Insert Dissertation Title Here

UP893057

School of Computing

Final Year Research Project

April 4, 2022

Abstract

No more than 300 words summarizing this dissertation.

Table of Contents

Acknowledgements

Thanks.

Chapter 1

Introduction

1.1 Overview

Stock markets are often thought to be irrational or in some cases random, which suggests that markets are not made based on fundamental factors alone. In a 1996 speech, former Federal Reserve chair, Alan Greenspan questioned "... how do we know when irrational exuberance has unduly escalated asset values ..." (Greenspan, 1996), suggesting that there may be psychological factors that affect markets such as positive feedback loops.

This project intends to identify if a stock index's price can be predicted ahead of time, and which factors contribute the most to an accurate prediction. The intended audience for such a project can vary from the institutional to the retail investor to identify opportunities in the market by using only publicly available data rather than proprietary information.

1.2 Aims

This project aims to produce an artificial intelligence model that attempts to predict the following day's price of a stock index. This will allow the

project to research if a stock index's price can be predicted from the day prior; the factors which affect a stock index's price and which factors are most important in predicting a stock index's price.

1.3 Objectives

The objectives are as follows:

- Identifying key factors that can affect stock market performance.

- Collecting above mentioned factors, where the historical data is publicly available and free to use (either public domain, non-commercial licenses, etc).

- Transforming collected data into normalised, or otherwise modified, data that can easily be used to identify patterns, either by humans or by computational neural networks.

- Identifying and applying correct techniques to train a computational neural network (or other suitable technology) to understand which factors are most important when predicting stock market performance.

- Filtering through training data to ensure there are no biases in play.

- Ensuring that the network will be efficient enough to be trained on hardware that is available today, and if possible on consumer hardware such as Nvidia RTX 3000 series GPUs.

- Understanding outputs of the computational neural networks and presenting the output data learned in a way that is clear.

- Interpreting overall results to identify if it is possible to predict a stock market's index direction and/or price for any given day

- Evauluating the results to identify accuracy, limitations and biases

1.4 Constraints

- May be difficult / expensive to get historical data

- May be difficult / expensive to get reliable data

- Access to currently publicly available data may be removed, or otherwise unavailable

- May not have the correct technical skills to sufficiently carry out the projects

- May not have sufficient time to carry out the full objectives of the project

- Hardware available may not be feasible to run / train complex neural networks on

Even if the project was successful for now, it may not be suitable for these reasons:

- May not work in extreme market events

- May not work if new market measures are placed

- May not work if new economic measures are placed

- Access to reliable data may be restricted, removed or otherwise unavailable in the future

Chapter 2

Literature Review

2.1 Introduction

As the world advances in technology, an increasing number of decisions are dependent on computational models; including financial decisions. This includes, but is not limited to, decisions for money management, risk management, economics, and investing. This literature review will discuss: the feasibility of predicting stock market prices; different artificial intelligence research methods used in financial modelling, particular those that relate to investing and stock market indices; and the different input features used in these models.

2.2 Feasibility of predicting stock market index prices

2.2.1 Efficient Market Hypothesis

Over the year, many people have attempted to predict stock markets in order to profit from price movements as they occur. However, some have theorised that it is not possible to accurately predict prices due to the there being an innumerable number of factors that can affect a stock market, and these factors would already be a function of the current market's price, therefore make it impossible to forecast the price. (Bachelor, 1900, p. 1) This is the basis of the Efficient Market Hypothesis, where an asset price fully reflects the information that is available at any given time.

2.2.2 Empirical evidence of market (in)efficiencies

A 1973 book explains the term 'random walk' as "A random walk is one in which future steps or directions cannot be predicted on the basis of past actions", and continues to describe how when applied to the stock market that "short-run changes in stock prices are unpredictable. [and] Investment advisory services, earnings forecasts, and complicated chart patterns are useless" (Malkiel, 1973, p. 24). A reflection paper by the same author 30 years on, attempts to validate that the efficient market hypothesis still holds true in the 21st century by comparing actively managed funds to tracker funds: suggesting that markets are likely efficient, due to the fact that most actively managed funds are outperformed by a index tracker fund that follows the S&P 500 index; the data presented suggests that most experts (68% to 90% over various time periods) cannot outperform the index whether it be from a one year, three year, five year, ten year or twenty year holding period leading up to 31st December, 2003 (Malkiel, 2005). The author suggests that if markets were inefficient, fund managers should easily be able to outperform the index.

