Modelling Gains and Falls in Financial Asset Returns

An Extreme Value Theory and Machine Learning Approach

Prepared by:

Olatomiwa Akinlaja

201628555

Supervisor: **Dr. M.S. Mosia**

PROJECT PROPOSAL

September 10, 2019

1 Introduction & Background

The use of an intelligent trading system plays a major role when it comes to monitoring and determining the rise and fall of stock value within the market.Abdul-Aziz et al. (2018) This helps businesses, organizations, investors and stock traders to stay ahead of the competition, when making key decisions regarding financial asset returns. The unpredictability of the market leads decision makers to acquire systems that can help predict prices based on various situations and conditions, which assists with making key investment decisions.Abdul-Aziz et al. (2018)

Thus, in making these key decisions, a model that is able to accurately predict rises and falls within financial value returns (stock data) would provide intelligence and would generate profit to the user, whether a business, organization or stock trader. The underlying nature of the financial domain parameters leads to the dynamism of stock prices and their susceptibility to rapid changes; Mandelbrot et al. (1997). resulting to random rises and falls within the data, some of which are very high and very low.

By using Extreme Value Theory to target extreme prices within the stock data in question, this ensures that important aspects of the data is targeted by fitting the data with the extreme value distribution that compensates for the fat tails. suggested by McNeil (1999). This project seeks to explore the combination of extreme value theory and machine learning, in order to provide a model that can accurately predict rises and fall within stock data. in order to determine the occurence of financial spikes. Herera et al. (2017) through a deep neural network with a probabilistic layer to ensure trustworthy stock predictions.

2 Problem Statement / Motivation for the project

The fears for most financial analysts and risk administrators are that the events of abnormal market conditions produce enormous and unforeseen stock losses that could upset and likely leads to liquidations and absolute risk. Abdul-Aziz et al. (2018)

The importance of modeling gains and falls regarding stock data is to prevent unforseen losses. By analysing the trends within the stock data, forecasts and recommendations can be made. The stochatic nature of stock data introduces a factor of unpredictability. This introduces challenges that arise when modelling gains and falls of stock data. Feature engineering would help determine how to classify a fall and gain within stock data. Thus providing a target variable which can be fed into the predictive model.

Generally, it can be difficult to estimate these extreme events due to the scarcity of past data available. De Haan and Ferreira (2007) mentions the scarcity of data to help determine these instances. It is due to this fact that estimates would be uncertain. The level of uncertainty extends not only within a certain range but beyond. This research project seeks to strengthen the generic stock intelligent trading system by including additional layers of pre-processing, feature engineering and probabilistic predictive modelling.

This research project uses machine-learning techniques to predict rises and falls within financial asset returns before the actual event occurs within the stock market. A lot of interesting work has been done in the area of applying machine learning algorithms to analyse price patterns and predicting stock prices. Hsu and Lin (2002), Liu and Zio (2018) and Tang et al. (2018) are some of the research studies that employ the use of support vector machines as a means for classification. Essentially, they fail to parametrise the stock data through extreme value theory.

Furthermore, they also fail to provide a level of confidence in which their results are based upon. Hence, why this project includes a probabilistic aspect, in order to show the level of confidence of the model accuracy.

3 Aim

This research project aims to model gains and falls in financial asset returns through the integration of extreme value theory and machine learning.

4 Objectives

- Test for normality to show that stock data is not normally distributed as determined by statistical tests for normality (Kolmogorov-Smirnov Test and the Shapiro-Wilk Test). Based on the fat tail distributions, they consist of very high and low values.

- Measure the tail behaviour of stock data returns with Extreme value distributions such as weibull, gumbell and frechet. In order to properly analyse the fat tails of the returns.

- Provide a trading strategy based on a predictive model which integrates of Extreme Value Theory and Probabilistic neural networks. EVT will provide the analytical framework that will be used to identify extreme values within the financial data in order to determine risk assessment.

5 Literature Review

5.1 Financial Asset?

Investopedia (2018) defines a financial asset as a liquid asset that gets its value from a contractual right or ownership claim. Investopedia (2018) provides examples of financial assets, namely: Cash, Stocks, bonds, mutual funds and bank deposits. Financial assets generally have neither physical worth or form. Its value within the marketplace in which they trade reflects factors of supply and demand as well as the risks in which they carry.

