Constructing Peer Benchmarks for Mutual Funds: *A Style Analysis-Based Approach*

ARIK BEN DOR, VERNON BUDINGER, LEV DYNKIN, AND KENNETH LEECH

ARIK BEN DOR
is a senior analyst in Quantitative Portfolio Strategy at Lehman Brothers in New York City, NY.
abendor@lehman.com

VERNON BUDINGER
is a quantitative analyst at Western Asset Management Company in Pasadena, CA.
vbudinger@westernasset.com

LEV DYNKIN
is the head of Quantitative Portfolio Strategy at Lehman Brothers in New York City, NY.
ldynkin@lehman.com

KENNETH LEECH
is the chief investment officer of Western Asset Management Company in Pasadena, CA.
kleech@westernasset.com

Evaluating the investment strategy and performance of mutual funds is increasingly important in light of the continued growth in their number and the intense competition among them for new capital.[1] The most common approach relies on classifying funds into one of several style categories based on their investment practices and risk characteristics.[2] Their performance is then evaluated against that of a certain benchmark—typically a broad-based index that is supposed to be representative of the investment style of the funds.

For styles with a narrow, well-defined investment mandate, such an approach may be appropriate. Yet, some style categories are characterized by flexible mandates that allow investments in multiple asset classes with fairly few restrictions on the portfolio allocation among them. Using an index as a representative benchmark in this case will not capture the large variation across funds of the same style category for at least two reasons. First, a single index that spans the entire investment universe available to the funds may not exist. Second, even if a benchmark can be constructed by say, combining several indices, its fixed weighting scheme is likely to deviate substantially from that employed by the funds over time as well as cross-sectionally.

Proper identification of investment style and performance measurement of mutual funds has been examined primarily to benefit investors and plan sponsors. However, this is also of great importance to mutual fund managers. For example, suppose a manager of a bond fund has a bullish view on interest rates. Such a manager may structure his portfolio to have a duration longer than the benchmark by half a year, but at the same time be actually bearish as compared to peer managers (i.e., if their portfolios exhibit even higher durations).

Some managers are also explicitly subject to a risk budget that is defined in terms of tracking error volatility relative to the aggregate peer group. The challenge for portfolio managers in such a case is to use the risk budget optimally. This requires a reasonable understanding of the peer group portfolio composition. It also requires an understanding of how sensitive these portfolios are to many factors.

Given the broad array of opportunity sets available to some managers (especially managers of multisector funds), how can this information be inferred? How reliable are the conclusions we can draw?

One possible approach is to construct benchmarks based on the quarterly disclosures of holdings reported by mutual funds.[3] This approach suffers from several drawbacks, however.[4] First, holdings data may be difficult to obtain for the entire universe of peers and are available only at a lag. Second, certain reported holdings may be difficult to evaluate (such as private placements and highly structured securities) and may require substantial judgment calls. Third, 'window-dressing' around reporting dates can affect the accuracy of reported holdings and the resulting benchmark.[5]

In this article we demonstrate another approach for constructing peer manager style benchmarks that relies on Sharpe's [1988, 1992] 'Return-based Style Analysis' methodology. Return-based Style Analysis seeks to replicate the performance of a fund over a specified time period as best as possible by constructing a passively managed portfolio of market factors. It identifies the actual investment style a manager uses as compared with the stated style, and helps identify the main drivers of the manager's performance.[6]

We use data on the performance of multisector bond funds to illustrate how style analysis can be used to construct peer manager style benchmarks that allow proper measurement of a fund's risk relative to its competitors. Our approach offers more timely comparisons and results that are easier to interpret than those obtained using holdings-based style analysis. In addition, the approach requires only historical return data which is more readily available than security holdings data.

The article is organized as follows. We first demonstrate that a single index may not serve as an adequate passive benchmark for certain mutual fund categories, even if the index includes all relevant asset classes. Following a short review of the return-based style analysis methodology, we illustrate how it is applied in practice. We discuss the choice of factors, formulation of constraints and estimation procedure. The next section shows the estimation results focusing on their predictive power out-of-sample. The last section provides an example of assessing a fund's risk relative to its peers based on the style analysis results derived for a composite peer index.

PERFORMANCE EVALUATION USING A STATED BENCHMARK

The performance of mutual fund managers is often evaluated against a particular manager-specific passive benchmark. Funds focusing on sectors like technology, health care, or finance would be evaluated against the appropriate sector index. The same is true for managers of equity funds that focus on stocks with certain characteristics (e.g., the S&P 500 for a large-cap core manager). Managers are generally expected to outperform the benchmark while at the same meeting a risk budget typically defined in terms of tracking error volatility relative to the benchmark.[7]

Whether a single index can serve as an appropriate passive benchmark depends on the investment style. In certain categories, managers focus primarily on security selection, over- and underweighting securities in the index; in other categories, funds can invest in multiple asset classes, sectors, and regions.

Exhibit 1 illustrates the variation in the ability of a single index to capture the performance profile of funds in different styles. We assign to each style an index that would serve as a (passive) benchmark, based on the investment guidelines of the funds. Corporate bond A-rated funds invest at least 65% of their assets in U.S. government and corporate debt issues rated A or better, so we therefore select the Lehman U.S. Gov/Credit Index as the benchmark. For international and emerging market funds that focus on government and investment-grade securities outside the U.S. and in emerging markets, we use the Lehman Global Aggregate ex-U.S. and the Global Emerging Markets (dollar-denominated) indexes. The last category—Multisector bond funds—is assigned the Lehman Global Aggregate Index (unhedged).

We use three measures to gauge the adequacy of the selected benchmark. The first column in Exhibit 1 shows the R^2 from estimating Equation (1) separately for each fund:

$$R_{i,t} = \alpha_i + \beta_i R_{J,t}^{bm} + \varepsilon_{i,t} \quad \text{for } i \in J,\ t = 1, 2, 3, \ldots, T \quad (1)$$

where $R_{J,t}^{bm}$ denotes the return of the index assigned to style J, and $R_{i,t}$ is the contemporaneous return of fund i in that style.

