While established performance criteria and passive index benchmarks exist for long-only traditional funds that are evaluated primarily on selectivity, more actively managed funds and alternative investments with changing fund risk and skewed distributions pose new challenges to performance techniques. Quantitative style analysis applied to traditional mutual funds is relatively straightforward [Sharpe, 1992], but strategy classifications for actively managed funds remain largely as qualitative industry descriptions. The methodology presented in this paper provides a quantitative classification technique for commodity trading advisors (CTAs) within a performance analysis framework. This contribution is particularly important because the distinction between traditional and alternative investments is fading as traditional fund managers venture into derivative products and include alternative investments in their portfolios.

This article introduces a new methodology for performance evaluation which incorporates multiple criteria and classifies funds or other securities according to these criteria through benchmarking. This methodology is distinctly different from multifactor analysis where benchmarks are risk factors. In contrast, benchmarks in this article are efficient portfolios defined in n dimensions where each dimension represents risk and return criteria. This approach has the advantage of simultaneously affording a classification scheme and performance evaluation.

This article is organized as follows. The next section reviews traditional portfolio performance measures and analysis techniques. By highlighting the goals and limitations of the existing measures, the review establishes the basis for the new methodology. The data envelopment analysis (DEA) methodology is presented in the third section. The fourth section illustrates the application of DEA to a small data set of commodity trading advisor returns. The final section concludes with a summary and directions for further research.

TRADITIONAL PORTFOLIO PERFORMANCE MEASURES AND ANALYSIS TECHNIQUES

Performance measures based on early or “modern” portfolio theory, models that incorporate time-varying risk and higher moments, multi-index models and factor analysis are reviewed briefly below. While not a comprehensive review, this discussion highlights traditional assumptions regarding forms of the return generating process, both linear and nonlinear, concerns over mean-variance sufficiency, and the interest in both style and strategy classifications.

Early Portfolio Theory

While risk has long been recognized as an important consideration for investment decisions, performance was commonly measured only in terms of returns before “modern” portfolio theory was developed a half a century ago.
Within the Markowitz [1952] framework, total risk is quantified by the standard deviation of returns. Tobin [1958] extended the Markowitz efficient frontier by adding the risk-free asset, resulting in the capital market line (CML) and paving the way for the development of the capital asset pricing model (CAPM), developed by Sharpe [1964], Lintner [1965], and Mossin [1966]. The CAPM defines systematic risk, measured by beta ($\beta$), as the relevant portion of total risk since investors can diversify away the remaining portion.

Two of the earliest composite performance measures, still in common use today, are the Treynor [1965] measure of excess return per unit of systematic risk, and the Sharpe [1966] measure of excess return per unit of total risk. These two measures are used to rank financial investments according to risk-adjusted performance. Since a fully diversified portfolio’s total risk equals its systematic risk, the Treynor and Sharpe measures will provide identical rankings for well-diversified funds. Jensen’s [1968] alpha is a CAPM-based performance measure of an asset’s average return in excess of that predicted by the CAPM, given its beta and the market return. These early performance measures and related models are summarized in Exhibit 1.