In order to prove, or disprove, the Efficient Market Hypothesis, many studies have been carried out with varying results. In this literature review we will focus on the US markets and the different results the studies have found. Using data from 1964 to 2003 for the US stock market, one study has found markets to be efficient using a threshold autoregressive model and a unit root process (Narayan, 2006). Another empirical look into equity premium prediction, that looks specifically at the US S&P 500 index, using various regression models with additional variables described such as dividends, earnings price ratio, book value, net issuing activity and more has concluded that "the equity premium has not been predictable" (Goval & Welch, 2004, abstract) which would agree with the efficient market hypothesis.

However, other recent studies have suggested the efficient market hypothesis may not hold true. A paper that looks into returns of the US Dow Jones Industrial Average (DJIA) index from 1928 to 2012 showed that while for autocorrelation tests on daily and weekly intervals suggest market efficiency, the same cannot be said for monthly and annual intervals (although the degree of correlation is low). Furthermore, the paper posits that autocorrelation tests are not sufficient to determine dependencies, and puts forwards another type of test called 'run tests' to show that markets are also inefficient in daily and weekly intervals (Sewell, 2012). Another paper that analyses data from 1999 to 2007 of various markets, including the US as well as other developed and developing countries, using a unit root process has stated that their findings show that "real stock price indices are stationary processes that are inconsistent with the efficient market hypothesis" (Lee et al., 2010, p. 1).

2.3 Types of artificial intelligence methods used

Different studies have utilised different artificial intelligence models to predict the price of stock market indices, with varying degrees of accuracy.

One study using multilayer perceptrons (MLP) to predict the daily direction of an exchange traded fund (ETF) that tracks the US S&P 500 stock market index suggests that the model has an accuracy of up to 60% (Zhong & Enke, 2019). Whereas another study has suggested they can predict the price, not just direction, of the same stock market index with an accuracy of 76% using support vector machines (SVMs) and reinforcement learning (Shen et al., 2012). As there is a difference of 16%+ in of accuracy in the same stock market index (S&P 500), and the higher accuracy model suggests that support vector machines may provide an improvement over feed-forward neural networks such as MLPs. Another study comparing the performance of MLPs and SVMs agrees that SVMs are superior as, even though both models were able to predict the direction of the S&P 500 index, the MLP model had a maximum error difference of 15% over a 45 day period, whilst the SVM model had a maximum error difference of 6% in their test cases(Sheta et al., 2015). This is further supported by another study comparing various artificial intelligence methods on the Indian Nifty50 index with the SVM model outperforming a back-propagation neural network model, which is somewhat similar to the MLP model, by 5.51% (Kumar & Thenmozhi, 2006).

Further studies have been done on using hybrid approaches to predict stock markets. One study using an approach with a convolutional neural network (CNN) and three long short-term memory (LSTM) networks found that the accuracy of predicting weekly directions of the S&P 500 index was 66.6%, which was greater than with SVMs or CNNs alone, with those models achieving 62.0% and 59.3% respectively (Hao & Gao, 2020). Another study agrees with these findings, with their results also suggesting a model built with CNNs and bi-directional LSTMs are able to outperform SVMs (Eapen et al., 2019). However, the same cannot be said for all types of hybrid neural network models; one study comparing MLPs and hybrid networks on the US Nasdaq index has found that the MLP performs better, albeit slightly with a difference of 0.26% in mean absolute deviation, than an approach with a hybrid model of MLP and Generalized Auto-Regressive Conditional

Heteroskedasticity (GARCH) (Guresen et al., 2011).

2.4 Input features for intelligence methods

The studies mentioned in the previous section use different input features for their studies; where the amount of input features ranges from just a single type of input to many various macro-economic factors. One of the hybrid approaches has only stated using the 'daily closing price dataset of the S&P 500 index' in their CNN-LSTM model (Hao & Gao, 2020). Another study using the hybrid approach slightly expands on this by using additional factors related to the index: the opening and closing prices of the S&P 500 index, the low and high prices of the day, and the trading volume. However, these studies have not considered any macro-economic factors. One study expands on this further by also including other indexes across the world as inputs, as well as various currency rates in addition to commodities such as oil and metals (Shen et al., 2012). Other studies expand on this further, for example one study with 27 input features also includes US Treasury yield rates and bond yields (Sheta et al., 2015) and another looks into additional factors for a total of 60 input features including certificate of deposit rates, and term and default spreads (Zhong & Enke, 2019) - though the latter study explains that with principal component analysis, the model peaked in terms of accuracy with 31 input features.

