared to the variability of that value-added return would make it difficult to reject the null hypothesis.

Conversely, piercing the confidence interval on the downside might lead us to reject the null hypothesis that the manager's expected value-added return is zero. The unstated implication is that the manager is systematically incapable of recapturing the costs of doing business and should be classified as an "underperformer."

THE PRACTICE OF PERFORMANCE EVALUATION

The three components of performance evaluation provide the quantitative inputs required to evaluate the investment skill of an account's manager. However, regardless of the amount of performance data compiled, the process of performance evaluation is fraught with imprecision. Performance evaluation is ultimately a forward-looking decision, and the connection between past performance and future performance is tenuous at best.[32] Indiscriminate use of quantitative data can lead to counterproductive decisions.

As a result, in evaluating investment managers, most fund sponsors follow a procedure that incorporates both quantitative and qualitative elements, with the latter typically receiving more weight than the former. For example, in selecting investment managers, many fund sponsors follow a relatively standard set of procedures. For the sake of exposition, we consider a "typical" fund sponsor. The fund sponsor has a several-person staff that carries out the fund's day-to-day operations. The fund sponsor may retain a consultant to assist in the search for new managers. The staff continually scans the marketplace for promising investment managers. The staff may become aware of a manager through such means as visits from the manager to the staff's office, attendance at various conferences, discussions with peers at other fund sponsor organizations, meetings with consulting firms, and the financial press. The staff maintains files on those managers who have attracted interest, collecting historical return data, portfolio compositions, manager investment process descriptions, and other pertinent data. Upon deciding to hire a new manager, the staff will research its files and select a group of managers for extensive review. This initial cut is an informal decision based on the staff's ongoing survey of the manager marketplace.

The review of the "finalist" group is a much more formal and extensive process. The staff requests that each finalist submit detailed data concerning virtually all aspects of its organization and operations. We broadly group this data into six categories, as shown in Table 1.11.

The staff assigns weights or relative importance to each of these criteria. Table 1.11 shows one possible set of weights. The staff does not apply these criteria and weights in a

TABLE 1.11 Criteria for Manager Selection

Criteria	Importance
Physical	5%
• Organizational structure, size, experience, other resources	
People	25
• Investment professionals, compensation	
Process	30
• Investment philosophy, style, decision making	
Procedures	15
• Benchmarks, trading, quality control	
Performance	20
• Results relative to an appropriate benchmark	
Price	5
• Investment management fees	

mechanical manner. Its ultimate decisions are actually quite subjective. The important point is that the staff considers a broad range of quantitative and qualitative factors in arriving at a selection recommendation. No single factor dominates the decision: Performance data are only one component in the ultimate evaluation decision.

In addition to collecting written information, the staff meets personally with the key decision makers from each of the finalist managers. In those meetings, the staff engages in a broad discussion, the purpose of which is to focus on specific aspects of the managers' operations as highlighted by the selection criteria.

After meeting with all of the finalists, the staff compares notes and selects a manager (or managers) to recommend to the fund sponsor's investment committee, which makes the final decision. The committee members are much more performance-oriented than the staff. Nevertheless, they usually support the staff's well-researched recommendations.

Noisiness of Performance Data

The goal of evaluating prospective or existing managers is to hire or keep the best managers and to eliminate managers likely to produce inferior future results. If past performance were closely tied to future performance, then it would be desirable to rely heavily on past performance in evaluating managers. The problem is that empirical evidence generally does not support such a relationship.

The confusion results from the uncertain, or stochastic, nature of active management. Active managers are highly fallible. While we may expect a superior manager to perform well over any given time period, we will observe that the superior manager's actual performance is quite variable. Even sophisticated investors tend to focus on expected returns and ignore this risk element.

EXAMPLE 1.17 The Influence of Noise on Performance Appraisal

Suppose that we know in advance that a manager is superior and will produce an annual value-added return of 2 percent, on average. The variability of that superior performance is 5 percent per year. Our hypothetical manager has an information ratio of 0.40 (2% ÷ 5%), which by our experience is a high figure. (Hence our assertion that this manager is a superior manager.) Table 1.12 shows the probability of managers outperforming their benchmarks over various evaluation periods, given the information ratios.