The last sets of columns in Exhibit 1 show for each fund the standard deviation of returns in excess of the benchmark (the difference between a fund's return and the contemporaneous benchmark return over time and then averaged across funds, TS Std) and the cross-sectional standard deviation (calculated monthly across funds and averaged over time, CS Std). TS Std indicates how well the passive benchmark represents the active investment decisions of the median (or average) fund. CS Std measures the degree of dispersion across funds in the same style.

Notice that performance divergence between funds in a certain style and that of the benchmark does not necessarily translate into large cross-sectional standard

EXHIBIT 1

Performance of Mutual Funds Relative to a Representative Index by Category

Based on complete return histories of funds in each category. Number of funds varies by style. The R^2 value is based on estimating Equation (1) separately for each fund. TS Std. is the standard deviation of returns in excess of the benchmark computed for each fund and then averaged across funds. CS Std. is the cross-sectional standard deviation of excess returns calculated monthly across funds and averaged over time.

Fund Category	Sample Period	Benchmark	Benchmark (bp/month) Return Std.	Adj. R^2			TS Std (bp/month)			CS Std (bp/month)		
				Avg.	Med.	Std.	Avg.	Med.	Std.	Avg.	Med.	Std.
Corporate Bonds A-rated	June 04–Oct 06	U.S. Gov/Credit	94	91%	95%	17%	31	25	23	28	25	16
International Bonds	Mar 01–Feb 07	Global Aggregate Ex USA (Unhedged)	237	93%	95%	8%	58	59	30	49	40	34
Emerging Markets Bonds	Jan 01–Dec 06	Emerging Markets (USD denominated)	256	88%	91%	8%	99	90	29	71	62	41
Multisector Fixed Income Bonds	Jan 99–July 06	Global Aggregate (Unhedged)	163	28%	27%	11%	155	153	28	75	67	34
		Universal (Unhedged)	158	32%	32%	11%	148	145	27	74	67	33

Source: Lehman Brothers, Lipper.

deviations if the difference is due to common investment decisions. Hence, even if the benchmark is incorrectly specified, this alone would not materially affect the measurement of commonality among funds.

The results in Exhibit 1 suggest a clear distinction between the multisector bond category and the other styles. The average R^2 for multisector bond funds is only 28%, compared with 85% or higher for the other three styles. The two measures of standard deviation exhibit the same pattern; both are much higher for multisector bond funds than for funds in the other categories. For example, the average time-series volatility of multisector bond funds versus their benchmark (Lehman Global Index) is 155 basis points per month compared with only 58 bp per month in the case of international bond funds; their cross-sectional volatilities are similarly 75 versus 49 bp per month. The similarity between the average and median figures for all three measures indicate that these results are not driven by a few outlier funds (or months), but rather reflect the entire fund population across the sample period.

Clearly, the Global Aggregate Index fares poorly as a benchmark for multisector bond funds. This should come as no surprise, as these funds may invest in asset classes that are not reflected in the composition of the benchmark such as high-yield, emerging markets and convertible bonds, currencies, and even equities. Yet even using a broader benchmark such as the Lehman Universal Index (which incorporates the Global High-Yield and Emerging Markets indexes in addition to the Global Aggregate Index) hardly affects the results (the R^2 increases by only 4%, and the cross-sectional volatility is unchanged). This demonstrates that an index may not be an adequate benchmark not just because of incomplete asset class coverage but also because of its passive nature. Multisector bond funds can rotate across sectors dynamically to improve performance, but the index weighting scheme is typically determined by the relative (market) weights of the asset classes.

The key to managing a multisector bond fund is setting up a replicating portfolio or a composite index that one can use to understand the aggregate allocations of funds in the sector. In the remaining sections of the article we describe how to construct such a portfolio using style analysis, and demonstrate its predictive power.

DERIVING FUND EXPOSURES USING RETURN-BASED STYLE ANALYSIS

Factor models are commonly used to characterize how industry and economywide factors affect the return on individual securities and portfolios of securities. Sharpe's [1988, 1992] return-based style analysis can be considered a special case of a generic factor model. The objective is to replicate the performance of a managed portfolio over a specified period by using a passively managed collection of marketwide factors:

$$\tilde{R}_{p,t} = [\delta_{1,p} x_{1,t} + \delta_{2,p} x_{2,t} + \cdots + \delta_{n,p} x_{n,t}] + \tilde{\varepsilon}_{t,p} \qquad t = 1, 2, 3, \ldots, T \tag{2}$$

where $\tilde{R}_{p,t}$ represents the managed portfolio return at time t, and x_1, x_2, …, x_n are the returns on marketwide factors. The coefficients δ_1, δ_2, …, δ_n represent the managed portfolio average allocation among the different factors, or asset classes, during the period. The sum of the terms in the brackets is that part of the managed portfolio return that can be explained by its exposure to the different factors, which is termed the style of the manager. The residual component of the portfolio return, $\tilde{\varepsilon}_{t,p}$, is the part of the return attributable to the manager's security selection ability; it reflects the manager's decision to deviate from the composition within each factor.

An active manager is looking for ways to improve performance by investing in asset classes as well as individual securities within each asset class that the manager considers underpriced. Hence, the manager will typically have different exposure to the asset classes compared with the performance benchmark exposure. The manager will also be holding a different portfolio of securities within each asset class. As a result, the security selection component will be lower for passively managed funds than for actively managed funds.[8]

Given a set of returns for a managed fund, along with comparable returns for a selected set of factors, we can estimate the portfolio weights, δ_1, δ_2, …, δ_n in Equation (2). In order to get coefficient estimates that closely reflect the fund's actual investment policy, it is also important to incorporate restrictions on the factor weights. The following two restrictions are often imposed:

$$\delta_{j,p} \geq 0 \quad \forall\, j \in \{1, 2, \ldots, \mathrm{n}\} \tag{3}$$

$$\delta_{1,p} + \delta_{2,p} + \cdots + \delta_{n,p} = 1 \tag{4}$$

The first restriction represents the constraint prohibiting short positions in securities. The second requirement that the asset class exposures sum to one prohibits the use of leverage and allows interpreting exposures as portfolio weights. For funds known to use leverage or short-selling, such as hedge funds, other constraints may be imposed.