2.5 Conclusion

To conclude, there are differing opinions on whether or not markets are efficient, and therefore whether or not they are predictable. Some economists use the fact that experienced fund managers are rarely able to beat the market as proof that markets are efficient, and there are studies that agree with this hypothesis. However, there are also other studies that disagree presenting their findings that show there is some evidence of market inefficiencies in

either the short term or long term.

Regardless of markets being efficient or not, numerous artificial intelligence methods – including MLPs, SVMs, CNNs as well as hybrid approaches using CNNs and LSTMs – have been employed to varying degrees of success. While some studies look at predicting just the direction, and others attempt to predict the price, most have claimed to be able to do so with above 50% accuracy. From the studies researched here, it can be assumed that in general hybrid approaches involving CNNs and LSTMs outperform SVMs, which outperform MLPs.

These studies use a diverse set of input features in order to build their models, with some using features only directly related to the index and some using extended datasets including inputs from macroeconomic sources in order to build models that take account external factors that may affect the index's price, though there is evidence to suggest that not all of the input features are required to build a good model to predict the price of an index.

Chapter 3

Research Questions

Chapter 4

Research Design

Chapter 5

Artefact Design

Chapter 6

Results

Chapter 7

Evaluation

Chapter 8

Project Management

Chapter 9

Conclusion

Chapter 10

Future Work

References

Bachelor, L. (1900). Speculation theory. *3rd series, 17*, 21–86. https://doi.org/10.24033/asens.476

Eapen, J., Bein, D., & Verma, A. (2019). Novel deep learning model with cnn and bi-directional lstm for improved stock market index prediction. *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, 0264–0270. https://doi.org/10.1109/CCWC.2019.8666592

Goval, A., & Welch, I. (2004). A comprehensive look at the empirical performance of equity premium prediction. *NBER Working Paper Series, 10483*. https://doi.org/10.3386/w10483

Greenspan, A. (1996). The challenge of central banking in a democratic society [Remarks by Chairman Alan Greenspan at the Annual Dinner and Francis Boyer Lecture of The American Enterprise Institute for Public Policy Research, Washington, D.C. [Accessed: 2021 10 29]]. http://www.federalreserve.gov/boarddocs/speeches/1996/19961205.htm

Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications, 38*(8), 10389–10397. https://doi.org/https://doi.org/10.1016/j.eswa.2011.02.068

Hao, Y., & Gao, Q. (2020). Predicting the trend of stock market index using the hybrid neural network based on multiple time scale feature learning. *Applied Sciences, 10*(11), 3961. https://doi.org/10.3390/app10113961

Kumar, M., & Thenmozhi, M. (2006). Forecasting stock index movement: A comparison of support vector machines and random forest. *Indian institute of capital markets 9th capital markets conference paper.*

Lee, C.-C., Lee, J.-D., & Lee, C.-C. (2010). Stock prices and the efficient market hypothesis: Evidence from a panel stationary test with structural breaks. *Japan and the World Economy, 22*(1), 49–58. https://doi.org/https://doi.org/10.1016/j.japwor.2009.04.002

Malkiel, B. G. (2005). Reflections on the efficient market hypothesis: 30 years later. *Financial Review, 40*(1), 1–9. https://doi.org/https://doi.org/10.1111/j.0732-8516.2005.00090.x

Malkiel, B. G. (1973). *A random walk down wall street : The time-tested strategy for successful investing.* W.W. Norton; Company.

Narayan, P. K. (2006). The behaviour of us stock prices: Evidence from a threshold autoregressive model. *Mathematics and computers in simulation, 71*(2), 103–108.

Sewell, M. V. (2012). The efficient market hypothesis: Empirical evidence. *International Journal of Statistics and Probability, 1*(2). https://doi.org/10.5539/ijsp.v1n2p164

Shen, S., Jiang, H., & Zhang, T. (2012). Stock market forecasting using machine learning algorithms. Retrieved December 8, 2021, from http://cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearning.pdf

Sheta, A., Ahmed, S., & Faris, H. (2015). A comparison between regression, artificial neural networks and support vector machines for predicting stock market index. *International Journal of Advanced Research in Artificial Intelligence, 4,* 55–63. https://doi.org/10.14569/IJARAI.2015.040710

Zhong, X., & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation, 5.* https://doi.org/10.1186/s40854-019-0138-0

Appendix A

First Appendix