TABLE 1.12 Probability of a Manager Outperforming a Benchmark Given Various Levels of Investment Skill

	Information Ratio					
Years	0.20	0.30	0.40	0.67	0.80	1.00
0.5	55.63%	58.40%	61.14%	68.13%	71.42%	76.02%
1.0	57.93	61.79	65.54	74.75	78.81	84.03
3.0	63.81	69.83	75.58	87.59	91.71	95.84
5.0	67.26	74.88	81.45	93.20	96.32	98.73
10.0	73.65	82.86	89.70	98.25	99.43	99.92
20.0	81.70	91.01	96.32	99.86	99.98	99.99

Perhaps surprisingly, Table 1.12 shows that the manager has a one-in-four chance of underperforming the benchmark over a period as long as three years, as seen by the boxed cell in the exhibit. Remember, we have defined this manager in advance to be a superior manager. Other value-added managers with less skill than this one have a greater chance of underperforming their benchmarks over typical evaluation periods.

Most fund sponsors hire more than one manager. Consider a group of 10 superior managers whose investment skills equal those of the manager in Example 1.17 (who has an information ratio of 0.40) and assume independence of decision-making processes. Table 1.13 shows the probability of a given number of this group simultaneously underperforming their benchmarks over a three-year period. As we can see, a fund sponsor using a simple decision rule of firing any manager who underperforms his or her benchmark over a three-year period can expect to follow a busy manager search schedule. Moreover, these probabilities are conservatively low. Few of the fund sponsor's managers will have the investment skill with which we have endowed our hypothetical managers.

In summary, using past performance to evaluate existing managers is statistically problematic. In the long run, superior managers will outperform inferior managers. However, due to the inherent uncertainty of investment management, over typical evaluation periods (three to five years) the odds that superior managers will underperform their benchmarks (and, conversely, that inferior managers will outperform their benchmarks) are disturbingly high. Expensive, incorrect decisions may frequently result from relying on past performance to evaluate investment managers.

Manager Continuation Policy

Frequent manager firings based on recent performance might seem to be merely a waste of a fund sponsor's time if not for the expenses associated with manager transitions. Fired managers'

TABLE 1.13 Probability of Superior Managers Jointly Underperforming Their Benchmarks over a Three-Year Period

Managers Below Benchmark	Probability
0	6.10%
1	19.68
2	28.59
3	24.60
4	13.90
5	5.38
6	1.45
7	0.27
8	0.03
9	0.00
10	0.00

portfolios must be converted to the hired managers' portfolios. This conversion requires buying and selling securities, which in turn involves trading costs. Making assumptions about the cost of trading securities is a tenuous business at best, because many factors influence that cost. For U.S. large-capitalization common stocks, it is reasonable to assume transaction costs of 0.50 percent (one way), and for small company stocks and stocks of companies traded in less liquid markets, those costs can be much higher. A substantial percentage of the fired manager's portfolio may need to be liquidated in the process of moving the assets to a new manager, particularly when the managers' styles are not closely similar. Moreover, this tally of the expenses of converting a manager's portfolio considers only direct monetary costs. For most fund sponsors, replacing managers involves significant time and effort.[33]

In an attempt to reduce the costs of manager turnover yet systematically act on indications of future poor performance, some fund sponsors have adopted formal, written **manager continuation policies** (MCP) to guide their manager evaluations. The purpose of an MCP is severalfold:

- To retain superior managers and to remove inferior managers, preferably before the latter can produce adverse results
- To ensure that relevant nonperformance information is given significant weight in the evaluation process
- To minimize manager turnover
- To develop procedures that will be consistently applied regardless of investment committee and staff changes

An MCP can be viewed as a two-part process. The first part we refer to as **manager monitoring**, while the second part we call **manager review**. Figure 1.8 displays a flow chart description of an MCP.