Investopedia (2018) categorizes financial assets as real, intangible or financial.Real assets are physical assets that draw their value from substances or properties, such as real estate. Intangible assets are valuable and non physical. such as intellectual property and patents. Financial assets can be seen as a comination of both real and intangible as it fall within a seet spot right in between. It may seem intangible with only the stated value on a piece of paper, such as Money or a listing. The piece of paper provides a claim of ownership of an entity.

5.2 Stock Data Prediction Methodologies

Shah (2007) decribes two stock prediction methologies. **Fundamental Analysis** was performed by the fundamental analysts, this method is concerned more with the company rather than the actual stock. The analysts make their decisions based on the past performance of the company. Shah (2007) also described **Technical Analysis**, which was performed by the technical analysts. it deals with the determination of the stock price based on the past patterns of the stock.

The application of machine learning to stock data results in a technical analysis. This determines if the predictive model can accurately learn from the underlying patterns in the stock data. It is also possible to apply machine learning in order to evaluate the performance of a company through fundamental analysis. Furthermore the most succesful automated stock prediction and recommendation systems use a hybrid involving both fundamental and technical analysis.

5.3 Extreme Value Theory

Extreme Value Theory (EVT) is a statistical theory of extreme events, which can categorized as very unlikely to occur. In terms of creating financial damage, it can be costly to handle due to its low frequency and high severity. EVT is applicable to almost all univariate external problems. It seeks to assess a random variable from a given order of sample, the probability of events that are more extreme than any observed. . EVT extremes form a very large domain of stochastic processes which follow one of the three generalised extreme value distribution types, Gumbel, frechet/pareto, or Weibull. These distribution types referred to as heavy, median and short tailed respectively Friederichs (2007). Only these three types characterise the behaviour of extremes shown in figure 1.

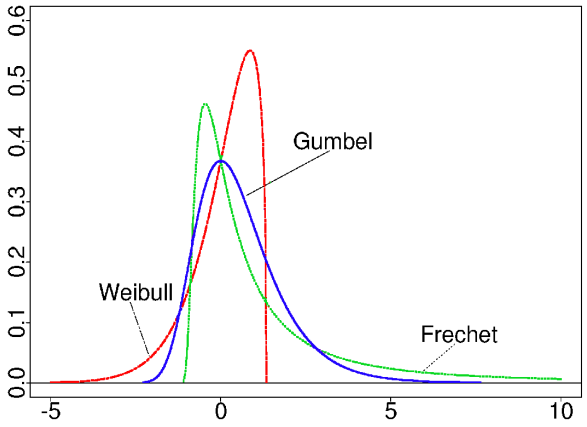


Figure 1: Extreme Value Theory Distributions

5.4 Artificial Neural Network

In its simplest form, an artificial neural network (ANN) is an imitation of the human brain. A natural brain has the ability to learn new things, adapt to new and changing environment. The brain has the most amazing capability to analyse incomplete and unclear, fuzzy information, and makes its own judgment out of it. Chen et al. (2003)

AndrejKrenker et al. (2011)defines an Artificial Neural Network (ANN) a.k.a multi-layer perceptron as a mathematical model that tries to simulate the structure and functionalities of biological neural networks. The basic building blocks of every ANN is the artificial neuron. This neuron performs three kinds of operations namely: mul-

tiplication, summation and activation. For example, given an ANN that has an input layer, a single hidden layer and an output layer. The process involves feeding inputs into the neuron within the input layer; the next step involves multiplying each input neuron to form a weighted sum before feeding it into the middle layer, called the hidden layer. After another weighted sum, the output gets fed into an activation function a.k.a transfer function, before generating the final output within the output layer. figure 2 illustrates.

