Testing the Hypothesis "Hedge Fund Performance Is Good"

Published in the Spring, 2005 issue of the *Journal of Wealth Management*

Synopsis

The article offers a solution to the problem of evaluating the performance of long-short equity hedge funds. While peer groups and indexes do not work for a variety of reasons discussed in the article, a modern-day application of classical statistics does work. Taking the view that performance evaluation is a hypothesis test, where the hypothesis is "performance is good," the article recommends comparing what could have happened with what actually happened. Portfolio simulations create random portfolios that conform to an individual manager's approach, thereby customizing performance evaluation to each manager. This simulation technology is not new or hypothetical. It has been used to evaluate traditional long-only managers for more than a decade and has recently been extended to hedge funds. Examples are provided that show how hedge funds are evaluated using portfolio simulations as the backdrop.

Ronald J. Surz, 78 Marbella, San Clemente, CA 92673, 949/488-8339, Ron@PPCA-Inc.com is president of PPCA, a software firm providing advanced analytics, primarily to financial consultants, for performance evaluation and attribution. Mr. Surz is also active in the investment community, serving on numerous boards, including AIMR's Investment Performance Council and after-tax subcommittee, the Investment Management Consultants Association (IMCA) board of directors, and several advisory boards of financial organizations.

Testing the Hypothesis "Hedge Fund Performance Is Good"

Ron Surz, Ron@PPCA-Inc.com, 949/488-8339

We are what we repeatedly do.
Excellence therefore is not an act, but a habit.
------ Aristotle

It's time for some new improved habits.

Despite the high costs involved, investors have been flying blind when attempting to choose a hedge fund. Even investment consultants who help investors select among these funds readily admit that there is no way to scrutinize hedge funds the way they would a less complex money management firm—that is, until now. The heart of the problem has been the fact that traditional performance evaluation approaches simply don't work for hedge funds. That's attributable to two basic truths:

- Traditional peer groups and indexes are poor benchmarks for traditional performance evaluation.
- **Traditional peer groups and indexes are simply a joke for non-traditional performance evaluation.**

In this article, we describe the reasons traditional performance evaluation approaches do not work—for traditional investments as well as hedge funds. However, unlike previous articles that have simply documented the problems, we offer a solution: Namely, performance evaluation in general, and hedge fund performance evaluation in particular, should be viewed as a hypothesis test where the validity of the hypothesis "Performance is good" is assessed. To accept or reject this hypothesis, we construct all of the possible outcomes and see where the actual performance result falls. If the observed performance is toward the top of all of the

possibilities, the hypothesis is correct, and performance is good. Otherwise, it is not. In other words, the hypothesis test compares what could have happened with what actually happened. By doing so, we determine how much of the available return was actually earned by the manager as well the significance of the manager's value added or subtracted. That is, the available return is the middle of the simulated distribution, and the manager's location within the entire opportunity set is the significance of how much or little of the available return was captured.

The Problems with Peer Groups and Indexes

Everyone who has earned the CFA (Chartered Financial Analyst) designation has learned the problems with peer groups, i.e., they are loaded with biases. Ankrim[1998], Bailey[1992], and Bleiberg[1986] have identified these biases. The popular financial press frequently writes about manifestations of these biases. For example, Eley [2004] documents material differences in performance rankings. The same manager that appeared to be successful against one peer group provider's universe was unsuccessful against a comparable peer group supplied by another provider. Numerous articles of this ilk have appeared in the press over the years. The sell side knows how to deal with these problems: Use the peer group that makes them look best. It's the buy side—the investor—that suffers most from the problems with peer groups.

The CFA Institute (formerly AIMR) website on benchmarks[1] features a discussion of peer groups that acknowledges their broad usage, but that also cautions about the groups' drawbacks and biases. The website seems to promote the use of custom benchmarks, which solve the bias problems, but which come with two new problems of their own: defining the benchmark and waiting. While custom benchmarks, or "normal portfolios," were very hard to define prior to the introduction of returns-based style analysis, that job has now become somewhat easier. However, we are still confronted with the fact that it takes decades to develop statistical confidence in a manager's outperformance of a custom benchmark

A move from the traditional world to the non-traditional reveals that the problems with peer groups and indexes are even worse for hedge funds, primarily because each hedge fund is unique. If we consider, for example, "market-neutral," we find that this very popular strategy comes in many forms—dollar, beta, style, sector—the list goes on, and many funds that call themselves market-neutral should not. Accordingly, it is virtually impossible to construct an appropriate peer group or index for a specific market-neutral manager. As a result, many hedge fund managers win or lose the evaluation game based on approach rather than skill. Because they don't take into account the unique specifications of each fund, the peer groups and indexes that are constructed are faulty, which only serves to compound the problems inherent in these approaches.