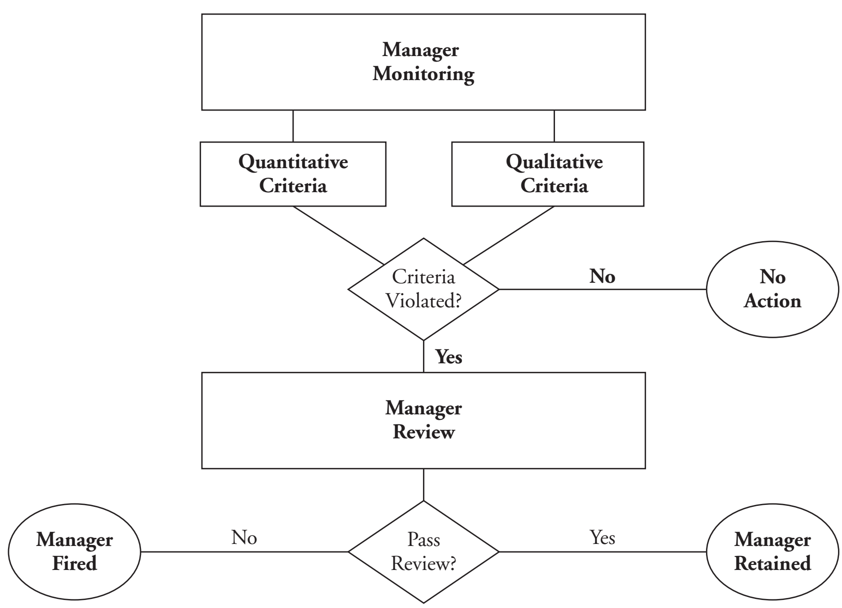


FIGURE 1.8 Manager Continuation Policy

Manager Monitoring

The ongoing phase of an MCP is manager monitoring. The goal of MCP manager monitoring is to identify warning signs of adverse changes in existing managers' organizations. It is a formal, documented procedure that assists fund sponsors in consistently collecting information relevant to evaluating the state of their managers' operations. The key is that the fund sponsor regularly asks the same important questions, both in written correspondence and in face-to-face meetings.

There is no firm set of appropriate manager monitoring criteria. Each fund sponsor must determine for itself the issues that are relevant to its own particular circumstances. Monitoring criteria may even vary from manager to manager. Regardless, the fund sponsor should clearly articulate its established criteria at the time a manager is hired, rather than formulate them later in a haphazard manner.

As part of the manager monitoring process, the fund sponsor periodically receives information from the managers, either in written form or through face-to-face meetings. This information is divided into two parts. The first part covers operational matters, such as personnel changes, account growth, litigation, and so on. The staff should flag significant items and discuss them in a timely manner with the respective managers.

The second part of the responses contains a discussion of the managers' investment strategies, on both a retrospective and a prospective basis. The fund sponsor should instruct the managers to explain their recent investment strategies relative to their respective benchmarks and how those strategies performed. The managers should follow this review with a discussion of their current strategies relative to the benchmark and why they believe that those strategies are appropriate. The goal of these discussions is to assure the fund sponsor that the manager is continuing to pursue a coherent, consistent investment approach. Unsatisfactory manager responses may be interpreted as warning signs that the manager's investment approach may be less well-defined or less consistently implemented than the staff had previously believed.

As part of the manager monitoring process, the staff should regularly collect portfolio return and composition data for a performance attribution analysis. The purpose of such a periodic analysis is to evaluate not how well the managers have performed, but whether that performance has been consistent with the managers' stated investment styles. The staff should address questions arising from this analysis directly to the managers.

Typically, the results of the MCP manager monitoring stage reveal nothing of serious concern. That is, the managers' organizations remain stable, and the managers continue to follow their stated investment approaches regardless of the near-term success or failure of their particular active strategies. While the managers should be able to explain why particular strategies failed, the mere occurrence of isolated periods of poor performance should typically not be a cause for concern, unless the staff finds related nonperformance problems.

