Correlation and Peer Group Composition

By Dan diBartolomeo, CEO, Northfield Information Services

March 25, 2008

This note follows up our article last week on "Measuring Funds by a Jury of their Peers," and addresses the question of whether correlation is a valid yardstick for assessing the homogeneity of a peer group.

Correlation is a useful measure for determining whether a fund belongs to a particular peer group (i.e. correlations among peers should be high), but it is an insufficient measure for deciding how to put together a peer group from a large universe. If you have a pre-defined set of peer groups, lack of correlation is probably a good way to find "an apple mixed in with the oranges." But correlation is a measure between two sets of data. It is not in any way appropriate for simultaneous measurement among a large number of sets of data (i.e. among all the members of a group). There are other statistics, such as the Kendall W, called a concordance coefficient, that are meant for that sort of thing.

When forming peer groups from a large universe, the usual method would be some form of "cluster analysis," which uses correlation in specific ways to form groups. There are a number of "flavors" to cluster analysis (single linkage, complete, average, centroid, etc.). With financial return data, we consistently have the most intuitive looking results from the "complete linkage" method.

The other issue here is what measure of correlation you use. The standard measure, called Pearson correlation, is subject to a lot of influence from outliers in the data, which is a problem when looking at returns (there are a lot of outliers). You can easily get nonsense results. There are other measures of correlation, such as Spearman rank correlation and Kendall's Tau measure that are less subject to influence from outliers, but are not BLUE (best linear unbiased estimator). Being BLUE just means that when the estimated value is not the exact true value, it is equally likely to be too high or too low.