Unlike the traditional world, where the problems with peer groups and indexes are only occasionally discussed, the hedge fund world routinely bemoans the inadequacies of current approaches for evaluating hedge fund performance. The financial press (e.g., Lukomnik [2003]) has reported extensively on this issue. The problem is exacerbated by the fact that billions of dollars are pouring into hedge funds without the benefit of proper due diligence, since performance evaluation is misleading at best.

However, no one to date has come forth with a solution, despite the extensive documentation and painful awareness of the problem. As stated in the introduction, we believe the solution lies in a contemporary application of classical statistics, namely hypothesis testing. For ease of understanding, one can view hypothesis testing as a dartboard game.

Can Hedge Fund Managers Win the Dartboard Game?

Most readers will have seen *The Wall Street Journal*'s "The Dartboard Game," which challenges professional investors to outperform a portfolio chosen at random by figuratively throwing darts. Surprisingly, or maybe not so surprisingly, this has been a tough game to win. We

recommend a similar challenge for hedge fund managers, not just for amusement but as a practical solution to the problem with current performance evaluation approaches.

Using the dartboard analogy, each individual hedge fund should have its own unique dartboard: some round, some square, some with concentric circles, some with random shapes. In other words, the game is played to each manager's unique specifications, because each hedge fund is unique and therefore without peers, and the game is played at random.

The dartboard game has a real-world application in evaluating investment performance that is directly related to hypothesis testing. This application is not new. Monte Carlo simulation (MCS), as the application is known, has been used to evaluate traditional investing for more than a decade (see Surz [1994, 1996, and 1998] and Bridgeland [2001]). While the application of MCS to traditional portfolios has not caught on (see Chernoff [2003] and Picerno[2003]), John O'Brien [2] explains this lack of acceptance with the observation that "It's hard to get someone to understand an idea when their livelihood relies on not understanding it." Traditional peer groups, not MCS, are financial consultants' bread-and-butter.

Recently, however, MCS technology has been extended to hedge funds, where we believe it will gain acceptance due to the lack of good alternatives and, therefore, no inherent barriers to comprehension. MCS addresses the uniqueness challenge of evaluating hedge fund performance by creating at random all of the possible portfolios that a hedge fund manager could have conceivably held following his/her unique investment process, thereby applying the scientific principles of modern statistics to the problem of hedge fund evaluation. The answer to the question "What hedge funds are in a MCS peer group?" is "All of them that matter."

How MCS Works

In constructing a specific peer group, MCS follows the same rules that the individual hedge fund manager follows in constructing his/her portfolio:

Hedge Fund Attributes

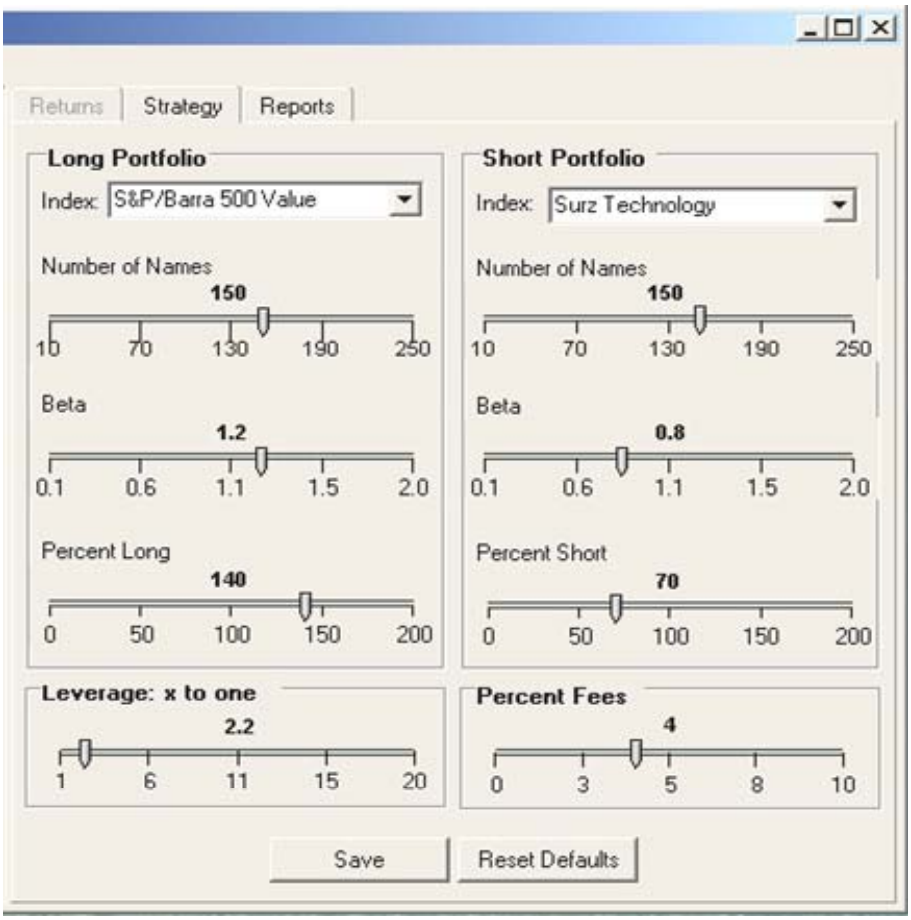
- Investment style: What approach is employed for the long portfolio? For the short portfolio?
- Breadth: How many securities are held long and short?
- Risk: What are the betas of the individual long and short portfolios?
- Direction: How much is held long, and how much is held short?
- Leverage: Is money borrowed to amplify performance?
- Fees

Exhibit 1 shows a typical control panel used to operate MCS. This is where an individual hedge fund's attributes are specified. The simulator uses these inputs to randomly create 10,000 portfolios that conform to the same portfolio construction parameters followed by the actual hedge fund manager. The result is a scientific and unbiased backdrop for evaluating that manager's performance.

Academics have performed such portfolio simulations for years. Fisher and Lorie [1970], for example, used portfolio simulations to determine the risk reduction benefits of increasing the number of holdings in a portfolio. However, in performing simulations, academics typically make two significant mistakes that must be avoided if the resulting opportunity sets are to reflect the real world, i.e., equal weighting and buy-and-hold. Simulating equal-weighted portfolios conforms to the common management practice of holding roughly equal positions in each stock, but the macro consequences are a "world" that holds as much Penny Stock as it does General Electric. The solution to this serious error is to use value-weighting sampling so that the probability of choosing a given stock is proportionate to its outstanding capitalization. As a result, GE is in a lot of equal-weighted portfolios, and Penny is in just a few. Importantly, the sum of the dollar holdings of a given stock across all portfolios is that stock's outstanding capitalization, so the simulations preserve macroeconomic consistency.

Similarly, academics make the mistake of creating buy-and-hold portfolios. For example, Fisher and Lorie's simulated portfolios buy stocks at the beginning of 1926 and hold them through the end of 1965. This is also not real world. Companies go in and out of business, and real portfolios transact. We recommend that simulated portfolios transact every month so they reflect the real world.

