**Deep Insights into Market Efficiency: A Neural Network Based Approach**

*A Dissertation submitted to CHRIST (Deemed to be University) in partial fulfillment of the requirements for the Award of the Degree of*

MASTER OF ARTS

IN

APPLIED ECONOMICS

by

**Yeshwant S Nair**

(2239216)

Under the Supervision of

Dr. Vineeth Mohandas

Assistant Professor of Economics



DEPARTMENT OF ECONOMICS

CHRIST (Deemed to be University)

YESHWANTHPUR CAMPUS

BENGALURU

APRIL 2024

**TABLE OF CONTENTS**

[List of Figures 4](#_Toc164153398)

[CHAPTER 1 - INTRODUCTION 2](#_Toc164153399)

[1.1 Prelude to the Study 2](#_Toc164153400)

[1.2 Problem Statement 3](#_Toc164153401)

[1.3 Research Questions 3](#_Toc164153402)

[1.4 Objectives of the study 4](#_Toc164153403)

[1.5 Research Design and Methodology 4](#_Toc164153404)

[1.6 Organization of the Chapters 5](#_Toc164153405)

[1.7 Limitations of your Study 5](#_Toc164153406)

[1.8 Contribution 6](#_Toc164153407)

[CHAPTER 2 7](#_Toc164153408)

[2.1 Introduction 7](#_Toc164153409)

[2.2 Literature Review 7](#_Toc164153410)

[2.3 Methodology 9](#_Toc164153411)

[2.4 Results and Findings 11](#_Toc164153412)

[2.5 CONCLUSION 13](#_Toc164153413)

[CHAPTER 3 14](#_Toc164153414)

[3.1 Introduction 14](#_Toc164153415)

[3.2 Literature Review 14](#_Toc164153416)

[3.3 Methodology 16](#_Toc164153417)

[3.4 Results and Findings 18](#_Toc164153418)

[3.5 Conclusion 20](#_Toc164153419)

[CHAPTER 4 21](#_Toc164153420)

[4.1 Introduction 21](#_Toc164153421)

[4.2 Methodology 21](#_Toc164153422)

[4.2.1 Mean Equation 22](#_Toc164153423)

[4.2.2 Variance Equation 24](#_Toc164153424)

[4.3 Results and Findings 26](#_Toc164153425)

[4.4 Conclusion 27](#_Toc164153426)

[CHAPTER 5 - CONCLUSION 28](#_Toc164153427)

[5.1 Introduction 28](#_Toc164153428)

[5.2 Summary of the Thesis 28](#_Toc164153429)

[5.3 Major Findings 29](#_Toc164153430)

[5.4 Conclusion of the Study 30](#_Toc164153431)

[5.5 Directions for Further Research 30](#_Toc164153432)

[**References** 32](#_Toc164153433)

# List of Figures

Figure 1 Forecast of Neural Network for Testing the Weak Form of EMH 11

Figure 2 Close up of Neural Networks Movement against the Actual Price for Weak Form of EMH 12

Figure 3 Forecast of Neural Network for Testing the Semi-Strong Form of EMH 18

Figure 4 Close up of Neural Networks Movement against the Actual Price for Semi-Strong Form of EMH 19

Figure 5 Checking for Stationarity of Closing Price 22

Figure 6 Showing the Stationarity of Return Variable 23

Figure 7 Forecast Accuracy of the ARCH forecast 26

# CHAPTER 1 - INTRODUCTION

## 1.1 Prelude to the Study

The most influential theory in the field of financial economics is the Efficient Market Hypothesis (EMH), which suggests financial markets are efficient, as they reflect and incorporate all relevant information into the securities prices. Eugene Fama and Paul Samuelson developed this concept in the 1960s and 1970s. An Efficient market implies that it is near impossible for an investor to achieve higher-than-average returns on a consistent basis just by merely analyzing all the available information,as asset prices already incorporate all this information.

A fundamental premise of EMH is that stock prices tend to follow a random walk, implying that the securities prices vary randomly and unpredictably. If the prices of the securities incorporate all the available information in the financial market, then any fluctuation in the prices of the securities must be an instantaneous reaction to any new sources of information, and the new information must be unpredictable in nature. Thus, the historic movement of prices cannot be used to predict future movement of the securities.

However, the question as to whether the markets are really efficient has not been conclusively proved. Critics of the EMH argue that there are several instances in which the market appears to be inefficient, one notable criticism associated with the EMH is the behavioral factors, such as investors psychology and cognitive biases, which often has led to deviations in the market efficiency.

Embracing the EMH, Andrew Lo (2004) propounded the Adaptive Market Hypothesis, which expands the idea of market efficiency by incorporating insights from behavioral finance and acknowledges the role of varying market conditions. The basic tenets of the adaptive market are that (i) individuals operate in their self-interest, (ii) individuals make mistakes, (iii) Individuals learn from those mistakes, and they learn, adapt, and innovate, (iv) natural selection operates on the individual's institutions and markets as they experiment and succeed or fail (v) the dynamics of the financial markets are determined by this evolutionary process. According to AMH, investors are assumed to exhibit rational behavior majority of the time, but not entirely.Implying that investors often participate in satisficing behavior in lieu of maximizing, resulting in them often developing certain market heuristics which is based on profit and loss.This often results in markets primarily functioning rationally majority of the time, unless major fluctuations or economics shocks occur which renders the previously adaptive heuristics to become maladaptive due to evolutionary changes in the market’s dynamics. Hence resulting in EMH to lose its validity during scenarios of rapid transitions , increased stress and abnormal situations.s. Most often, it has been discovered that market conditions such as liquidity and crisis have an influence on the evolution of market efficiency and momentum profits. Periods associated with higher liquidity are often related to high market inefficiencies and significant momentum profits. Similarly, during the periods of uncertainty, it can be observed that market inefficiencies are lower and t momentum profits are non-existent because of market correction. AMH can better describe the real situations of financial and related markets than EMH. However, it also has a limitation as it is predominantly qualitative in nature, and there is a need to establish a quantitative model of AMH.

