Kandidatuppsats

Bachelor's thesis

Företagsekonomi
Business Administration

Predicting the unpredictable - Can Artificial Neural Network replace ARIMA for prediction of the Swedish Stock Market (OMXS30)?

Alberto Ferreira de Melo Filho

Mittuniversitetet
MID SWEDEN UNIVERSITY

Mittuniversitetet

Avdelningen för Ekonomivetenskap och Juridik

Kandidatuppsats i företagsekonomi, 15hp

Examinator: Helene Lundberg

Handledare: Darush Yazdanfar

Författare: Alberto Ferreira de Melo Filho

(albertomelo.business@gmail.com)

Datum: 01 July 2019

Abstract

During several decades the stock market has been an area of interest for researchers due to its complexity, noise, uncertainty and nonlinearity of the data. Most of the studies regarding this area use a classical stochastics method, an example of this is ARIMA which is a standard approach for time series prediction. There is however another method for prediction of the stock market that is gaining traction in the recent years; Artificial Neural Network (ANN). This method has mostly been used in research on the American and Asian stock markets so far. Therefore, the purpose of this essay was to explore if Artificial Neural Network could be used instead of ARIMA to predict the Swedish stock market (OMXS30). The study used data from the Swedish Stock Market between 1991-07-09 to 2018-12-28 for the training of the ARIMA model and a forecast data that ranged between 2019-01-02 to 2019-04-26. The forecast data of the ANN was composed of 80% of the data between 1991-07-09 to 2019-04-26 and the evaluation data was composed of the remaining 20%. The ANN architecture had one input layer with chunks of 20 consecutive days as input, followed by three Long Short-Term Memory (LSTM) hidden layers with 128 neurons in each layer, followed by another hidden layer with Rectified Linear Unit (ReLU) containing 32 neurons, followed by the output layer containing 2 neurons with softmax activation. The results showed that the ANN, with an accuracy of 0,9892, could be a successful method to forecast the Swedish stock market instead of ARIMA.

Keywords: Artificial Neural Network, ARIMA, LSTM, stock market

Acknowledgements

I would like to take this opportunity to thank my wife for the countless hours that I spent talking to her about Artificial Neural Network, for her support and suggestions that made this essay possible and for my daughter, who is my everyday inspiration.

I would also like to thank Professor Darush Yazdanfar for all his great suggestions in every step of this essay

July 2019, Sundsvall, Sweden

Alberto Ferreira de Melo Filho

Albertomelo.business@gmail.com

Nomenclature

Adam – Adaptive Moment Estimation

AI – Artificial Intelligence

ANN - Artificial Neural Network

AR - Autoregressive Model

ARIMA - autoregressive integrated moving average

FFT – Fast Fourier Transform

GZP – Gaussian Zeo-Phase Filter

LSTM - Long Short-Term Memory

MA - Moving Average

ML - Machine Learning

MAPE - Mean Absolute Percentage Error

ME - Mean Error

MAE - Mean Absolute Error

MPE - Mean Percentage Error

OMXS30 - 30 most traded shares on Nasdaq Stockholm

ReLU - Rectified Linear Unit

RMSE - Root-Mean-Square Error

Table of Contents

1. Introduction

This chapter introduces the subject of the essay, followed by the research questions, purpose and the limitation of the essay. It also arguments as to why this study is relevant and how this study differs from previous research.

1.1 Introduction

The stock market has been an interesting area of research for many decades due to its profitability and the difficulty to accurately predict the future stock price. Many researchers and investors have tried to find ways to reduce risk of investments and today most of the papers written about prediction of the stock market utilize two approaches to analyze the stock market: the fundamental analysis and the technical analysis (Law, 2019).

The fundamental analysis consists in gathering all information that is available on the market, financial or nonfinancial, to analyze the stocks (Contreras, 2012). A few examples of fundamental analysis are overall profitability of the company, risk of the company (Mohanram, et al., 2018) and signals from financial statements (Yan & Zheng, 2017).

The other method typically used by researchers is the technical analysis which is based on the data that the stocks produce (Moanta & Ioana, 2018). The complexity of the fundamental analysis (Xuemin & Lingling, 2017) discourages statistical analysis, which can be seen with the lack of studies in this area (Bin, et al., 2017), making the technical analysis a more direct choice for prediction. In practice, the technical analysis has been used and supported by many top traders (Smith, et al., 2016). Therefore, the technical analysis was used in this essay.