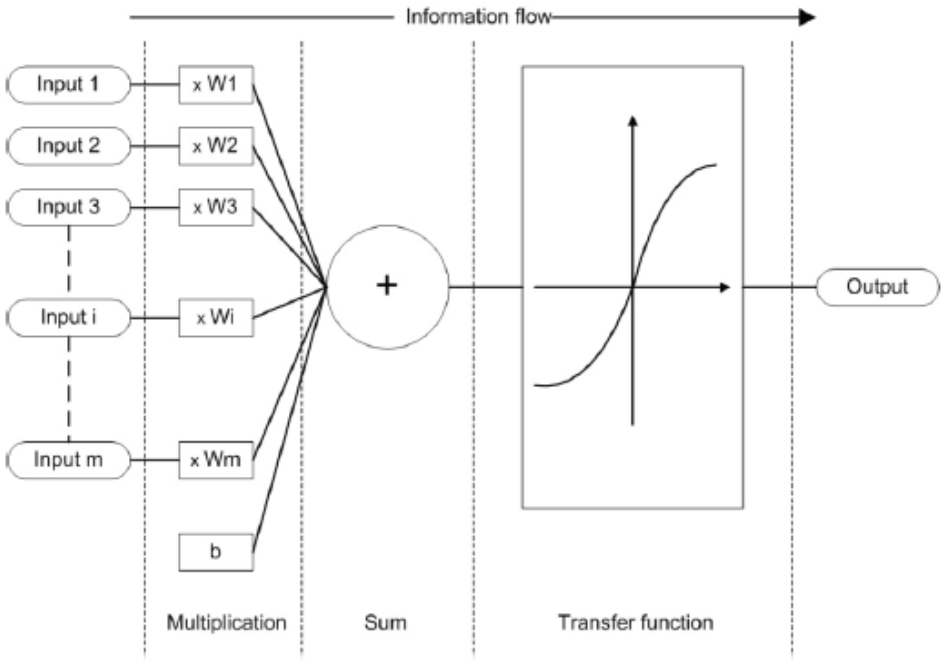


Figure 2: Overview of neural network

Areas where ANNs are applied includes chemistry, gaming, radar systems, automotive industry, space industry, fraud detection, etc. they are also used for the solving of problems like function approximation, regression analysis, time series prediction, pattern recognition, clustering, etc. for the basis of this report they will be used for the purpose of classification and decision making.

5.5 Probabilistic Programming

Chen et al. (2003) conducted a study based on the application of neural networks in an emerging financial market. Chen et al. (2003) uses a Probabilistic Neural Network (PNN) to forecast the direction of index return after it is trained by historical data.

Chen et al. (2003) decribes a number of appealing properties of the PNN in otder to justify the adoption of this type of neural network in the study. Chen et al. (2003) decibed the purpose of a PNN in three stages;

- Firstly, the training of PNN is rapid, enabling the development of a frequently updated training scheme. The network is re-trained each time the dataset is updated and thus the most current information can be incorporated into the network.
- Secondly, the network is able to identify outliers and questionable data points and thereby the extra effort on scrutinizing the training data can be reduced.
- Thirdly, and the most important, the PNN provides the Bayesian probability of the class that a particular input belongs to. This valuable information is utilized by a trading strategy to determine the possible action.

6 Methodology, Datasets & Data Analysis Techniques

The S&P 500, or just the S&P is an American stock market index based on the market capitalizations of 500 large companies. Figure 3 is a visual representation of a company's stock data

Date	High	Low	Open	Close	Volume	Adj Close
2014-09-10	19.680000	19.430000	19.650000	19.610001	4309400.0	18.881136
2014-09-11	19.540001	19.200001	19.469999	19.410000	6268000.0	18.688570
2014-09-12	19.530001	19.100000	19.530001	19.120001	6563400.0	18.409348
2014-09-15	19.209999	18.780001	19.180000	18.860001	7353800.0	18.159008
2014-09-16	19.240000	18.750000	18.809999	19.139999	5498400.0	18.428604

Figure 3: view of company's stock data

The integration of the Yahoo-Finance API and Beautiful Soup library within python for webscraping. The datasets for all companies within the S&P 500 can be aquired and saved as csv files. This provides access to a wide variety of stock data that can be used for analysis.

Ethical Considerations

There are no ethical implications involved with using the stock data of the companies within this project. The data is open source and free to be used by the public. Google and Yahoo! provide free access to download stock quotes using an API. All companies have stock tickers. E.g. Apple has the stock ticker AAPL. Stock data can be scraped through the use of the provided API by specifying the ticker and tag within the beautiful-soup python module for web scraping, in order to acquire information such as trading data. Index (2019) provides an in depth explanantion of Yahoo finance API.