Manager Review

Occasionally, manager monitoring may identify an item of sufficient concern to trigger a manager review. For example, a recently successful small manager might experience excessive growth in accounts and assets. Despite discussions with the manager, the staff might be convinced that such growth threatens the manager's ability to produce superior returns in the future. At this point, a formal manager review becomes necessary.

The manager review closely resembles the manager selection process, in both the information considered and the comprehensiveness of the analysis. The staff should review all phases of the manager's operations, just as if the manager were being initially hired. We can view this manager review as a zero-based budgeting process (a budgeting process in which all expenditures must be justified each new period). We want to answer the question, "Would we hire the manager again today?"

As with the initial selection of a manager, the fund sponsor should collect the same comprehensive data and meet face-to-face with the manager in a formal interview setting. The manager's key personnel should attend, with the advance understanding that they must persuade the staff to "rehire" them. On conclusion of the interview, the staff should meet to compare observations, weighing the evaluation criteria in the same manner that it would if it were initially considering the manager. As part of these deliberations, the fund sponsor should also review the information that led to the manager's hiring in the first place.

The primary differences between hiring a new manager and retaining a manager under review are that the fund sponsor once had enough confidence in the manager to entrust a large sum of money to the manager's judgment and that there is a sizable cost associated with firing the manager. Thus, the fund sponsor should address the following questions:

- What has fundamentally changed in the manager's operation?
- Is the change significant?
- What are the likely ramifications of the change?
- Are the costs of firing the manager outweighed by the potential benefits?

Manager Continuation Policy as a Filter

For many reasons, investment skill does not readily lend itself to rigid "good" or "bad" interpretations. For discussion purposes, however, we will arbitrarily divide the investment manager community into three categories: positive-, zero-, and negative-value-added managers. Assume that positive-value-added managers beat their benchmarks (after all fees and expenses) by 2 percent per year, on average. Zero-value-added managers exhibit just enough skill to cover their fees and expenses and thereby match the performance of their benchmarks. Finally, negative-value-added managers lose to their benchmarks by 1 percent per year, on average, due primarily to the impact of fees and expenses.

We have no firm evidence as to how the manager community is apportioned among these three categories, although if we follow the logic of Grossman and Stiglitz (1980) and Sharpe (1994), the zero- and negative-value-added managers must predominate, with the former outnumbering the latter. Nevertheless, we speculate that out of five managers hired, a fund sponsor would be fortunate to hire two positive-value-added managers, two zero-value-added managers, and one negative-value-added manager. Therefore, in aggregate, this successful fund sponsor's managers are expected to outperform their benchmarks by 60 basis points per year, net of all costs [$0.6\% = (2\% \times 0.4) + (0\% \times 0.4) + (-1\% \times 0.2)$].

We can view a MCP as a statistical filter designed to remove negative-value-added managers and retain positive-value-added managers. Zero-value-added managers, much to the consternation of fund sponsors, always present a problem for a MCP, because they are so numerous and because they are statistically difficult to distinguish from positive- and negative-value-added managers.

We begin our MCP analysis with the null hypothesis that the managers under evaluation are at best zero-value-added managers. Then, as with any filter, two types of decision errors may occur:

- **Type I error**—keeping (or hiring) managers with zero value added. (Rejecting the null hypothesis when it is correct.)

- **Type II error**—firing (or not hiring) managers with positive value added. (Not rejecting the null hypothesis when it is incorrect.)

In implementing a MCP, the fund sponsor must determine how fine a filter to construct. A coarse filter will be conducive to Type I errors. For example, a fund sponsor may choose to overlook many violations of its manager monitoring guidelines, with the expectation that most problems experienced by managers are temporary and that they will eventually work themselves out. While this policy will avoid firing some positive-value-added managers, the fund sponsor could have identified in advance some managers who will provide mediocre long-term performance.

Conversely, a fine filter will lead the sponsor to commit more Type II errors. For example, a fund sponsor might apply its manager monitoring guidelines rigidly and automatically fire any manager who loses a key decision maker. While this policy will remove some managers whose operations will be disrupted by personnel turnover, it will also eliminate some managers possibly anticipated to recover from that turnover and to continue with superior results.

