NORTHFIELD | modeling financial markets worldwide

October 2005

A Unified Approach to Monitoring and Evaluating Investment Managers

Northfield Research

Dan diBartolomeo and Sandy Warrick

Introduction

Using specific techniques from our ART software (as have been described in previous editions of our newsletter), we have formulated a new process for monitoring and evaluating investment managers. Extensive empirical studies on this technique suggest that it is effective in predicting one-year end relative manager performance to a degree which is both statistically and economically significant.

Investors are constantly looking to invest with superior active managers, but have a hard time finding the managers that will be superior in the future. Typically, active managers are evaluated by looking at simple performance measures over fixed past time periods. In our process, we use a combination of returns based style analysis, CUSUM analysis and a Bayesian framework for past excess returns.

To hire active managers we must believe at least one of three things;

(a) The average professional investment manager outperforms passive index funds because individual investors have below index performance.

(b) Manager active returns persist. We can predict with reasonable reliability which managers are going to outperform in the future, even if the average manager is just average.

(c) We are doing a societal good because if all investors were passive, there would be no functional mechanism to ration capital in the economy. Our economy would break down over time

We believe there are some serious analytical problems in the typical evaluation process today. The first problem is that much manager evaluation occurs relative to benchmarks that are often not suitable for the manager's investment approach. The second problem is that evaluation of past performance is based on standardized periods (i.e. 5 years) rather than periods that are relevant to the manager in question. Third, many commonly used evaluation measures such as Sharpe ratio or information ratio correspond to meaningful investor utility for only a small fraction of investors. Finally, the statistical significance of ex-post performance is measured in a simple time

series fashion. The analysis does not include the context of whether the manager exists among a tightly bunched set of peers or a widely dispersed set, which is a critical issue in examining the "luck versus skill" issue.

It is our view that there are three key elements for a successful manager evaluation process. The first is to ensure that each fund is being measured against the right benchmark and the right group of peer funds. In this regard, we use an augmented method of returns-based style analysis. The second element is to evaluate each manager over the length of history that is optimal for that particular manager. Our CUSUM analysis is used to accomplish this. Finally, we evaluate each fund using a direct measure of value-added to investors. This measure is constructed in a Bayesian framework that adjusts for the dispersion of contemporaneous return dispersion across managers within a fund's peer group.

There is a large amount of academic literature as to whether persistence is demonstrated in active manager returns. If markets are very efficient, there should be no persistence patterns in active management returns. While there are there are innumerable studies showing markets are relatively efficient, many fund studies show that some persistence does exist. Most of these studies such as Hendricks, Patel and Zeckahuaser (1993), Elton, Gruber, and Blake (1996), Goetzmann and Ibbotson (1994) suggest that persistence exists over relatively short time frames, generally strongest around a one-year horizon. Carhart (1997), Stewart (1998), Brown and Goetzmann (1995) and Detzel and Weigand (1998) all have varying degrees of short-term persistence but find that most of the persistence can be explained by style effects and fund cash flows rather than investment management skill.

One possible explanation for the observed persistence in active manager returns is that funds are being misclassified and hence, measured against the wrong benchmarks. As it was put in diBartolomeo and Witkowski (1997), "The best way to win a contest for the largest tomato is to paint a cantaloupe red and hope the judges don't notice". For example, imagine you operate two funds, one that is aggressive and other conservative; to persistently outperform your peer groups you actually have to be more skillful than competitors (which is not easy). You might choose the easy way out. You could mischaracterize both funds: market the aggressive fund as conservative, market the conservative one as aggressive. Depending on market conditions, *one of the two will always compare well* to the intentionally wrong peer group.

