A Fresh Look at Investment Performance Evaluation

Unifying best practices to improve timeliness and reliability.

Ronald J. Surz

Impermanence is eternal.
—Buddha

The jury is still out on the use of active investment management versus passive index funds. Research indicates that active managers don't add value, but common practice suggests that investors either do not believe this research or else derive other benefits beyond performance from active management. Consultants try to give clients what they want, so they look for skillful investment managers, but this is a difficult task complicated by the use of performance evaluation tools that suffer from documented deficiencies.

An integral part of the search for skill is getting the numbers to tell their most important stories, confirming subjective judgments about the talent of the people and the wisdom of their processes and philosophy. This story is virtually always told by comparing a manager's investment return with the return of an appropriate peer group and index.

It's time to recognize the deficiencies of these popular approaches so that we can open up consideration of new alternative methodologies.

The investment industry has been using peer groups and indexes for such a long time that no one thinks to question them, except for the occasional discussion about peer group survivor bias. This is just the way it's always been done, so it must be right. In actuality, peer groups suffer from a collection of biases, of which survivor bias is only one, and each peer group has its own unique set of idiosyncratic distortions. As a result, the exact same performance number will rank differently against different peer groups, even when all the peer groups operate on the same management mandate, such as large-cap growth.

Ronald J. Surz
is president of PPCA, a software firm in San Clemente, CA.
Ron@PPCA-Inc.com

If we use peer group comparisons, it is difficult to truly determine which manager has succeeded and which has failed, because success against one peer group can easily be failure against another comparable peer group. A manager, in fact, can be both a success and a failure at the same time. This is a serious problem that can lead to faulty decisions, hiring unsuccessful managers and firing successful ones. Similarly, indexes suffer from problems of appropriateness and significance.

There is an alternative, although it is not currently in the mainstream. I describe the reasons that traditional performance evaluation approaches do not work for traditional investments or for hedge funds either. Beyond documenting the problems, I offer a solution: Namely, performance evaluation should be viewed as a hypothesis test that assesses the validity of the hypothesis, "Performance is good."

To accept or reject this hypothesis, all the possible performance outcomes are constructed to determine where the actual performance result falls. If the observed performance appears toward the top of all the possibilities, the hypothesis is correct, and performance is good. Otherwise, it is not.

In other words, the hypothesis test compares what actually happened with what could have happened. As a practical matter, the actual implementation approximates all the possible outcomes by using simulation techniques.

I first describe the problems with peer groups and indexes, and then present a solution to these problems, along with answers to criticisms and some examples.

PROBLEMS WITH PEER GROUPS AND INDEXES

The CFA (Chartered Financial Analyst®) Institute (formerly AIMR) has appointed a Subcommittee on Benchmarks and Performance Attribution. In August 1998, the subcommittee reported:

> Universes do offer certain benefits in performance comparisons. Universes represent achieved results of manager portfolios which are effectively available as investment alternatives for investment fiduciaries (fund sponsors etc.), they take full account of transactions and other trading costs and they reflect decisions taken by investors across the board (e.g. to underweight Japan relative to the index.) But as the sole benchmark for comparing performance, universes are subject to certain drawbacks. These include:
>
> 1. They are not available real time, resulting in a time lag for comparison.
> 2. There is no established oversight process for determining universe participants and whether the universe accurately represents the entire asset class or style of management.
> 3. Survivor bias will develop over time as some managers are deleted from the universe.
> 4. They are not replicable or investible.
> 5. They do not permit the manager to move to a known neutral position.

Everyone who has earned the CFA designation has learned the problems with universes, otherwise known as peer groups. They have several biases that cannot be corrected. Bleiberg [1986], Bailey [1992], and Ankrim [1998], among others, have identified the problems with peer groups. They document various biases that undermine evaluations based on traditional peer groups. Three of these biases are classification, composition, and survivorship.

- *Classification* bias results from the practice of forcing every manager into a prespecified pigeonhole, such as growth or value. It is now commonly understood that most managers use a blend of styles, so that pigeonhole classifications misrepresent the manager's actual style as well as those employed by peers.
- *Composition* biases result both from concentrations of certain fund types in databases, such as bank commingled funds, and from small sample sizes. International managers and socially responsible managers cannot be evaluated properly because there are no databases of adequate size.
- *Survivorship* bias causes performance results to be overstated because accounts that have been terminated, some of which may have underperformed, are no longer in the database. This is the most documented and best understood source of peer group bias. Eliminating an unsuccessful management product from current peer groups results in an overstatement of past performance. A simple illustration is the "marathon analogy," which asks: If only 100 runners of a 1,000-contestant marathon actually finish, is the 100th runner the last? Or in the top 10%?

Manifestations of these biases are frequently noted in the popular financial press. For example, Eley [2004] observes material differences in performance rankings. The same manager who appears to be successful against

one peer group provider's universe is unsuccessful against a comparable peer group supplied by another provider. Numerous articles of this ilk have appeared over the years.