## 1.2 Problem Statement

This begs us to ask the question whether markets are truly efficient or not. If not, can we use Artificial Intelligence and Machine Learning models to predict future movement in prices using Historical prices and information to check the soundness of the Efficient market hypothesis in both the weak and Semi-strong forms.

## 1.3 Research Questions

This research seeks to answer the following questions:

1. Can deep learning models be used to predict future movement in prices of the NIFTY 50 index? If yes, how do they compare to traditional econometric models?
2. Does incorporating historical news information into the deep learning models serve to improve the predictability of future price movements?
3. Are the profit opportunities that appear from time to time have any correlation with the investor sentiments at that point in time?
4. Based on the findings from the above three questions, do we validate or invalidate the weak and semi-strong forms of EMH?

## 1.4 Objectives of the study

The main objective of this research is to critically evaluate the validity of the Efficient Market Hypothesis by leveraging various deep learning techniques and also accounting for the investor sentiment in the modeling process.

The study specifically aims to achieve the following research objectives:

1. To evaluate the scope of deep learning models in predicting the future movement of the NIFTY 50 index using past prices and to compare their performance against the traditional econometric models such as ARIMA ARCH and GARCH.
2. To use Natural Language Processing, to investigate the role of past news formations in predicting the future prices of NIFTY 50.
3. To assess the validity or invalidity of EMH forms in the NIFTY 50 Index based on the above objectives

## 1.5 Research Design and Methodology

This research aims to determine the validity of the Efficient Market Hypothesis(EMH), particularly in the context of the Nifty 50 index. In order to do this, a comparative analysis will be done between the budding Deep learning models with the traditional econometric methods to see which method yields the slightest error in predicting the future movement of the index. The nature of the research adopted is quantitative, which involves leveraging the various statistical and machine learning techniques available to analyze the financial data available in order to facilitate a rigorous examination of the existence of market efficiency and provide a solid empirical basis for assessing the effectiveness of various modeling approaches to index price prediction.

The time period of the study is from January 2018 onwards till the December of 2023.The data sources for the study involves the weekly Opening and Closing prices of the NIFTY 50 index which is obtained from the National Stock Exchange (NSE), while the news articles pertaining to the financial news are obtained from Money Control, a financial portal. The obtained underwent preprocessing and relevant feature extraction. To incorporate the historical news information for analysis, the FINBERT NLP model was utilized to categorize the sentiments associated with the news information into three broad categories : positive, neutral and negative.

In order to assess the weak form of EMH, the historical prices will be incorporated into the Feedforward Neural Network (FNN) to predict future movement in index movement. Further to check the validity of Semi-strong form EMH, news sentiment will be incorporated in addition to the historical prices to assess the performance of the future prediction.Finally a comparison of FNN with the traditional ARCH model will be done on the basis of Mean Absolute Error (MAE) value to compare their performance in price prediction. The primary metric that is used to evaluate the different model is MAE as it provides a measure of the average absolute error between the actual and the predicted values, enabling a straightforward comparison of the forecast accuracy.

## 1.6 Organization of the Chapters

The organization of the chapters for the study will involve in total 5 chapters. The first chapter of the study relates to the introduction to the study which will provide a prelude to the study and highlight the objectives, methodology and limitations of the study. The second chapter of the study corresponds to the first empirical chapter of the study. This chapter will look into the Validity of the weak form of EMH using deep learning models and will be divided into four parts i.e. the introduction, literature review, methodology, results and discussion and finally the conclusion. The third chapter of the study will follow the same format as the second chapter and will look into the validity of the Semi-strong form of EMH. Chapter 4 of the study will also be in line with the format followed in Chapter 2 and 3 and will involve the comparison of the traditional ARCH model with the Deep learning model. The final chapter of the study will bring together all the findings and summarize the entire study

## 1.7 Limitations of your Study

The limitations of the study centers around the quality and quantity of news available in the public domain. Since the study confines only to the financial news available on the Money Control portal, it does not take into account news information available in other domains. Moreover, the study only focuses on the historical prices and news information’s associated with the NIFTY 50 index and the 50 companies contained within, it does not take into considerations the past financial statement of the individual companies contained within the NIFTY 50.

## 1.8 Contribution

The primary goal of this study is to provide a better understanding of the EMH in the context of NIFTY 50 index. To do so the study will be relying upon deep learning techniques to assess the validity of EMH in the context of Indian markets. This method of adopting deep learning techniques to assess the validity of EMH is a departure from the traditional econometric methods which are prevalent in the existing literature. The concept of incorporating news sentiments into the modeling process is something that falls outside the purview of traditional econometric methods. Thus, this study will go a long way in broadening horizons when it comes to EMH.

# CHAPTER 2

## 2.1 Introduction

This chapter will focus on looking into the validity of the weak form of EMH using Deep learning models. The weak form of EMH states that the present prices of a security reflect all the attributes associated with the past prices, therefore preventing investors from reaping any sort of abnormal profits from these types of securities. This in short implies that the movement of security prices follows a random walk. The advent of deep learning models opens up a new avenue for exploration into the validity of the weak form of EMH. This is attributed to the fact that deep learning models are advanced enough to pick up on various underlying patterns that are exhibited by the movement of prices.