There have been several studies on the issue of fund misclassification. The first was diBartolomeo and Witkowski (1997). It studied 748 mutual funds from 1990 through 1995, about 40% of which appeared to be misclassified in terms of objective. This study also finds that misclassifications did not occur at random and appear to be intentional to a meaningful degree. The resulting confusion diminished investors' ability to diversify fund types, with an annual associated cost in the billions of dollars. Very similar results on different data sets using different methods were obtained in Brown and Goetzmann (1997) and Kim, Shukla and Thomas (2000). Comparable problems with institutional money managers are less severe but still present to a material degree.

The most rigorous approach to classification would come through the formation of "normal portfolios" by comparing actual portfolio holdings to benchmarks across time as the best method, described in Kritzman (1987). Unfortunately, the required data is not available to large universes of managers, and even if the data is available the process is too labor intensive to be applied to large fund universes.

A Unified Approach to Monitoring and Evaluating Investment Managers

As an alternative, we use an augmented form of returns-based style analysis to match funds to a broad range of benchmarks as proposed in Sharpe (1992). This procedure forms a portfolio of market indices that mimics fund behavior over time. Inferences about fund composition can then be drawn from the index representation. Our key improvement to Sharpe's method is to utilize a calculation of confidence intervals on the style weights as derived in diBartolomeo and Lobosco (1997). We want to avoid drawing meaningless conclusions from information like "a fund is 10% small cap value" if it turns out to be "10% plus or minus 30%."
Our next step is to establish time samples over which different managers may be evaluated. Practitioner tradition in the investment industry is to evaluate active manager track records over a long period of at least 3 to 5 years. Some will argue a full "market cycle" is needed (although there is no apparent agreement over what constitutes such as a cycle. As we've seen previously, all the academic studies refute this. Those studies find no evidence that long-term past performance is predictive of future performance. The academic research argues strongly that if there is any meaning to past performance at all, it's short-lived, probably only the last year.

The key question we must ask ourselves is "What time portion of a track record do we really need to evaluate as part of our monitoring of manager quality control?" We need a procedure to draw the line between getting enough meaningful data within a manager's record and older, stale data that should be ignored. Our choice for this analytical task is a statistical process control technique called CUSUM. This technique was described in detail in our February, 2005 newsletter. With it, we are able to define the best observation period over which to review fund performance for each fund. The CUSUM article is posted at http://www.northinfo.com/documents/72.pdf.

Now that we know the period over which we want to evaluate performance for a manager, we can move on to defining a performance metric. deGroot and Plantinga (2001) studied many popular performance measures and concluded that many measures are congruent to value-added for investors for only a small fraction of investors. One such widely used measure is the information ratio (alpha/tracking error). Consider a manager that adds exactly one basis point of return in every time period. The information ratio is infinite, but very little investor wealth is added.

For our empirical study, we chose to measure excess return above a carefully chosen benchmark that should reflect both risk and investing style. This directly measures added value for investors. To the extent that an efficiency measure such as the information ratio is also desirable, our version of CUSUM analysis already incorporates the trend in the information ratio over time.

The most obvious reason why persistence may occur among fund active returns is that some managers are more skillful than others. Therefore, our metric of performance must consider the statistical significance of a manager's active return record, so as to identify those who are skillful from those who have been merely lucky. To do this we must incorporate information about the dispersion of performance of peer funds into the evaluation of each individual fund. If manager returns are widely dispersed within a peer group, it's easier to have a high excess return or a high information ratio. If the dispersion of returns is low, it's harder.

We adopted a Bayesian framework of a "precision weighted" excess return estimate that incorporates information

about the dispersion of peer fund returns during the evaluation period for each fund. This is similar to the method in Shanken and Jones (2004), but without the elaborate Monte-Carlo simulations. Here's a simple example of the math:

- Lets assume Manager X has an excess return (A) of 4% per year with a standard deviation (S) of 4%
- Over the same time period, the average peer manager had an annual excess return of .25% (Mean), and the dispersion (CSD) of the excess returns across the peer group is 1.5%

$E = (A/S^2 + Mean/CSD^2) / (1/S^2 + 1/CSD^2)$

$A = 4, S = 4, MEAN = 0.25, CSD = 1.5$

$E \text{ (precision weighted)} = (0.361) / (0.5069) = 0.712$

So in this case, we would believe that our best estimate of the active return derived from manager skill rather than luck is 71 basis points, rather than the 4% active return that the manager actually achieved during the sample period studied.