A move from the traditional world to the non-traditional reveals that the problems with peer groups are exacerbated for hedge funds, primarily because each hedge fund is unique. Kat [2003] documents the lack of correlation among funds in the same peer group, as well as between different indexes for the same strategies. If we consider, for example, "market-neutral," we find that this very popular strategy comes in many forms—dollar, beta, style, sector—and many funds that call themselves market-neutral should not. Kat finds correlations to be a mere 0.23 among funds in market-neutral peer groups, indicating that these funds are different from one another. This 0.23 is the R of the R^2, so only 5% (0.23 squared) of the variance in one hedge fund can be explained by that of any another fund in the same peer group.

Accordingly, it is virtually impossible to construct an appropriate peer group for a specific market-neutral manager. As a result, many hedge fund managers win or lose the evaluation game because of approach rather than skill. Because it is impossible for peer groups to take into account the unique specifications of each fund, the resulting constructions are faulty, which only serves to compound the problems inherent in this approach. Hedge fund databases serve a useful purpose in providing access to hedge fund managers, but they should not be used as ways to evaluate investment performance.

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

For an extensive discussion of an advanced approach to evaluating hedge fund performance, see Surz [2005].

The CFA Institute's Subcommittee on Benchmarks and Performance Attribution acknowledges the widespread use of peer groups, but cautions as to their drawbacks. Indexes, especially blends of indexes customized to reflect "portfolio objectives and risk constraints," appear to be the subcommittee's preferred approach, but custom benchmarks come with two challenges of their own: first, correctly defining them, and then waiting long enough to develop statistical confidence. The first challenge has been simplified by the introduction of returns-based style analysis, but the waiting time problem remains.

Returns-based style analysis, introduced by Sharpe [1988], has made customization practical. Custom benchmarks can be created as blends of recognized indexes. Sharpe recommends the use of style indexes that are mutually exclusive and exhaustive, but this is rarely the case in practice, so caution is advised. Most managers are best represented as a blend of style indexes, as opposed to the common practice of comparing all managers using a single index such as the Standard & Poor's 500. It is easy to confuse style with skill, but difficult to make good decisions in the face of this mistake. Witness the 1990s' growth bubble.

As for waiting time, researchers have estimated it takes many decades to determine the statistical significance of outperforming a benchmark, even a benchmark that is customized to an individual manager. While combining indexes into custom benchmarks may have become more practical, waiting decades to determine the significance of a manager's performance remains highly impractical.

A SOLUTION

A solution to the problems with traditional peer groups and indexes is actually quite simple in concept, but has only recently been made practical when the requisite computing power became available. As I have noted, performance evaluation can be viewed as a test that assesses the validity of the hypothesis, "performance is good." To accept or reject this hypothesis, we construct an approximation of all the possible outcomes, and determine where the actual performance result falls.

We identify the best benchmark possible and then expand it into a peer group by creating thousands of portfolios that could have been formed from stocks in the benchmark, following reasonable portfolio construction rules. This approach combines the better characteristics of both peer groups and indexes, while diminishing the deficiencies of each. Most important, statistical significance is determined much more quickly than with indexes because inferences are drawn in the cross-section rather than across time. In other words, the ranking of actual performance against all possible portfolios is a measure of statistical confidence.

Most readers have seen *The Wall Street Journal*'s "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. Our recommended

solution to the problems with traditional peer groups and indexes presents a similar challenge, not just for amusement, but as a practical answer.

Using the dartboard analogy, each individual manager should have his or her own unique dartboard: some round, some square, some with concentric circles, some with random shapes. This customization for traditional managers is tied to the manager's investment style, such as large growth or small value. For non-traditional managers, the customization extends to characteristics such as styles long and short, amounts long and short (otherwise known as *direction*), betas long and short, leverage, and fees.

In other words, the game is played to each manager's unique specifications, thereby eliminating the biases inherent in traditional peer groups. This removes the confusion about ranking well in one peer group but poorly in another, since there is only one customized ranking.

Two questions are central to this process: What does this manager do (style, e.g.), and does the manager do it well? The first question addresses the form of the investment, and the second identifies the substance, or skill. Hedge fund managers obfuscate the first question, and answer the second with credentials and absolute return targets rather than real performance. To obtain the correct answers, use good old-fashioned due diligence for the first question and hypothesis testing for the second.

Performance evaluation is all about making a judgment as to whether performance is good or bad. This judgment should be made relative to a pure and unbiased backdrop based on a passive alternative. It doesn't matter how other managers in a particular peer group have fared, since they too should each be evaluated against their particular passive alternatives. What does matter is the degree of success or failure the manager has experienced relative to the benchmark.

The benchmark in this context is the answer to the first question: What does this manager do? The ranking within the manager's customized opportunity set answers the question, Does the manager do it well? Peer groups still serve the useful purpose of access, providing candidates for investment management positions, but they can be improved upon when it comes to performance evaluation.

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 simulations (MCS), as the application is known, have been used to evaluate traditional investing for more than a decade (see Surz [1994, 1996, and 1998], Bridgeland [2001], Haltiner and Surz [2004], and Burns [2004]). While the application of MCS to traditional portfolios has not yet been accepted as standard practice (see Chernoff [2003] and Picerno [2003]), that doesn't mean that the idea is faulty. Modern portfolio theory (MPT) took 30 years to become established.

