

Project Proposal (Group 8)

1. Overview

Our strategy is inspired by the paper "[Lazy Prices](#)". Briefly speaking, our strategy is an equity long/short strategy that tends to maximize the exposure on text-based quality factor and minimize the portfolio exposure on common risk factors, such as size, momentum and value (P/E, P/B, P/CF, EV/EBITDA).

Text-based quality factor means that we first download 10/K and 10/Q from EDGAR and parse the filings into different sections (MD&A, Risk Factors etc.). This step has already been finished. Then we conduct text analysis on these sections and calculate some alternative quality factors. (Traditional quality factors include leverage, dilution, earnings quality, profitability).

By default, the factor in the "Lazy Prices" is to calculate the similarity for the same 10-K section on t and $t-1$. The basic idea is that firms typically repeat what they most recently reported. Thus, changes to the language and construction of financial reports may have strong implications for firms' future returns. Empirically, great changes lead to negative future excess return. In our project, we will use different methods to calculate distance. Methods include cosine distance, Jaccard similarity, word-embedding and BERT.

The second factor will be trained by supervised learning algorithm. The basic idea is that for each firm's 10-K, we can set some labels based on their financial variables and future returns. Then we train ML models with text features to predict these labels. When new 10-K is released, we can get a predicted score from the model. We then orthogonal this text factor to common risk factors and sort portfolios based on the residual factor.

The last factor will be a sentiment factor based on commonly used financial dictionary.

2. Universe

We will cover S&P500 and also test strategies on sub-sector universe.

3. Data Frequency

Most financial data is annually or quarterly. Factors like lottery, momentum and value will be monthly.

4. Source

We already get 10-K from EDGAR. Firms' Characteristics are from Bryan Kelly's Website and WRDS.

5. Frequency

We will test our strategy assuming monthly and quarterly rebalance.