

Financial Data Forecasting

Project Plan

Sumer Malhotra (3035347457)

FYP19062

Project Background

Over the last couple of years, Quantitative investing has become extremely prevalent in the world of financial markets and institutions, which can be seen with the mass adoption of the techniques involved with this type of investing by major hedge funds and investment banks. This type of technology has become household staples in these institutions due to the ability of computers to analyse large amounts of data with ease. To give some perspective on this technique, we can look at some statistics related to it. Approximately 90% of public market trading volume in the United States is executed using quantitative analysis. Furthermore, this market has a capitalization of roughly \$1 trillion dollars. Finally, around 6 of the top 10 performing hedge funds in the world majorly employ quantitative techniques in their investment process [1].

Quantitative analysis is a process used to understand market behaviour by analysing a large number of variables including asset prices, trading volume, technical indicators, fundamental indicators, market sentiment, day and time, etc. and how these variables may be correlated with each other. Following this, a mathematical model is developed which assumes these variables as input and outputs either a predicted direction in which the market may move or a signal (buy or sell) which serves as the building blocks of a trading strategy.

One of the main reasons owing to the increasing popularity of this kind of investing, is the removal of human emotion from the investment process. Numerous analysts are of the impression that fear, greed and the inability to handle and process large datasets serves in stifling logical thinking and reasoning, usually leading to losses. The absence of these biases

in quantitative investing is what sets it apart, giving it such an esteemed position in the world of finance.

A fundamental rule of economics is that when supply overwhelms demand, prices decline. This unfortunate rule serves as a potential drawback for quantitative analysis. According to a CNBC report from July, 2018, less than 40% of institutional managers have outperformed their benchmarks [1]. This figure may be overestimated due to the skewness from the inclusion of both qualitative and quantitative managers in the statistic. However the point is obvious, majority of quant managers no longer provide superior returns for their clients. In financial jargon, this term is called crowding, resulting from a rapid inflow of capital into these strategies in recent years which can be seen from an increasing adoption of these techniques by not only institutional managers but also individual investors.

To solve this problem of crowding, investors are now at the peak of their creativity, employing traditionally absurd techniques and looking at unconventional data sources. Furthermore, with the introduction of cryptocurrencies, a new asset class with no intrinsic value, investment techniques have become more complex, serving as the main reason for my interest in this project. Furthermore, past industrial experience in hedge funds and crypto trading strategies helped finalise my decision in choosing this project for my final year thesis.

Project Objective

The main objective of this project would be to predict the direction in which prices may move for certain asset classes. As far as assets go, interesting areas include cryptocurrencies as well as equities, options and futures listed on the Hong Kong exchange. The expected accuracy for prediction of these assets is between 60% and 70%, subjective on the time series/machine learning model used in prediction. This analysis will also help in the understanding of the advantages and limitations of various models; the impact of variables, investment time-horizon, scale and number of data sources on model accuracy.

Apart from just quantitative objectives, I also hope to gain qualitative insights into investing. Market psychology of investors is a key insight I hope to gain while progressing with this project. Prediction of company growth, bankruptcy, market bubbles, inflated asset prices, impact of financing decisions and importance of accounting variables are other factors prevalent within qualitative investing that I hope to gain insight on.

Project Methodology

The initial phase of the project will revolve around performing literature review from various sources, including IEEE in order to gain insight on the past usage and effectiveness of various models. The key takeaway from each paper will be areas of exploration that were left undeveloped and how my project can capitalise by taking advantage of these inefficiencies.

In addition to direct data retrieval with the usage of RESTful APIs, the application of various data retrieval techniques like web and data scraping is imperative in the preparation of a large enough dataset for optimal training and testing. Following this, data pre-processing will be conducted through cleaning, categorization, normalization, aggregation, PCA transformation and splitting [2].

The algorithm development will begin with feature selection. The process will be carried out by fitting the dataset with various researched models while simultaneously analysing the effect of various asset-classes and variables on prediction accuracy. This process will be accompanied by the back testing of various data sources, aggregations for different time periods and for numerous investment horizons, thereby helping understand the time-periods demanded for most optimal predictions.

If time permits and accuracy thresholds are satisfied, an application will be developed which will allow manual selection of data sources, variables and prediction models.

Project Schedule and Milestones

Sept 29, 2019	Deliverable of Phase 1 <ul style="list-style-type: none">• Detailed Project Plan• Project Website
Oct – Nov 2019	Literature Review <ul style="list-style-type: none">• Research Machine Learning models• Research Time series models• Research on Natural Language processing• Research on any other interesting prediction techniques
Dec 2019	Data collection and pre-processing <ul style="list-style-type: none">• Web-scraping with ScraPy and RESTful APIs for data retrieval• Data pre-processing using Pandas, NumPy, scikit-learn
Jan – Feb 2020	Algorithm Development and Testing <ul style="list-style-type: none">• Feature selection• Model development and testing• Model optimization for various asset-classes
Mar 2020	Application development and deployment

References

- 1) Henry, G. (2019). The Rise of Quantitative Investing. Retrieved 26 September 2019, from <https://medium.com/@garethhenry/the-rise-of-quantitative-investing-93c4229e9704>
- 2) Rajaratne, M. (2019). Data Pre Processing Techniques You Should Know. Retrieved 26 September 2019, from <https://towardsdatascience.com/data-pre-processing-techniques-you-should-know-8954662716d6>