Further improving the likelihood of acceptance, MCS technology has been extended to hedge funds, where recognition of the fact that peer groups don't work has eliminated any inherent barriers to comprehension and adoption. MCS addresses the uniqueness in evaluating both long-only and hedge fund performance by creating at random a good representation of all the possible portfolios that a manager could conceivably have held, following a unique investment process, thereby applying the scientific principles of modern statistics to the problem of performance evaluation. An MCS universe is readily produced in risk-adjusted as well as fee-adjusted forms, a necessary requirement for hedge fund investing.

Regardless of the form, the answer to the question of what funds are in an MCS peer group is: all of them that matter.

CONTRAST OF MCS TO PEER GROUPS AND BENCHMARKS

There are many ways to implement an MCS approach, some better than others, and some worse. Importantly, MCS does not solve the problem of identifying the appropriate benchmark, but it does solve the waiting time problem associated with benchmarks. As the benchmark might be misspecified, MCS is subject to the classification and composition biases I have described above. In other words, if the benchmark is wrong, MCS is wrong, and even the best MCS approach can't remedy this. As we will see in the examples below, a ranking outside an MCS universe is likely to be an indication of an incorrect benchmark.

Another potential shortcoming is that properly constructed MCS universes need to adhere to certain fundamental rules of portfolio construction and macroeconomic consistency in order to be valid. We advocate MCS approaches that meet these rules, recognizing that there are implementations that do not.

The following list revisits the peer group drawbacks noted by the CFA Institute's Subcommittee on Benchmarks and Performance Attribution, and comments on the applicability to benchmarks and MCS:

1. **Peer groups are not available real time**. Benchmarks are usually available in real time. Since benchmarks are the building blocks for Monte Carlo Simulations, MCS could be made available in real time, but as a practical matter MCS is available monthly or quarterly.
2. **There is no established oversight process for establishing peer groups**. Establishing benchmarks, and therefore MCS, also lacks an established oversight process. Although the CFA Institute has established criteria for good benchmarks, one can still easily pick the wrong benchmark. MCS may provide a clue to benchmark misspecification if the fund's performance falls outside the MCS opportunity set. There is also no established process for creating MCS universes.
3. **Survivor bias will develop over time as some managers are dropped from the universe**. Benchmarks generally do not suffer from survivor bias. Properly constructed MCS universes should also be free of survivor bias. Among other things, this requires that the MCS process transacts.
4. **Peer groups are not replicable or investable**. Some argue that passive benchmarks are also not investable because of transaction costs or turnover. This may be the case, but low-cost index funds do represent a legitimate investable alternative. Also, MCS universes drawn from indexes can be created with various costs netted out from performance, including transactions, management fees, and taxes.
5. **Peer groups do not permit the manager to move to a known neutral position**. Benchmarks do provide such a neutral position. Since properly constructed MCS universes have the benchmark as the median, MCS also provides a neutral position.

ANSWERS TO CRITICISMS

Nothing is perfect, including MCS, although MCS is a legitimate alternative to peer groups and indexes. Like peer groups and indexes, MCS has its own detractors and critics.

MCS Isn't Real

The most common criticism of Monte Carlo Simulation is that it isn't "real." "My clients want to know how other managers have performed," a consultant might say; that is, clients want to see the real horse race.

One must first recognize that performance evaluation is a hypothesis test, whether one is using MCS or traditional peer groups. The question "Is performance good?" can only be answered relative to something, such as a sample. The choice between MCS and peer groups is essentially a choice between sampling approaches.

What sample should be used to test the hypothesis, "performance is good"? When traditional peer groups are used, the sample is a group of portfolios that is presumably managed in a manner similar to the portfolio that is being evaluated, so the hypothesis is tested relative to the stock picks of similar professionals. This makes sense, except someone has to define *similar*, and then collect data on the funds that fit this particular definition of similar.

Each peer group provider has its own definitions and its own collection of funds, so each provider has a different sample for the same investment mandate. Large-cap growth describes one set of funds in one provider's peer group and another set of funds in the next provider's peer group.

These sampling idiosyncrasies are the source of peer group biases. If a manager doesn't like his ranking, he can pick another sample, i.e., choose another peer group provider. The hypothesis test is only as good as the sample—the poorer the sample, the wider the lack of confidence band.

A related criticism is that surely a sophisticated investor can pick an appropriate peer group. This misses the key fact that there are inherent biases in all peer groups, just by their very nature. Can you pick the puppy that won't wet on the floor?

So it becomes a matter of selecting the least biased, but how do we know when we have done so? How is a good sample distinguished from a bad one?

Consider a peer group provider that wants to prove its peer groups are representative. One way to make this proof statement is to determine all of the portfolios that the investing public could hold, and compare this opportunity set to the traditional peer group.

In other words, MCS is the proof statement, or standard, to use in validating traditional peer group samples. Doing this correctly requires rigorous sampling techniques.

MCS Creates Portfolios that No One Would Ever Hold

Another real-world concern is that MCS will create some random portfolios that no investor in her right mind

would hold. It's important that the sampling rules used in an MCS approach preserve real-world practice. Like traditional peer groups, not all MCS approaches are alike. A good MCS approach provides something close to what the statistician calls *perfect information*, which has a cost.