In this chapter the weekly opening and closing prices of the NIFTY 50 index from 2018 to 2023 is used in the deep learning models to predict future movement of the NIFTY 50 index.

## 2.2 Literature Review

The question of whether or not the markets are efficient to date still remains to be answered. However, the hypothesis related to efficient capital markets can be traced back to the work of Fama (1965) and Samuelson (1965). Eugene Fama's article "Efficient Capital Markets: A Review of Theory and Empirical Work" (1970) defined efficient markets and all its three forms. Fama defines *efficient markets* as markets consisting of large number of rational, profit-maximizing individuals who compete with each other to predict the future trend of market values of individual securities, where current information is readily available to all the participants. Fama divided the EMH into three forms i.e. Weak, Semi-strong and Strong forms of EMH.

The weak form of EMH is attributed to a market in which the present price of the securities exhibits all the historical financial information. Due to this, investors will find it difficult to obtain abnormal profits on a consistent basis from investing in these types of securities. This EMH implies that prices tend to follow a random walk.

The Semi-strong EMH on the other hand assumes that the securities reflect all the publicly available information in the market at any point in time, as well as historical prices and information. Moreover, the prices associated with the securities change rapidly and without bias to incorporate any new information that is made available. In such a scenario, neither Technical nor Fundamental analysis will help investors determine where to park their funds.

The Strong EMH incorporates the weak and semi-strong EMH and all private information regarding a particular financial asset.

However, the subsequent studies spanning over decades invalidated the hypothesis regarding the Semi-strong and strong forms of EMH, while opinions are split regarding the weak form of EMH, including the random walk theory, as various literature has found that stock returns do have predictive power and the EMH does not always hold.

Borges (2010), while looking into the EMH in European stock markets, found that Greece and Portugal initially showed strong positive first-order autocorrelation in daily return, which is indicative of the return persistence, but this declined from 2003 to 2007. This observed drop corresponded with the findings of fewer runs than expected and the tendency of variance ratios (VRs) to grow with k, implying better market efficiency. This corresponds to these countries' transitions from emerging to developed markets in the early twenty-first century.

In order to incorporate the inefficiencies that are associated with the market, Andrew Lo (2004) propounded the Adaptive Market Hypothesis to enable market efficiency and inefficiencies to co-exist in an intellectually consistent manner by applying the principles of - competition, adaptation, and natural selection.

Kohzadi et al. (1996) compared the ANN with the ARIMA model for forecasting monthly data. The results showed that ANN forecasts provided a considerably more accurate prediction consistently and were able to capture more turning points compared to the ARIMA model. Sharda and Patil (1992), in their research, determined in their experiment that seasonality of the time series is not a factor hindering the performance of ANNs and that ANNs inherently accounted for seasonality.

Several empirical studies suggested that ANNs were found to be better fit at forecasting monthly and quarterly series as these forms of data often tend to exhibit irregularities, and ANNs are good at detecting these irregular patterns that often get concealed by such noisy factors in the complex system.

Balkin and Ord (2000) stated that ANN usually performs better when we have a sufficiently long series to detect the non-linearity and provide reliable estimates of the parameters. When these conditions are satisfied, the ANN model outperforms more straightforward methods.

Hansson, M., & Olsson, C. (2017) discussed the pivotal role of the Rectified Linear Unit (ReLU) activation function particularly in the Feedforward neural Networks. Their study underscored the simplicity of ReLU in the modeling process and how it helps overcome the vanishing gradient problem making it a really good fit for deep learning models.

Hansson, M., & Olsson, C. (2017) in their work used ReLU activation function to introduce non-linearity in their modeling process, they did so because the ReLU activation function helps prevent saturation of the gradient when the neural networks are deep thereby preventing the problems associated with the vanishing gradient problem. This ability of ReLU proves to be beneficial when dealing with multi-layers models.

Zaheer, R., & Shaziya, H. (2019, January) in their work showed the superiority of the Adam operator in the testing phase across all their datasets during which they compared the performance of Adam with various other optimization methods such as SGD, Momentum, Adagra and RMSProp and concluded that Adam outperformed all the other methods in the training phase and established the highest testing results.

Jais, I. K. M., Ismail, A. R., & Nisa, S. Q. (2019) discussed the impact of Adam on wide and deep neural networks. They found that the use of Adam resulted in improved performance of the neural network which was backed by higher accuracy and better metrics as opposed to when Adam was not used. Their work suggested that the Adam optimization function was ideal for wide and deep neural networks.

## 2.3 Methodology

To assess the validity of the weak form of EMH in the context of the Indian Market, the weekly opening and closing price of NIFTY 50 index were obtained from the National Stock Exchange portal. The time period for the study ranges from 2018 January till 2023 December. The rationale behind choosing the beginning date to start from 2018 has to do with the fact that 1st January falls on Monday.

The statistical software used to model the deep learning model is Python, in which tensorflow, a free and open source software library was used to train the neural networks for future predictions in the movement of prices.

The first step in the modeling process consisted of dividing the dataset consisting of the opening and closing price into a training set and test set. The training set consisted of the first 80% of the dataset i.e the first 249 weeks, whereas the testing set consisted of the remaining 20 % of the dataset.The logic behind this is to train the neural network on the training set and assess the model's validity against the testing set. The metric that is used to assess the accuracy of the model is confined to the mean absolute error (MAE) and Mean Average Percentage Error (MAPE).