Figure 1.9 presents the four possible results from testing the null hypothesis that a manager has no investment skill. Referring back to the quality control chart, if in truth the manager has no skill and we reject the null hypothesis because the manager's value-added returns fall outside of the confidence band (particularly, in this case, on the upside), then we have committed a Type I error. Conversely, if the manager is indeed skillful yet we fail to reject the null hypothesis because the manager's value-added returns fall inside the confidence band, then we have committed a Type II error.

Both Type I and Type II errors are expensive. The art of a MCP is to strike a cost-effective balance between the two that avoids excessive manager turnover yet eliminates managers likely to produce inferior performance in the future. We can control the probabilities of committing Type I and Type II errors by adjusting the width of the confidence band within the quality control chart. For example, suppose that we widened the confidence band to encompass 95 percent of a manager's possible value-added return outcomes. Now it will be less likely than in our earlier examples that a zero-value-added return manager will generate returns that lie outside the confidence band. We thus reduce the chances of a Type I error. However, it will also now be less likely that a truly skillful manager will come to our attention by generating returns that fall outside that manager's confidence band. By continuing not to reject the null hypothesis for such a manager, we commit a Type II error.

Due to the high costs and uncertain benefits of replacing managers, it would seem advisable for fund sponsors to develop manager evaluation procedures that are tolerant toward Type I errors in order to reduce the probability of Type II errors. That is, it may be preferable to endure the discomfort of keeping several unskillful managers to avoid the expense of firing a truly superior manager. However, there is no right answer to this dilemma, and fund

	Reality	
	Value Added = 0	Value Added > 0
Reject	Type I	Correct
Do Not Reject	Correct	Type II

FIGURE 1.9 Null Hypothesis: Manager Has No Skill. Alternative Hypothesis: Manager Is Skillful.

sponsors must undertake their own cost–benefit analyses, weighing the chances of committing one type of error versus the other. The quality control chart approach, however, provides fund sponsors with an objective framework with which to address this issue.

NOTES

1. The evaluation period in this sense can also be called the measurement period.
2. From the fund sponsor's perspective, the account's market value should reflect the impact of all fees and expenses associated with investing the account's assets. Many managers report the return on accounts that they manage without including the effect of various fees and expenses. This practice is often justified based on the fact that fees vary among clients.
3. Note that the account's reported return was lower when the contribution took place at the start of the month than at the end. This result occurs because the account had both a positive return and proportionately more assets to invest over the month when the contribution was received at the beginning as opposed to the end. If the account's return had been negative, then, given the same ending value, a contribution at the beginning of the month would have resulted in a less negative reported return than would have resulted from a contribution that occurred at the end of the month.
4. For a discussion of the use of the MWR as a performance measure for opportunistic investments, see Tierney and Bailey (1997).
5. Nevertheless, for periods beginning January 1, 2010, firms will be required to value portfolios on the date of all large external cash flows to claim compliance with the GIPS standards. In the interim, the GIPS standards admit the use of acceptable daily weighted methods for estimating the time-weighted rate of return. These methods are presented in Chapter 13 of *Managing Investment Portfolios.*
6. Bank Administration Institute (1968, p. 22).
7. The variables used in this section can be interpreted as either rates of return or weights assigned to securities that make up a portfolio.
8. As we have used the term, a benchmark is a means to differentiate managers or fund sponsors who add value through investment insights from those who do not. In this sense, a sponsor's liabilities may also be treated as a type of benchmark. That is, institutional investors such as defined-benefit pension plan sponsors and endowment and foundation executives seek to achieve rates of return enabling them, at a minimum, to meet liabilities as they come due without making greater-than-planned additions to fund assets. (Another way to express this financial objective is to say that institutional investors seek at least to maintain a stated level of fund surplus, defined as the present value of assets less the present value of liabilities.) In terms of asset-liability management, or surplus management, the fund's investment objective may be to achieve a rate of return on assets that meets or exceeds the "return" on liabilities—that is, the percentage change in the present value of the liabilities over the evaluation period. Moreover, because a liability, or a stream of liabilities, may be considered a financial asset held short, it is possible, in principle, to construct a custom index representing the fund's liabilities and to use that index as a benchmark at the level of the total fund.
9. Factor models are discussed in DeFusco, McLeavey, Pinto, and Runkle (2004) as well as in standard investment textbooks such as Sharpe, Alexander, and Bailey (1999).