Empirical Tests

In order to test the predictive efficacy of our method, we undertook to study a large body of US equity mutual funds from the early 1990s through 2005. Further tests were then conducted on international mutual funds and on a large selection of hedge funds. The method showed predictive power that was both statistically and economically significant in all cases.

For the study of domestic funds, our initial universe was all equity mutual funds in existence as of April, 2005. We then removed a variety of funds including multiple classes of shares in the same fund, index funds, tax efficient funds, and other highly specialized strategies. Our data consisted of monthly returns (as reported by Standard and Poor's Micropal) on the remaining funds.

A set of fifteen benchmark indices were chosen for the domestic funds. These included the Russell 1000, 2000 and 3000 (plus their Value and Growth subsets for a total of nine indices), the S&P 500 and the S&P Midcap (plus their Value and Growth subsets for a total of six indices). We tested two approaches to benchmark assignments. The first approach was to use style analysis (with confidence intervals) to assign funds to benchmarks based on a set of rules. The second approach was to simply assign each fund to whichever benchmark index it was most correlated over the past sixty months. Both approaches provided similar results, so the latter was chosen as more tractable in operation. **Table 1** summarizes the benchmark assignments for the overall period of 1995 through 2004.

A Unified Approach to Monitoring and Evaluating Investment Managers

Table 1: Assignment of Benchmarks						
Style\Size	S&P 500	Russell 1000	S&P Mid	Russell 2000	Russell 3000	Total
Value	8.3%	7.7%	2.5%	3.9%	9.2%	31.6%
Blend	11.1%	4.9%	2.8%	2.8%	14.3%	35.8%
Growth	1.3%	2.8%	3.5%	10.7%	14.3%	32.5%
Total	20.7%	15.4%	8.9%	17.4%	37.7%	100.0%

If mutual fund managers behaved like institutional managers, then the choice of a single benchmark for each fund would be simple – the appropriate benchmark would be highly correlated with the fund's returns and the appropriate benchmark would not change over time. In retail funds, the ambiguity of benchmarks allow for wide latitude by fund managers. There was a high degree of benchmark switching based on the rolling 60 month correlation values, as shown in Figure 1. It should be noted that this analysis may overstate the seriousness of the problem in that many of the benchmark indices are highly correlated with one another (e.g. the Russell 1000 and the Russell 3000 typically have correlation over 99%), so many of the switches may arise from statistical noise in the data.

Method of Analysis

The goal is to determine if mutual fund (or other funds, such as institutional or hedge funds) returns demonstrate a persistence that can be anticipated by the CUSUM statistic.

Currently, we believe that a reasonable approach to this is to do a set of non-overlapping cross sectional regressions. If we do an 8 year estimation period, we will have 16 independent cross sectional regressions of the sort:

$$r = \alpha + \beta \times \text{Average Annual Returns}$$

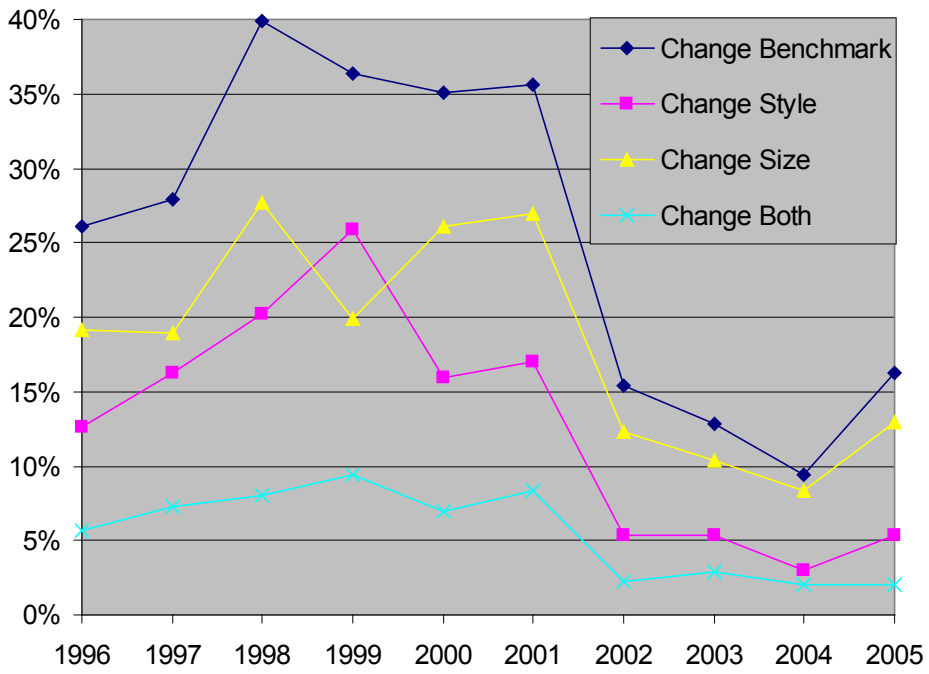
where:

r = return over following 12 months

α = regression intercept

β = regression coefficient, presumably positive and statistically significant, that represents persistence of returns from the inflection point forward.

Figure 1: Portion of Mutual Funds that Change Maximum Correlation 60 Month Rolling Correlations



The initial regression used pooled results on 9550 fund-years from 1993 through the third quarter of 2005. The dependent variable was the annual active returns and the independent variable as the active returns from the CUSUM inflection point to the month preceding the prediction period.

These results show that the previous returns are related to the future returns at a statistically and economically significant level, with about 20% of the years' returns being predicted by the returns during the preceding period. These regressions were done using both pooled and annual regressions, and are shown in Table 2 (next page), which shows that two years (2000 and 2003) had negative persistence, and the negative persistence was statistically significant only in 2003, the year that reversed the three-year decline in most market cap and sector benchmarks.

Table 2: Annual Regressions of Active Return vs. Previous Return, Domestic Funds			
	Intercept	Previous	t-stat
1993	0.237	0.288	6.5
1994	0.364	0.110	2.2
1995	-1.398	0.063	0.7
1996	-0.041	0.002	0.0
1997	-2.691	0.327	4.2
1998	-1.528	0.409	3.0
1999	1.204	0.758	5.9
2000	4.067	-0.028	-0.4
2001	-1.027	0.151	2.5
2002	-0.410	0.254	5.0
2003	-2.388	-0.382	-7.0
2004	-1.896	0.352	10.2
2005	-0.381	0.268	3.6
Average	-0.453	0.198	2.8
Pooled	-0.604	0.192	9.6

It is interesting to note that in using the CUSUM technique to evaluate the appropriate time period to evaluate for funds, the average "best evaluation period" was around two years for the bulk of the time sample. Since 2001, this figure as trended upward illustrating the pervasive effect of the collapse of the tech bubble on the performance of domestic mutual funds.

Given that our data sample contained only surviving funds, there were concerns about the potential bias that this could induce in the results. This issue has been extensively researched in academia in papers such as Elton, Gruber and Blake (1996) and Hallahan and Faff (2001). It was our initial judgment that the rate of extinction among domestic funds was not sufficient to have meaningfully biased our result. Typically, funds with poor performance are the ones which become extinct by being merged with more successful funds. As such, one simple test for potential survivorship bias is to determine whether our results change when looking at subsets of funds with only positive or negative returns.