---

**E X H I B I T  1**

Early Performance Measures, Related Models, and Notation

<table>
<thead>
<tr>
<th>Performance Measures</th>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treynor (T)</td>
<td>$T_i = (R_i - r_f)/\beta_i$</td>
<td>An asset with a Treynor ratio higher than the market’s Treynor ratio will plot above the SML, indicating superior risk-adjusted performance.</td>
</tr>
<tr>
<td>Sharpe (S)</td>
<td>$S_i = (R_i - r_f)/\sigma_i$</td>
<td>An asset with a Sharpe ratio higher than the market’s Sharpe ratio will plot above the CML, indicating superior risk-adjusted performance.</td>
</tr>
<tr>
<td>Jensen’s alpha (α)</td>
<td>$\alpha = R_i - [r_f + \beta_i(R_m - r_f)]$</td>
<td>A positive alpha indicates superior performance.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Related Models</th>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Covariance Model</td>
<td>Minimize $x'Vx$ subject to $x'E = r^*_p$, $x'1 = 1$.</td>
<td>This model finds the portfolio with the least variance in a universe of $n$ assets, given a target return, $r^<em>_p$. $V$ is the $n \times n$ variance-covariance matrix of asset returns, $E$ is the $n \times 1$ vector of expected returns, and $x$ is the resultant vector of optimal weights. Varying $r^</em>_p$ and repeating the minimization process would trace out an efficient frontier of the given assets.</td>
</tr>
<tr>
<td>Capital Market Line (CML)</td>
<td>$R_i = r_f + \beta_i[R_m - r_f]\sigma_i/\sigma_m$</td>
<td>The capital allocation line (CAL) shows all feasible risk-return combinations of a risky and risk-free asset. The CML is a CAL with the risky asset replaced by the market portfolio. The market portfolio lies on the CML where it is tangent to the efficient frontier of all possible risky assets.</td>
</tr>
<tr>
<td>CAPM’s Security Market Line (SML)</td>
<td>$E(R_i) = r_f + \beta_i[E(R_m) - r_f']$</td>
<td>Ex-ante equilibrium relationship. Derived from the CML.</td>
</tr>
<tr>
<td>Index Model</td>
<td>$R_i - r_f = \alpha + \beta_i(R_m - r_f)$</td>
<td>Ex-post regression estimates. If the CAPM holds, $\alpha$ is expected to be zero.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_i$</td>
<td>Beta, $\beta_i = \sigma_{im}/\sigma_m^2$, a measure of systematic or market risk</td>
</tr>
<tr>
<td>$R_i$</td>
<td>is the return on the individual fund or asset being evaluated</td>
</tr>
<tr>
<td>$\sigma_i$</td>
<td>is the standard deviation of asset $i$’s returns</td>
</tr>
<tr>
<td>$r_f$</td>
<td>is the risk-free rate</td>
</tr>
<tr>
<td>$E(R_m)$</td>
<td>is the expected return on the market portfolio</td>
</tr>
<tr>
<td>$\sigma_{im}$</td>
<td>is the covariance between an asset’s returns with the returns on the market portfolio</td>
</tr>
<tr>
<td>$\sigma_m^2$</td>
<td>is the variance of the market returns</td>
</tr>
</tbody>
</table>
Time-Varying Risk

Single parameter risk measures are problematic if managers are changing fund betas over time, as they would if they were attempting to time the market. For example, when the market is expected to increase, the manager may increase the fund's beta and vice versa. In this situation, risk may be accurately measured only if the portfolio weights are known, and this information is generally not publicly available. Therefore, other techniques must be developed and employed.

Treynor and Mazuy [1966] suggest adding a quadratic term to the basic linear regression model to capture nonlinearities in beta resulting from market timing activities. Kon and Jen [1978, 1979] use a switching regression technique. Henriksson and Merton [1981] develop nonparametric and parametric option-based methods to test for directional market timing ability. The nonparametric approach requires knowledge of the managers' forecasts. The more commonly employed parametric approach involves adding an extra term to the usual linear regression model and is CAPM based. More recently, Ferson and Schadt [1996] note that fund betas may change in response to changes in betas of the underlying assets as well as from changing portfolio weights. They modify the classic CAPM performance evaluation techniques to account for time variation in risk premiums by using a conditional CAPM framework. This method removes the perverse negative performance often found in earlier tests, and suggests that including information variables in performance analysis is important. Some of these alternative models that focus on addressing the effects of nonstationary betas are summarized in Exhibit 2.

Higher Moments

The standard CAPM framework assumes that investors are concerned with only the mean and variance of returns. Ang and Chau [1979] argue that skewness in return distributions should be incorporated into the performance measurement process. Even if the returns of the risky assets within a portfolio are normally distributed, dynamic trading strategies may produce non-normal distributions in portfolio returns. Both Prakash and Bear [1986] and Stephens and Proffitt [1991] also develop higher-moment performance measurements. These models are summarized in Exhibit 3.

Sortino and van der Meer [1991], Fishburn [1977], Sortino and Price [1994], Marmer and Ng [1993], Merriken [1994], and others have also developed measures for investors that are more concerned with downside risk (or semivariance) than the standard deviation of returns. Although some differences exist among these measures, the popular Sortino ratio captures their essence. Whereas the Sharpe ratio is defined as return divided by standard deviation, the Sortino ratio is defined as return divided by downside deviation. Downside deviation (DD) measures the deviations below some minimal accepted return (MAR). Of course, when the MAR is the average return and returns are normally distributed, the Sharpe and Sortino ratios will yield identical rankings. Martin and Spurgin [1998] illustrate that even if individual asset or fund returns