The idea is that samples are imperfect, since the next sample may give substantially different results. The ultimate goal is to not sample at all, but rather to identify all the possibilities, which would represent perfect information. The acquisition of perfect information comes at a cost that frequently can be estimated. Properly constructed MCS universes are close to this concept of perfection and come at a cost that is generally far less than traditional peer groups.

One can simplify the challenges of constructing solid MCS universes by separating the process into the security set that is drawn from and the sampling rules that are used to make the draws. The ideal security set is the manager's normal portfolio, including neutral weights to each stock in the normal. MCS uses these as target holdings so that the universe sums up to this composition, and the middle of an MCS universe is the return on the normal. The ideal sampling rules obey the manager's portfolio construction disciplines, including security and sector guidelines, and number of names. The modeling process must also include periodic trading.

This form of MCS, which can be and has been implemented in the past, works very well, but not many have mastered its nuances. The manager can benefit by a properly constructed MCS universe.

Academicians have performed portfolio simulations for years, and they set an example for how not to create a good MCS universe for purposes of performance evaluation. Fisher and Lorie [1970], for example, use portfolio simulations to determine the risk reduction benefits of increasing the number of holdings in a portfolio.

Scholars typically make two significant mistakes that must be avoided if the resulting opportunity sets are to reflect the real world: 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 macroeconomic consequences are a world that holds as much penny stock as it does General Electric. The solution to this error is to use value-weighted sampling, so that the probability of choosing a given stock is proportionate to its outstanding capitalization. With this sampling technique, GE is held 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.

Buy-and-hold is also not real world. Fisher and Lorie's simulated portfolios bought stocks at the beginning of 1926 and held them through the end of 1965. Yet companies go in and out of business, and real portfolios transact. Properly constructed simulated portfolios transact periodically so that they reflect the real world.

MCS Doesn't Tell Me How My Peers are Doing

The next criticism of MCS is that the evaluator wants to know how other similar entities like other foundations or other pension plans have performed. It is important to recognize that this desire is aimed more at evaluating the competition than at evaluating an investment manager. Consider, for example, the possibility that a few large value managers have penetrated the pension fund marketplace, giving them the lion's share of this market. An evaluation of one of these managers against a pension fund/large value peer group is essentially a comparison against a very small sample that tells the client only that his peers and competitors have succumbed to the same sales efforts as he has, while revealing very little about the skill of the manager evaluated. The manager could perform very well and still be below the median of this elite club, or vice versa.

As the client is not limited to this club when it allocates assets, the hire/fire decisions should address all the opportunities available, not just choices made by a select handful of managers with good marketing efforts. For competitive reasons, the client may still want to look at a peer group of similar investors, but the manager should certainly not be evaluated in this framework.

This competitive focus may even lead to a second-order game: Regardless of manager skill, do I want to be that different from my competitors and venture outside the club? This is fair enough, but it is not performance evaluation.

MCS is Too Different

The final criticism is that MCS is too radical a departure from common practice, and that there is no standardization of MCS approaches. The standardization criticism is valid. There are only a few implementations

of MCS at this time, and they are indeed different from one another.

In most of this article I have described one particular implementation that I can contrast with common practice. The proof of the pudding is in the tasting. How do MCS universes compare with traditional peer groups?

Unconstrained MCS universes, constructed from all companies in the United States, are literally identical to large traditional all-funds equity databases for periods of up to about a year. Longer than that, results from the MCS universes begin to lag due to survivor bias in the traditional peer groups (see Surz [1989]). An advantage of this replicating ability is that MCS universes are available many weeks before traditional peer groups are assembled.

Stylized universes, like small-cap value or large-cap growth, are harder to compare because of definitional differences, but one thing is clear: MCS is the only way to tie performance rankings back to the benchmark. Because MCS includes all possible funds that could be created from a specific benchmark, like the S&P 500, the benchmark return always ranks median. In traditional peer groups, the benchmark never ranks median, indicating that the group of funds is not properly benchmarked or that some funds have been excluded.

If the peer group is complete and accurate, "the arithmetic of active management" (Sharpe [1991]) dictates that in aggregate the managers will earn the benchmark return; that is, the expected return is the benchmark return. There should be no surprises in properly constructed MCS universes.

It is critical to view the choice of universe—whether it is MCS or peer group—as a sampling decision for testing the hypothesis, "performance is good." The better the sample, the better the inference.

EXAMPLES

MCS has been used in a variety of ways in practice to evaluate the performance of both long-only traditional managers and long-short hedge fund managers. It can also be used in investment performance attribution.

Getting in Style

Exhibit 1 presents a Monte Carlo simulation performance evaluation for a mystery manager against an S&P 500 universe for periods ended December 31, 2004. The mystery will soon be revealed. MCS creates all the portfolios that a manager could have potentially held when selecting stocks from a particular index, in this case the S&P 500.