The next step in the modeling process involves building the neural network. The neural network will take the input of weekly opening price and predict the output of closing price. The neural network developed consists of one input layer, 3 hidden layers and one output layer. The first hidden layer consists of 100 nodes, followed by the second hidden layer containing 50 nodes, which is further followed by the 3rd hidden layer consisting of 10 nodes. The activation function involved in all these hidden layers is Rectified Linear Unit (ReLU), this is because the ReLU

activation function helps prevent saturation of the gradient when the neural networks are deep thereby ensuring the prevention of the vanishing gradient problem.The neural network developed takes the form of Feedforward Neural Network (FNN). The prime characteristic that is associated with neural networks is its ability to learn the hidden patterns contained within the data set that it does through forward and back propagation and adjusting its weights and biases accordingly. The optimizer function adopted for the neural network to learn in this research is the Adam, short for “adaptive moments”, which will be designed to update the network iteratively at a rate of 0.001. In order to ensure that the best weight is stored a callback function will be used in the modeling process.

After developing the neural network with one input layer, three hidden layers and one output layer and setting its activation and optimization function, the neural network is fitted to the training dataset where in the input layer will be feeded with the weekly opening prices and try to predict the weekly closing prices.The said model will be trained for 1000 epochs, based on the actual weekly closing prices and predicted closing prices the model will adjust its weights and biases accordingly in order to save the model with the least MAPE.

After fitting the model with the training dataset,the model's accuracy will be assessed based on its performance on the testing dataset. The metric for assessing the accuracy of the neural network model against the testing data set will be the same as the training dataset i.e, MAE and MAPE

## 2.4 Results and Findings

Diagram 1, shows the visual representation of the training and testing set for the neural network. The region following to the left of the dotted line, represented by the yellow shaded region depicts the data on which the neural network was trained on. The region to the right of the dotted line , represented by the pink shaded region, depicts the testing data against which the neural network is evaluated.The X-axis represents the weeks whereas the Y-axis represents the Weekly closing price of the NIFTY 50 index.

|  |
| --- |
| Figure 1 Forecast of Neural Network for Testing the Weak Form of EMH |

In diagram 2, the dotted orange line represents the movement of the closing price predicted by the Feedforward neural network. The diagram depicts that the predicted price more or less moves along the same trend line of that of the actual closing price.

Dwelling deeper into the metrics paints a clearer picture. The neural network after being evaluated against the testing data yields a MAE of 212.311, which indicates that on an average the model’s prediction deviates from the actual values by 212.311. To understand the accuracy associated with the model MAPE is a better metric, the neural network exhibited MAPE of 1.1353, implying that on an average the absolute percentage difference between the predicted and actual value is approximately 1.1353%.

|  |
| --- |
|  |

Figure 2 Close up of Neural Networks Movement against the Actual Price for Weak Form of EMH

## 2.5 CONCLUSION

The objective of this chapter was to look into the validity of the weak form of EMH in the context of the NIFTY 50 index using neural networks. Based on the finding represented in

section, it can be concluded that the by utilizing the past movements in the NIFTY 50 index, it was possible to predict future movements in NIFTY 50 index prices by a Mean Absolute Percentage Error (MAPE) of 1.1353 and a Mean Absolute Error of 212.311.

This finding concludes the absence of a weak form of EMH in the NIFTY 50 index, as the neural network developed outperformed the random walk hypothesis, which is a central concept of the weak form of EMH. Thus, challenging the notion that past price movements contain no information for predicting future movements in the NIFTY 50 index.

# CHAPTER 3

## 3.1 Introduction

The semi-strong form of EMH states that past information and historical prices are accurately reflected in the current prices of securities. This chapter will assess the presence of Semi-strong EMH in the context of the NIFTY 50 index, a leading benchmark for the Indian Equity Market, using neural networks.

This chapter will entail looking into how the Weekly opening prices, along with news information related to the 50 companies incorporated within the NIFTY 50 index dating from 2018 to 2023, can be incorporated into the deep learning model to predict weekly closing prices of the NIFTY 50 index.

By incorporating news sentiments along with historical prices, the chapter will attempt to imitate the decision-making process of the market participants, who rely on both quantitative and qualitative information in their trading strategies. This will be done with the help of neural networks, which will attempt to unravel the hidden patterns, nonlinear relationships, and latent information contained in the data, thereby assessing the validity of semi-strong EMH in the context of the NIFTY 50 index.

## 3.2 Literature Review

Lo (2004) stated that the market's predictability fluctuates over time, which is often attributed to the changing market conditions and environment. However, Lo did not specify the relationship that often tends to exist between the market conditions and its predictability. Andrew Urquhart and Frank McGroarty (2016) found that the return predictability fluctuated over time in each market, with each of the return series going through a period of significant predictability and periods associated with no predictability. Furthermore, they found that different markets often experience significant predictability at different points of time, suggesting that markets evolve differently with time. This implies that markets adapt over time and interact differently with varying market conditions.

Moreover, it was found that profit opportunities tend to appear from time to time and vanish once the investors exploit them. Interestingly, it was found that the momentum profits, more often than not, was highly correlated with the degree of market efficiency and the market conditions. Thus implying that investors have the potential to capitalize on these inefficiencies and specific market conditions that arise from time to time using various trading strategies.

Groenewold, N., & Kang, K. C. (1993), in their analysis of the Australian stock market data from 1980, proved the presence of a semi-strong form of EMH by proving that the past returns or lagged values of unexpected variables such as inflation, money supply, and exchange rates did not have any bearing on the share returns. Their work suggested market efficiency but concluded that more work was needed to resolve the question of semi-strong efficiency in the Australian stock market.