---

**Exhibit 2**

Standard Regression-Based Models Incorporating Time-Varying Risk

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treynor and Mazuy [1966]</td>
<td>( r_p - r_t = a + b(r_m - r_p) + c(r_m - r_p)^2 + \epsilon_p )</td>
</tr>
<tr>
<td>Kon and Jen [1978, 1979]</td>
<td>( Y_1 = X_1\alpha_1 + u_1; Y_2 = X_2\alpha_2 + u_2; )</td>
</tr>
<tr>
<td>Henriksson and Merton [1981]</td>
<td>( \begin{align*} r_p - r_t &amp;= \alpha + \beta(r_m - r_t) + c(r_m - r_t)D + \epsilon_p \ D &amp;= 1 \text{ for } r_m &gt; r_t \text{ and } 0 \text{ otherwise} \end{align*} )</td>
</tr>
<tr>
<td>Ferson and Schadt [1996]</td>
<td>Conditional versions of Treynor and Mazuy and Henriksson and Merton</td>
</tr>
</tbody>
</table>

**Exhibit 3**

Standard Regression-Based Models Incorporating Higher Moments

<table>
<thead>
<tr>
<th>Model</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ang and Chau [1979]</td>
<td>( R_i - r_t = E(R_i) + C_1(R_m - R_p) + C_2(R_m - E(R_m))^2 + \epsilon_i )</td>
</tr>
<tr>
<td>Prakash and Bear [1986]</td>
<td>( R_{it} = C_0 + C_1\text{COV}(R_{it},R_{mt}) + C_2\text{COSKEW}(R_{it},R_{mt}) + \epsilon_{it} )</td>
</tr>
<tr>
<td>Stephens and Proffitt [1991]</td>
<td>( R_{it} = C_0 + C_1\text{COV}(R_{it},R_{mt}) + C_2\text{COSKEW}(R_{it},R_{mt}) + C_3\text{COKURT}(R_{it},R_{mt}) + \epsilon_{it} )</td>
</tr>
</tbody>
</table>
are skewed, investors may not care because the returns of a diversified portfolio will not be skewed. However, they also illustrate that managers may choose to follow strategies that produce skewed returns as a form of signaling their skill. Note that coskewness remains irrelevant if it can be diversified away, but skewness may have some signaling value. Additionally, the popularity of the related Value at Risk (VaR) measure and the common practice of reporting drawdown information for various alternative investments suggest that skewness may be important, whether it be in terms of investor utility or skill signaling.

Multi-Index Models and Factor Analysis

Whereas various single-index models are based on the CAPM and assume that security returns are a function of their comovements with the market portfolio, multi-index (or multifactor) models assume that returns are also a function of additional influences. For example, Chen, Roll, and Ross [1986] develop a model where returns are a function of factors related to cash flows and discount rates such as gross national product and inflation. The purposes of multi-index models are varied and, in addition to performance attribution, include forming expectations about returns and identifying sources of returns.

Sharpe [1992] decomposes stock portfolio returns into several “style” factors (more narrowly defined asset classes such as growth and income stocks, value stocks, high-yield bonds) and show that the portfolio’s mix accounts for up to 98% of portfolio returns. Similarly, Brinson, Singer, and Beebower [1991] show that rather than selectivity or market timing abilities, it is the portfolio mix (allocation to stocks, bonds, and cash) that determines over 90% of portfolio returns. However, Brown and Goetzmann [1995] identify a tendency for fund returns to be correlated across managers, suggesting performance is due to common strategies that are not captured in style analysis.

Schneeweis and Spurgin [1998] use various published indexes (Goldman Sachs Commodity Index, Standard and Poor’s 500 stock index, the Salomon Brothers government bond index, and U.S. dollar trade-weighted currency index) with absolute return, intramonth volatility, and the Mount Lucas management (MLM) index in a multifactor regression analysis to describe the sources of return to hedge funds, managed futures, and mutual funds. The index returns are employed to capture the source of natural returns. The use of absolute return is designed to capture the source of return that derives from the ability to go short or long. Returns from the use of options or intramonth timing strategies are proxied by the intramonth standard deviation. The MLM index, an active index designed to mimic trend-following strategies, is used to capture sources of returns from market inefficiencies in the form of temporary trends.