10. Although the market model has some resemblances to the capital asset pricing model (CAPM), the market model is not an equilibrium model of asset pricing, as is the CAPM. Under a set of specific assumptions, the CAPM states that investors will act in a manner that generates a unique relationship between the beta of a security or portfolio and the return on the market portfolio. Any security or portfolio with the same beta is expected to produce the same return. The market model, on the other hand, is an empirical relationship between the return on a security or portfolio and a particular market index (as opposed to the market portfolio). See Markowitz (1984) for a discussion of this distinction.
11. The ability to track the account's returns is typically measured by the standard deviation of the monthly return differences of the account and the benchmark, called the tracking error.
12. Bailey (1992a) critiques in detail the use of manager universes as benchmarks. Beyond the failure to possess the properties of a valid benchmark and the issue of survivor bias, Bailey also discusses the failure of manager universes to pass tests of benchmark quality. The tests of benchmark quality are summarized in the section titled Tests of Benchmark Quality.
13. The historical beta of the account relative to the benchmark is derived from a regression of the account's past returns on the past returns of the benchmark. The resulting slope of the regression line, termed the beta of the regression, indicates the sensitivity of the account's returns to those of the benchmark. Note that a benchmark may fail this test because the manager holds cash in the account, typically for transaction purposes, while the benchmark may reflect a zero cash position. If the account's beta relative to the benchmark would be 1.0 excluding the positive cash position, the overall beta of the account (including the cash position) will be less than 1.0. As a result, the account will have an unfavorable performance bias in an up market and a favorable bias in a down market. The simple solution is to hold cash in the benchmark at a level reflective of the manager's "neutral" cash position.
14. Risk characteristics refer to factors that systematically affect the returns on many securities. We will return to the issue later in the discussion on performance attribution.
15. Violations of this quality criterion often occur when a benchmark is market capitalization weighted. Because many managers do not utilize a market-capitalization weighting scheme in building their portfolios, the possibility of negative active positions can arise when a capitalization-weighted benchmark is assigned.
16. See Koh, Lee, and Fai (2002).
17. Another approach to determining a rate of return for a long-short portfolio would be to specify the numerator in Equation 1.1 as the profit and/or loss resulting from the particular hedge fund strategy. The denominator could be specified as the asset base over which the strategy applies. This could be defined as the amount of assets at risk and could be approximated by the absolute value of all the long positions plus the absolute value of all the short positions.
18. Rather than using broad market indexes as asset category benchmarks, some fund sponsors and consultants construct asset category benchmarks by weighting the managers' benchmarks in accordance with their policy allocations. Under this approach, using the data given in Table 1.2 and Table 1.3, the blended asset category benchmark for domestic equities would consist of a 65 percent weighting in Large-Cap Growth Index and a 35 percent weighting in Large-Cap Value Index. However, this approach impairs the sponsor's ability to evaluate the impact of "misfit returns" or "style bias" as described later in this chapter.