The goal of return-based style analysis is to find the set of non-negative style asset class exposures δ_1, δ_2, …, δ_n that sum to one and minimize the unexplained variation in returns (i.e., the variance of $\tilde{\varepsilon}_{t,p}$), referred to as the fund's tracking error over the style benchmark. The objective is not to choose style benchmarks that make the fund appear good or bad; rather, it is to be able to infer as much as possible about a fund's exposures to variations in the returns of the given style benchmarks over a particular period.

Ben Dor, Jagannathan, and Meier [2003] provide a detailed exposition of the subject and discuss the potential pitfalls in implementing the approach.

Effective Style of a Multi-Manager Composite Index

An attractive property of return-based style analysis is the straightforward extension from a single manager to an index of multiple managers. Consider the group of funds representing a certain investment style. Denote by w_p the (relative) size of assets managed by manager p. In total there are $p = 1, 2, \ldots, P$ mutual funds, and R_t is the funds' universe aggregate return.

The aggregation involves two steps. First, identify the style of each fund manager separately. Second, sum up the exposures of all managers for each asset class, weighted by the proportion of overall assets managed by each manager.

Formally, the effective mix of an index comprised of these funds is described by:

$$R_t = \sum_{p=1}^{P} w_p r_{p,t} = \left[\sum_p w_p b_{p,1}\right] x_{1,t} + \left[\sum_p w_p b_{p,2}\right] x_{2,t} + \cdots + \left[\sum_p w_p b_{p,N}\right] x_{N,t} + \left[\sum_p w_p e_{p,t}\right] \tag{5}$$

The brackets are the asset-weighted exposures to the asset classes. Consider the first asset class with returns $x_{1,t}$. Each fund p has an exposure of $b_{p,1}$ to that asset class. Fund p has a weight w_p in the overall index. The summation over all funds measures the exposure of the composite peer index to the first asset class.

Data and Model Formulation

To demonstrate how return-based style analysis can be applied in practice to construct a peer manager benchmark, we use 38 multisector bond funds (based on Morningstar classifications). The monthly return data at our disposal span almost eight years, from January 1999 to July 2006.[9]

In addition to U.S. investment-grade debt securities, multisector bond funds can hold high-yield, emerging market and international bonds. The funds can have exposure to various currencies and even limited equity positions. Selecting the appropriate set of asset classes is important in order to obtain an informative style description and provide the ability to map a broad spectrum of portfolios.

We follow standard practice and use one-month U.S. dollar LIBOR as the cash equivalent. The U.S. fixed-income universe is divided into treasuries, credit (investment-grade and separately high-yield), and mortgage-backed securities (MBS). All four asset classes are represented by their respective Lehman Brothers indexes. Investments in fixed-income securities outside the U.S. are captured by the Lehman Brothers Euro Aggregate, Euro High Yield, Japanese Aggregate, and Global Emerging Markets (dollar-denominated) Indexes. To reflect the funds' ability to take currency and equity positions, we include the returns on the S&P 500 Index as well as the monthly percentage changes in the spot rate of the dollar against the euro and the yen.

Employing an extensive set of assets automatically increases the explanatory power of the style benchmark, measured by R^2, but may also introduce multicollinearity. If some of the asset classes are highly correlated, the optimization algorithm will have difficulty in identifying asset class weights. The estimated weights of two highly correlated asset classes may vary over time and make exposures uninformative and difficult to interpret. For example, the correlations between the return of the Treasury index and the returns of the credit and mortgage-backed securities indexes are about 0.9, on average, due to the common yield curve exposure of the assets.

One solution to the high correlation would be to use excess returns over a duration-matched portfolio of Treasuries for the credit and MBS benchmarks. In this case, however, the Treasury index coefficient δ_T would represent the aggregate yield curve sensitivity of the three assets.[10]

To understand why, notice that the total returns on the credit and MBS indexes may be written as the sum of excess returns and the Treasury index returns scaled by the ratio of durations:

$$R_i^{TR} = R_i^{ER} + R_i^{Treas} \cong R_i^{ER} + \left({}^{D_i}\!/_{D_T}\right) R_T \quad i = Credit,\ MBS \tag{6}$$

and therefore δ_T (based on total returns) becomes δ'_T where $\delta'_T = \delta_T + \delta_C\left(\frac{D_C}{D_T}\right) + \delta_{MBS}\left(\frac{D_{MBS}}{D_T}\right)$. As the credit and MBS coefficients are unchanged ($\delta'_C = \delta_C$, $\delta'_{MBS} = \delta_{MBS}$) and the durations are known, we can back out the actual allocation to Treasuries, δ_T, from the estimated δ'_T.

Exhibit 2 displays the correlation matrix for the various assets once the yield curve exposure has been removed. The (absolute) correlations between Treasuries and credit and MBS have declined to 0.33 and 0.24, respectively. We still observe, however, high correlations between U.S. and European Treasury bonds (the European Aggregate Index returns are dominated by changes in the euro yield curve) and high-yield securities.

Before proceeding with estimation of the funds' actual style, we need to reconcile the interpretations of the currency exposures (the euro and the Japanese yen) with that of the coefficients from Equation (2). While δ_1, δ_2, ..., δ_n in Equation (2) represent portfolio allocations to asset classes, the exposures to the euro and the yen reflect the portfolio sensitivity (or beta) to a unit change in the spot rate of these currencies versus the U.S. dollar.

In discussion with several fund managers, we learned that currency views are typically not implemented by holding outright currency positions, but rather indirectly by partially hedging exposures resulting from non-USD-denominated securities. By definition, the return to any asset class of foreign securities is the total return in the local currency plus the returns generated by converting the investment back to the reference currency (e.g., the U.S. dollar). Therefore, the total return from investing in an asset class j denominated in currency i and hedging only

EXHIBIT 2

Asset Class Return Correlations

Computed over entire sample period January 1999–July 2006. Returns hedged into U.S. dollars. U.S. Credit and MBS returns are in excess of a duration-matched portfolio of Treasuries.