Mitev [1998] uses maximum likelihood factor analysis technique to classify commodity trading advisors (CTAs) according to unobservable factors. Similarly, Fung and Hsieh [1997] also use a factor-analytic approach to classify hedge funds. In both cases, the results identify general investment approaches or trading strategies (e.g., trend-following, spread strategies, or systems approaches) as sources of returns to these alternative investment classes. The difference between the factor analysis and multifactor regression analysis is in their approach to identifying the factors (benchmarks) that serve as proxies for risk. In multifactor regression analysis, the factors are pre-specified. Factor analysis will identify funds that have common yet unobservable factors, although the factors can be inferred from the qualitative descriptions of the funds. While this may seem redundant, the clustering of funds is done independently of the qualitative descriptions, is a formal data-driven process.

REVIEW SUMMARY

Sharpe, Treynor, and Sortino ratios, alpha, beta, drawdown statistics, Value at Risk (VaR), M-squared, a variety of regression-based models, and other performance models and metrics exist because some measures are more appropriate for certain purposes than others. For example, it may be argued that the Sharpe measure is appropriate when analyzing an entire portfolio, while the Treynor measure is appropriate when evaluating a security or investment that is part of a larger portfolio. However, the multitude of approaches also suggests that more than one measure of risk may be needed to accurately assess performance. Furthermore, many existing performance models can yield inaccurate results if managers are following active strategies and changing the funds’ risk or the riskiness of the assets comprising a fund is variable. Such limitations, combined with evidence that fund returns are correlated across fund managers, suggest that strategy as well as style classifications may be important in evaluating the risk-adjusted performance of managed investments.

Established performance criteria, passive index benchmarks, and style analysis may be appropriate for long-only
traditional funds. However, more actively managed funds and alternative investments with changing fund risk and skewed return distributions call for different techniques. While strategy classifications remain largely qualitative, efforts have been made through multifactor and factor analysis techniques as described above. The methodology presented below provides a quantitative classification technique within a performance analysis framework. The methodology incorporates multiple criteria and classifies (“benchmarks”) funds or other securities according to these criteria. This is distinctly different from multifactor analysis. Here, benchmarks are not risk factors but rather they are efficient securities as defined in n dimensions where each dimension represents risk and return criteria. This approach has the advantage of simultaneously affording both a classification scheme and performance evaluation.

A NEW APPROACH

It can clearly be seen that a traditional portfolio evaluation approach either incorporates only a single risk measure or assumes a certain functional relationship among various portfolio performance measures. For example, if (linear) regression-based models are used, then it is assumed that all portfolios obey the same underlying functional relationship among different measures. In particular, this functional form is usually unknown and needs to be estimated. However, if the relationship is not correctly expressed before the estimation, then the results are unreliable.

On the other hand, regression-based methods only describe average behavior. Consider Exhibit 4 where a set of portfolios is plotted with return as the y-axis and risk as the x-axis. The regression gives a line AB that is the average performance of this particular set of portfolios, whereas the true frontier (the best or efficient portfolio) is the piecewise linear line segments of CD and DF. In order to obtain CD and DF rather than AB, the coefficients in a regression need to be estimated differently. That is, for each segment of the efficient frontier, a new set of estimates on coefficients is required. However, regression models fail to do this. The current paper seeks help from mathematical programming techniques. Specifically, the current paper uses the data envelopment analysis (DEA) method to fulfill the new estimation task as presented by CD and DF in Exhibit 4.

DATA ENVELOPMENT ANALYSIS (DEA)

DEA is a mathematical programming approach developed to evaluate the relative efficiency or performance of a set of units that have multiple performance measures [Charnes, Cooper, and Rhodes, 1978]. DEA is particularly useful when the relationship among the multiple performance measures is unknown. Through the optimization for each individual unit, DEA yields an efficient frontier (such as CD and DF in Exhibit 4) that represents and estimates the relations among the multiple performance measures.

Suppose we have a set of n units (portfolios), \(U_j (j = 1, \ldots, n)\) and let \(x_i (i = 1, \ldots, m)\) be the m performance measures where smaller values are preferred, e.g., risk measures and \(y_r (r = 1, \ldots, s)\) be the s performance measures where larger values are preferred, e.g., returns. Thus, we have \(m + s\) performance measures for the n units. Further, we have \(x_{ij}\) as the observed value on the ith “small-preferred” performance measure and \(y_{rj}\) as the observed value on the rth “large-preferred” performance measure.