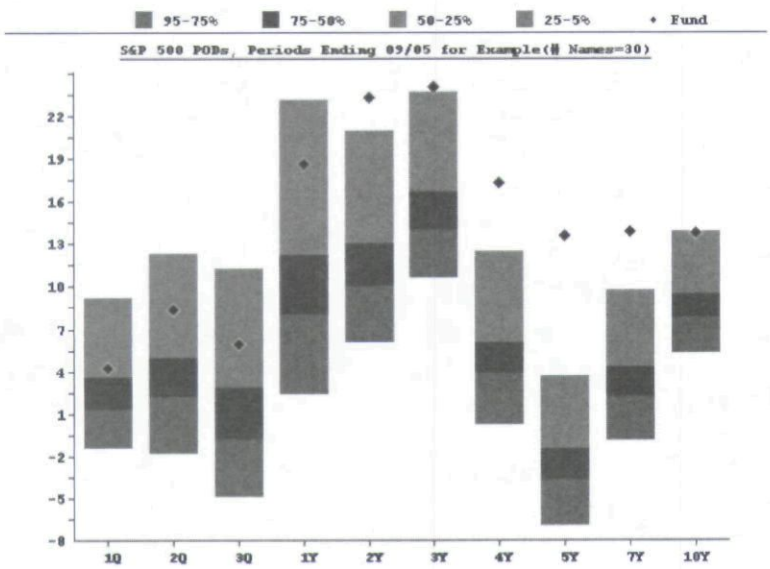
As can be seen in Exhibit 1, this manager has performed very well and is, in fact, off the charts for most of the periods of a year or longer. Should more assets be allocated to this good performer? The answer relies on unveiling the mystery.

There's a reason this manager is off the charts when compared against the S&P 500 opportunity set. She has a style that has been in favor, namely, small-cap value—the benchmark is misspecified in Exhibit 1.

MCS cannot stop the evaluator from choosing the wrong benchmark, but MCS can signal that this error has occurred. Exhibit 2 reveals that, when the manager is evaluated against the appropriate style rather than the broader index, she is just median (mediocre).

The lesson, of course, is that style matters. This may seem like an obvious lesson, but common practice suggests otherwise. Managers are frequently evaluated against

EXHIBIT 1
Rankings Against S&P 500



	1Q	2Q	3Q	1Y	2Y	3Y	4Y	5Y	7Y	10Y
Fund	4.2	8.4	5.9	18.6	23.3	24.1	17.3	13.6	13.9	13.8
5%	9.2	12.3	11.3	23.1	21.0	23.7	12.5	3.7	9.8	13.9
25%	5.5	8.0	6.3	16.2	15.8	19.2	8.4	.5	6.0	11.1
50%	3.6	5.0	2.8	12.2	13.0	16.7	6.0	-1.5	4.3	9.5
75%	1.3	2.2	-.8	8.0	10.0	14.0	3.9	-3.7	2.2	7.8
95%	-1.5	-1.8	-4.9	2.4	6.0	10.6	.2	-7.0	-.9	5.3

EXHIBIT 2
Rankings Against S&P 600 Small-Cap Value

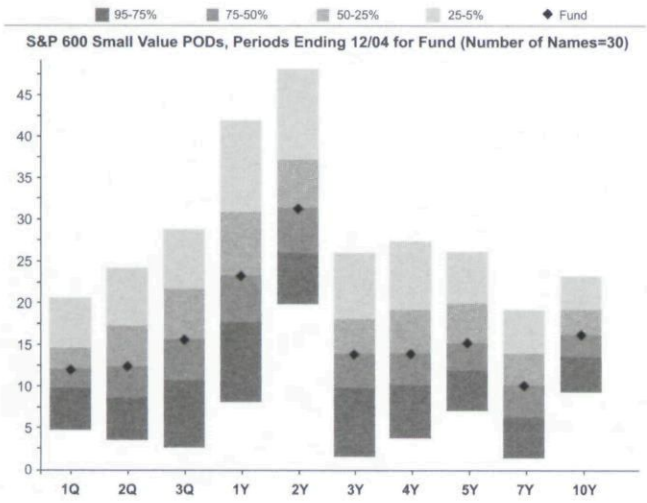
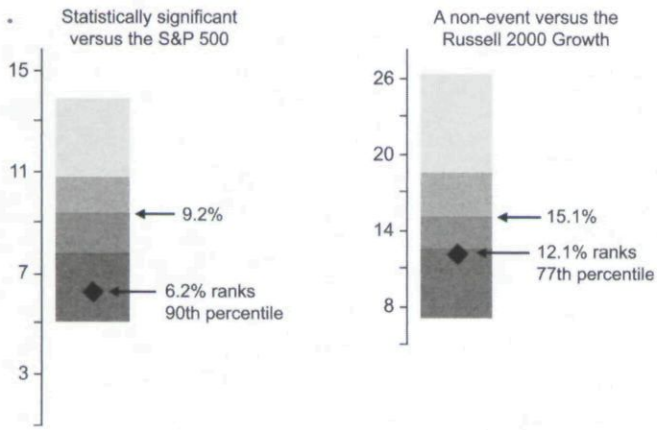


EXHIBIT 3
Using MCS to Determine Significance—Missing the Benchmark



the S&P 500, whatever their style. The fact that the dots are off the charts in Exhibit 1 demonstrates that the S&P 500 is the wrong benchmark for this manager because stocks outside this index had to be held in order to produce performance that was off the charts.