19. Alternatively, a pension fund might identify the risk-free asset as a portfolio of bonds that best hedges its liabilities.
20. The increment of \$575,474 cannot be replicated by multiplying \$187,369,405 by 0.31 percent because the \$950,000 net contribution (to obtain \$187,369,405) was not a single, beginning-of-the-month cash flow.
21. Note: $\sum_{j=1}^{M} w_{ij} = 1$ for all i and $\sum_{i=1}^{A} w_i = 1$
22. For simplicity we assume that the Portfolio's securities are chosen from among the securities in the benchmark. Otherwise n needs to represent the number of securities in the union of the benchmark and the Portfolio.
23. Note that the sum of the security weights in any portfolio must equal 1.0, or, equivalently, $\sum_{i=1}^{n}(w_{pi} - w_{Bi}) = 0$. Because zero multiplied by a constant equals zero, $\sum_{i=1}^{n}(w_{pi} - w_{Bi}) \times r_B = 0$, where r_B is the known return on the benchmark (the constant). Subtracting this expression from the right-hand side of the equation just given yields $r_v = \sum_{i=1}^{n}[(w_{pi} - w_{Bi}) \times (r_i - r_B)]$.
24. See Spaulding (2003). Transaction-based attribution analysis is outside the scope of the present discussion.
25. Equation 1.16 covers performance attribution in the single-period case. Multiperiod performance attribution, while an extension of the single-period approach, involves considerably more complexity. For a discussion of some of the issues involved in multiperiod performance attribution, see Menchero (2004) and Frongello and Bay (2002).
26. Exposure to a fundamental factor in this case is measured in terms of standard deviations from the mean, where the mean is determined by the average value of the particular factor for a group of capitalization-weighted stocks.
27. Although this type of performance attribution analysis provides valuable insights to investment practitioners, there is a serious limitation. It involves the ambiguity of the benchmark. If the benchmark is based solely on a set of exposures to investment risk factors, then the benchmark is ambiguous. That is, we can construct multiple portfolios that have the same risk characteristics, but they will not have the same investment return. For example, many portfolios might have the same beta, but they will have different investment returns. The solution to this limitation is to base the attribution analysis on the risk exposures of an appropriate benchmark portfolio, i.e., a portfolio with specified securities and weights. In this case, the benchmark portfolio will have a specific or unexplained return component. The difference between it and the portfolio's specific return is attributed to the investment manager.
28. A more rigorous treatment of this discussion of fixed-income micro attribution can be found in Fong, Pearson, and Vasicek (1983).
29. Fong, Pearson, and Vasicek (1983).
30. The *ex post* alpha relationship can be expanded to incorporate other sources of risk (for example, the three-factor model developed by Fama and French). See Carhart (1997) for further discussion.
31. Of course, the assumptions underlying the statistical test may not hold. For example, the manager's investment process may have become more aggressive, and hence the variability of his value-added returns may have increased.
32. See Carhart (1997).
33. The costs associated with manager hiring and firing decisions are discussed in Goyal and Wahal (2005).

REFERENCES

Ambachtsheer, Keith. 1986. *Pension Funds and the Bottom Line: Managing the Corporate Pension Fund as a Financial Business.* Homewood, IL: Dow Jones-Irwin.

Bailey, Jeffery V. 1992a. "Are Manager Universes Acceptable Performance Benchmarks?" *Journal of Portfolio Management*. Vol. 18, No. 3: 9–13.

Bailey, Jeffery V. 1992b. "Evaluating Benchmark Quality." *Financial Analysts Journal*. Vol. 48, No. 3: 33–39.

Bank Administration Institute. 1968. *Measuring the Investment Performance of Pension Funds*. Park Ridge, IL: Bank Administration Institute.

Bernstein, Peter L. 2003. "Points of Inflection: Investment Management Tomorrow." *Financial Analysts Journal*. Vol. 59, No. 4: 18–23.

Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance*. Vol. 51, No. 1: 57–82.

DeFusco, Richard, Dennis McLeavey, Jerald Pinto, and David Runkle. 2004. *Quantitative Methods for Investment Analysis*, 2nd edition. Charlottesville, VA: CFA Institute.

Dietz, Peter. 1966. *Pension Funds: Measuring Investment Performance*. New York: The Free Press.

Ellis, Charles. 1985. *Investment Policy*. Homewood, IL: Dow Jones-Irwin.

Fama, Eugene. 1972. "Components of Investment Performance." *Journal of Finance*. Vol. 27, No. 3: 551–567.

Fama, Eugene, and Kenneth French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance*. Vol. 47, No. 2: 427–465.