Exhibit 1: Defining the Hedge Fund



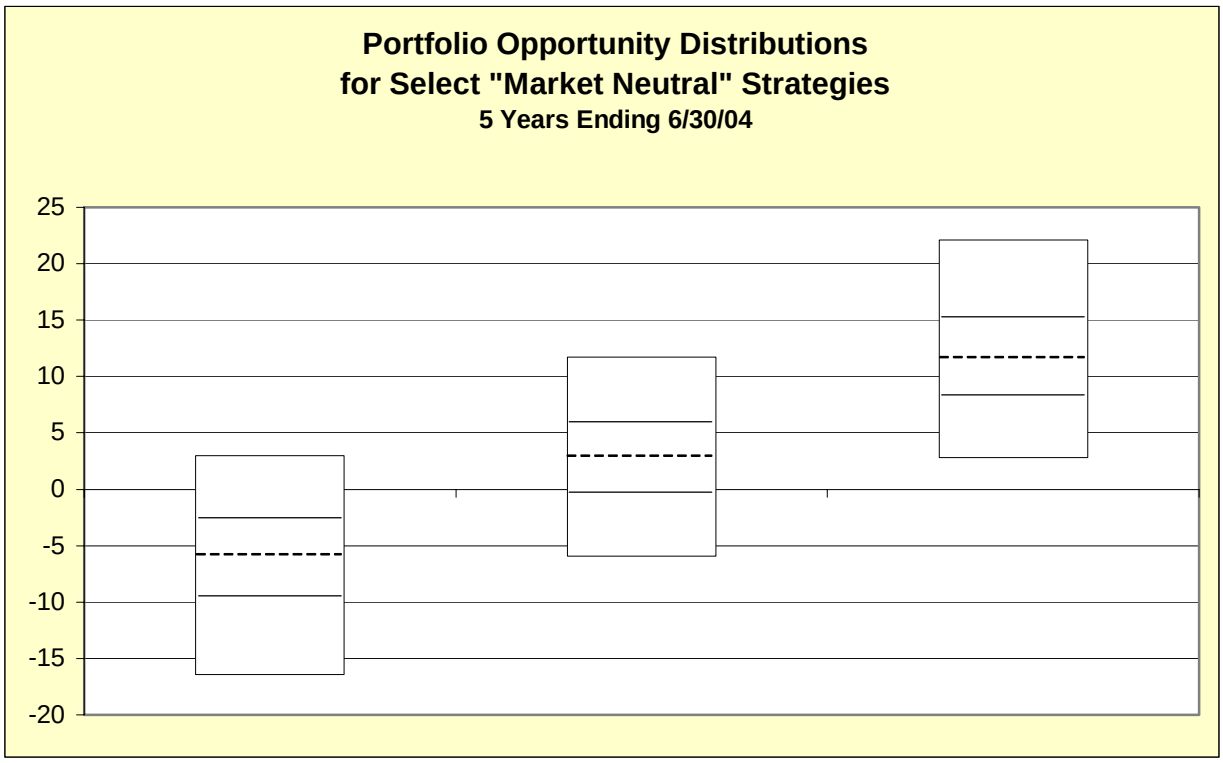
Looking at an Example: Market-Neutral

Now we can look at how this works using a single strategy. To remain consistent with our previous discussion, we'll use market-neutral. As stated earlier, market-neutral is a very popular hedge fund strategy, and there are as many types of market-neutral as there are

market-neutral funds. This creates a real problem for performance evaluators, who would like to extend traditional evaluation approaches to hedge funds but find that they cannot.

Big differences exist among the various flavors of market-neutral, and performance results reflect these differences. To illustrate this point, we'll use three market-neutral substrategies as examples. All three are dollar-neutral as well as beta-neutral, but they vary in their style neutrality. One is style-neutral, one is long value and short growth, and the third is long growth and short value. The following scientific peer groups show the dramatic differences in the opportunities available to these three substrategies over the five years ended June 30, 2004, even though all three are properly designated as market-neutral.

Exhibit 2: MCS Peer Groups for Different Market-Neutral Strategies



Long :	Growth	Market	Value
Short :	Value	Market	Growth
5	3.0	11.7	22.1
25	-2.5	6.0	15.3