To do this, we create a new independent variable in our regressions that is the product of a dummy variable, which is equal to zero if the active returns are positive and one if the returns are negative, and the average annual return in the prediction period.

If this variable has a positive coefficient, then one might conclude that, among surviving funds, that persistence

was stronger for funds with negative returns. This would reduce the value of the technique in identifying managers with superior returns if only unskilled (negative active returns) managers showed persistent performance. A pooled regression that includes a variable for negative active returns shows that this effect was negligible over our data sample, and not economically nor statistically significant.

Table 3 shows the regression results when using precision returns for the independent variable and actual returns for the dependent variables. Because the mean active return varies between the independent and dependent periods, a pooled regression will not be useful for this analysis.

Table 3: Annual Regressions of Active Return vs. Previous Return, with Precision Adjustment

	Intercept	Previous	t-stat
1993	2.364	1.675	4.3
1994	2.907	2.935	2.0
1995	-1.388	2.292	0.9
1996	2.907	2.935	2.0
1997	-0.428	4.259	2.2
1998	5.943	6.539	2.2
1999	3.798	6.007	2.4
2000	4.430	-0.408	-0.4
2001	-1.929	0.698	0.6
2002	-1.328	3.373	2.1
2003	-3.670	-4.267	-1.6
2004	1.439	10.101	5.7
2005	7.005	16.659	2.6
Average	1.696	4.061	1.9
Pooled	-0.513	0.918	8.0

We repeated this same analysis on a database of international funds managed, for the most part, by US based mutual fund companies. We assigned one of two different benchmarks, either the MSCI EAFE index (assigned to 79% of funds) or the IFCI Emerging Markets Investible index (assigned to 21% of funds).

In Table 4, we see that the average coefficient is similar (0.118 vs. 0.198) to the result for the domestic fund, but the t-statistics are somewhat lower, but still highly significant. On the other hand, Table 5 shows that the average persistence coefficient is 1.341 when using precision returns.

A Unified Approach to Monitoring and Evaluating Investment Managers

Table 4: Annual Regressions of Active Returns vs. Previous Return, International Funds								
	1998	1999	2000	2001	2002	2003	2004	Average
Intercept	-4.629	6.806	4.364	-0.159	-0.055	-0.639	0.894	0.940
Previous	0.046	-0.127	-0.218	0.154	0.394	0.241	0.332	0.118
t-stat	0.806	-1.106	-6.822	4.433	10.699	4.494	9.149	3.093

Table 5: Annual Regressions of Active Returns vs. Previous Return, with Precision Adjustment, International Funds								
	1998	1999	2000	2001	2002	2003	2004	Average
Intercept	-5.131	2.264	7.435	-3.216	-2.297	0.014	3.540	1.290
Previous	0.153	-0.741	-0.682	1.071	3.094	2.250	3.055	1.341
t-stat	0.542	-0.897	-7.068	6.224	9.396	3.376	6.688	2.953

Using the Hedge Fund.net database, we identified about 500 funds with a sufficiently long history to repeat this analysis on hedge funds. Following the technique used for domestic and international mutual funds, we estimated the correlation between hedge fund returns and several different indices, represented by the columns in Table 1. We then determined which index had the highest correlation with the fund returns over the study period and then compared this to alternate indices that would allow a more parsimonious choice of benchmark. This established that three indices (cash, MSCI World Index and the Russell 3000 index) would be reasonable benchmarks for the various hedge funds.