Hussin, B. M., Ahmed, A. D., & Ying, T. C. (2010) in their work analyzed the impact of announcements related to dividends and earnings on the stock prices listed on the Malaysian stock exchange. This was done to test the efficiency related to the semi-strong form of EMH. Their study revealed that announcements related to dividend increase correlated with positive abnormal return, which was also true for the inverse of it, suggesting some evidence of semi-strong efficiency in the Malaysian Stock market, with stock prices tending to adjust in line with dividend and earning announcements.

Mallikarjunappa, T., & Dsouza, J. J. (2013) looked into the reaction of the Indian Stock market to the quarterly earning announcements and found that the Indian stock market is not efficient when it comes to the semi-strong form as investors were able to reap returns post the earning announcements, which hints at the slow adjustments of the stock prices to publicly available information.

Mehtab and Sen (2019) incorporated public sentiment from social media into the neural network. They found that the sentiment-augmented neural network outperformed various other methods and conclusively proved that public sentiment acted as a significant input in predicting stock price movements.

Krauss, C., Do, X. A., & Huck, N. (2017) used advanced machine learning techniques, specifically deep neural networks, to explore their effectiveness in generating trading signals. Their findings showed that the model was able to generate significant profits, thus challenging the semi-strong EMH.

Araci, D. (2019) introduces a financial sentiment analysis, the FinBERT model based on the foundations set by the BERT, to tackle the NLP tasks that are associated with the financial domain. The FinBERT model showed an upper hand compared to other state-of-the-art machine learning models when it came to sentiment analysis relating to financial aspects.

Vargas, M. R., De Lima, B. S., & Evsukoff, A. G. (2017, June) highlighted the efficiency of neural networks in predicting the direction of stock prices using financial news. Their work emphasized incorporating a hybrid model, which incorporates the traditional technical analysis along with various text mining techniques to capture the various semantic and temporal patterns in the data.

## 3.3 Methodology

To assess the validity of the semi-strong EMH, the approach adopted follows the similar pattern that was followed in the chapter 1 that is building a neural network to assess the validity of the semi-strong EMH. The neural network developed consists of one input layer followed by 3 hidden layers and one output layer. Unlike chapter 1 where the input layer consisted of one node, here the input layer consists of 2 nodes. The two inputs to the neural network consist of weekly opening price of the NIFTY 50 index and weekly news sentiment.

In order to derive the weekly news sentiment, daily news for all the 50 companies listed within the NIFTY50 index was extracted from moneycontrol.com, a portal for news information related to stocks, business and the economy in general, from 2018 to 2023 . After the news is extracted, it is passed through FinBERT, a financial sentiment analysis model based on the BERT model, which categorizes the news sentiment into three categories of Positive, Neutral and Negative. If the news is positive, it is assigned a value of 1, if neutral 0, and if negative -1. These daily news sentiments are grouped into weekly format, to coincide with the format associated with the opening and closing price.

The next step in the modeling process involves fitting the weekly opening price of the NIFTY 50 index and the weekly News sentiment into the input layer of the neural network to predict the weekly closing price. Prior to this the dataset will be divided into a training and testing set, with the training set consisting of 80% of the dataset and the rest 20% as the testing set on which the accuracy of the model will be assessed on. The metric on which the model accuracy will be assessed are MAE and MAPE.

The next step in the modeling process will involve building the neural network, in which the neural network will consist of one input layer consisting of 2 nodes, followed by three hidden layers consisting of 100, 50 and 10 nodes respectively and finally an output layer consisting of one node. The activation function that will be used in all the three hidden layers will be Rectified Linear Unit (ReLU) as it will help overcome the vanishing gradient problem. Just like the model in chapter 2, the neural network in this chapter will also take the form of Feedforward Neural Network (FNN). The optimizer function that will be used to update the weights and biases of the model will be the Adam optimizer, which will update the weights of the model at a learning rate of 0.001. In order to ensure the model with the lowest MAE will be saved, a callback function will be incorporated into the modeling process

After the neural network is built, it will be fitted on the training dataset first, where in the input layer will be fed with weekly opening price and weekly news sentiment. This neural network will be trained for 1000 epochs and the model with the least MAE between the actual closing and predicted closing will be saved.

After training the neural network on the training dataset, the accuracy of the model will be assessed based on its performance against the testing dataset. Similar to chapter 2, the metric for assessing the accuracy of neural network on testing dataset will be MAE and MAPE

## 3.4 Results and Findings

Diagram 3, shown below shows the distinction between the training dataset and testing dataset. The region to the left of the dotted line, illustrated by the yellow shaded region, represents the training dataset, whereas the region to the right of the dotted line, characterized by the pink shaded region , represents the testing dataset. The X-axis of diagram 3 represents the weekly closing price of NIFTY 50 index and the Y-axis representing the Weeks respectively.

|  |
| --- |
|  |

Figure 3 Forecast of Neural Network for Testing the Semi-Strong Form of EMH

In both the diagrams, diagram 3 and diagram 4, the orange line represents the predicted movement of the closing price based on the FNN, whereas the black line represents the actual movement of the weekly closing price.

|  |
| --- |
|  |

Figure 4 Close up of Neural Networks Movement against the Actual Price for Semi-Strong Form of EMH

In diagram 4, we can see that the predicted price, i.e. the dotted orange line, follows the near same trend that is exhibited by the actual price, i.e. the solid black line. This newly modified neural network which incorporates news sentiments, when evaluated against the testing dataset exhibited a MAE of 213.9576 with the model showcasing an average absolute percentage difference between the predicted and actual price by approximately 1.1437 % MAPE.

## 

## 3.5 Conclusion

This chapter aimed to assess the validity of semi-strong EMH in the context of NIFTY 50 index using neural networks. Based on the findings represented in section 3.4, suggest that by incorporating the weekly opening prices and news sentiments it was possible to predict future movements in the weekly closing prices by a mean absolute percentage error of 1.1437 and a mean absolute error of 213.9576.