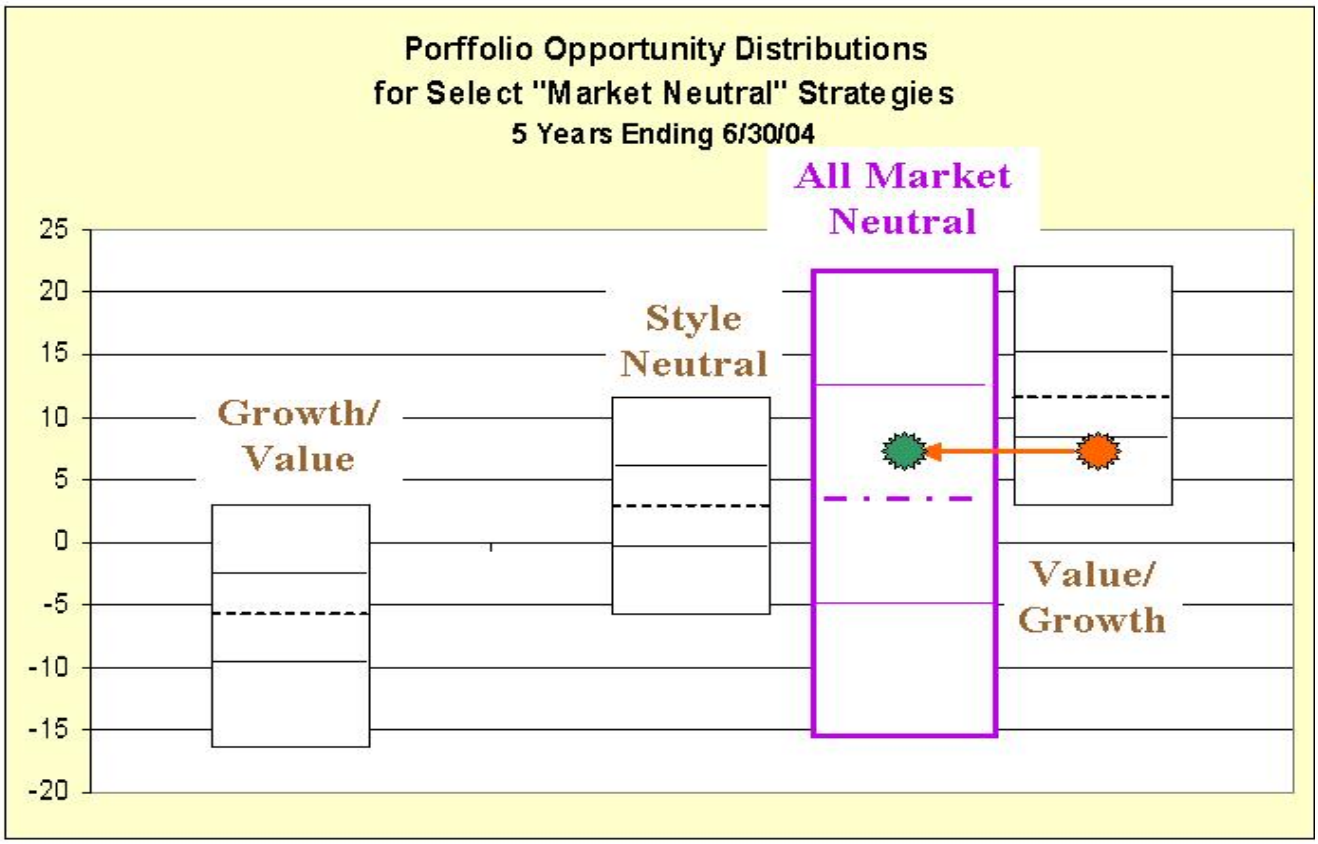
50	-5.8	2.9	11.6
75	-9.5	-0.3	8.3
95	-16.4	-5.9	2.8

As we can see in Exhibit 2, the best performing long growth/short value manager, with a 3.0% return, ranks among the worst when compared with the long value/short growth universe and barely breaks the median in the style-neutral universe. It's obvious from this exhibit that a skillful manager could very easily be adjudged as unskilled, and vice versa, if the wrong peer groups are used—a mistake that can be avoided with proper specification of the manager's approach in MCS.

This ability to identify skill is particularly important to fund-of-funds managers. The three substrategies in our example might very well be candidates for inclusion in the market-neutral sleeve of a fund of hedge funds, since these three variants on the market-neutral theme tend to exhibit low correlations relative to one another, making them good diversification partners. A fund-of-funds manager tries to gather the best talent within each strategy and diversify among that talent. As is the case in traditional investing, the search for talent must start with proper identification of the opportunities available to the specific management style.

To more fully illustrate the benefits of MCS, we can consider what happens currently without MCS. As we can see in Exhibit 3, a candidate for inclusion in a fund-of-funds has delivered above-median performance when contrasted with the traditional non-customized peer group of all market-neutral managers, making him an apparently viable candidate. However, when viewed with MCS technology, this same performance ranks in bottom quartile of the candidate's long value/short growth substrategy. Without the insights provided by MCS, even the most sophisticated investors, including fund-of-funds managers, can easily be led to mistake strategic dominance for skill. The question, of course, is will style dominance shift? Presumably skill, or lack thereof, will persist through style shifts.

Exhibit 3: The Difference Customization Makes



Why MCS?