Fong, H. Gifford, Charles Pearson, and Oldrich Vasicek. 1983. "Bond Performance: Analyzing Sources of Return." *Journal of Portfolio Management*. Vol. 9, No. 3: 46–50.

French, Dan, and Glenn Henderson, Jr. 1985. "How Well Does Performance Evaluation Perform?" *Journal of Portfolio Management*. Vol. 11, No. 2: 15–18.

Frongello, Andrew, and Scott Bay. 2002. "Linking Single Period Attribution Results." *Journal of Performance Measurement*. Vol. 6, No. 3: 10–22.

Goyal, Amit, and Sunil Wahal. 2005. "The Selection and Termination of Investment Management Firms by Plan Sponsors." Working Paper, Goizueta Business School, Emory University.

Grossman, Sanford, and Joseph Stiglitz. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review*. Vol. 70, No. 3: 393–408.

Jensen, Michael. 1968. "The Performance of Mutual Funds in the Period 1945–1964." *Journal of Finance*. Vol. 23, No. 2: 389–416.

Jensen, Michael. 1969. "Risk, the Pricing of Capital Assets, and the Evaluation of Investment Portfolios." *Journal of Business*. Vol. 42, No. 2: 167–185.

Koh, Francis, David Lee, and Phoon Kok Fai. 2002. "Investing in Hedge Funds: Risk, Return and Pitfalls." Working Paper.

Lintner, John. 1965. "Security Prices, Risk, and Maximal Gains from Diversification." *Journal of Finance*. Vol. 20, No. 4: 587–615.

Markowitz, Harry. 1984. "The Two Beta Trap." *Journal of Portfolio Management*. Vol. 10, No. 1: 12–20.

Menchero, Jose. 2004. "Multiperiod Arithmetic Attribution." *Financial Analysts Journal*. Vol. 60, No. 4: 76–91.

Modigliani, Franco, and Leah Modigliani. 1997. "Risk-Adjusted Performance." *Journal of Portfolio Management*. Vol. 23, No. 2: 45–54.

Mossin, Jan. 1966. "Equilibrium in a Capital Asset Market." *Econometrica*. Vol. 34, No. 4: 768–783.

Roll, Richard. 1978. "Ambiguity When Performance Is Measured by the Security Market Line." *Journal of Finance*. Vol. 33, No. 4: 1051–1069.

Rosenberg, Barr, and Vinay Marathe. 1975. "The Prediction of Investment Risk: Systematic and Residual Risk." *Proceedings of the Seminar on the Analysis of Security Prices*. Center for Research in Security Prices, Graduate School of Business, University of Chicago.

Sharpe, William. 1966. "Mutual Fund Performance." *Journal of Business*. Vol. 39, No. 1: 119–138.

Sharpe, William. 1982. "Factors in New York Stock Exchange Security Returns, 1931–1979." *Journal of Portfolio Management*. Vol. 8, No. 4: 5–19.

Sharpe, William. 1988. "Determining a Fund's Effective Asset Mix." *Investment Management Review*. Vol. 2, No. 6: 59–69.

Sharpe, William. 1992. "Asset Allocation, Management Style and Performance Measurement." *Journal of Portfolio Management*. Vol. 18, No. 2: 7–19.

Sharpe, William. 1994. "The Sharpe Ratio." *Journal of Portfolio Management*. Vol. 21, No. 1: 49–59.

Sharpe, William, Gordon Alexander, and Jeffery V. Bailey. 1999. *Investments*. UpperSaddle River, NJ: Prentice Hall.

Spaulding, David. 2003. "Holdings vs. Transaction-based Attribution, an Overview." *Journal of Performance Measurement*. Vol. 8, No. 1: 52–56.

Tierney, David, and Jeffery V. Bailey. 1997. "Opportunistic Investing." *Journal of Portfolio Management*. Vol. 23, No. 3: 69–78.

Treynor, Jack. 1965. "How to Rate Management of Investment Funds." *Harvard Business Review*. Vol. 43, No. 1: 63–75.