PORTFOLIO OPPORTUNITY DISTRIBUTIONS: AN INNOVATION IN PERFORMANCE EVALUATION

RONALD J. SURZ

is president of the Performance Presentation Consulting Alliance in Wheaton, Illinois, which audits manager compliance with performance reporting standards. He is also a vice president of Glenwood Trust Company, which specializes in oversight management of alternative strategy investment programs, and a member of the Alaska Retirement System Investment Advisory Council. Mr. Surz holds an M.S. in applied mathematics from the University of Illinois and an M.B.A. in finance from the University of Chicago.

To make sound decisions about the future, it helps to understand the past. In investment management, an understanding of the past encompasses knowledge of success and failure as well as the reasons behind either outcome. Before these judgments can be deemed valid, however, a meaningful reference point must be established. Achievement should always be measured relative to the opportunities available, and these opportunities are generally unique to the investment approach being evaluated.

PERFORMANCE EVALUATION

Performance evaluation is an assessment of whether investment management has added value. In this assessment, the evaluator should apply the science of inferential decision-making, which asks "What are the chances that the observed result is a significant occurrence?" The two evaluation practices currently in use — benchmark and peer group comparisons — fall short of answering this critical question, although they have some desirable features that should be maintained.

The benchmark approach revolves around a viable, low-cost alternative to active management. In its purest form, the benchmark is tailored to the manager's process and philosophy (or style), and the resulting benchmark is called a "normal portfolio." Purveyors of normal portfolios vehemently assert that their approach is the only correct way to evaluate a manager because it is the only way to capture the "essence" of the investment. Peer group comparisons, they argue, are contaminated by innumerable biases, not the least of which is the inability to identify a given manager's peers. After all, each manager is unique and therefore without peers.

Purveyors of peer group comparisons are quick to rebut that the uniqueness argument doesn't hold up when it comes to a manager search. Manager searches create feeding frenzies that attract hundreds of managers who will readily espouse the style at hand. Furthermore, these purveyors say, peer group comparisons benefit evaluators by providing a ranking that serves as a "report card," graded on a scale composed of the other participants in the peer group. The normal portfolio's counterpart to this grade, the "T-value," is an indicator of statistical significance that takes decades to produce reliable results.

AN INNOVATION

The shortcomings of normal portfolios and peer groups have been surmounted through an inno-

vation that combines the better aspects of both while eliminating their undesirable properties. The innovation expands the normal portfolio concept to encompass performance rankings heretofore only available through peer groups. The biases inherent in peer group comparisons have been eliminated through simulations of the true opportunity sets available to investment practitioners. The inability of a normal portfolio to produce a reliable grade in a timely fashion has been remedied by applying the laws of statistical inference. Exhibit 1 contrasts the current approaches to the innovation, which is called "Portfolio Opportunity Distribution," or POD.

The theory behind the innovation is based on the fact that, in common practice, the statistician constantly compares his results to those expected purely by chance. By applying this concept to performance evaluation, POD generates thousands of simulated, unmanaged portfolios at random, drawn from the manager's normal universe of stocks, using the manager's portfolio construction rules. This assures that the resulting opportunity distribution fairly reflects the manager's decision processes. For example, the average commitment to a stock, across all simulated portfolios, is the stock's weighting in the manager's normal universe.

The resulting distribution provides a grading system that shows the full range of results that could have been achieved by the manager while eliminating the biases inherent in peer group universes. These PODs have the normal portfolio as their median, with fractiles around the median representing degrees of success or failure. A ranking in the top decile of a POD universe gives the statistician 90% confidence that the return was not merely random, but a significant indication of success. Similarly, a ranking in the bottom decile is a significant indicator of failure. Exhibit 2 shows two such POD universes generated for 1993.

Evaluation against a POD universe tells the investor whether the observed performance result was good or bad relative to the unique opportunities available. As discussed earlier, no benchmark or peer group universe can provide this insight. Further investigations — such as attribution analyses and manager interviews — into the reasons for success or failure can reveal the manager's level of skill.

When used, POD universes can replace the need for peer groups and normal portfolios. Comparisons against POD universes are unbiased, and POD universes are available immediately. Normal portfolios are merely the medians of POD universes, with universe fractiles representing degrees of success or failure. Furthermore, the science can be extended beyond U.S. borders so superior evaluations can be achieved for international investment programs.

