Here is a brief summary of the projects and outline of the answers to the questions. For the details, please check the ipython notebook.

In this project, variable selection applies to each industry. Then factor beta is estimated by OLS. Rolling beta is calculated to capture the local effect. Then based on the factor beta, industries are categorized into 4 groups, which is in line with fundamental intuition. Finally two long only portfolios are constructed and rebalanced every month. The target weight is adjusted every 5 year and each time new return data is used to estimate the covariance matrix.

1. Which industry portfolios are closely related to each other  (how many groups exist – a fundamental analyst may say there are two groups - defensive and cyclical J)? Does this relationship/group composition change through time?
   1. What technique/s can you use to show this – why did you choose your technique over others?

4 Groups are found.

Money and Durable are close. They can be thought as cyclic.

Utilities, Telcom and Hlth are in one group. The can be treated as defensive. Others are between the two polars.

First LDA is applied to the beta, projecting data to most separated directions. Then KMeans suggests the categorization.

1. Construct minimum variance portfolio with constraints of not more than 50% allocation to each group that you found in no. 1. Compare it with equal weighted 12 industry portfolio. Do monthly rebalancing for both. Note: if you don’t attempt 1. just assume 4 industries in 3 groups each.

Much weight of the min variance portfolio falls into defensive, Utilities, Telcom and Hlth. The weight is determined by covariance matrix input, which tilts at each adjustment.

1. What are the FF beta for each of the industry portfolios? In regression does it help to use 5 or 6 (includes MOM) factors or was origin 3 or 4 (includes MOM) factors model adequate? Are these betas time varying?
   1. (Optional) What regression technique can one employ to find more stable beta? How does your results change?

For each industry, the selected factors are different. The new factors work for some industries.

Beta is time varying. But it does not make sense to expect factor exposure explode over time.

For out of sample filtering, a mean reversion process is assumed, while the convergence may be slow, and model calibration is challenging.

For in sample smoothing, Gaussian conjugate is utilized.