It's important not to confuse this lesson with picking an appropriate peer group. Surely a sophisticated investor will select a peer group with a name that sounds like what the manager does: large value, small growth, and the like. Yet another peer group with a similar name provided by another vendor will produce a different—and frequently substantially different—inference (or ranking), and the investor has no way of knowing which is right.

The lesson here is to pick the right benchmark and then use an unbiased MCS universe to determine the manager's degree of success or failure, thereby combining the benefits of benchmarks with those of peer groups.

Significance Now

If a manager underperformed his benchmark by 3% in the fourth quarter of 2004, would that be really bad or just sort of bad? How significant is that?

The answer depends on the volatility in the manager's style, as represented by his benchmark. Underperformance of 3% in a very conservative style would be more disappointing than the same underperformance in a very aggressive, volatile approach. But where are the lines drawn?

Exhibit 3 shows that in the quarter in question a 3% underperformance was significant versus the S&P 500, but not significant if the benchmark had been the Russell 2000 Growth Index. We use 90% confidence as a measure of significance.

Note that significance is determined in Exhibit 3 for a single quarter, demonstrating how MCS solves the waiting time problem inherent in benchmarks. Benchmarks by themselves require decades to draw similar inferences.

Wealth Scan: Attribution Analysis

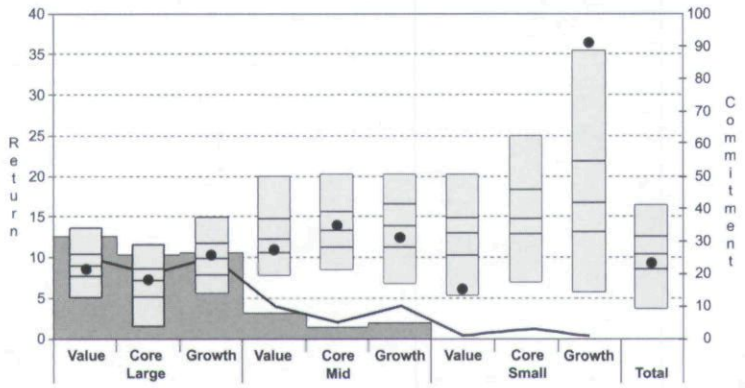
Exhibit 4 provides a quick overview of the stock market during the fourth quarter of 2004. In a nutshell, both size and orientation matter. Small and mid-sized companies outperformed large companies. Within size groups, growth outperformed value, a reversal of what occurred in the first nine months of the year.

Here's how to read the graph. The bars in the graph are MCS universes for each of nine styles as well as the total market. Note the middles of the bars, which are the medians. The median return for large value is 8.9%, for large core 7.2%, and so on. The best-performing median is small growth with a 16.7% return, and the worst is large core with a 7.2% return.

Also note the ranges of returns, representing the risk and opportunity for each style. As expected, the more volatile styles, like small growth, have a wider range.

EXHIBIT 4

Style Performance for S&P 500—Quarter Ended 12/31/2004



	Large-Cap			Mid-Cap			Small-Cap			Total
	Value	Core	Growth	Value	Core	Growth	Value	Core	Growth	
#S&P 500	8.60	7.21	10.21	10.93	13.86	12.35	6.01		36.41	9.23
S&P Commitment	31.47	25.52	26.65	7.83	3.55	4.91	0.01	0.00	0.08	
Total Mkt Return	8.90	7.20	9.90	12.40	13.30	13.90	13.00	14.70	16.70	10.44
Mkt Commitment	25.07	20.08	24.98	9.99	5.00	9.99	1.00	3.00	0.90	

Now note the shaded area at the bottom of the graph, which shows the style profile of the S&P 500. Each stock in the S&P 500 is classified into one of the nine style groups, and the aggregate dollar allocation is shown for each group as a percentage of the index's capitalization. In round numbers, the S&P 500 is 84% large/16% mid/0% small companies, representing a large-company tilt relative to a broad market that is 70%/25%/5%. The broad market style profile is shown as the solid line at the bottom of the exhibit. This tilt hurt S&P 500 performance in the fourth quarter since smaller companies were in favor.

Lastly, the dots in the bars illustrate how the style subportfolios within the S&P 500 fared against their respective style MCS universes. Note, for example, that the mid-cap value companies in the S&P 500, treated as a separate cap-weighted portfolio, performed at the third quartile of the mid-cap core MCS universe. That is, the stocks selected by the S&P committee in this style underperformed for the quarter. In fact, S&P 500 stocks generally performed at or somewhat below median for every style, except small growth, where there was a negligible 0.08% allocation.

In summary, the S&P 500's large-company orientation in the fourth quarter of 2004 hurt performance, and stock selection in the mid-cap value area also subtracted value. Yes, the S&P 500 is a managed portfolio; it's just managed by committee. The net result for the quarter is shown in the bar at the far right, where we can see that the S&P 500 has underperformed the broad market.

This analysis is sometimes called a *wealth scan* because of its similarity to a health scan. A health scan is an electronic comparison and evaluation of internal organs relative to a norm. A wealth scan compares and evaluates a portfolio's sector and style composition relative to an appropriate benchmark.