EXHIBIT 1
CONTRASTS OF PEER GROUPS AND NORMAL PORTFOLIOS TO PORTFOLIO OPPORTUNITY DISTRIBUTIONS (POD)

	Strengths	Weaknesses
Peer Group	Report Card	Biased
Normal Portfolio	Captures Essence	Evaluation Takes Decades
POD	Report Card Unbiased Captures Essence Immediate Evaluation	

REPLACING PEER GROUPS

A key feature of POD is its ability to replicate peer groups while eliminating biases. When applied to a large data base of U.S. securities, POD has been proven to consistently replicate the U.S. equity fund universes of large pension consulting firms over periods of two years or less (see Surz [1989]). For longer periods, POD medians are consistently lower than those of managed universes, because managed universes suffer from survivorship bias, as has been well-documented (see Raftopoulos [1987]). Survivorship bias results from the fact that only managers who have not been fired can be measured

EXHIBIT 2
SAMPLE 1993 POD UNIVERSES

Fractiles		Universe Returns	
Percentile Ranking	Interpretation	Generalist	Growth Stock Specialist
5	Significant Success	24.9%	19.1%
25	Moderate Success	17.4	14.2
50	Expected	12.2	10.1
75	Moderate Failure	7.8	5.9
95	Significant Failure	1.9	−0.1

EXHIBIT 3
SURVIVORSHIP BIAS
POD MEDIANS VERSUS MEDIANS FROM A LARGE MANAGED UNIVERSE (PERIODS ENDING MARCH 31, 1993)

	1 Year	3 Years	5 Years
POD Medians	15.3	13.6	14.6
Managed Universe Medians	15.3	14.7	15.8

in the current time period and through the cumulative time periods ending with the current. Consequently, the historical results of managers who were fired are eliminated, thus creating upwardly biased historical universe returns. An analogy can be drawn with a marathon race. If 1,000 runners start a marathon, but only 100 finish, is the 100th finisher in last place or in the top 10%?

Importantly, a benefit of replicating a peer group universe with POD technology is removal of this survivorship bias so a fair evaluation can be performed. Also, since POD universes are manufactured from security data bases, they can be created very quickly — long before portfolio data are collected for peer group universes. For extensive documentation of the replicating history of POD, see Surz [1989].

An example of survivorship bias is presented in Exhibit 3, which contrasts POD medians to managed universe medians. As can be seen, the one-year medians are identical, implying that neither clientele nor survivorship bias is an issue. However, over longer periods the POD medians are lower, revealing survivorship bias in the managed universe. This pattern of short-run matches but lower POD medians over longer periods is always the case whenever such comparisons are made, thus proving that survivorship bias is a material reality in managed universes. Clientele biases in managed universes, such as overrepresentation of bank managed funds, are occasionally manifested in these comparisons. Survivorship bias is always present in a universe of managed portfolios.

Furthermore, *any* peer group universe can be replicated through POD technology and a calibration process. For example, the median of an international peer group can be matched by appropriately weighting countries, especially Japan. Then the other fractiles can be replicated by properly matching characteristics such as capitalization. Once the appropriate calibration has been set using historical distributions, POD reliably predicts the results of future peer group universes well in advance of their release. In other words, the right number of draws from the appropriate "urns" provides reliable universes for evaluating performance.

REPLACING NORMALS

POD also improves upon normal portfolios by extending the concept to a ranking backdrop for evaluating success or failure. Ideally, the evaluator would like to enjoy this benefit without going through the laborious and expensive process of developing a normal portfolio. This can be achieved by constructing certain "standardized normals" used in combination to estimate the manager's "essence." For a given security universe, nine standardized normals are formed from independent security groupings based on size and orientation, as shown in Exhibit 4.

To construct standardized normals, we first break a large data base of stocks into size groups based on market capitalization, calculated by multiplying shares outstanding by price per share. Beginning with the largest-capitalization company, we add companies until 60% of the entire capitalization of the security data base is covered. This group of stocks is then categorized as "Large-Cap" (capitalization). For the U.S., this group currently comprises 200 stocks, all with capitalizations in excess of $5 billion. The next size group represents the next 35% of market capitalization and is called "Mid-Cap." Finally, the bottom 5% is called "Small-Cap."

