Warning ! : Peer Groups Are Hazardous to Our Wealth

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

January is drawing to an end as I write this, and most financial advisors are scheduling their end of year performance reviews. In a few weeks peer groups will be out and we'll know who has succeeded or failed. Or will we? The investment industry has been using peer groups for so long that no one thinks to question them, except for the occasional discussion about survivor bias, which seems to be the only bias that people know and understand. In actuality, peer groups suffer from a collection of biases, only one of which is survivor bias, 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 of the peer groups are for the same management mandate, such as large cap growth. **We really don't know who has succeeded or failed because success against one peer group can easily be failure against another comparable peer group, so a manager can be both a success and a failure.** It just doesn't make sense. In this article I document the problems with peer groups and offer a solution. In a nutshell, traditional peer groups are hazardous to our wealth because they are more likely to mislead than to inform, so we make the wrong decisions. Making hire/fire decisions on the basis of peer group rankings is fraught with peril. Fortunately there is a better way.

The Problems with Peer Groups

Everyone who has earned the CFA (Chartered Financial Analyst) designation has learned the problems with peer groups: they are loaded with biases.

Ankrim[1998], Bailey[1992], and Bleiberg[1986] have identified these biases. These authors 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 some pre-specified pigeonhole, such as growth or value. It is now commonly understood that most managers employ a blend of styles, so that pigeonhole classifications misrepresent the manager's actual style as well as that of peers.

- *Composition* biases result both from concentrations of certain fund types in databases, such as bank commingled funds, and from small sample size. International managers and socially responsible managers cannot be properly evaluated because there are no databases of adequate size.

- *Survivorship* bias causes performance results to be overstated because accounts that have been terminated, which may have underperformed, are no longer in the database. This is the most documented and best understood source of peer group bias. For example, an unsuccessful management product that was terminated in the past is excluded from current peer groups. This screening out of losers results in an overstatement of past performance. A good illustration of how survivor bias can skew things is the "marathon analogy", which asks: If only 100 runners out of a 1,000-contestant marathon actually finish, is the 100th the last? Or in the top ten percent?

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.

Even if a bias-free peer group existed, it still wouldn't qualify as a valid investment performance benchmark as defined by the criteria established by Richards & Tierney, a Chicago-based investment consulting firm specializing in custom benchmarks, and supported by the CFA Institute (formerly AIMR). These criteria can be summarized as follows:

1. Unambiguous: Names and weights of securities are clearly stated.
2. Investable: Investors have the option to forego active management in lieu of a passive alternative.
3. Measurable: The benchmark's return can be calculated on a reasonably regular basis.
4. Appropriate: The benchmark accurately represents the manager's approach.
5. Reflective of current investment opinions: The manager has current knowledge of the securities constituting the benchmark.
6. Specified in advance: The composition of the benchmark is agreed to and constructed prior to the start of all evaluation periods.

Traditional peer groups meet none of these criteria, except possibly number 3.

And it Gets Even Worse

A move from the traditional world to the non-traditional reveals that the problems with peer group get even worse 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—the list goes on, 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, substantiating the fact that these funds are different from one another. This 0.23

is the "r" of the "r-squared", so only 5% (.23 squared) of the variance in one hedge fund is 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 based on approach rather than skill. Because they don't take into account the unique specifications of each fund, the peer groups 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. For an extensive discussion of an advanced approach for evaluating hedge fund performance see Surz[2005].

A Solution

The solution to these problems is actually quite simple. Performance evaluation ought to 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 good. In other words, the hypothesis test compares what actually happened to what could have happened.

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 not just for amusement but as a practical solution to the problem with traditional peer groups.

Using the dartboard analogy, each individual manager should have his/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 several characteristics as follows: 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. There is no confusion about ranking well in one peer group but poorly in another, since there is only one right peer group. You rank how you rank, period. There are two questions involved in this process: (1) What does this manager do (style, etc.)? and (2) Does (s)he do it well? The first question addresses the form of the investment and the second identifies the substance, or skill.

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 respective passive alternatives. What does matter is the degree of success or failure you experience relative to your benchmark. The

benchmark in this context is the answer to the first question: What does this manager do? The ranking within his/her customized opportunity set answers the question "Does (s)he do it well?"

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], Haltiner and Surz [2004], Burns [2004] and Bridgeland [2001]). While the application of MCS to traditional portfolios has not caught on (see Chernoff [2003] and Picerno[2003], it doesn't mean that the idea is faulty. Modern Portfolio Theory (MPT) took 30 years to catch on. Importantly, the SEC recently approved the use of MCS in client reporting.

