

Title: MISPRICING FACTORS

Speaker: ROBERT F. STAMBAUGH
Wharton School, University of Pennsylvania

Importance: Why this matters:

Many factor-based and active quant strategies use multiple characteristics to select stocks and/or construct portfolios. Clustering has the potential to combine correlated factors in an intuitive and robust manner.

Investigation: "Speaker analyzed XXX data to address the questions yyy, zzz, etc."

Stambaugh demonstrated how to use clustering as a technique for combining stock characteristics (linked to return anomalies) into factors. He used the method to combine the 11 characteristics he used in his earlier papers into two factors. He then tested the combination of them plus a market and a size factor against the standard Fama-French three factor and five factor models. Using multiple metrics, Stambaugh's model did a superior job of "explaining" the anomalies. (I.e. including the two factors constructed using the clustering technique eliminated the significance of the anomalies based on the individual characteristics, while most were still significant using F-F.) His four factor model was also superior to F-F in explaining a longer list of 73 anomalies. His model also did a better job of explaining the F-F factors, than the F-F model did in explaining his factors.

Innovation: Are there new techniques of interest in the data or approach to the problem?

The first innovation was to first calculate residual returns, and then to clustering to combine the characteristics into factors.

The second innovation was to construct the SMB (small minus big) factor using only the stocks that were not in the extreme factor portfolio. That keeps the stocks' factor exposures from biasing the size portfolios.

Insights: 1-2-3, what are the three most important things the speaker offered?

1. Clustering permits identification of factors constructed from multiple characteristics in a way that principle components would not.
2. Creating the SMB factor using only stocks that did not have significant exposure to the other factors "improved" its performance.
3. Even without any explicit "value" characteristics, the model did a better job of explaining the anomalies observed in prior studies.

Audience rating: 3.35