Then within each size group, a further breakout is made on the basis of orientation. Value, core, and growth stock groupings within each size category are defined by establishing an aggressiveness measure. Aggressiveness is a proprietary measure that combines dividend yield and price/earnings ratio. The top 40% (by count) of stocks in aggressiveness

EXHIBIT 4
STANDARDIZED NORMALS

	Large	Middle	Small
Value			
Core			
Growth			

EXHIBIT 5
PORTFOLIO COMPOSITION AND PERFORMANCE ATTRIBUTION USING STANDARDIZED NORMALS CALENDAR 1992

		Portfolio		Market				
		% of Market Value A	Rate of Return B	% of Market Value C	Rate of Return D	Selection Stock	Selection Style	Total
Large	Value	21.5	7.4	26.3	8.5	–0.2	–0.0	–0.2
	Core	0.0	0.0	12.8	2.5	0.0	0.7	0.7
	Growth	0.0	0.0	20.9	–1.8	0.0	2.0	2.0
Mid	Value	73.4	23.6	16.3	20.4	2.3	7.3	9.6
	Core	0.0	0.0	7.2	18.5	0.0	–0.8	–0.8
	Growth	0.0	0.0	11.5	8.4	0.0	–0.1	–0.1
Small	Value	5.1	25.0	2.6	29.3	–0.2	0.5	0.3
	Core	0.0	0.0	1.3	4.7	0.0	0.0	0.0
	Growth	0.0	0.0	1.1	0.6	0.0	0.1	0.1
		100.0	20.2	100.0	8.6	1.9	9.7	11.6

Stock = A × (B – D)/100
Style = (A – C) × (D – Total Index Return)/100

are designated as "Growth," while the bottom 40% are called "Value," with the 20% in the middle falling into "Core."

The object of forming these nine standardized normals is to create building blocks that can be used to specify a manager's style. Any given manager is described as a weighted combination of standardized normals, where the weights sum to 100%. The weights can be derived by reviewing historical portfolio composition, or through regression analyses such as those introduced by Sharpe [1992]. An example of historical portfolio composition is shown in Exhibit 5. As can be seen, the breakout into standardized normals can also be used for performance attribution. In the example, the manager is predominantly a mid-cap value manager, with secondary allocations to large-cap and small-cap value. Because value was in favor for the period evaluated, this manager did very well, but most of the outperformance was due to style rather than stock selection. Evaluation against a POD universe weighted "21.5% large-value/73.4% mid-value/5.1% small-value" places the manager in the thirty-seventh percentile, which, while good, is not nearly as impressive as the seventh percentile ranking achieved against the broad universe (rankings derived from POD universes not shown in Exhibit 5).

To meet these building block criteria, the standardized normals must be statistically independent, mutually exclusive, and exhaustive. They are, by construction, mutually exclusive and exhaustive. A certain degree of statistical independence has been achieved, as shown in Exhibit 6.

The excess returns in Exhibit 6 are calculated by subtracting out the corresponding size benchmark of each normal. For example, a large company benchmark is constructed as the capitalization weighted return on the top 60% of stocks. This return is then subtracted from the large-value, large-core, and large-growth standardized normals. The table covers the eight-year period since the style classifications were begun.

With standardized normals, a style proxy is built, and a POD universe is constructed with the appropriate weights. In other words, customized universes are constructed as weighted combinations of standardized normal POD universes. In this way, the manager's success or failure is assessed relative to his unique opportunity distribution.

GLOBAL PERFORMANCE EVALUATION

A natural extension of POD technology is its application to global performance evaluation. In this context, global standardized normals are developed as the building blocks for customized universes.

EXHIBIT 6
CORRELATIONS OF EXCESS RETURNS FROM JANUARY 1, 1986-DECEMBER 31, 1993

		Large-Cap			Mid-Cap			Small-Cap	
		Value	Core	Growth	Value	Core	Growth	Value	Core
Small	Growth	−0.142	−0.112	0.231	−0.392	−0.112	0.467	−0.775	−0.168
	Core	−0.440	−0.024	0.527	−0.552	0.089	0.557	−0.492	
	Value	0.404	0.118	−0.539	0.706	0.045	−0.777		
Mid	Growth	−0.549	−0.055	0.670	−0.920	−0.033			
	Core	−0.332	0.426	0.135	−0.362				
	Value	0.640	−0.115	−0.675					
Large	Growth	−0.863	0.005						
	Core	−0.510							

