Portfolio Management Project

Javier Arevalo
DaTa Driven Methods for Finance

Predictions

The weekly investment decisions where made from the following models:

1. **Fundamental Model**
   - Consisted in factor models, namely the Fama French 3 Factor Model, Fama French 5 Factor model and the Barra Model
2. **Sentiment Model**
   - Used Natural Language Processing to analyze the market sentiment towards a particular stock
   - Used the Google BERT Model for the Sentiment Analysis Task
3. **Options Volatility Data**
   - Used Local Volatility Surface data to monitor trading activity for OTM Puts (i.e. "crash protection")

Fundamental Model 1: Fama French 3 Factor Model

- This model aims to describe stock returns through a combination of factors.
- Fama French 3 Factor Model aims to describe a stock's excess returns through 3 factors:
  - Market Risk
  - Excess returns of small cap over large cap
  - Excess returns of value stocks over growth stocks
- Due to the model being outdated, results were not as expected.
- Concluded that those 3 factors were helpful but not enough to explain a stock's excess returns.

Fundamental Model 2: Fama French 5 Factor Model

- Fama French 5 Factor Model expands on the FF3 Model by adding two additional factors:
  - Excess returns of the most profitable over least profitable firms
  - Excess returns of firms that invest conservatively over firms that invest aggressively
- Results were considerably better with Fama French 5 model.
- However, similarly to FF3 its predictive power was limited due to the fact that all traders know about it
  - Does not provide any additional information that others traders/market does not have access to

Investment Decision

- The previous models output a expected excess return of a given stock
- Thus needed a way to pass from the model output to one of 5 ranks

Approach used:

- Found the 20th, 40th, 60th and 80th percentile of the expected returns distributions.
- If the expected return of a stock fell below the 20th percentile, it was categorized in the first rank exclusively and so on for the rest of the ranks.
- Note: just assigned each stock to one of the ranks instead of probability of being in each of the ranks.

Barra Model

- Barra model is a multi factor model that incorporates more than 40 data metrics including earnings, growth, and many other ratios.
- The risk of using to many factors is overfitting which performs poorly out of sample.
- Key to using these model is factor picking

Two approaches to selecting factors:

- Combinatorial approach: brute force
- Pick 3-5 factors based on theory -> backtest -> repeat

Sentiment Model

- The second main model used in the predictions was the Sentiment Model.
- Used data from the news (Forbes, Analyst Reports, tweets from prominent investors, etc.)
- Feed data into pre trained model that outputs a Buy/Hold/Sell recommendation.
- Model used for prediction was Google's BERT Model: State-of-the-Art Pre-training for Natural Language Processing
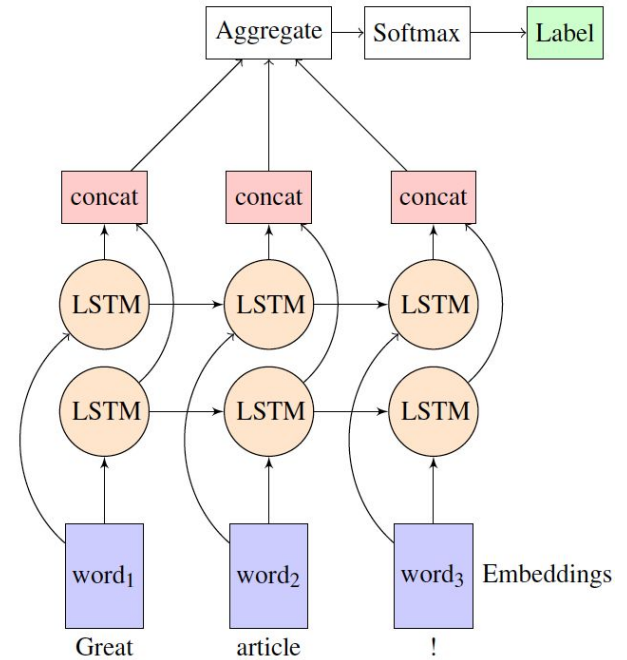
Options Data Intro

- Even though trading options was not allowed, options market data provide a really valuable source of information to analyze stocks.
- Volatility data across different maturiteis can be a indication of the demand/supply for a company's option and thus the demand/supply for the stock itself.

Crash protection:

- Out of the money Puts are considered "crash protection"
- When demand for OTM Puts is high, it might be an indication that experienced traders or a market in general expect the company's stock to go down in the near future.
- By buying the OTM Put, the traders minimize their losses if a particular stock goes down significantly.

Vol Surface