	Cash	U.S. Treasury	U.S. Credit (ER)	U.S. MBS (ER)	U.S. HY	EURO Agg	EURO HY	JAPAN Agg	USD EM	SP500	EUR	JPY
Cash	1.00											
U.S. Treasury	0.04	1.00										
U.S. Credit (ER)	-0.12	-0.33	1.00									
U.S. MBS (ER)	-0.04	0.24	0.27	1.00								
U.S. HY	-0.17	-0.05	0.79	0.25	1.00							
EURO Agg	-0.01	0.81	-0.17	0.19	0.03	1.00						
EURO HY	-0.16	-0.15	0.61	0.13	0.82	0.00	1.00					
JAPAN Agg	0.33	0.11	0.00	0.06	-0.04	0.25	-0.05	1.00				
USD EM	0.00	0.21	0.46	0.39	0.54	0.17	0.40	0.09	1.00			
SP500	-0.05	-0.32	0.51	0.10	0.46	-0.26	0.53	-0.15	0.46	1.00		
EUR	-0.24	0.27	0.03	0.02	0.09	0.18	0.03	0.04	0.06	-0.04	1.00	
JPY	-0.12	0.23	-0.04	0.11	0.05	0.11	0.02	-0.08	0.20	0.14	0.41	1.00

Source: Lehman Brothers.

a fraction h_i of the currency exposure can be written approximately as:

$$R_{j,usd}^{TR} \cong R_{j,local} + (1-h_i)R_{i,FX} \quad \text{for } 0 \le h_i \le 1 \quad i = Euro,\ JPY \tag{7}$$

where $R_{j,local}$ is the return on asset j in the local currency, and $R_{i,FX}$ is the percentage change in the spot rate between the local (foreign) currency and the U.S. dollar. A bearish view on the currency versus the dollar implies $h_i = 1$ (i.e., fully hedging the currency exposure), while $h_i = 0$ reflects the most bullish view possible.

After we incorporate Equations (6) and (7) in the basic formulation in Equation (2), we estimate the model below (time subscripts are dropped for clarity):

$$\begin{aligned}\tilde{R}_p = {} & \delta_{Cash}R_{Cash} + \delta_T' R_T + \delta_C R_C^{ER} + \delta_{MBS}R_{MBS}^{ER} + \delta_{HY}R_{HY} \\ & + \delta_{EuroAgg}R_{EuroAgg}^{Local} + \delta_{EuroHY}R_{EuroHY}^{Local} + \delta_{JapanAgg}R_{JapanAgg}^{Local} \\ & + \delta_{EM}R_{EM} + \delta_{EUR}R_{EUR} + \delta_{JPY}R_{JPY} + \delta_{SP500}R_{SP500}\end{aligned} \tag{8}$$

subject to:

Treasury allocation: $$\delta_T = \delta_T' - \delta_C\left(\frac{\bar{D}_C}{\bar{D}_T}\right) - \delta_{MBS}\left(\frac{\bar{D}_{MBS}}{\bar{D}_T}\right) \ge 0$$

Currency hedging: $$0 \le \delta_{EUR} \le \delta_{EuroAgg} + \delta_{EuroHY}$$
$$0 \le \delta_{JPY} \le \delta_{JapaneseAgg}$$

Equity allocation: $0 \le \delta_{SP500} \le 0.1$

Portfolio weights: $$\delta_{Cash} + \delta_T + \delta_C + \delta_{MBS} + \delta_{HY} + \delta_{EuroAgg} + \delta_{EuroHY} + \delta_{JapanAgg} + \delta_{EM} + \delta_{SP500} = 1$$

The first constraint reflects the requirement that the direct allocation to U.S. Treasuries is non-negative (allocations to all other asset classes are non-negative as well).[11] Notice that the average durations of the Treasury, Credit and MBS indexes during the sample period (denoted by bars over the variable) are used to back out the treasury holding. The bounds on the currency coefficients guarantee that the hedged proportion of the investment in euro- and yen-denominated securities lies between zero and one.[12]

The allocation constraint to the S&P 500 prevents a large and unrealistic weighting to equities that may be caused by spurious correlations. Finally, the last restriction reflects our goal to approximate mutual funds' returns as closely as possible by the return on a portfolio of market factors without the use of leverage.

ESTIMATION RESULTS

Exhibit 3 presents the style analysis results for the model in Equation (8) using all available data. The factor weights are estimated for each fund separately and for a composite peer index (based on equally weighted fund returns). This allows an understanding of the aggregate allocation across asset classes in this category as well as the dispersion in holdings and skill across funds.

The adjusted R^2, time series, and cross-sectional standard deviation figures are strikingly different from those reported for multisector bond funds in Exhibit 1 (using the Universal Index as the benchmark). The average R^2 increases from 32% to 84%. Similarly, the average magnitude of deviation over time between actual returns and the style benchmark returns declines from 148 to 52 basis points per month. Using a fund-specific style benchmark over a single uniform benchmark (i.e., index) also reduces the cross-sectional standard deviation of the difference between the funds' and style benchmark's returns from 74 to 49 bp per month.

The extent of variation in the three statistics is reduced as well (as reflected by the lower standard deviation of the three measures), which provides further support for using style analysis to construct the appropriate peer benchmark. Finally, the explanatory power of the style benchmark for the composite peer index (96%) is significantly higher than the average R^2 for individual funds (84%). This should come as no surprise, because averaging returns across funds likely eliminates most of the

EXHIBIT 3

Summary of Style Analysis Results

Based on entire available history for each of the 38 funds in the sample January 1999–July 2006. The R^2 value is based on estimating Equation (8). Asset allocation figures indicate asset holdings except for Euro and JPY where they reflect the proportion of the currency exposure that is being hedged.

		Individual Funds			Composite Fund Universe
		Mean	Median	Std.	(Equal Weight)
Adj. R^2		84%	86%	10%	96%
TS Std. (bp/month)		52	50	19	26
CS Std. (bp/month)		49	48	14	NA
			Asset Allocation		
Alpha (bp/month)		2	1	9	2
Cash		6%	0%	14%	0%
U.S.	Treasury	8%	0%	11%	11%
	Credit	16%	9%	20%	11%
	MBS	12%	6%	14%	15%
	HY	21%	20%	12%	25%
EURO	Aggregate	12%	12%	8%	16%
	HY	4%	4%	0%	4%
JAPAN	Aggregate	1%	0%	2%	0%
EM		15%	14%	12%	15%
SP500		4%	2%	3%	3%
Total Allocation		100%			100%
FX	EUR	41%	51%	35%	48%
	JPY	13%	0%	33%	-----

Source: Lehman Brothers.

idiosyncratic bets they take while maintaining their systematic exposures.