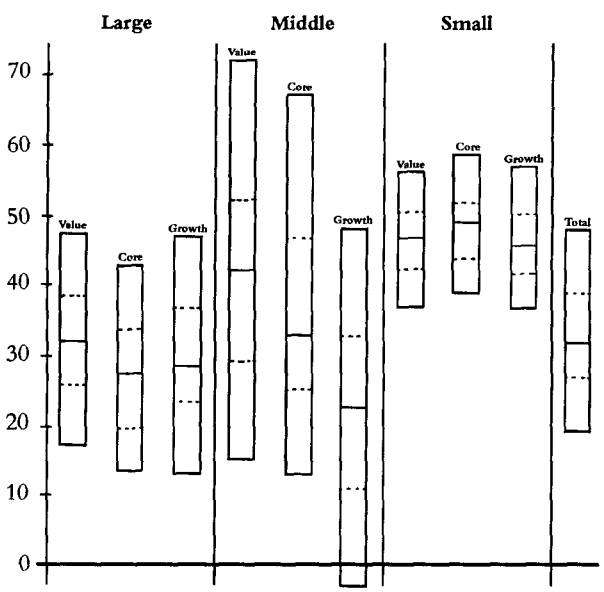
Hundreds of such normals have been constructed by forming the nine style groups described in Exhibit 4 for various countries and regions of the world, using the size and aggressiveness classification rules described in conjunction with Exhibit 4. In other words, a security data base is assembled for each region of the world and then broken into standardized normal stock groupings. Some of these regions are listed in Exhibit 7.

The same rules of mutual exclusivity and exhaustiveness apply, and a satisfactory degree of independence is achieved. Weighted combinations of standardized normals within a region determine a customized POD universe that is used to evaluate performance within that region. Examples of such universes for non-U.S. investing are shown in Exhibit 8, which graphically depicts pertinent fractiles of non-U.S. style universes in a floating bar format. The fractiles graphed are the fifth, twenty-fifth, fiftieth, seventy-fifth, and ninety-fifth percentiles; the returns associated with these fractiles are listed at the bottom of Exhibit 8. Non-U.S. universes are constructed from the Compustat Global Vantage data base, which contains more than 4,000 companies from over thirty countries with total market capitalization of $6.5 trillion. The U.S. practice in non-

EXHIBIT 7
SAMPLE REGIONS FOR STANDARDIZED NORMALS

Global
Non-U.S.
Non-Japanese
Non-U.S. and Non-Japanese
Far East
Far East ex Japan
Europe
Emerging Markets
Ten Individual Countries

EXHIBIT 8
NON-U.S. POD UNIVERSES FOR THE YEAR ENDING DECEMBER 31, 1993



	Large			Middle			Small			
	Value	Core	Growth	Value	Core	Growth	Value	Core	Growth	Total
5	47.7	43.1	47.2	72.2	67.4	48.5	56.3	59.2	57.5	48.3
25	38.4	33.9	36.8	52.5	47.2	33.1	51.5	52.4	50.4	39.0
50	32.4	27.6	28.6	42.3	32.9	22.7	47.2	49.1	45.9	32.0
75	25.9	19.7	23.4	29.2	25.1	11.0	42.5	44.0	41.6	26.6
95	17.0	13.5	13.3	14.8	12.9	−3.5	36.9	38.6	36.7	19.4

U.S. investing has generally been conservative and therefore most resembles the large-value and large-core approaches. Accordingly, these universes closely resemble the universes of international peer groups prepared by consulting firms.

WHAT NEXT?

The two current approaches to evaluation — comparison to a normal portfolio and peer group ranking — have become entrenched through years of usage. The POD approach combines the best attributes of both and eliminates the worst. Only time will tell if this innovation will be accepted.

ENDNOTE

The author is grateful to Laurence B. Siegel of Ibbotson Associates for his invaluable assistance in accurately describing the benefits of the POD technology. In other words, Larry kept me honest.

REFERENCES

Raftopoulos, Dimitri. "The Universe Exposed." *Financial Analysts Journal*, March/April 1987, pp. 12-14.

Sharpe, William F. "Asset Allocation: Management Style and Performance Measurement." *Journal of Portfolio Management*, Winter 1992, pp. 7-19.

Surz, Ronald J. "Customized Performance Standards." In Frank Fabozzi, ed., *The Institutional Investor Focus on Investment Management*. Ballinger Publishing Co., 1989, pp. 223-237.