FINM 35000 Final Project Proposal

[redacted]

Data: We will obtain monthly returns of individual equity listed in the NYSE, AMEX and NASDAQ from CRSP or Bloomberg. We will also get as many factors as possible (around 50-70) from both resources. The major factor categories are: value, quality, leverage, growth, momentum, volatility and sentiment. Topics mentioned in class including ESG scores and monetary policy indicators will also be explored as potential factors. The time period for our dataset will be from 2000 to present. Data before 2016 will be used for getting factors and training our strategy. Data from 2016 and later will function as testing sets.

Trading strategy:

PCA:

After getting the set of potential factors, we will apply Principal Component Analysis to reduce dimensions by selecting a few factors that could explain most of the variation in our dataset. The reason we will use PCA as our first step is because our dataset will be large and contain a high number of dimensions. PCA will help increase the interpretability of our data. In addition, some financial factors could be highly correlated. For instance, some factors might be ratios of similar components or changes in one factor could also lead to change in another factor. PCA will also help us solve the multicollinearity problem. We will then build our trading strategy on this reduced dataset.

Penalized Regression:

When we are getting a large multivariate dataset containing a number of variables superior to the number of samples, the appropriate way to figure out the relationship between portfolio return and factors is to use penalized regression. Penalized regression, which penalizes for having too many variables by adding a constraint in the equation, would produce more stable results for correlated data. We will try both Ridge Regression and Lasso Regression to incorporate a large number of independent variables and dependent variables. These two models have different regularizations and limitations, which would result in various sets of predictor variables and prediction errors. We will compare these models based on stability, prediction accuracy, computational efficiency, and diagnostic assumptions. Based on our analysis, we will figure out a set of factors that have statistically significant relationships with the portfolio return. According to their relationships and coefficients from our regression, we are able to generate a trading strategy and backtest our portfolio to get the performance measurement.