Further improving the likelihood of acceptance, MCS technology has been extended to hedge funds, where peer groups clearly don't work, so there are no inherent barriers to comprehension and adoption. MCS addresses the uniqueness challenge of evaluating both long-only and hedge fund performance by creating at random all of the possible portfolios that a manager could have conceivably held following his/her unique investment process, thereby applying the scientific principles of modern statistics to the problem of performance evaluation. An MCS universe is easily produced in risk-adjusted as well as fee-adjusted form, a must for hedge fund investing. Regardless of the form, the answer to the question "What funds are in an MCS peer group?" is "All of them that matter."

Getting Real

The most common criticism of MCS is that it isn't "real": "My clients want to know how other managers have done." Recognize that performance evaluation

is a hypothesis test, regardless of whether you use MCS or traditional peer groups. We want to know if performance is good, and this can only be determined relative to something, a sample. So the choice between MCS and peer groups is essentially a choice between sampling approaches. What sample should we use to test the hypothesis "performance is good"? When we use traditional peer groups the sample is a group of portfolios that is presumably managed in a manner similar to the portfolio we want to evaluate, so the hypothesis is tested relative to the stock picks of other similar professionals. This makes sense, except someone has to define "similar" and go out and collect data on these similar funds. 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" is 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 sources of the biases described at the beginning of this article. Don't like your ranking? 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.

But how do we distinguish a good sample from a bad one? Let's say that you are a peer group provider and you want to prove that your peer group is representative. I personally was challenged with this proof statement some years ago when I worked for a peer group provider, the largest in the country at the time. Here's what we did. After much discussion and deliberation we decided that our peer groups should represent all of the portfolios that the investing public could hold, and that we could create this opportunity set by drawing portfolios at random. This in fact was the beginning of MCS universes for me. We used MCS as the proof statement that our peer groups were representative.

In other words, our peer groups looked like the MCS universe we created a couple of months earlier, so our peer groups passed the representativeness test we designed. The only problem was that we couldn't tell the outside world because competitors would see that all they needed was a simulator and securities data, not our thousands of funds. However, we did use MCS to calibrate our sample, so we signed off and released peer groups when they became reasonably close to what we knew to be the reality generated by MCS. MCS was the standard we used to validate our samples, knowing full well that we could have come out with the reality many weeks earlier, but that doing so would actually undermine our business. This little secret is being told for the first time here. To do this right, we had to develop rigorous sampling techniques, which speak to the next criticism.

Another "real world" concern is that MCS will create some random portfolios that no one in their 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". In hypothesis testing there is a measure called "the cost of perfect information". The idea is that samples are imperfect, since the next sample may give substantially different results. The ultimate is to not sample at all, but rather to identify all of the possibilities, which is perfect information. The acquisition of perfect information comes at a cost that can frequently be estimated. Properly constructed MCS universes are close to this concept of perfect and come at a cost that is generally far less than the seriously flawed samples generated by traditional peer group providers.

The final criticism of MCS is that the evaluator wants to know how other similar entities have done, for example other foundations or other union plans. Recognize that this desire is designed to evaluate the competition more than it is to evaluate the investment manager. Let's say for example that a handful of large value managers have penetrated the union fund marketplace, so they have the lion's share of this market. Then an evaluation of one of these managers against a union fund large value peer group is essentially a comparison against a very small sample that only tells you that your competitors have succumbed to the same sales efforts you have. It tells you very little about the skill of the manager you're evaluating. Your manager could perform very well and still be below the median of this elite club, or vice versa. Since you are not limited to this club when you allocate assets, your hire/fire decisions should address all of the opportunities available, not just choices made by a select handful of managers with good marketing. For competitive reasons you may still want to look at a peer group of similar investors, but you certainly don't want to evaluate the manager 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.

View the choice of universe, MCS or peer group, as a sampling decision for testing the hypothesis "performance is good." The better the sample, the better the inference.

Summary

MCS performance evaluations offer a new standard appropriate to the ongoing technological revolution that characterizes our entrance into the twenty-first century. This innovation is only recently possible due to the fact that the

computing power necessary to run MCS simply did not exist as little as 15 years ago. The introduction of MCS coincides with growing recognition of the inadequacies of the old approaches we've used for the last three decades. Sure, peer groups are used by just about everybody, and there's comfort in being part of the herd. But smoking was also common practice not too long ago, demonstrating that popularity is not synonymous with good judgment.

Monte Carlo simulations take the guesswork out of performance evaluation by comparing what actually happened to what could have happened. And as an extra added bonus MCS ranks are available within days after the end of a performance measurement period, versus many weeks with traditional approaches. Once you get it you get it, and it's available at www.ppca-inc.com/hedgepods.html.

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.

Burns, Patrick. "Performance Measurement via Random Portfolios." Newsletter for Professional Risk Managers' International Association (PRMIA), December, 2004

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.

Haltiner, James and Surz, Ronald. "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 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. "Testing The Hypothesis "Hedge Fund Performance is Good"." *The Journal of Wealth Management, Vol 7, No. 4*, Spring 2005, pp 78-83.
_____. "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.