There are many ways to do a technical analysis and many studies have used this analysis for a short- and medium-term investment strategies (Garcia, et al., 2018). One example of this technique is the traditional statistics method, like the simple regression or the autoregressive integrated moving average (ARIMA) method. Even though many researchers use this technique for prediction, there are some studies that argue for a lower performance in predicting the stock market (Dzikevicius, et al., 2010) compared to another technique called Artificial Neural Network (ANN). Because of this motive, ANN has grown in popularity. The improvement of

the prediction using ANN can be due to its performance in difficult prediction problems (Garcia, et al., 2018; Galeshchuk & Mukherjee, 2017). Some examples of difficulties for accurate prediction of the stock market are uncertainty, noise, nonlinearity and complexity (Rababaah & Sharma, 2015; Safa & Panahian, 2018) and ANN has been showing a considerable progress into this area (Safa & Panahian, 2018).

Many researchers have used ANN in the recent years to analyze the Asian and US indices like S&P 500, SET50 (Thailand) and KLCO Composite (Malaysia) (Garcia, et al., 2018), which present a lack of studies in analyzation of the European stock market using ANN.

1.2 Research Questions

Because of the gap in research regarding the European stock market and prediction using ANN, the following research questions were made:

- Is ANN a satisfactory method to predict a European stock market, specifically the Swedish Stock Market (OMXS30)?

- Can the computing technique ANN substitute a more statistical approach as ARIMA for prediction of the Stock Market?

1.3 Purpose of the study

The main purpose of this essay is to analyze if Artificial Neural Network can be used to predict the Swedish Stock Market (OMXS30) instead of ARIMA.

1.4 Limitation

The limitation of this essay is that the data analyzed comes only from one source, the Swedish Stock Market (OMX30) and it cannot directly generalize to other stocks and the ANN algorithm used in this essay cannot be generalized to other methods.

2. Theoretical framework and previous studies

This chapter describes the background of the Stock Market, the Swedish Stock Market, Time series data, Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN). This chapter also discusses previous research in the area of Artificial Neural Network, it presents the analyze model used for ARIMA and ANN and introduces the hypotheses generated for this study.

2.1 Stock Market

To be able to analyze the stock market, it is important to know what the stock market is. The stock market has grown together with capitalism since the 17th century and it is nothing more than allowing investors to buy and sell public companies, governments, local authorities and other incorporated bodies securities (Law, 2014). The stock market acts as an intermediary agent between the buyers and the sellers (Atack & Neals, 2009) and a few reasons for a company to sell stocks is to increase liquidity for the company and as a measure of the value of the company (Atack & Neals, 2009). While the motive for buyers is to invest in a company through risk analysis of the stocks for future gains (Schreder, 1962). One way for an investor to reduce the risk while maintaining the liquidity is to use the capitalization-weighted index, or market-capitalization-weighted, which is a group of the most influential stock markets that share some common traits (Bolla, 2017). A few examples of such index are S&P 500, NASDAQ-100 and the Swedish Stock Market (OMXS30).

To calculate the capitalization-weighted index it is necessary to multiply the total number of current stock prices and their weights in the index. Here is the formula to calculate the weight of each index (Corporate Finance Institute, 2019):

Equation 1- Capitalization-weighted index formula

$$Weight = \frac{Market\ Capitalization\ (Component)}{Total\ Index\ Market\ Capitalization}\ x\ 100\%$$

2.2 Swedish Stock Market (OMXS30)

The Swedish Stock market is an index that contains the 30 most traded shares on Nasdaq Stockholm (Nasdaq, 2019). To be able to be added to OMXS30 a company must rank under the 15 most traded shares, and for a company to be removed from OMXS30 the company must rank over the top 45 companies shared on Nasdaq Stockholm to avoid a volatility on the list. The index of the Swedish Stock Market is calculated by summing the total value of shares of all the companies that belong to OMXS30. The list of the companies on OMXS30 can be seen on Appendix A.