As this example has illustrated, MCS offers investors the ability to distinguish between the form and the substance of an individual hedge fund. This ability has become more critical as hedge funds attract increasing interest based on their offer of protection (hedge) against the downside, while potentially delivering value added on the upside through investment management skill. The downside protection is the form, and the presumed skill is the substance. This has led some to proclaim that the industry is "buying beta, not alpha," that is, investors are buying the downside protection and not getting the value added through skill. Downside protection can be purchased cheaply in a variety of ways. The average fund created by the MCS technology represents the form of the strategy being simulated. In most cases, this average fund's returns could be purchased passively as long and short index funds. An individual fund's rank in its

MCS universe is its substance, that is, the likelihood that performance is good, evidencing skill or absence thereof.

Investors have literally been throwing money at hedge funds because of their mystique, rather than any real verifiable confidence in manager skill. Hedge fund managers have gladly accepted these riches while denying access to the information that might reveal the emperor's clothes, or lack thereof. MCS circumvents this lack of transparency by using returns to determine the likelihood of skill, an insight that cannot be achieved with other evaluation approaches currently available. MCS also provides a credibility check on manager performance. A reported return is suspect if it exceeds or lags all of the possible returns in its strategy.

For investment firms offering hedge funds, MCS technology also provides an ideal way to backtest new product ideas. By creating all of the possible hedge funds in the new strategy and examining them in various market environments, MCS enables an investment firm to observe the natural return to the strategy (the middle of the distribution of returns) as well as the implementation risk (the range of potential returns), thereby providing invaluable guidance for product development decisions.

In response to the question "Why MCS?," the acronym SAVE provides the answer:

- ✓ **S**ubstantiate new product ideas
- ✓ **A**ccentuate skill, versus hedge (alpha versus beta)
- ✓ **V**alidate performance reports
- ✓ **E**valuate hedge fund performance

Conclusion

Because they can reflect the unique specifications of each individual hedge fund, Monte Carlo simulations offer an important breakthrough in hedge fund evaluation. With billions of dollars pouring into hedge funds every year, it's imperative that investors have the opportunity to

access the insights provided by this new technology. Unlike the *Journal*'s Dartboard Game, MCS is a game that skillful hedge fund managers should win.

FOOTNOTES

[1] http://www.cfainstitute.org/standards/pps/benchmark_incl.html

[2] John O'Brien currently serves as executive director of the Master's in Financial Engineering program at the Haas School of Business, University of California, Berkeley. He is also a co-founder of Wilshire Associates, a pension consulting firm in Santa Monica, California, and Leland O'Brien Rubinstein, also known as LOR portfolio insurance.

REFERENCES

Ankrim, Ernest M. "Peer-Relative Active Portfolio Performance: It's Even Worse Than We Thought." *The Journal of Performance Measurement*, Summer 1998, pp 6-11

Bailey, Jeffrey V. "Are Manager Universes Acceptable Performance Benchmarks?" *Journal of Portfolio Management*, Spring 1992, pp 9-13

Bleiberg, Steve. "The Nature of the Universe." *Financial Analysts Journal*, March/April 1986, pp 13-14

Bridgeland, Sally. "Process Attribution – A New Way to Measure Skill in Performance Construction." *Journal of Asset Management*, December 2001.

Chernoff, Joel. "Consultant Offers a Way to End Classification Bias." *Pensions & Investments*, August 18, 2003, page 3.

Eley, Randall R. "Database Dysfunction." *Pensions & Investments*, September 6, 2004, page 12.

Fisher, Lawrence and Lorie, James H. "Some Studies of Variability of Returns on Investments in Common Stocks." *Journal of Business Vol 43, No. 2,* April 1970.

Lukomnik, Jon. "Doing Diligence." *Plan Sponsor*, July 2003.
_____. "Hedge Funds." *Plan Sponsor*, April 2003
_____. "Hedge Fund Indices: Apply With Caution." *Plan Sponsor*, January 2003

Picerno, James. "In the Quest to Identify Investment Skill, Ron Surz Claims He Has the Better Mousetrap." *Bloomberg Wealth Manager*, June 2003, pp 80-82.

Surz, Ronald J. "Cyberclone Peer Groups." *Journal of Investing,* Winter 1998, pp 63-67.

_____. "Portfolio Opportunity Distributions: A Solution to the Problems with Benchmarks and Peer Groups." *Journal of Performance Measurement,* Winter 1996, pp 24-30.
_____. "Portfolio Opportunity Distributions: An Innovation in Performance Evaluation." *Journal of Investing,* Summer 1994.