

A New Choice For Handling NA Values In Ranking Computations

In an ideal world, data would be available for every factor for every company. Unfortunately, that's not the case in reality. We very often encounter situations where important data items are nonexistent for some companies. These are reported in portfolio123 as NA, which stands for Not Available. A good example is PE, which is defined for little over 40% of the whole universe of stocks (8,000+).

When a screening rule encounters an NA item, it is assumed that the company failed the test and, hence, is omitted from the screen. Ranking systems present a more complex challenge since, unlike screens, the factors are not all-or-nothing. All companies are ranked somewhere on a best to worst scale. We therefore need to decide what NA means in terms of best-to-worst.

The Traditional Approach: NA's have a negative impact

The traditional portfolio123 approach puts NA values at the bottom of the sort, the worst position. But while we do penalize companies for NA items, we do so with a sense of moderation. When it comes to translating NA to a number, we don't assign it a value of zero or one. Instead, for each factor in the ranking system, we consider all companies with NA as being in a tie, and the number assigned is just below the rank of the worst company that had meaningful data. The extent of the NA penalty varies from item to item. If NA is a rarity for a particular factor, a company that has one will be penalized sharply. If NA is commonplace, the penalty will be mild, but still a penalty.

Lets see an example of a ratio ranked "higher is better" for 5 companies:

Ticker	Ratio	Rank	Comments
A	20	80	Rank increment is 20 since there are 5 values: $100/5=20$
B	5	60	
C	2	40	
D	NA	20	Rank never reaches 0 because of tie.
E	NA	20	

The Implications

NA values count against a company when it comes to computing its rank. This can be important when rankings are computed relative to industries, some of which may be very small.

It can also have a more general impact. Financial companies in particular are prone to being pushed downward in ranking systems that use several factors that don't exist for them because of the way they report (turnover, operating margin, etc.). And where

financials are still able to pass user models (because they are sufficiently strong in the non-NA factors), they're more likely to hover near the borderline and fall out of portfolios more easily when rebalancing takes place.

A New Choice: NA's have a neutral impact

Portfolio123 users can now choose to have NA play a neutral role in rank computations.

Where users opt for this approach, all NA values will, when the computation process begins, be put to the side. Percentile rankings from zero to 100 will be computed for all firms that have the necessary data. Then, at the end of the process, all NA companies will be assigned a rank in the middle of the valid ranks, a perfectly neutral score. The rank assigned to NA's will usually be around 50, except when there are very few ranked stocks (such as in a small universe or industry).

The more NA rank factors a company has, the harder it will be for the firm to appear in any of your rank-driven result sets. It will be harder for the company to come in near the top and harder for the company to come in near the bottom (as you might seek if you are looking to short poorly ranked stocks).

Lets look how the example above changes.

Ticker	Ratio	Rank	Comments
A	20	66.66	Rank increment is 33.3 since there are 3 values: $100/3=33.3$
B	5	33.3	
D	NA	33.3	NA's are neutral
E	NA	33.3	
C	2	0	Lowest rank value is reached

This new approach should not change long systems since they pick from the top, where it will mostly be stocks with valid values. Where this could become useful is in a long/short system that buys top ranked stocks and shorts lower ranked stocks using the same ranking system. In this example a long/short system of the best/worst stock would go long A and short C, all with valid values. In the first example the long would be A, and the short either D or C (each with NA).

A quick test to see if this will lead to better long/short systems is to run a Rank Performance analysis. What you should see is lower returns in the low ranked stocks, thereby improving the hedge performance.