These results point towards the fact that the FNN model developed, which incorporates both weekly opening prices and news sentiments, outperforms the expectations associated with the efficient market hypothesis. Therefore, implying that the semi-strong form of EMH does not hold true in the context of the NIFTY 50 index, as evidenced by the predictive capabilities exhibited by the neural network model. This finding challenges the long-standing notion that all publicly available information, i.e. past prices and news information, are instantaneously reflected in asset prices, as proposed by the semi-strong form of EMH.

# CHAPTER 4

## 4.1 Introduction

In recent years, we have witnessed the revolutionization of predictive modeling processes thanks to the introduction and advancement of machine learning techniques, particularly in the field of deep learning and neural networks. The preceding chapters have already demonstrated the capability of neural networks in predicting the weekly closing prices of the NIFTY 50 index by using two different sets of predictors, i.e., weekly opening prices and weekly opening prices augmented with News sentiments. Therefore, an analysis must be made to compare the traditional econometric methods with upcoming neural networks.

The primary objective of this chapter is to assess the predictive capacity of neural networks by stacking them up against the traditional econometric methods.

## 4.2 Methodology

In order to assess the effectiveness of neural networks over the traditional econometric methods, this chapter will compare the neural networks developed in Chapters 2 and 3 with the Autoregressive Conditional Heteroskedasticity (ARCH) model. The methodology for developing the neural networks will be the same as those contained in sections 2 and 3.

In order to carry out the ARCH forecast, the dataset used in this analysis will involve the weekly closing price of the NIFTY 50 index. Similar to splitting the dataset into training and testing sets, the dataset will be split into two parts, where the estimated ARCH modeling will be done on the first 249 observations and forecasted ahead and evaluated against the next 63 observations of the actual closing price. Before moving ahead with the ARCH forecasting, tests need to be performed to assess the presence of ARCH effects in the dataset. The estimation of the ARCH model can be categorized into two parts, i.e., i) Mean Equation ii) Variance Equation.

### 4.2.1 Mean Equation

The first step in estimating the mean Equation involves checking the stationarity of the variable and the weekly closing price. To check the stationarity of the variable, graph analysis and an Augmented Dicky Fuller test will be conducted.

|  |
| --- |
|  |

Figure 5 Checking for Stationarity of Closing Price

From diagram 5, it is visible that the variable close\_price has a positive trend and intercept, hinting at the variable being non-stationary.

In order to confirm the existence of non-stationarity, an ADF Test will be conducted at the level. For variable close\_price, the ADF test at level reveals that the variable close\_price is non-stationary. However, when checked at the first difference, the variable close\_price becomes stationary, revealing that the variable close\_price is i1 in nature.

The second step is to estimate the mean Equation. For this, we will estimate an ARIMA (p,d,q) model and, based on the correlograms, select the order of “p” and “q”. In order to determine the correct order of “p” and “q” for the ARIMA model, the variable close\_price needs to be stationary. This is achieved by taking the first difference of variable close\_price. After taking the first difference of variable close\_price, we can see from Diagram 6 that the statistical properties of the series, i.e., Mean, variance, and Autocorrelation, do not change over time.

|  |
| --- |
|  |

Figure 6 Showing the Stationarity of Return Variable

In order to ensure parsimony and to make the mode simple, based on the results from the correlogram, an AR(6) model will be estimated.

After estimating the AR(6) model, it is evident that the AR components are statistically significant.

### 4.2.2 Variance Equation

The first step to developing the variance equation will be to check for the presence of ARCH effects in the AR(6) model. The ARCH heteroskedasticity at lag level 1 yields an Obs\*R-squared value of 21.35 and a p-value less than 0.01, which results in the rejection of the null hypothesis of homoscedasticity and concludes that ARCH(1) effects are present in the return variable

The next step in the process is to estimate the ARCH(1) model. As the mean equation model, including the constant (“c”) and an AR(6) component, is already determined, an estimate of the same using 1 ARCH effect will be done.

The estimated ARCH(1) equation will take the following form:

= 39.48458 - 0.112107 + (i)

Variance

= 68485.08 + 0.171429 (ii)

Where

is the returns at time t

is the conditional variance at time t

Here the equation (i) indicates that the closing price is dependent on the closing price six periods ago and exhibits a negative relationship.While equation (ii) suggests that the conditional variance is dependent on the conditional variance six periods prior and exhibits a positive relationship.

After estimating the ARCH (1), the next step will be to look into the model diagnostics, i.e., to check whether the ARCH effects still persist. To do this, an ARCH test will be conducted on the estimated output. The ARCH LM test reveals that the lag level 1, Obs\*R-squared value of 3.24, and a p-value greater than 0.05 result in accepting the null hypothesis of homoscedasticity.

The final step in this process is to forecast for the following 63 observations, compare the ARCH model forecasts with the actual observations, and compute their MAE and MAPE values.