The use of the very versatile, python & R programming languages and the modules and packages they offer makes it possible to demostrate an integration of machine learning and extreme value theory. This would help to predict gains and falls within

stock data returns in order for investors or companies to gain an advantage based on the business intelligence provided.

Scipy, a statistical library provides multiple statistical distributions and tests for normality, will be used to test the stock data for normality before fitting with the weibull, gumbell and frechet extreme value theory distributions. The predictive neural network will be developed within the tensorflow and keras frameworks. Other major libraries that will be used are matplotlib for data visualization and pandas for data manipulation.

Once the back-testing has been performed, a form of accuracy or performance measure would need to be incorporated within the model through the means of probabilistic programming. A credibility interval would help determine which level, be it positive, negative normal provides the most reliable predictions for financial extreme value spikes. The use of pymc3 and tensorflow-probability will be used to provide a probabilistic aspect to the model's predictions.

Analysis Process

- Scrap Data from Yahoo Finance.
- Cleaning, Aggregating and pre-processing.
- Calculate returns
- Prove that stok data is not normally distributed through tests for normality.
- Fit data with EVT distributions.
- Feature Engineering (Gains and Falls)
- Neural Network Predictive Model.
- Backtesting and forecasting.

Layout of Chapters

- Introduction
- Literature Review & Review of Methods

    Financial Asset

    Stock Data Prediction Methodologies

    Motivation for the Project & Problem Statement

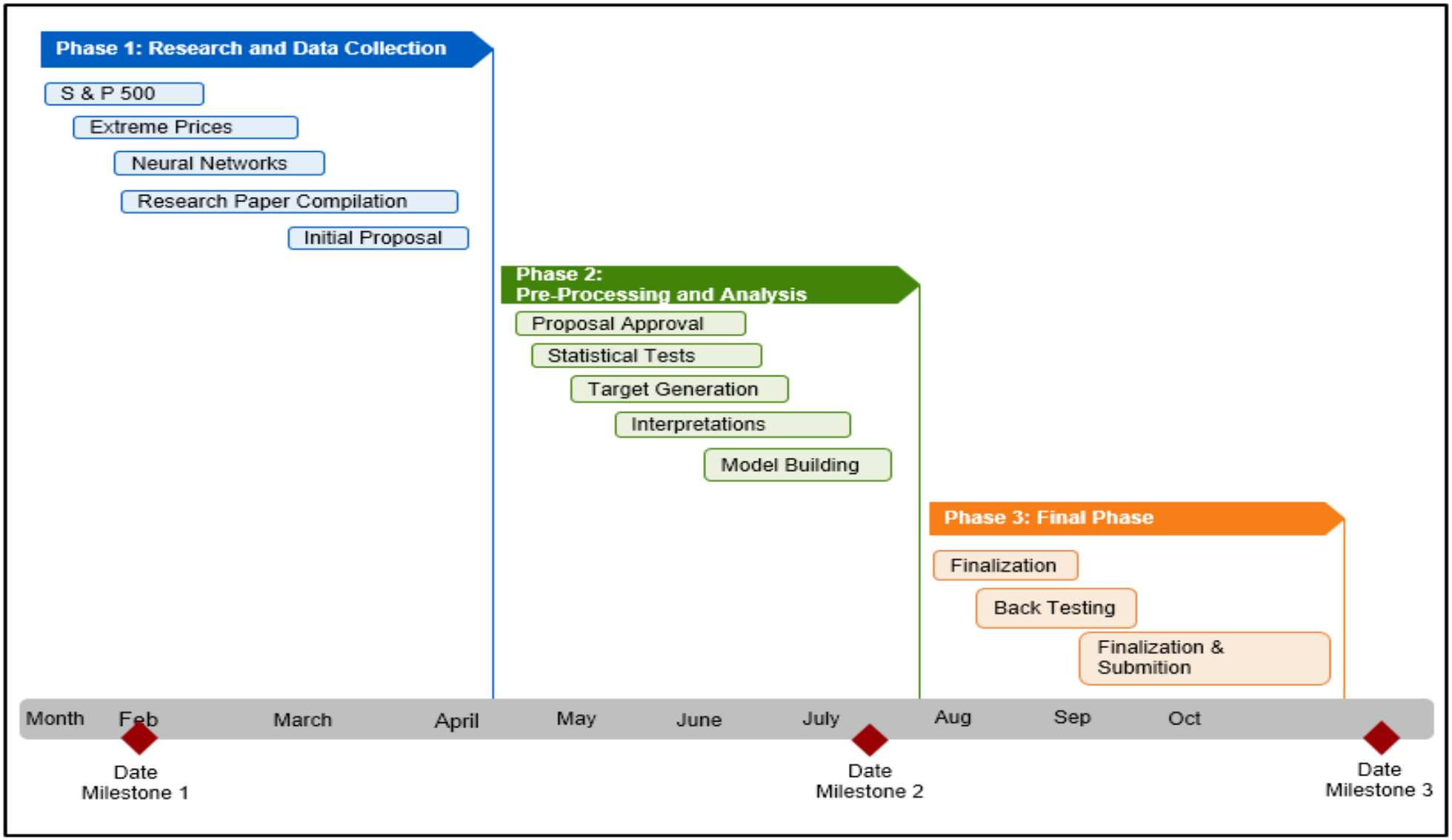
    Data

    Extreme Value Theory

    Artificial Neural Network

- Analysis & Results
- Conclusion & Recommendations

In conclusion, a corporation or investor might want to determine when the best window period to buy or sell stock within the stock market is. A peek into the possible future would gain an advantage within the stock market. This research project seeks to answer those questions by providing a possible foresight into the future of stocks within the market.


Phase 1: Research and Data Collection
