BY TIMOTHY P. RYAN, CIPM

Time Matters in Performance Analysis

High-frequency performance reporting, if left unchecked, can facilitate well-intentioned but ultimately inaccurate performance analysis, especially if investment performance professionals stop after calculating and reporting performance and do not go on to analyzing investment results. In this article, I will present two practical examples, discuss the analytical and statistical concepts embedded within each, and explain the appropriate treatment.

EXAMPLE 1

Can the information contained in peer-relative percentile rankings done at regular reporting intervals (quarterly, for instance) be aggregated to provide insight into performance through the overarching period? For example, if a fund ranks in the 25th, 25th, and 50th percentiles for the first three quarters of the year, is it reasonable to conclude that its year-to-date performance ranks in the top 33 percent of comparable funds? After all, 33 percent is the average of the first three observations.

ANALYTICAL AND STATISTICAL CONCEPTS: EXAMPLE 1

The first concept packed inside this example is return compounding. Compound returns are affected by the order, direction, and magnitude of subperiod returns.[1] Averages mask the factors that contribute to a fund's compound return over multiple measurement periods.

The second concept contained in this example is that of a collective whole. Essentially, each period is an entity unto itself and representative only of itself. In this case, each quarter is informative only of itself. Similarly, the year-to-date period through the third quarter would be another episode, albeit larger, but distinct and no less singular. As such, the year-to-date period through the third quarter can be thought of as one singular collective whole. This singular collective whole reflects individual fund returns having already been compounded into each fund's singular year-to-date through the third-quarter return. It is this singular compounded return number for each fund that peer group providers use in their multiperiod relative ranking percentiles.

Relatedly, the third concept in this example is that of identity. Reichmann covers this point in a worked example and remarks that "the individual percentages are calculated to different bases and therefore have different identities" (p. 82).[2] Taking an average of an average fosters confusion and produces incorrect results.

The fourth concept—or bundle of concepts—represented in this example deals with ranges, distributions, and the limitations of averages. In Example 1, the range of any quarter or of the year is unknown; whether the return distributions in any quarter or the year-to-date period are normal is also unknown. If the distribution is not normal, then knowing the degree of kurtosis and skewness is all the more important. This issue works its way into return compounding and the makeup and structure of the peer universe being analyzed. In return compounding, the range and distribution are important. As I have noted elsewhere, "In return space, it takes an exceptionally large positive return to regain initial ground lost, +400 percent return after losing 80 percent initially, versus a comparably small negative return to lose all ground initially gained, <44.44 percent> after initially gained 80 percent.…" (p. 45).[3] The makeup and structure of the peer universe are also important. For instance, if the peer universe has a constituent count over 100, the returns are low, and the range of returns is very narrow (consider, for example, the current state of money market funds in this low-to-zero interest rate environment), then the difference between top- and bottom-quartile performance may be fractions of a basis point.

The fifth concept is high-frequency performance monitoring. This concept may be a larger issue in manager continuation policy decisions, but it also begins to infiltrate this analysis. Dimson and Jackson explore continuous, monthly, quarterly, and yearly performance monitoring. Their findings "illustrate how monitoring frequency can affect the distribution of observed results" (p. 44).[4] Depending on the intended uses of the output, less may actually be more—that is, less data may be more useful. Similarly, diBartolomeo writes about the low-information and high-noise content of daily data.[5]

Clearly, the concepts enmeshed in this example are overlapping issues rather than isolated, individual aspects. These concepts, too, are part of the collective whole. So, to directly address the question posed in Example 1, the answer is no: It is neither reasonable nor accurate to suppose that a fund's percentile ranking over an extended period can be deduced from its subperiod ranks. Now, let's cover these concepts in Table 1's worked example.

The top panel of Table 1 presents the quarterly and the year-to-date returns for a hypothetical four-fund universe. The distribution of returns is fairly normal and uniform. Notice that although the average return of the universe for each quarter is zero, the average return for the year is −11 percent. Notice further that the (incorrect) average-of-average year-to-date return is −3 percent, which confirms Reichmann's statement. Notice also that the standard deviation of returns for the year, 0.49, is greater than any standard deviation for the component quarters.

The bottom panel of Table 1 presents the peer rankings and percentile rankings for this four-fund universe. The top portion of this panel presents the actual and correct calculations for these values. The bottom portion presents the incorrect average-of-average versions of these numbers. On the left, the average-of-average value uses the results in individual quarters. On the right, the average-of-average uses the year-to-date values, which at least reflect some, but not full, return compounding. As this bottom panel shows, the average-of-average approach fundamentally does not provide correct numbers. As we can see, the error is larger using quarterly numbers than year-to-date numbers. Wrong, however, is wrong, and the average-of-averages should not be used at all. I now cover an extension of the Example 1 scenario in Example 2.

EXAMPLE 2

If my fund is in the bottom quartile of performers for a given rolling period—for instance, three years—when can I reasonably expect my fund to return to the top half for a rolling three-year period? If the fund experienced noticeable underperformance in a subperiod, answering the question, "What happens once that bad period rolls off?" provides a good analytical double check.