## 4.3 Results and Findings

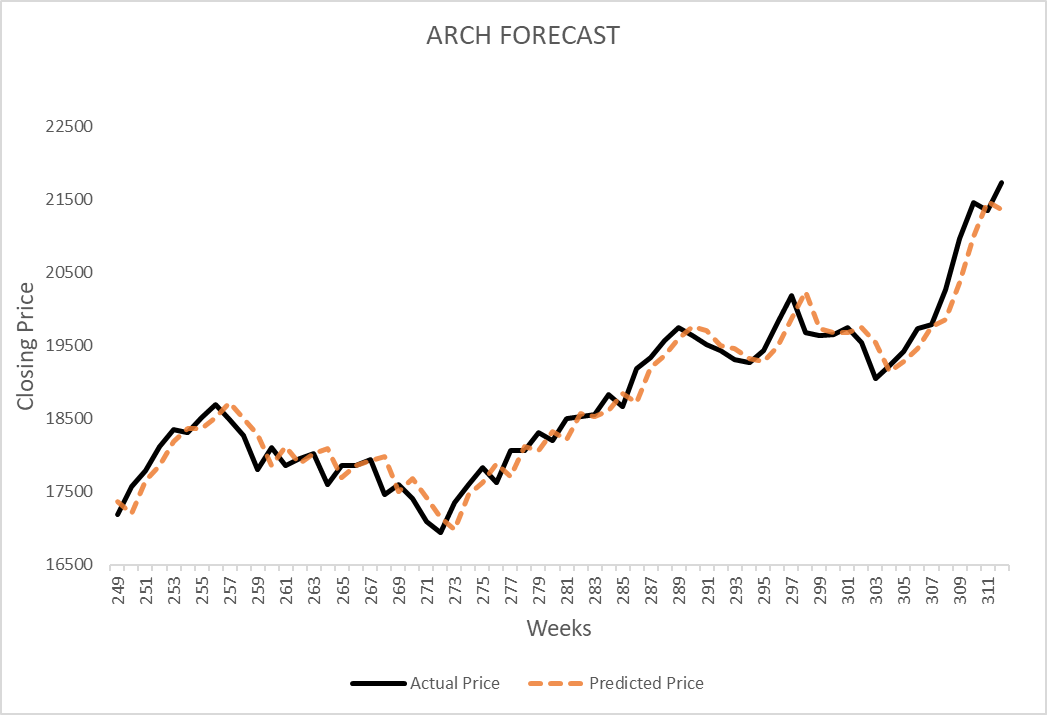
****

Figure 7 Forecast Accuracy of the ARCH forecast

The dotted orange line represents the movement of the predicted closing price as estimated by the ARCH forecast. In contrast, the solid black line represents the movement of the actual closing price. The ARCH forecast evaluation against the testing dataset yields an MAE of 217.468 and a MAPE of 1.15907%.

## 4.4 Conclusion

The objective of this chapter was to assess the predictability power of the ARCH model with the Neural Network model. Based on the findings from section 4, we can conclude that the Neural Network model with MAE 212.311 and MAPE 1.1353% in the case of one input model, as well as the sentiment augmented model with MAE 213.9576 and MAPE 1.1437 %, outperformed the traditional ARCH model with MAE 217.468 and MAPE of 1.15907%. Although the difference is minute, the Neural Network model successfully captured the movement in weekly closing price better than the ARCH model.

# CHAPTER 5 - CONCLUSION

## 5.1 Introduction

The Efficient Market Hypothesis for time immemorial has been the cornerstone theory in the field of financial economics. This particular theory, in general, has gone a long way in shaping the investors' understanding of how prices reflect all the available information in the market. Ever since the theory was propounded, empirical researchers across the globe have continuously sought to evaluate the validity of all forms of EMH in the context of different financial markets and across varying time periods. In this thesis, similar empirical research was conducted to examine the validity of EMH in the context of the NIFTY 50 index. The empirical research contained within this thesis sought to set itself apart by deviating away from the traditional econometric methods and adopting advanced deep learning techniques, which account for not only the past prices of the NIFTY 50 index movements but also the impact of news sentiments on the movement of the prices.

The research was guided by three main objectives. Firstly, it aimed to assess the efficacy of the deep learning models in projecting the future movements of the NIFTY 50 index as opposed to the traditional econometric models such as ARCH/GARCH. Secondly, it delved into the realm of Natural Language Processing (NLP) to assess the impact of news information on future price predictions. Finally, based on the insights from the prior two objectives, the validity of EMH was examined in the context of the NIFTY 50 index.

## 5.2 Summary of the Thesis

In order to achieve the objectives contained within the thesis, the research can be categorized into three broad stages. The first stage was associated with the pre-processing of the data before the data was integrated into the neural networks. This process involved extracting the weekly opening and closing price of the NIFTY 50 index from the National Stock Exchange (NSE) portal. While extracting the weekly opening and closing prices, significant care was taken to account for the week's beginning and closing as various events, such as holidays, market closures, etc., affected the week's length. Aside from collecting the weekly opening and closing prices, news information associated with all the NIFTY 50 companies was extracted from the money control website. Furthermore, the dataset was divided into two parts, the training set, i.e. the set on which the neural network will be trained on and a testing set , i.e. the set on which the respective models accuracy will be assessed on.

The second stage of the research involved building the neural network and carrying out Natural Language Processing (NLP). Building the neural network involved deciding how many hidden layers to incorporate along with the nodes contained within and deciding the optimal activation and optimizer functions. In order to ensure news information was integrated into the neural network, NLP was undertaken to convert them into sentiments, which ranged from -1 to 1.

The third stage involved evaluating the neural network's prowess in predicting future movements and how it fares against the traditional econometric methods. This stage involved looking into the model diagnostics, such as MAE and MAPE, to determine which model is the best fit; along with this, an ARCH model was built against which the neural networks would be evaluated.