Table 1: Assignment of Benchmarks								
Row	Bonds	Cash	EAFE	IFCI	MSCIW	R3000	SP500	Total
Cash	18.8%	15.7%						34.6%
MSCIW			11.7%	3.7%	5.2%			20.6%
R3000						39.8%	5.0%	44.9%

Because hedge funds normally employ a long-short strategy, we also need to adjust each hedge fund by its appropriate beta for the appropriate period when estimating active returns. This is particularly important because of the serial correlation of index returns (1995-1999 were positive while 2000-2002 were negative for domestic equities).

$$\alpha = r_{fund} - r_f - \beta \times (r_{index} - r_f)$$

Table 6 shows the range of betas for the different indices:

Table 6: Average and Standard Deviation of b for Highest Correlation Equity Indices		
Benchmark	$\beta_{average}$	$\beta_{standard\ Deviation}$
EAFE	0.37	0.36
IFCI	0.14	0.12
MSCIW	0.60	0.35
R3000	0.54	0.39
SP500	0.42	0.43

Comparing Table 7 and Table 8 show that, in the case of hedge funds, use of Precision-Weighted Returns increases both the average regression coefficient and its statistical significance.

Table 7 : Annual Regressions of Active Return vs. Previous Return, Hedge Funds								
	1998	1999	2000	2001	2002	2003	2004	Average
Intercept	-0.941	10.043	10.151	8.417	5.806	9.674	3.491	6.663
Previous	0.198	-0.132	0.166	-0.013	0.069	0.059	0.114	0.066
t-stat	2.224	-2.856	5.314	-0.378	2.257	1.601	3.625	1.684

Table 8 : Annual Regressions of Active Return vs. Previous Return with Precision Adjustment, Hedge Funds								
	1998	1999	2000	2001	2002	2003	2004	Average
Intercept	-7.512	10.244	6.114	-0.409	-0.207	4.458	-4.406	1.183
Previous	0.630	-0.059	0.629	0.692	0.730	0.654	0.907	0.598
t-stat	1.624	-0.264	4.484	4.055	3.988	2.947	4.850	3.098

Conclusions

We have demonstrated a methodology for the monitoring and evaluation of active managers. This methodology incorporates three stages. We begin by using augmented returns based style analysis to help us select the most appropriate benchmark index. We then use CUSUM analysis to determine the most meaningful period of past

performance to evaluate. Finally, we use a precision-weighted estimate of excess returns as the metric of performance that incorporates contemporaneous information about returns achieved across the peer group of managers.

The results of large scale empirical tests of US domestic mutual funds, international mutual funds and hedge funds all support this methodology as being able to provide predictive rankings of managers that are both statistically and economically significant over a one calendar year time horizon.

References

Bauman, W. Scott and R. Miller, "Portfolio Performance Rankings in Stock Market Cycles," *Financial Analysts Journal*, March–April 1995, pp.79–87.

Brown, S. and W. Goetzmann, "Performance Persistence," *The Journal of Finance*, Vol. 50, No. 2, June 1995, pp. 679–698.

Brown, Stephen J. and William N. Goetzmann. "Mutual Fund Styles," *Journal of Financial Economics*, 1997, v43(3,Mar), 373-399.

Brown, Stephen J., William Goetzmann, Roger G. Ibbotson and Stephen A. Ross. "Survivorship Bias in Performance Studies," *Review of Financial Studies*, 1992, v5(4), 553-580.

Brown, Stephen J., William N. Goetzmann, Robert G. Ibbotson and Stephen A. Ross. "Rejoinder: The J-Shape of Performance Persistence Given Survivorship Bias," *Review of Economics and Statistics*, 1997, v79(2,May), 167-170.

Carhart, M. "On Persistence in Mutual Fund Performance," *The Journal of Finance*, Vol. 52, No. 1, March 1997, pp. 57–82.

Carhart, M. M., J. N. Carpenter, A. W. Lynch and D. K. Musto. "Mutual Fund Survivorship," *Review of Financial Studies*, 2002, v15(5), 1439-1463.