Exhibit 3 also reports estimates of the funds' asset allocation and the proportion of currency exposure they choose to hedge. Several results are clear. First, the average allocation across funds to each asset class is very similar to the results for the composite peer index. This follows from the linear formulation of the estimated model that allows for aggregation, and implies that a single estimation for the composite peer index is equivalent to averaging allocations across funds [see Equation (5)].[13]

Second, as the term *multisector* suggests, funds' capital is deployed across the entire spectrum of assets. Roughly 40% is invested in below-investment-grade securities (high-yield and emerging markets bonds), and an additional 3% is allocated to equities compared to only 26% invested in U.S. credit and mortgage-backed securities. Given their flexible investment mandate, it is not surprising that multisector funds did not find the low-yielding Japanese market attractive and avoided it altogether. Yet the funds allocated almost one-fifth of their assets to the euro market, hedging only about half the exposure to changes in the spot rate between the dollar and the euro.

Third, there is wide dispersion in allocation across funds to the same asset class (as reflected in the standard deviation figure). For example, many funds do not hold U.S. investment-grade bonds, while others have a high allocation to them. This again illustrates the need for customized style benchmarks.

Fourth, net of fees, the funds exhibited some security selection skill on average (2 bp per month), although as implied by the high standard deviation of alpha some of them did outperform their style indexes substantially (11 of the 38 funds had an alpha higher than 5 bp per month).

The results in Exhibit 3 reflect the average allocations of the universe of multisector bond fund managers over the period January 1999–July 2006, but not how allocations varied over time. Exhibit 4 displays the changes in the style benchmark for the composite peer index in terms of a rolling 36-month window. For a comparison of a fund's current asset allocation to the historical allocation of the composite peer index to be informative, the aggregate asset allocation should exhibit stability.

Exhibit 4 reveals that although allocations to some of the assets varied substantially during the period, the changes were fairly gradual, allowing the composite style benchmark weights to change in a smooth fashion. The weight of U.S. Treasuries, for example, increased from about 15% at the end of 2002 to about 30% until 2004, and then gradually dropped back to 12% toward the end of the period. The increase in Treasuries was accompanied by a corresponding decline in the weight of U.S. credit, likely reflecting the credit market turmoil in 2002. In contrast, the allocation to U.S. high-yield was little changed during the period, ranging between 25% and 30%.

While the results in Exhibit 4 are encouraging, they are based on in-sample computations. Essentially, the optimizer is looking for the set of allocations that best explains the observed pattern of performance. Since the results rely on averaging historical data, their predictive power for future allocations and consequently performance is unclear.

Exhibit 5 plots the tracking errors between the actual returns of the composite fund universe (equal-weighted) and the style analysis-based model forecasts out-of-sample. Tracking errors are computed starting in January 2002 based on a 36-month trailing historical period.[14] In addition, the difference between the composite fund universe return and the contemporaneous return of the Lehman Brothers Universal Index (unhedged) is also shown.

The results clearly indicate how inappropriate the Universal Index is as a benchmark for multisector bond funds. Often, the tracking errors are in excess of 100 basis points per month. In contrast, the tracking errors based on the style analysis model are quite low and distributed fairly symmetrically around zero.

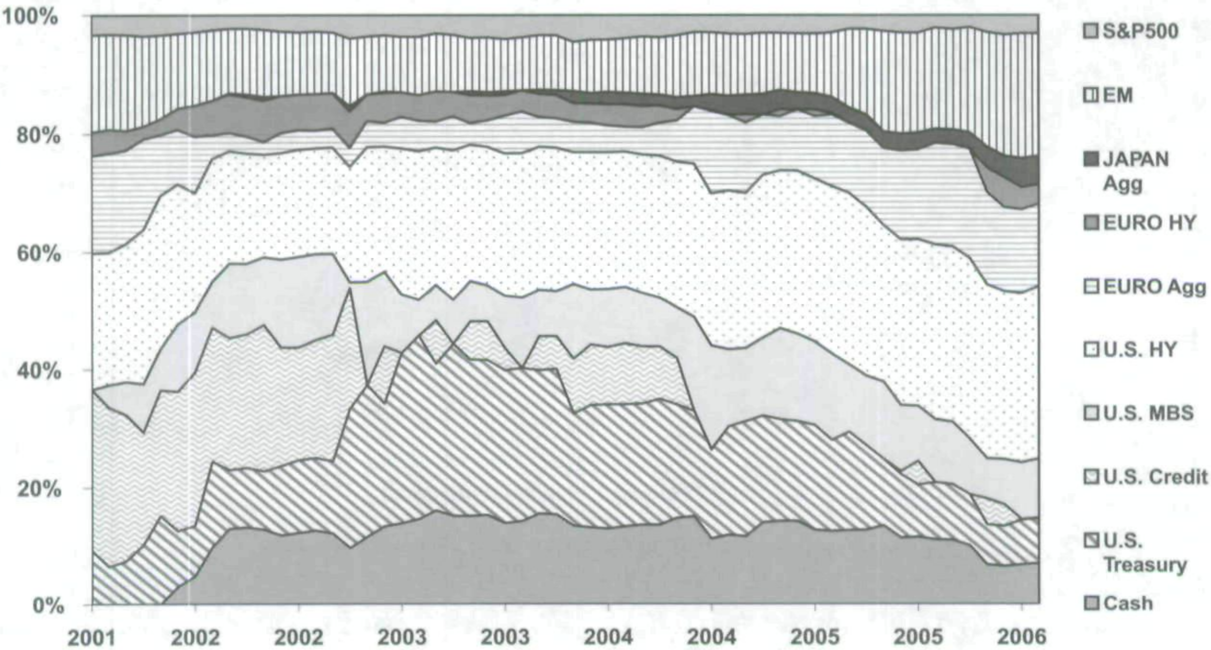
One further issue to address before we assess the risk of a fund versus its peers based on the style analysis model allocations is the length of the estimation window. The optimal estimation window is affected by how often (and how much) funds reallocate their capital among various asset classes. If the aggregate asset allocation changes fairly slowly, a longer estimation window would result in more accurate estimates of the asset weights as compared to using more recent data. This in turn would lead to better out-of-sample predictions.