## 5.3 Major Findings

In Chapter 2, a feedforward neural network was utilized to predict the weekly closing price of the NIFTY 50 index using just the weekly opening price as the input. This FNN developed yielded an MAE of 212.311 and a MAPE of 1.1353%, indicating a fairly accurate model for predicting the closing price. However, in Chapter 3, the FNN built in Chapter 2 was enhanced to incorporate news sentiments alongside the historical prices. This model provided an MAE of 213.9576 and MAPE of 1.1437%. Though it was expected that the news sentiment integrated FNN model would perform better than the FNN that only integrated weekly opening price, it is evident from the results that it performed slightly less accurately than the latter. In Chapter 4, an Autoregressive Conditional Heteroskedasticity (ARCH) forecast model was employed to predict the closing price, the main intent behind building the ARCH model was to compare its MAE and MAPE values against those of the neural network models. The ARCH forecast yielded a result which is more or less comparable to the neural network model, with an MAE of 217.468 and a MAPE of 1.15907%.

Overall, it can be understood that the neural network models and traditional econometric methods exhibited predictive capabilities more or less on par with each other, with the neural network models exhibiting a slightly better performance.

## 5.4 Conclusion of the Study

Throughout the course of the study, the main objective of the study was to assess the validity of the Efficient Market Hypothesis (EMH) primarily in the context of the NIFTY 50 index. In order to do this the study relied on using neural networks which were designed to integrate not only historical price but also historical information.

Chapter 2 of the study aimed to gain insights into the existence of the weak form of EMH in NIFTY 50 index. By leveraging the FNN it was found that based on the past prices it was possible for the neural network to outperform the random walk hypothesis, this finding indicates the absence of a weak form of EMH in the NIFTY 50 index as past prices were able to predict future movements. Chapter 3 went one step further by incorporating the past news information along with the past prices to assess the validity of the semi-strong form of EMH in the NIFTY 50 index. The FNN model incorporating both these factors exhibited superior performance compared to the EMH expectations, thus challenging the notion that all publicly available information is instantaneously incorporated into the asset prices. In Chapter 4, the neural networks predictive capabilities were put to test against the ARCH model. Despite a minute difference, the neural network model was successful in consistently outperforming the traditional ARCH model. This underscored the significance of advanced computational methodologies in assessing the market conditions as opposed to the traditional methods.

## 5.5 Directions for Further Research

This study confined its period of study from 2018 onwards till 2023. Expanding the period of study can go a long way in improving the predictive capability of the neural network, which, in turn, will help provide even more conclusive proof regarding the existence of EMH. In order to assess the Semi-strong form of EMH, the scope of the study was confined only to past financial news relating to the 50 companies contained within the NIFTY 50 index. The model's predictive power can be enhanced by incorporating news relating to the economy and past financial statements of respective companies, that is expanding upon the domain of public information available.

# **References**

Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. arXiv preprint arXiv:1908.10063.

Balkin, S. D., & Ord, J. K. (2000). Automatic neural network modeling for univariate time series. *International Journal of Forecasting*, *16*(4), 509-515.

Borges, M. R. (2010). Efficient market hypothesis in European stock markets. The European Journal of Finance, 16(7), 711-726.

Connor, J. T., Martin, R. D., & Atlas, L. E. (1994). Recurrent neural networks and robust time series prediction. IEEE transactions on neural networks, 5(2), 240-254.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The Journal of Finance, 25(2), 383-417.

Groenewold, N., & Kang, K. C. (1993). The semi‐strong efficiency of the Australian share market. Economic Record, 69(4), 405-410.

Hansson, M., & Olsson, C. (2017). Feedforward neural networks with ReLU activation functions are linear splines. Bachelor's Theses in Mathematical Sciences.

Hussin, B. M., Ahmed, A. D., & Ying, T. C. (2010). Semi-Strong Form Efficiency: Market Reaction to Dividend and Earnings Announcements in Malaysian Stock Exchange. IUP Journal of Applied Finance, 16(5).

Jais, I. K. M., Ismail, A. R., & Nisa, S. Q. (2019). Adam optimization algorithm for wide and deep neural network. Knowl. Eng. Data Sci., 2(1), 41-46.

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Research, 259(2), 689-702.

Lekhal, M., & El Oubani, A. (2020). Does the Adaptive Market Hypothesis explain the evolution of emerging markets efficiency? Evidence from the Moroccan financial market. Heliyon, 6(7).

Lo, A. W. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. Journal of Portfolio Management, Forthcoming.

Mallikarjunappa, T., & Dsouza, J. J. (2013). A Study of Semi-Strong Form of Market Efficiency of Indian Stock Market. Amity Global Business Review, 8.

Mehtab, S., & Sen, J. (2019). A robust predictive model for stock price prediction using deep learning and natural language processing. arXiv preprint arXiv:1912.07700.

Timmermann, A., & Granger, C. W. (2004). Efficient market hypothesis and forecasting. International Journal of forecasting, 20(1), 15-27.

Ţiţan, A. G. (2015). The efficient market hypothesis: Review of specialized literature and empirical research. Procedia Economics and Finance, 32, 442-449.

Urquhart, A., & McGroarty, F. (2016). Are stock markets really efficient? Evidence of the adaptive market hypothesis. International Review of Financial Analysis, 47, 39-49.

Vargas, M. R., De Lima, B. S., & Evsukoff, A. G. (2017, June). Deep learning for stock market prediction from financial news articles. In 2017 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA) (pp. 60-65). IEEE.

Wang, Z., Yan, W., & Oates, T. (2017, May). Time series classification from scratch with deep neural networks: A strong baseline. In 2017 International joint conference on neural networks (IJCNN) (pp. 1578-1585). IEEE.

Zaheer, R., & Shaziya, H. (2019, January). A study of the optimization algorithms in deep learning. In 2019 third international conference on inventive systems and control (ICISC) (pp. 536-539). IEEE.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. International journal of forecasting, 14(1), 35-62.