Hedge Funds

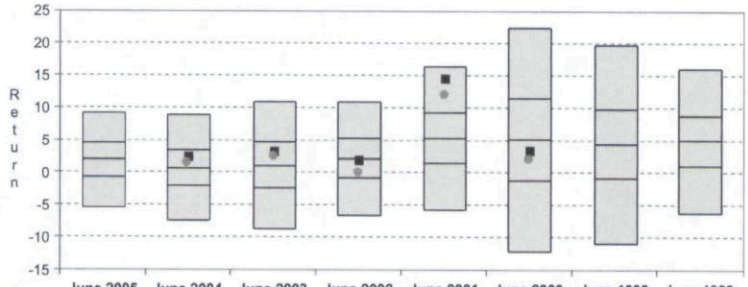
Disappointing markets breed interest in hedge funds because hedge funds promise good performance regardless of market conditions. While the original interest in hedge funds came primarily from individual investors, institutional investors began to participate in 2000, when the stock markets corrected. By 2005, hedge funds had reached the $1 trillion mark in assets and were becoming an increasingly important sector of the market based on continued forecasts for rapid growth.

Yet investors were disappointed with the performance of hedge funds in 2005. Some of this disappointment was the result of unrealistic expectations. For example, true market-neutral funds have an expected return close to that of Treasury bills, a return that is commensurate with the risk. This is the return that will be delivered if no value is added by the investment manager's decisions. With Treasury bill returns running between 1% and 2% during 2002-2005, a skillful market-neutral manager adding, for example, 2 percentage points would have returned a mere 3% or 4% in a year. Even investors looking for transportable alpha wouldn't look at a fund delivering a mere 3% or 4% annually, although pure market-neutral is just right for transporting.

Another source of disappointment was misrepresentation by hedge fund managers. Many, if not most, funds that call themselves market-neutral are not market-neutral at all; that is, they routinely make style, sector, and directional bets. As a result, investors were purchasing beta, or market exposure, for a very high price and generally getting very little, if any, alpha, or value-added. We use market-neutral here as an example for the entire hedge fund industry because of its popularity and simplicity. Investors should be buying alpha, or skill, not beta.

EXHIBIT 5

Opportunities in Market-Neutral Hedge Funds for Annual Periods Ending in June



Total Market Percentiles	June 2005 Return Rnk	June 2004 Return Rnk	June 2003 Return Rnk	June 2002 Return Rnk	June 2001 Return Rnk	June 2000 Return Rnk	June 1999 Return Rnk	June 1998 Return Rnk
5th	9.19	8.82	10.88	10.79	16.36	22.24	19.64	16.16
25th	4.54	3.47	4.76	5.18	9.24	11.43	9.71	8.65
50th	1.93	0.70	1.19	2.14	5.32	5.14	4.44	4.92
75th	-0.68	-2.07	-2.38	-0.90	1.40	-1.15	-0.83	1.19
95th	-5.33	-7.42	-8.50	-6.51	-5.72	-11.96	-10.76	-6.32
S&P 500	6.30 *18*	19.11 *1*	0.26 *57*	-17.99 *99*	-14.83 *99*	7.24 *42*	22.76 *1*	30.16 *1*

The third reason for dissatisfaction with hedge funds in 2005 was the realization that hedge funds were receiving investment dollars without being fully understood. Due diligence was non-existent or predominantly qualitative. The challenge for hedge fund investors is identifying managers who have skill. This challenge is heightened by the need to be critical of performance that appears to be too good. Investors need both a scientific and unbiased backdrop for evaluating a manager's performance and a credibility check on the manager's reported return.

MCS serves both purposes. A good MCS ranking evidences skill, but a reported return outside the realm of Monte Carlo possibilities is suspicious. No one can afford to make decisions on the basis of unquestioned suspicious performance.

Exhibit 5 shows how a scientific approach can be applied to an actual pure market-neutral manager. This manager "pair trades"; that is, for every stock that she buys long, she takes a short position in a comparable company. The manager also employs rigorous risk controls to keep the portfolio neutral in multiple dimensions: style, sector, dollar, beta, and so on. She holds 200 long positions and 200 short positions, and all companies are in the S&P 500.

The box in the middle of the return legend shows the expected annual returns for a pure market-neutral strategy, which are fairly low. The manager represented in Exhibit 5 has generally added value, even earning 15% (or 13% net, as represented by the lower circle) in the year ended June 30, 2001. In most years, however, her returns have been in the 2% to 5% range. One probably wouldn't hire this manager, although the decision might change if the alternatives, such as traditional long-only investing, were unattractive.

The boxes at the bottom of the exhibit highlight periods of good and poor performance for the S&P 500. The boxes on the left and right were generally bad periods for interest in hedge fund investments because the traditional markets were performing well, while the middle years were good for hedge funds because traditional markets performed poorly. Through June 2005, stock markets were good for traditional investing and bad for hedge funds, which resulted in a slowdown in hedge fund investing.

Looking at market-neutral hedge funds in general, there are about as many types of market-neutral strategies 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. There are great differences among the various flavors of market-neutral, and the performance results reflect these differences.