S & P 500
Extreme Prices
Neural Networks
Research Paper Compilation
Initial Proposal
Phase 2:
Pre-Processing and Analysis
Proposal Approval
Statistical Tests
Target Generation
Interpretations
Model Building
Phase 3: Final Phase
Finalization
Back Testing
Finalization &
Submition
Month
Feb
March
April
May
June
July
Aug
Sep
Oct
Date
Milestone 1
Date
Milestone 2
Date
Milestone 3

References

Abdul-Aziz, I. M., Jianguo, D., Khan, u. d., Salah, H., Alhassan, A., Abdul-Rasheed, A., et al. (2018). The asymptotic decision scenarios of an emerging stock exchange market: Extreme value theory and artificial neural network. *Risks*, 6(4):132.

AndrejKrenker, Bester, J., and Kos, A. (2011). An introduction to artificial neural networks. *Methodological Advances and Biomedical Applications*.

Chen, A.-S., Leung, M. T., and Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the taiwan stock index. *Computers & Operations Research*, 30(6):901–923.

De Haan, L. and Ferreira, A. (2007). *Extreme value theory: an introduction*. Springer Science & Business Media.

Friederichs, P. (2007). An introduction to extreme value theory. Meteorological Institute; University of Bonn.

Herera, R., Rodriguez, A., and Pino, G. (2017). Modelling and forecasting extreme commodity prices: a markov-switching based extreme value model. *Energy Economics*.

Hsu, C.-W. and Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE transactions on Neural Networks*, 13(2):415–425.

Index, M. (2019). Yahoo Finance API. `https://www.marketindex.com.au/yahoo-finance-api`. [Online; accessed Aug-2019].

Investopedia (2018). Financial Asset. `https://www.investopedia.com/terms/f/financialasset.asp`. [Online; accessed 01-Aug-2019].

Liu, J. and Zio, E. (2018). Integration of feature vector selection and support vector machine for classification of imbalanced data. *Applied Soft Computing Journal*.

Mandelbrot, B. B., Fisher, A. J., and Calvet, L. E. (1997). A multifractal model of asset returns. *Cowles Foundation Discussion Paper*.

McNeil, A. J. (1999). Extreme value theory for risk managers. *Departement Mathematik ETH Zentrum*.

Shah, V. H. (2007). Machine learning techniques for stock prediction. *Foundations of Machine Learning— Spring*, 1(1):6–12.

Tang, F., Adam, L., and Si, B. (2018). Group feature selection with multiclass support vector machine. *Neurocomputing*.