ANALYTICAL AND STATISTICAL CONCEPTS: EXAMPLE 2

Each of the five analytical and statistical concepts noted earlier also apply here. So, let's first talk a little bit about peer universes. They have well-known limitations, such as survivorship bias, among many others. If the system is an "open" one, the portfolios newly migrating to a given universe may have their histories of returns backfilled, which could change newly run historical rankings for this universe. Net-of-fee returns may also experience changing histories if new expense ratios, including any fee waiver or performance fees, are retroactively applied. Such changes should be disclosed in the resulting presentation.

There are many ways to address the question about reasonable expectations posed in Example 2. The approach with the fewest assumptions would be to look backward repeatedly across a series of multiple months (up to 36 months) to find the earliest rolling period within the desired target band. In this case, start with the past rolling 36 months, then the past rolling 35 months, the past rolling 34 months, and so on. The periods in this iterative process could be quarters as well. To be complete, it is necessary to repeat this analysis beyond the earliest rolling period to see if this desired ranking is transient and disappears just a few periods later. In this approach, the analytical double check—a similar question, but not merely another way of asking the same question—is useful.

Another way to address this question, with more explicit assumptions and projections, is to take returns from the existing universe and project similar performance and return distributions, as long as the fund has the median return in each future period. This procedure may give a rough indication of the time frame over which the fund might return to the top half. The approach must be used cautiously, however, because it may presuppose a much greater degree of performance persistence and consistency than empirical evidence supports.[6]

Table 1

Top Panel

					Year-to-Date Returns via			
	Q1 Return	Q2 Return	Q3 Return	Q4 Return	Q1	Q2	Q3	Q4
Fund A	−75%	0%	20%	5%	−75%	−75%	−70%	−69%
Fund B	25%	20%	0%	−15%	25%	50%	50%	28%
Fund C	50%	40%	−30%	0%	50%	110%	47%	47%
Fund D	0%	−60%	10%	10%	0%	−60%	−56%	−52%
Average return	0%	0%	0%	0%	0%	6%	−7%	−11%
Avg. of avg.		0%	0%	0%		3%	0%	−3%
Standard dev.	0.467707	0.374166	0.187083	0.093541	0.467707	0.769233	0.559793	0.49498

Bottom Panel: Correct Calculations

Return Rankings

	Individual Quarters				Year-to-Date via			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Fund A	4	3	1	2	4	4	4	4
Fund B	2	2	3	4	2	2	1	2
Fund C	1	1	4	3	1	1	2	1
Fund D	3	4	2	1	3	3	3	3
Total count	4	4	4	4	4	4	4	4

Return Percentile Rankings

	Individual Quarters				Year-to-Date via			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Fund A	80	60	20	40	80	80	80	80
Fund B	40	40	60	80	40	40	20	40
Fund C	20	20	80	60	20	20	40	20
Fund D	60	80	40	20	60	60	60	60
Average percentile	50	50	50	50				

Bottom Panel: Incorrect Average-of-Averaging Results

Avg. of Avg. (*return percentile rankings*)

	Individual Quarters				Year-to-Date via			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Fund A		70	53	50		80	80	80
Fund B		40	47	55		40	33	35
Fund C		20	40	45		20	27	25
Fund D		70	60	50		60	60	60

Error of Avg. of Avg. (*return percentile rankings*)

	Individual Quarters				Year-to-Date via			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Fund A		−10	−27	−30		0	0	0
Fund B		0	27	15		0	13	−5
Fund C		0	0	25		0	−13	5
Fund D		10	0	−10		0	0	0

Extra care and disclosures should be considered if this hypothetical analysis is to go beyond internal use only.[7] Additionally, the U.S. Securities and Exchange Commission's Clover No-Action Letter cites 11 specifically prohibited representations and emphasizes that "it is the responsibility of every adviser using model or actual results to ensure that the advertisement is not false or misleading."[8]

NOTES

1. Timothy P. Ryan, "Return Compounding: Essential Insights and Practical Implications," *Journal of Performance Measurement*, vol. 7, no. 3 (Spring 2003):42–46.
2. W.J. Reichmann, *Use and Abuse of Statistics* (New York: Oxford University Press, 1962).
3. Timothy P. Ryan, *op. cit*.
4. Elroy Dimson and Andrew Jackson, "High-Frequency Performance Monitoring," *Journal of Portfolio Management*, vol. 28, no. 1 (Fall 2001): 33–44.
5. Dan diBartolomeo, "Just Because We Can Doesn't Mean We Should: Why Daily Observation Frequency in Performance Attribution Is Not Better," *Journal of Performance Measurement*, vol. 7, no. 3 (Spring 2003):30–36.
6. Ronald N. Kahn and Andrew Rudd, "Does Historical Performance Predict Future Performance?" *Financial Analysts Journal*, vol. 51, no. 6 (November/December 1995):43–52.
7. David D. Spaulding, "Best Practices for Model/Hypothetical Performance," Presentation given at GIPS Standards Annual Conference, Boston (25–26 September 2008).
8. SEC Staff No-Action Letter, Clover Capital Management, Inc., Ref # 86-264-CC (28 October 1986).

Timothy P. Ryan, CIPM, is Vice President, Performance Management, at Hartford Investment Management Company.