To see this point, consider the three market-neutral substrategies shown as examples in Exhibit 6. All three are dollar-neutral as well as beta-neutral, but they vary in their style neutrality. The first is long growth and short value, the second style-neutral, and the last long value and short growth. A traditional market-neutral peer group is included for comparison.

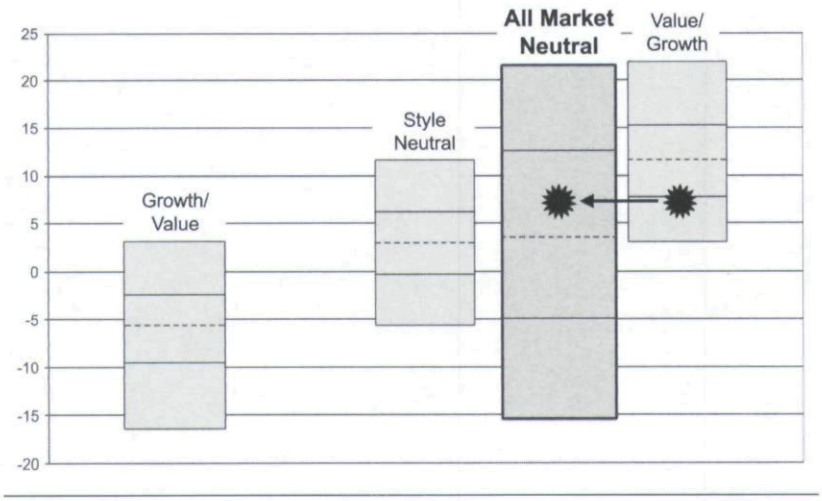
The MCS peer groups demonstrate the dramatic differences in the opportunities that were available to these three substrategies over the five years ended June 30, 2004, even though all three are properly described as market-neutral.

As can be seen in the bar representing all market-neutral managers, a candidate for inclusion in a portfolio has delivered above-median performance when contrasted with this traditional non-customized peer group, making him an apparently viable candidate. Yet when MCS technology is applied to the same performance, the candidate ranks in the bottom quartile of the 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.

EXHIBIT 6

Portfolio Opportunity Distributions for Select Market-Neutral Strategies—5 Years Ending 6/30/04



SUMMARY

Monte Carlo simulation performance evaluations offer an alternative in keeping with the ongoing technological revolution that has characterized our entrance into the 21st century. This innovation has only recently become possible—the computing power necessary to run MCS simply did not exist as little as 15 years ago. The introduction of MCS coincides with a growing recognition of the inadequacies of the approaches we have used for the last three decades.

Although this alternative is not perfect, it is worth consideration and exploration. To be sure, peer groups are used by just about everybody, and there's comfort in the commonplace. Yet common practice changes over time—impermanence is eternal.

Monte Carlo simulations take much of the guesswork out of performance evaluation, by comparing what actually happened with what could have happened. In addition, MCS rankings are available within days after the end of a performance measurement period, not the many weeks it takes to obtain rankings using traditional peer groups.

It's been said that anything not worth doing is not worth doing well. Investment managers care, though, because performance rankings are their report cards. A rate of return doesn't win or lose business. The evaluation of that return is what matters. Ranking well against one peer group and poorly against another, as frequently occurs, means a manager has both passed and failed, a predicament that can be settled by MCS.

Investors who believe that they and their consultants can find skillful managers care. Because the majority of assets in this country and abroad are actively invested, most investors care. Performance evaluation is indeed worth doing well.

ENDNOTE

Many thanks to Gale Morgan Adams for her remarkable editing.

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?" *The 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.

Burns, Patrick. "Performance Measurement via Random Portfolios." Newsletter of Professional Risk Managers International Association, December 2004.

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

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

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

Haltiner, James, and Ronald Surz. "Using Virtual Reality to Assess Performance Under Uncertainty." *The Journal of Investing*, Vol. 13, No. 4 (Winter 2004).

Kat, Harry M. "10 Things That Investors Should Know About Hedge Funds." *The Journal of Wealth Management*, Vol. 5, No. 4 (Spring 2003), pp. 72-81.

Lukomnik, Jon. "Doing Diligence." *Plan Sponsor*, July 2003a.

——. "Hedge Fund Indices: Apply With Caution." *Plan Sponsor*, January 2003b.

——. "Hedge Funds." *Plan Sponsor*, April 2003c.

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.

——. "The Arithmetic of Active Management." *Financial Analysts Journal*, Vol. 47, No. 1 (January/February 1991), pp. 7-9.

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

Surz, Ronald J. "Customized Performance Standards." *The Institutional Investor Focus on Investment Management*, 1989, pp. 223-237.

——. "Cyberclone Peer Groups." *The Journal of Investing*, Winter 1998, pp. 63-67.

——. "Portfolio Opportunity Distributions: An Innovation in Performance Evaluation." *The Journal of Investing*, Summer 1994.

——. "Portfolio Opportunity Distributions: A Solution to the Problems with Benchmarks and Peer Groups." *Journal of Performance Measurement*, Winter 1996, pp. 24-30.

——. "Testing The Hypothesis 'Hedge Fund Performance is Good'." *The Journal of Wealth Management*, Vol. 7, No. 4 (Spring 2005), pp. 78-83.