Exhibit 6 reports out-of-sample means and standard deviations of tracking error between projected and actual returns for several estimation windows over January 2003–July 2006. Tracking error volatility measures the dispersion of tracking errors around the mean error, which can be quite different from zero out-of-sample. Therefore, Exhibit 6 also reports the root mean squared error

EXHIBIT 4

Style Benchmark for Composite Fund Universe (equal-weighted)

Asset class allocations computed using a rolling 36-month window for January 1999–July 2006.



Source: Lehman Brothers.

(RMSE), which accounts for both the mean and the dispersion of the tracking errors.

Although the results are fairly similar for the three estimation windows, they suggest that using a 36-month period for computing the style benchmark generates the lowest TEV of 20.9 bp per month (and lowest RMSE). To put this figure into perspective, the absolute return volatility of the peer index over the period was 124 bp per month. Note also that in the aggregate multisector bond funds outperformed the style benchmark by 2 to 4 bp per month. This is due to security selection skills and market timing skills not captured by the trailing window analysis.

Is a TEV of 20.9 bp per month an artifact of the specific sample period, or can we reasonably expect it to persist in the future?

The TEV has two main sources: the difference between the actual and estimated asset weights, and the security selection volatility. The first stems from the fact that our estimates reflect averages across time and may depart from the exact current weights, while the second reflects security selection, the difference between the asset class composition and the actual funds holdings in each asset class.

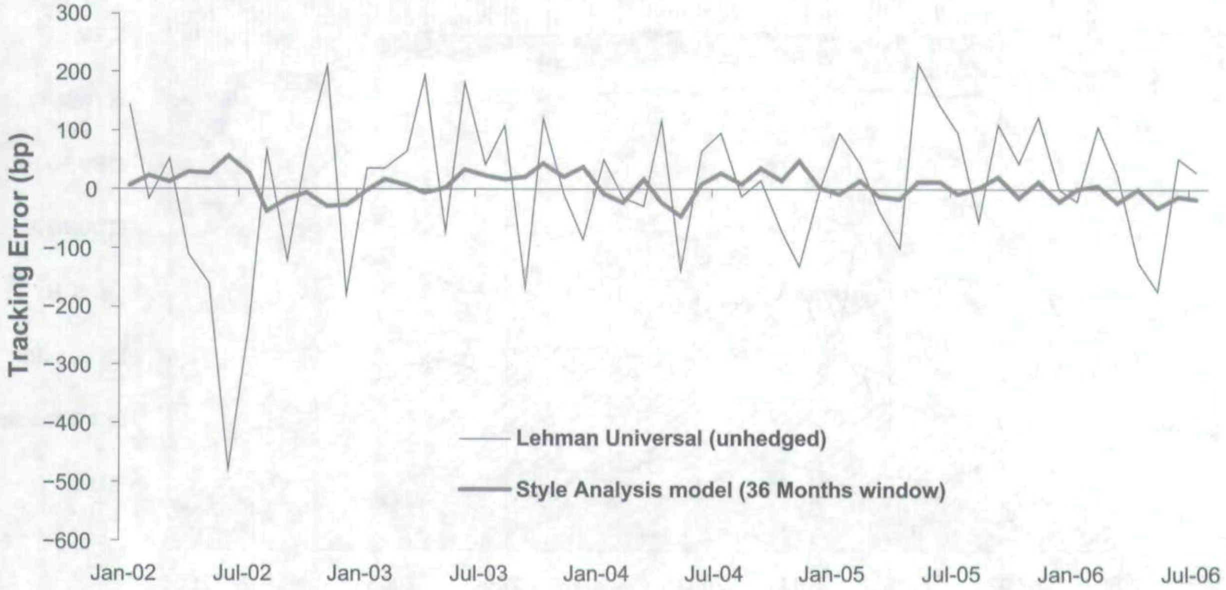
In 2005 Lehman Brothers launched RBI® (replicating bond indexes)—baskets of liquid derivatives constructed to closely replicate the performance of various U.S. and Global Lehman Brothers indexes.[15] Since the asset allocation and characteristics of the index being replicated, such as key rate durations and spread durations, etc. are known, they are matched exactly using the appropriate mix of interest and credit default swaps, mortgage TBAs, and interest rate futures. As a result, the component of index returns due to systematic broad market movements is almost perfectly replicated, and the TEV arises mainly from mismatches between the composition of asset classes and the composition of the instruments representing them.[16]

For the Lehman U.S. Aggregate (which experienced roughly the same total return volatility as the composite funds index since 1999) the tracking error volatility of the RBI is about 9 bp per month since January 2003 (with an average return differential of zero). If we take this figure as a measure of the security selection volatility we can

EXHIBIT 5

Comparison of Universal Index and Style Analysis Model Predictive Power

Tracking errors computed monthly starting in January 2002 as difference between equal-weighted composite multisector fund universe and the style benchmark. Style benchmark returns are product of the current month asset returns and asset weights computed over previous 36 months (style analysis) or contemporaneous return (Lehman Brothers Universal Index).



Source: Lehman Brothers.

anticipate in our case as well, it implies that the component of TEV due to allocation mismatches between the composite funds index and model portfolio is 18.8 bp per month. This last figure reflects the product of volatilities of the mismatches in asset weights as well as their return realizations.

Another approach to gauge whether the TEV we find is representative beyond the sample period analyzed is to think of the forecast and realized returns as two random variables with certain volatilities and correlation between them. Given this information, we can compute the expected volatility of the difference between the two (which corresponds to the TEV).[17]

In our case, the in-sample correlation between the model fitted and actual returns during January 2002–June 2006 was 98.6%. The volatilities were 118 and 124 bp per month for the forecast and the realized returns, respectively. If we use these numbers, we should expect that the tracking error volatility would be 18.1 bp per month, very similar to the TEV we observed in practice.

EXHIBIT 6

Out-of-Sample Performance Statistics by Estimation Window

Estimation Window (months)	24	36	48
Mean Outperformance	2.2	4.2	3.5
TEV	22.1	20.9	23.6
RMSE	22.2	21.3	23.9

Source: Lehman Brothers.