We then use the following procedure. First, for each unit define

\[
h_j = \frac{\alpha + \sum_{i=1}^{m} v_i x_{ij}}{\sum_{r=1}^{s} u_r y_{rj}}
\]

(1)

where \(\alpha, v_i \geq 0\), and \(u_r \geq 0\) are unknown variables. Next, consider one of the units, say the oth unit, \(U_o\) with \(x_{io}\) and \(y_{ro}\) and set up the following mathematical programming problem

---

Copyright © Institutional Investor, Inc. All rights reserved.
subject to

\[
\begin{align*}
\alpha + \sum_{i=1}^{m} v_i x_{ij} & \geq \gamma_j, \quad j = 1, \ldots, n, \\
\sum_{i=1}^{n} u_i y_{ro} & = 1
\end{align*}
\]

It is clear from model (2) that smaller value of \( h^*_o \) is preferred since we prefer larger values of \( y_{ro} \) and smaller values of \( x_{io} \). Therefore, model (2) tries to find a set of weights \( v_i \) and \( u_i \) so that the ratio of aggregated \( x_{io} \) to aggregated \( y_{ro} \) reaches the minimum. Because of the constraints \( h_j \geq 1 \) in model (2), the optimal value to (2) or the minimum \( h^*_o \), must be equal to or greater than one. Obviously, a score of one represents the best, i.e., if the unity value is achieved for \( U_o(h^*_o = 1) \), then \( U_o \) is efficient in terms of the given multiple performance measures. Note that model (2) is solved for each unit. Therefore, model (2) does not seek the average best performance, but the efficient or best performance achievable by a proper set of optimized weights.

Note that when \( h^*_o = 1 \), we have

\[
\sum_{r=1}^{s} u_r y_{ro} = \alpha^* + \sum_{i=1}^{m} v_i x_{io}
\]

where \( (\*) \) represents the optimal values in model (2). It can be seen that (3) is similar to the regression model with \( \alpha^* \) the intercept on the y-axis. The implicit difference between model (2) and the regression model lies in the fact that 1) model (2) deals with more than one dependent variables \( (y_{ro}) \) at the same time and 2) Equation (3) is obtained for each unit with a score of one. Further, (3) represents the efficient frontier. Since different units with score of one in model (2) may not be on the same frontier, the resulting efficient frontier is a piecewise linear one as shown in Exhibit 4.9

### Calculation of DEA Models and Benchmarks

In mathematical programming, model (2) is a linear fractional programming problem and is difficult to solve. However, model (2) can be transformed into a linear programming problem that can be easily solved.

Set \( \sum_{r=1}^{s} u_r y_{ro} = 1 \) in model (2) and model (2) is now a linear programming problem

\[
\begin{align*}
h^*_o & = \min_{\alpha, v_i, u_i} \alpha + \sum_{i=1}^{m} v_i x_{io} \\
\text{subject to}
\end{align*}
\]

\[
\alpha + \sum_{i=1}^{m} v_i x_{ij} \geq \gamma_j, \quad j = 1, \ldots, n
\]

\[
\sum_{i=1}^{n} u_i y_{ro} = 1
\]

The dual to (4) is

\[
\varphi^*_o = \max \varphi_o \\
\text{subject to}
\]

\[
\sum_{j=1}^{n} \lambda_j x_{ij} \leq x_{io}, \quad i = 1, 2, \ldots, m;
\]

\[
\sum_{j=1}^{n} \lambda_j y_{ro} \geq \varphi_o y_{ro}, \quad r = 1, 2, \ldots, s;
\]

\[
\sum_{j=1}^{n} \lambda_j = 1
\]

\[
\lambda_j \geq 0, \quad j = 1, \ldots, n
\]

Based upon Charnes et al. [1978], we have 1) if the optimal value to (5) \( \varphi^*_o = 1 \) then a particular unit (portfolio) \( U_o \) is efficient and \( \lambda^*_o = 1 \) and \( \lambda^*_j = 0(j \neq o) \) and 2) if \( \varphi^*_o \neq 1 \) (i.e., \( \varphi^*_o > 1 \)), then \( U_o \) is inefficient, and \( \lambda^*_o = 0 \) and some \( \lambda^*_j \neq 0(j \neq o) \). Further, only the efficient units will be associated with \( \lambda^*_j \neq 0 \). Thus, for an inefficient unit, the following gives a benchmark value

\[
\begin{align*}
\sum_{j=1}^{n} \lambda^*_j x_{ij} \\
\sum_{j=1}^{n} \lambda^*_j y_{ro}
\end{align*}
\]

where \( \sum_{j=1}^{n} \lambda^*_j = 1 \).
It is clear from (5) that 1) the composition of the benchmark is clearly delineated; 2) (6) is an investable option; 3) the return on the benchmark can be calculated; 4) since (6) is obtained via model (5), the benchmark is consistent with Uo’s investment style.