Carpenter, Jennifer N. and Anthony W. Lynch. "Survivorship Bias and Attrition Effects in Measures of Performance Persistence," *Journal of Financial Economics*, 1999, v54(3,Dec), 337-374.

deGroot, Sebastien and Auke Plantinga, "Risk Adjusted Performance Measures and Implied Risk Attitudes", *Journal of Performance Measurement*, Winter 2001/2002

Detzel, F. Larry and R. Weigand, "Explaining Persistence in Mutual Fund Performance," *Financial Services Review*, Vol. 7, No. 1, 1998, pp. 45–55.

diBartolomeo, Dan and Erik Witkowski. "Mutual Fund Misclassification: Evidence Based On Style Analysis," *Financial Analyst Journal*, 1997, v53(5,Sep/Oct), 32-43.

Elton, E., M. Gruber, and C. Blake, "The Persistence of Risk-Adjusted Mutual Fund Performance," *Journal of Business*, Vol. 69, No. 2, 1996, pp. 133–157.

Elton, Edwin J., Martin J. Gruber and Christopher R. Blake. "Survivorship Bias and Mutual Fund Performance," *Review of Financial Studies*, 1996, v9 (4,Winter), 1097-1120.

Fortin, R. and S. Michelson, "What Mutual Funds Really Return After Taxes," *Journal of Financial Planning*, April 1996, pp. 60–66.

Goetzmann, W., and R. Ibbotson, "Do Winners Repeat?" *The Journal of Portfolio Management*, Winter 1994, pp. 9–18.

Golec, J., "The Effects of Mutual Fund Managers' Characteristics on Their Portfolio Performance, Risk and Fees," *Financial Services Review*, Vol. 5, No. 2, 1996, pp. 133–148.

Hallahan, Terrence A. and Robert W. Faff. "Induced Persistence or Reversals in Fund Performance?: Effect Of Survivorship Bias," *Applied Financial Economics*, 2001, v11(2,Apr), 119-126.

Hendricks, D., J. Patel and R. Zeckhauser, "Hot Hands in Mutual Funds: Short–Run Persistence of Performance, 1974–88" *Journal of Finance*, March 1993, pp. 93–130.

Hendricks, Darryll, Jayendu Patel and Richard Zeckhauser. "The J-Shape Of Performance Persistence Given Survivorship Bias," *Review of Economics and Statistics*, 1997, v79(2,May), 161-166.

Jones, Christopher and Jay Shanken, "Mutual Fund Performance with Learning Across Funds", *Financial Management Association Annual Meeting Proceedings*, 2003

Khorana, A., "Top Management Turnover: An Empirical Investigation of Mutual Fund Managers," *Journal of Financial Economics*, Vol. 40 (1996), pp. 403–427.

Kim, Moon, Ravi Shukla and Michael Thomas. "Mutual Fund Objective Misclassification," *Journal of Economics and Business*, 2000, v52(4,Jul/Aug), 309-324.

Kritzman, Mark. "How To Build A Normal Portfolio In Three Easy Steps," *Journal of Portfolio Management*, 1987, v13(4), 21-23.

Lemak D., and P. Satish, "Mutual Fund Performance and Managers' Terms of Service: Are There Performance Differences?" *The Journal of Investing*, Winter 1996, pp. 59–63.

Lobosco, Angelo and Dan DiBartolomeo. "Approximately The Confidence Intervals For Sharpe Style Weights," Financial Analyst Journal, 1997, v53(4,Jul/Aug), 80-85.

Page, E.S. "Continuous Inspection Schemes", *Biometrika*, 1954

Philips, Thomas K., David Stein and Emmanuel Yashchin, "Using Statistical Process Control to Monitor Active Managers", *Journal of Portfolio Management*, 2003

Porter G., and J. Trifts, "Performance Persistence of Experienced Mutual Fund Managers," *Financial Services Review*, Vol. 7, No. 1, 1998, pp. 57–68.

Sharpe, William F. "Asset Allocation: Management Style And Performance Measurement," *Journal of Portfolio Management*, 1992, v18(2), 7-19.

Stewart, S., "Is Consistency of Performance a Good Measure of Manager Skill?" *Journal of Portfolio Management*, Spring 1998, pp. 22–32.