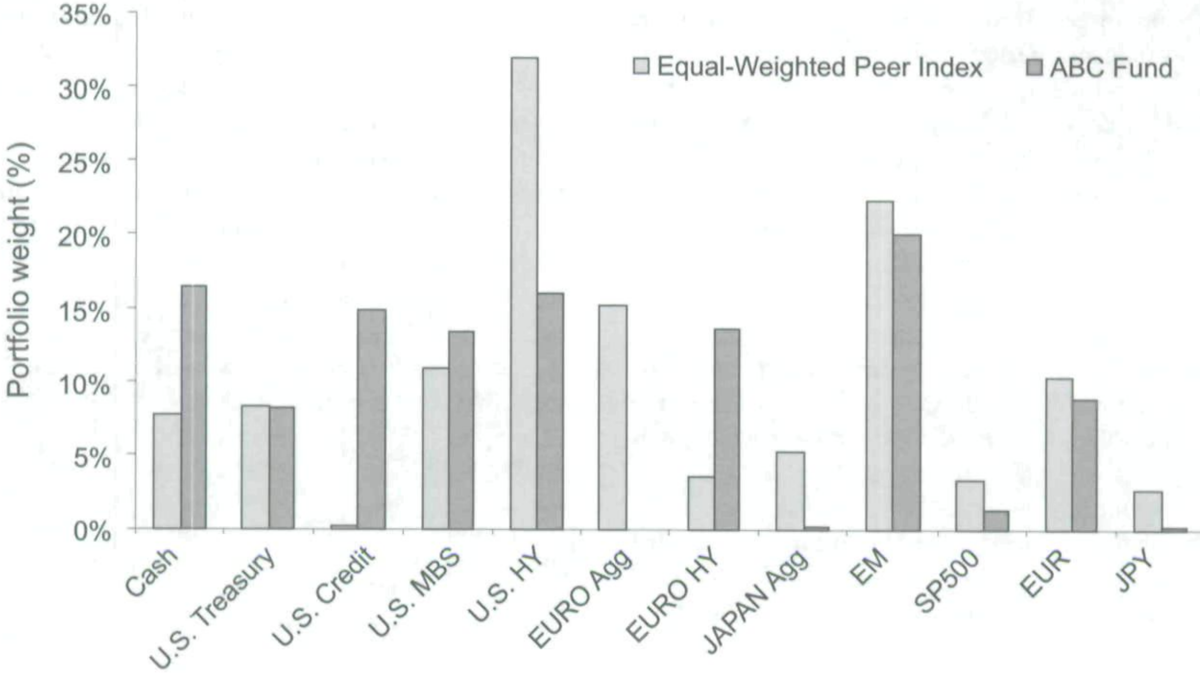
RISK MANAGEMENT USING A COMPOSITE INDEX

Having established the predictive power of the results derived from style analysis, we can now use the model index allocations in practice for assessing the risk implicit in a manager's bets versus the universe of funds in that manager category.

Exhibit 7 shows the allocations considered by the manager of ABC multisector bond fund as of July 2006 in comparison to the manager's peers. The peer allocations are based on the model style benchmark for the equal-weighted universe of multisector bond funds estimated

EXHIBIT 7
Asset Allocation of ABC Fund and Equal-Weighted Peer Index

As of July 2006. Euro and JPY weights reflect the product of the allocation to assets denominated in these currencies and the unhedged proportion of the currency exposure.



Source: Lehman Brothers.

over the previous 36 months. Notice that now the euro and yen weights are not the fraction of currency exposures hedged but the portfolio sensitivity (i.e., the product of the allocation to assets denominated in these currencies and the unhedged proportion of the currency exposure).

While ABC fund manager considers similar positions compared to his peers in terms of U.S. Treasuries, MBS, and euros (currency), he holds very different views concerning the credit markets. With respect to high-yield, the manager equally likes European and U.S. issues with roughly 13%–16% weight in the portfolio in sharp contrast to peers, who allocate a very high stake (31%) to U.S. high-yield. At the same time, the manager contemplates holding a large position in U.S. credit (14%), which is practically absent from the peers' composite portfolio.

Consider also emerging markets to which ABC's manager plans to allocate 20% of the portfolio. With no knowledge of peers' composite allocations, the manager may think this is an expression of a bullish view on EM. Yet when compared to other funds in the sector, that stake actually reflects a relatively bearish view.

If we express the difference between the weight of asset j in the portfolios of ABC fund and the composite as $\Delta_j = \delta_{p,j} - \delta_{ABC,j}$, the expected tracking error volatility of ABC (versus its peers) is given by (where Ω is the variance-covariance matrix of the asset returns):

$$\text{Expected TEV vs. peers} = \Delta'\Omega\Delta \tag{9}$$

Providing the manager with an analytic expression for the relative risk (e.g., TEV) the fund is exposed to (by having views that differ from those of the peer universe) allows evaluation of whether the expected outperformance $\Delta'\alpha$ is sufficient to compensate for the level of risk (for example by looking at the resulting information ratio).

CONCLUSION

Mutual fund managers are typically evaluated against a particular benchmark, often a broad-based index. The actual investment style employed by the manager, however, may be different from the style reflected in the composition of the index. The manager operates subject to a set of investment guidelines that may allow the portfolio's asset allocation to deviate substantially from the index allocation and even invest in assets that are not represented in the index. The stated index may therefore no longer capture the risk-return profile of the fund, and manager performance may often differ considerably from index performance.

Fund managers thus find themselves in a tough spot. They are judged ex post primarily by their ability to outperform their peers rather than the index, but at the same time have no information ex ante on the investment decisions made by their competitors. We try to address the issue by using return-based style analysis to extract the peer universe composite asset allocation. Understanding the aggregate holdings of peers (and the dispersion among them) allows the manager to assess the trade-off between the risk the fund is taking versus peers and the benefit (i.e., outperformance) the manager expects to collect.

We believe not only that this approach is appropriate conceptually, but also that mutual funds can implement it in practice. For example, after a model has been formulated, it is estimated regularly every period (e.g., month, quarter). The model portfolio for the peer universe can then be represented in the fund's internal risk systems according to the weights estimated from style analysis using a collection of asset class indexes. The fund manager's suggested portfolio can then be analyzed against the model portfolio to evaluate its (relative) risk ex ante and conduct a performance attribution ex post.

