

Quantitative Strategy 2016/2017 Assignment

Introduction

Risk-tolerant retail investors are frequently advised to invest in equities, as that is the asset class which has generated the greatest returns historically. In the early 90ies, Eugene Fama and Kenneth French noted that yield can be increased further still by adding stylistic biases to your portfolio. In short, they found that small stocks outperform large ones and value stocks outperform ones pricing in rapid growth.

Twenty years after the publication of the [original Fama-French paper](#), Jeremiah Green, John Hand, and Frank Zhang carried out an [extraordinary study](#) of subsequent academic literature and identified over 300 different 'return predictive signals' analysed over the years.

A common problem with studies of individual return predictive signals is that *in the absence of context, it is difficult to understand why a certain signal stopped working or why it is performing particularly well*. For this assignment, we challenge you to establish that context by modelling joint behaviour of 102 return predictive signals kindly made available for this program by Professor Jeremiah Green.

Note on terminology

Before we proceed further, it is important to establish that 'return predictive signals' are more generally referred to as 'investment factors'. You may also see them referred to as 'investment styles'; the widely used [Morningstar Style Box](#) borrows heavily from Fama-French.

Your assignment

The data

For any given point in time, each return predictive signal (RPS) generates a descriptive value for each stock. Stocks are then ranked by that value; RPS returns are calculated by taking the difference in returns between stocks in the top and bottom deciles in this list.

Monthly returns for 102 RPS are provided [here](#). This zip file contains the following:

- Copy of the paper for which this data was initially prepared
- Time series of RPS returns – equally weighted
- Time series of RPS returns – value weighted
- Variable definitions

Weighing stocks within deciles based on their descriptive values can lead to a few names becoming dominant, but weighing them equally may obfuscate the risk dynamics the RPS is intended to capture. This is why we are including both value and equally weighted returns.

Instructions

While having 100+ dimensions makes this seem like a big data problem, having about 400 data points means that you can't just run machine learning methods on this dataset and generate strong results. You need to explore and understand the data-set.

We would suggest doing the following:

- **Categorise the RPS**
Clustering methods such as k-means would allow you to extract subpopulations of RPS which you could then build intuition around. Just knowing what the subpopulations are is helpful. How has average performance varied for each subpopulation through time?
- **Assess relationships among RPS**
Once you have identified subpopulations of RPS, it might be helpful to look at correlations within them through time; there are **three ways to do this**. Try to understand what might be behind major changes in those correlations.

It might also be helpful to run PCA on the entire RPS population and look at variance explained through time. [This document](#) from last year's competition is a short, practical guide to PCA.
- **Try to define regimes**
Once you know how returns and other statistical properties have evolved for these RPS through time, the next step is to try and define regimes. This kind of analysis was carried out by one of last year's students ([recap](#)) and allowed him to create a model of the economic cycle. Understanding which regime we are in allows you predict which regime is likely to follow – and which RPS are likely to do best.

You are by no means obligated to follow the above process. Just explore the dataset, and do whatever makes the most sense to you.

Your report

The output from your analysis should be a 1-3 page write-up summarising key conclusions. Your analysis should be neither exhaustive nor complete – we are looking for preliminary findings only, and would encourage you to include a Next Steps section in your report. If you are approved for inclusion in our mentorship program, one of our industry mentors will help you build on your initial work.

We expect this assignment to take 3-5 hours to complete, but you are welcome to spend as much time on it as you deem appropriate. Feel free to [contact us](#) if you feel you have hit a dead end.

This is not an offer to sell or the solicitation of any offer to buy any securities.