Zeyuan Pan's Week2 note

Paper reading thoughts

SMA MRF - Alpha - Total Document, MD&A, Risk Factors Alpha

1. Intoduce how to create the GMRF dataset and the composition of the dataset
2. The filings focuse on word counts, sentiment factors, and the change in those factors
3. Make assumptions based on existing data and backtest them

- factor1:raw change in word count based on Lazy Price
- factor2:raw change in Subsection Count
- factor3:Total Document Average Sentiment
- factor4:Percentage of Sentiment Hits
- factor5:Sentiment per Word
- factor6:Positive and Negative Hits Difference

4. Through these outcomes of backtesting. The paper expresses the Machine Readable Filings is the most advanced and thorough product on the market for drilling into the un-tapped value of textual data in regulatory filings

problem

1. What's the difference between Sentiment Hits and Sentiment words. How to judge the word hits or not.
2. I found at the end of 2019 all these six factors have a large retracement. WHat could be the retracement occur? Covid Influence?

Contribute

Read the paper that I show on the up. Also read some important parts about the Lazy Price. Contributing to some pages of the slides.

Meeting note

This weeks' meeting talks about some problems that we proposed after reading the papers. And we also learnt from Mr koby that how the dataset would be look like. Also I found the backtesting system is not as strictly as I thought before and is easy to set up.