While style analysis can serve as an effective tool for constructing a customized and dynamic peer benchmark for the universe of peer managers, it also has limitations. First, the technique is sensitive to the asset class specification, which should span the investment universe of the funds in question. Inappropriate or inadequate choice of asset classes (or factors) may lead to wrong inferences. For example, when funds' performance is highly sensitive to reshaping of the yield curve, the model should include not only factor returns due to the level (e.g., shift), but also factors reflecting twist and butterfly changes of the curve.

Second, a parsimonious model would have less explanatory power in-sample but require a shorter return history to estimate and allow for more dynamic updating. An extensive set of factors are more likely to include highly correlated asset classes, and the optimization procedure may have difficulty attributing exposures to asset classes. Even though the in-sample tracking improves, interpretation of the results may become more difficult and the parameters less robust.

Third, formulating a set of constraints that reflects the investment practices of the funds is important, as it reduces the likelihood of estimation errors reflecting spurious correlations among the asset classes. Fourth, what is the right estimation window for the model? Clearly there is a trade-off between statistical accuracy using a longer return history and the speed of adjustment.

Even after a model has been formulated, tested, and put to use, one needs to monitor its predictions versus the actual performance of the underlying funds and make adjustments if necessary. Furthermore, actual holdings data can also be combined with the model outputs to improve the peer benchmark construction process.

ENDNOTES

[1]An estimated 52% of U.S. households invest in mutual funds which as a group control assets in excess of $10.5 trillion as of January 2007 (Source: Investment Company Institute's official survey of the mutual fund industry, February 2007).

[2]Morningstar reports returns on four broad categories (domestic stock funds, international stock funds, fixed-income funds, and municipal bond funds) which are further divided into 48 sub-categories. The Investment Company Institute enumerates 33 investment objective categories.

[3]Lipper and Morningstar offer tools that allow analysis of the average sector composition for peer groups.

[4]The authors would like to thank Mary Miller of T. Rowe Price for a thorough discussion of this aspect of the study.

[5]For example, Morey and O'Neal [2006] find that bond funds on average tilt their portfolios toward higher-quality credit and government instruments near holdings disclosure dates. Musto [1999] shows that funds allocating between government and private issues hold more in government issues around disclosures than at other times.

[6]Return-based style analysis has also been adapted for analysis of hedge fund performance. See Ben Dor, Dynkin and Gould [2006] and Ben Dor, Jagannathan, and Meier [2003] for more details.

[7]Tracking error volatility (TEV) is the standard deviation of the difference in the returns earned by the fund and the corresponding benchmark.

[8]A relatively straightforward way to evaluate the level of active management is to examine the R^2, defined as $R^2 = 1 - Var(\tilde{\varepsilon}_p)/Var(\tilde{R}_p)$. The resulting R^2 value thus indicates the proportion of the portfolio return variance "explained" by the n asset classes. Notice that since the vector of residuals is

not necessarily orthogonal to the matrix of benchmark returns, as is the case in multivariate regression, the alternative definition $R^2 = Var(\delta_{1,p}x_{1,t} + \cdots + \delta_{n,p}x_{n,t})/Var(R_p)$ is not in general equivalent.

[9]Many mutual funds offer investors different share classes with different fees and expenses and thus different performance results even though they represent the same underlying investment pool. Whenever possible, we use the returns to class A shares in order to get comparable results across funds.

[10]Excess returns are calculated as follows. The yield curve exposure of any single security is described by a set of six key rate durations (KRD). Then, a hypothetical Treasury portfolio is created that matches this KRD profile. At the end of the period, the excess return for the security is calculated as the difference between its total return and the return on the Treasury portfolio.

[11]While mutual funds can have short positions in certain asset classes by using derivatives (e.g., buying protection via CDX), it is unlikely we will observe a negative weight in an asset over a prolonged time period, particularly not for the fund universe composite allocations.

[12]Because the hedging proportion should be the same for all securities denominated in the same currency, we get:

$$\delta_{EUR} = (1 - h_{Euro})(\delta_{EuroAgg} + \delta_{EuroHY})$$

[13]Differences between the average allocation of assets across funds and allocations estimated for the composite reflect estimation noise as well as the fact that some funds have partial return histories.

[14]Tracking errors are computed as follows: Each month we estimate the asset weights for the aggregate fund universe. The product of the asset weights and their subsequent month returns is subtracted from the actual performance in that month.

[15]For a detailed description of the RBI products see "Replicating Bond Index (RBI℠) Baskets—A Synthetic Beta for Fixed Income," Lehman Brothers, August 23, 2006.

[16]For example, the credit sector with over 600 individual issuers is represented by the CDX.IG, which includes only 125 names.

[17]If x, y are random variables, then $\sigma^2(x - y) = \sigma^2(x) + \sigma^2(y) - 2\rho(x,y)\ \sigma(x)\ \sigma(y)$.

REFERENCES

Ben Dor, A., L. Dynkin, and A. Gould. "Style Analysis and Classification of Hedge Funds." *The Journal of Alternative Investments*, Vol. 9, No. 2 (2006), pp. 10–29.

Ben Dor, A., R. Jagannathan, and I. Meier. "Understanding Mutual Fund and Hedge Fund Styles Using Return—Based Style Analysis." *Journal of Investment Management*, 1 (2003), pp. 94–134.

Morey, M.R., and E.S. O'Neal. "Window Dressing in Bond Mutual Funds." *Journal of Financial Research*, Vol. 24, No. 3 (2006), pp. 325–347.

Musto, D.K. "Investment Decisions Depend on Portfolio Disclosures." *Journal of Finance*, Vol. 54, No. 3 (1999), pp. 935–952.

"Replicating Bond Index (RBI℠) Baskets—A Synthetic Beta for Fixed Income." Lehman Brothers, (August 23, 2006.)

Sharpe, W.F. "Determining a Fund's Effective Asset Mix." *Investment Management Review*, 1988, pp. 59–69.

——. "Asset Allocation: Management Style and Performance Measurement." *The Journal of Portfolio Management*, Vol. 18, No. 2 (1992), pp. 7–9.