In DEA literature, model (5) is called an output-oriented model. We can have an input-oriented model, which yields the same efficient frontier.

\[ \theta_o^* = \min \theta_o \]
subject to
\[ \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta_o x_{io} \quad i = 1, 2, \ldots, m; \]
\[ \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro} \quad r = 1, 2, \ldots, s; \]
\[ \sum_{j=1}^{n} \lambda_j = 1 \]
\[ \lambda_j \geq 0 \quad j = 1, \ldots, n \]

We have 1) if the optimal value to (7) \( \theta_o^* = 1 \), then a particular unit (portfolio) \( U_o \) is efficient and \( \lambda_j^* = 1 \) and \( \lambda_j^* = 0 (j \neq o) \) and 2) if \( \theta_o^* \neq 1 (i.e., \theta_o^* < 1) \), then \( U_o \) is inefficient, and \( \lambda_o^* = 0 \) and some \( \lambda_j^* \neq 0 (j \neq o) \).

\[ \sum_{j=1}^{n} \lambda_j = 1 \]
\[ \lambda_j \geq 0 \quad j = 1, \ldots, n \]

\[ x_{ij} = \text{Standard Deviation} \]
\[ x_{2} = \text{PropNeg (proportion of negative monthly returns during the year)} \]
\[ y_{1} = \text{Return (average monthly return)} \]
\[ y_{2} = \text{Skewness} \]
\[ y_{3} = \text{Min (minimum return)} \]

AN ILLUSTRATIVE EMPIRICAL APPLICATION

As a simple illustration, we use one year (1999) of monthly return data for the eleven commodity trading advisors (CTAs) that are classified as financial/metal in the International Traders Research (ITR) database. The following inputs to the model were calculated from the monthly returns:

\[ x_{1} = \text{Standard Deviation} \]
\[ x_{2} = \text{PropNeg (proportion of negative monthly returns during the year)} \]
\[ y_{1} = \text{Return (average monthly return)} \]
\[ y_{2} = \text{Skewness} \]
\[ y_{3} = \text{Min (minimum return)} \]

Exhibit 5 reports the data on the above five measures. Note that some values on return, skewness, and Min are negative. Therefore the average monthly return, skewness, and minimum return are displaced by 0.04, 4, and 1, respectively. Also, model (7) is used because we have negative outputs. Since there are 11 CTAs, we will solve model (7) 11 times, each time for a CTA. To illustrate the solution for one CTA, the last column of Exhibit 5 shows the solution to model (7) when CTA183 is under evaluation.

Exhibit 6 reports the detailed results from model (7), when each CTA is under evaluation. Model (7) yields a unity score for five CTAs (339, 363, 365, 389, and 1508). Thus, these five CTAs are efficient and can serve as benchmarks for the remaining inefficient ones. Since model (7) is a best-practice technique and does not reflect average performance, for each underperforming CTA, a different benchmark may be identified. For example, CTA384 uses CTA339 as the benchmark. That is, under the current risk level of CTA384, CTA384 is expected to have a return as high as that of CTA339. On the other hand, CTA183 uses a mix of CTA363 and CTA339 as its bench-
**EXHIBIT 5**
Data, Transformations, and Results for CTA183

<table>
<thead>
<tr>
<th>CTA Code</th>
<th>Standard Deviation</th>
<th>x1 = Proportion Negative</th>
<th>x2 = Monthly Av.</th>
<th>y1 = Minimum Return</th>
<th>y2 = Average Return</th>
<th>y3 = Skewness</th>
<th>( \bar{y}<em>r ) = ( y</em>{r,1} * \pi )</th>
<th>( \lambda_j^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>183</td>
<td>6.80%</td>
<td>58.30%</td>
<td>0.10%</td>
<td>-8.10%</td>
<td>4.10%</td>
<td>5.13</td>
<td>91.90%</td>
<td>0.00</td>
</tr>
<tr>
<td>321</td>
<td>4.00%</td>
<td>41.70%</td>
<td>0.70%</td>
<td>-7.90%</td>
<td>4.70%</td>
<td>4.61</td>
<td>92.10%</td>
<td>0.51</td>
</tr>
<tr>
<td>339</td>
<td>3.40%</td>
<td>37.50%</td>
<td>0.90%</td>
<td>-4.00%</td>
<td>4.90%</td>
<td>4.58</td>
<td>96.00%</td>
<td>0.49</td>
</tr>
<tr>
<td>363</td>
<td>5.00%</td>
<td>50.00%</td>
<td>0.60%</td>
<td>-5.60%</td>
<td>4.60%</td>
<td>5.7</td>
<td>94.40%</td>
<td>0.00</td>
</tr>
<tr>
<td>365</td>
<td>4.70%</td>
<td>37.50%</td>
<td>1.10%</td>
<td>-8.20%</td>
<td>5.10%</td>
<td>4.28</td>
<td>91.80%</td>
<td>0.00</td>
</tr>
<tr>
<td>384</td>
<td>3.30%</td>
<td>50.00%</td>
<td>-0.10%</td>
<td>-6.30%</td>
<td>3.90%</td>
<td>4.08</td>
<td>93.80%</td>
<td>0.00</td>
</tr>
<tr>
<td>389</td>
<td>11.20%</td>
<td>45.80%</td>
<td>3.20%</td>
<td>-17.10%</td>
<td>7.20%</td>
<td>4.39</td>
<td>83.00%</td>
<td>0.00</td>
</tr>
<tr>
<td>438</td>
<td>12.80%</td>
<td>58.30%</td>
<td>-1.00%</td>
<td>-25.70%</td>
<td>3.00%</td>
<td>4.46</td>
<td>74.30%</td>
<td>0.00</td>
</tr>
<tr>
<td>633</td>
<td>8.40%</td>
<td>52.20%</td>
<td>-2.20%</td>
<td>-17.10%</td>
<td>2.80%</td>
<td>3.74</td>
<td>82.90%</td>
<td>0.00</td>
</tr>
<tr>
<td>867</td>
<td>5.00%</td>
<td>54.50%</td>
<td>0.40%</td>
<td>-6.70%</td>
<td>4.40%</td>
<td>5.1</td>
<td>93.30%</td>
<td>0.00</td>
</tr>
<tr>
<td>1508</td>
<td>8.60%</td>
<td>25.00%</td>
<td>-3.60%</td>
<td>-16.50%</td>
<td>0.40%</td>
<td>2.02</td>
<td>83.50%</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The original data for each CTA is shown in the first set of columns for “inputs” \( x_1 \) and \( x_2 \) and “outputs” \( y_1 \), \( y_2 \), and \( y_3 \). Because some outputs are negative, we use the input-oriented model (7), and we replace \( y_r \) with \( \hat{y}_r \), \( r = 1, 2, \ldots, s \), \( s = 3 \), using the translation invariance property to transform the outputs and obtain \( \hat{y}_r \). The resulting performance value, \( \theta^*_0 \), is less than one and indicates inferior performance relative to CTA339 and CTA336. The resulting vector, \( \lambda^* \), is used to create a benchmark for CTA183, as illustrated in Exhibit 7.

**EXHIBIT 6**
Model (7) Results for Each CTA

<table>
<thead>
<tr>
<th>CTA</th>
<th>Performance (( \theta_j^* ))</th>
<th>CTA_A (( \lambda_{CTA_A} ))</th>
<th>CTA_B (( \lambda_{CTA_B} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>183</td>
<td>0.75</td>
<td>363</td>
<td>0.49</td>
</tr>
<tr>
<td>321</td>
<td>0.91</td>
<td>363</td>
<td>0.03</td>
</tr>
<tr>
<td>339</td>
<td>1.00</td>
<td>363</td>
<td>1.00</td>
</tr>
<tr>
<td>363</td>
<td>1.00</td>
<td>363</td>
<td>1.00</td>
</tr>
<tr>
<td>365</td>
<td>1.00</td>
<td>363</td>
<td>1.00</td>
</tr>
<tr>
<td>384</td>
<td>0.89</td>
<td>339</td>
<td>1.00</td>
</tr>
<tr>
<td>389</td>
<td>1.00</td>
<td>339</td>
<td>1.00</td>
</tr>
<tr>
<td>438</td>
<td>0.63</td>
<td>339</td>
<td>0.95</td>
</tr>
<tr>
<td>633</td>
<td>0.64</td>
<td>339</td>
<td>0.67</td>
</tr>
<tr>
<td>867</td>
<td>0.82</td>
<td>339</td>
<td>0.53</td>
</tr>
<tr>
<td>1508</td>
<td>1.00</td>
<td>339</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Notes: In this example, there were no more than two nonzero elements in the resulting \( \lambda^* \) vectors. CTA_A is the CTA that corresponds to the first nonzero element, and CTA_B is the CTA that corresponds to the second nonzero element.
mark, i.e., 50% of CTA363 and 50% of CTA339 can be used as a benchmark for CTA339.

By using equation (6), we can calculate the exact values for the benchmark. Calculations for CTA183 are illustrated in Exhibit 7. According to the benchmark, the expected monthly return for CTA183 should be 0.745%, which exceeds its actual return of 0.118%.

CONCLUSIONS AND DIRECTIONS FOR FUTURE RESEARCH

Common presentations for individual commodity trading advisor (CTA) performance results include the average return, standard deviation of returns, Sharpe ratio, and drawdown information. This paper developed a methodology that combines these factors related to performance into a single measure. It is based on data envelopment analysis (DEA), a technique that has been used extensively in the industrial engineering literature as a performance analysis tool [Zhu, 2000]. This method identifies a piecewise linear efficient frontier that considers multiple factors and identifies benchmarks for underperforming funds.

Directions for future research include various empirical analyses. In particular, it would be interesting to see how well this measure predicts returns for traditional and nontraditional funds, if the performance of funds as measured by the DEA is stable over time or correlated with market conditions, and how sensitive the measures are to different inputs.

ENDNOTES

1We use a data envelopment analysis (DEA) methodology. It is well known in the fields of operations research and industrial engineering, but has not yet been widely applied to portfolio analysis and performance evaluation. Recently, however, this application to mutual fund data has appeared in operations research journals (Murthi, Choi, and Desai [1997] and Morey and Morey [1999]). In contrast, we use commodity trading advisor (CTA) return data. A DEA methodology is particularly appropriate for analyzing nonlinear returns generated by CTA strategies.

2Originally called RAP (risk-adjusted performance), M-squared (or M²) is essentially a new variant of the Sharpe ratio developed by Modigliani and Modigliani in 1997. It will provide the same rankings as the Sharpe ratio, yet is expressed in the more intuitive basis points as a volatility-equivalent return. Similar volatility-adjusted return measures, GH1 and GH2, were developed by Graham and Harvey [1996, 1997].

3See Chung [1999] for a concise review of Value at Risk methodologies.

<table>
<thead>
<tr>
<th>x, y Values for CTA363</th>
<th>x, y Values for CTA339</th>
<th>λ^* CTA363</th>
<th>λ^* CTA339</th>
<th>β_1</th>
<th>β_2</th>
<th>Benchmark Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1 = Standard Deviation</td>
<td>0.05 × 0.49 + 0.03 × 0.51</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x_2 = Proportion Negative</td>
<td>0.50 × 0.49 + 0.38 × 0.51</td>
<td>0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{y}_1 = \text{Return} )</td>
<td>0.05 × 0.49 + 0.05 × 0.51</td>
<td>0.01</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{y}_2 = \text{Skewness} )</td>
<td>5.70 × 0.49 + 4.58 × 0.51</td>
<td>1.14</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \hat{y}_3 = \text{Minimum Return} )</td>
<td>0.94 × 0.49 + 0.96 × 0.51</td>
<td>-0.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
See Glazier and Wilkens [1999] for a manager performance methodology that distinguishes skill from luck, ties performance to increased turnover (in the presence of skill and including transaction costs), and is robust to alternative distributional assumptions.

Usually CAPM-based performance models describe covariance with the market portfolio; however, as noted earlier, they can attempt to describe coskewness and cokurtosis as well.

Arbitrage pricing theory (APT) establishes the conditions under which a multi-index model can be an equilibrium description [Ross, 1976].

Tracking error is important when security market benchmark indices are used to provide a performance index that reflects the particular style of an investment manager. Tracking error is the difference between the return on a portfolio and the return on a market index. Among indices that have similar risk/return characteristics, the index that minimizes tracking error with a given portfolio may be regarded as the more suitable benchmark.

For more discussions on DEA and its applications, please refer to Charnes et al. [1978] and Zhu [2000].

CTAs are sometimes closed to outside investment, however.

The ITR database contains monthly return data and descriptive information on over 400 CTAs. For 1999, there were 68 CTAs trading financials; 11 of these CTAs traded financials and metals. This group was chosen for its small size to facilitate a simple illustration.

REFERENCES