2.3 Time-series data

Data from the stock market are a typical example of time series (Merh, et al., 2011). The time-series data is the data collected over discrete intervals of time (Hill, et al., 2018; McCleary, et al., 2017). The representation of a complete time series can mathematically be described as:

Equation 2- Time-series formula representation

$$... Y_{-1}, Y_0, Y_1, Y_2, ..., Y_N, Y_{N+1}, Y_{N+2}, ...$$

A few examples of time-series data are the annual real gross domestic product (GDP) for the United States from 1948 to 2016, the daily closing value of the stocks of Apple between 2009 and 2012 or the number of incidences of Chagas disease in South America between 1996 and 2005.

What is important to note is that time-series data is usually correlated between different observations (Hill, et al., 2018). An example of such correlation is an expectation that the price of a product should reflect its own previous observations, if no external factor drastically effected the price.

Many uses of a time-series data can be seen in the literature for analysis or forecasting. Lee, Wu and Tsai (2014) mention how important forecasting is to estimate unforeseen events or scenarios. To do such forecasting a vast amount of studies have been performed and this essay will explore two technical approaches commonly used for prediction: the ARIMA model and the ANN model.

2.4 Markets efficiency

Market efficiency is a central concept in finance (Rösch, et al., 2017), which got a lot of track in the beginning of the 70's when Eugene F. Fama wrote that the market is efficient when the prices of the market "fully reflect" all available information (Fama, 1970). With his statement, Fama describes that only corporate insiders and specialists are the two groups whose monopolistic access to information has been documented. He also describes that the market is efficient, meaning that the gain through only the public information should be close to zero (Fama, 1970). Evidence, however, shows that the market is not efficient, and that profit can be obtained through data analysis (Palma & Sartoris, 2016; Galeshchuk, 2017; Nigam, 2018).

2.5 Autoregressive integrated moving average (ARIMA)

A time series model is an abstract representation of Y (McCleary, et al., 2017) and in the case of time series model of the stock market, this representation means the prediction of a future stock price. To be able to do a prediction, researchers and mathematicians created many models and one of these models is ARIMA. ARIMA have been one of the linear models most commonly used in time series forecasting (Aras, et al., 2017). To be able to understand what ARIMA is, the model can be divided into two parts: Autoregressive Model (AR) and Moving Average (MA), which combined becomes the model ARMA, and adding a differentiation, the model becomes ARIMA, as demonstrated down below.

The AR of ARIMA is "one where a variable y depends on past values of itself" (Hill, et al., 2018), or, in other words, when a value can be explained by previous values of the same time series. The representation of AR of order p (p = number of past values) is described as:

Equation 3- Autoregressive formula

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + e_t$$

An example of such equation is to use the last 2 months of sales data to predict this months' sales. In this case the equation would be:

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + e_t$$

Where y_{t-1} is last month values and y_{t-2} is the values from two months ago. δ is a constant which represents the interception of y-axis, Θ is the lag variable, or the slope and e_t is the randomness (white noise) created by the values that are not explained by the model.

The advantage of AR is that some prediction of the future can be made using only the past data, but its weakness is that the data cannot differentiate anomalies or seasonality in the data.

The next part of ARIMA is the Moving Average (MA). MA is used for capturing the essence of the observed sample autocorrelations (Hill, et al., 2018). In simple terms, a moving average describes a data set as removing the volatility of the data by averaging the data. With the help of the moving average, the data can be "filtered" and remove some of the noise created by randomness. The MA is described as:

Equation 4- Moving Average formula

$$e_t = \theta v_{t-1} + v_t$$

This part of ARIMA is to eliminate noise and anomalies from the data.

Adding these equations will give us the Autoregressive Moving Average model (ARMA). ARMA is a model that is commonly used to predict the momentum and mean effect while observing the white noise terms. ARMA can be described as:

Equation 5- ARMA formula

$$y_t = \delta + \theta_1 y_{t-1} + \theta_2 y_{t-2} + \cdots + \theta_p y_{t-p} + e_t + \theta_1 e_{t-1} + \theta_q e_{t-q}$$

The last part of ARIMA adds another variable called differencing of order d, meaning that the first thing to do in ARIMA is to apply differencing, which is the seasonality and trend that cannot be explained through the ARMA model. After the differentiation is applied, then the ARMA model can be used. Compared to ARMA, this method gets better results in some time series data by differentiating, making the data stationary which is a requirement for such type of analysis. An example of such seasoning in the data is the temperature of the year of a city. With enough data, the ARIMA model will take in consideration that summer has warmer days and winter has colder days. ARIMA are generally denoted as ARIMA(p,d,q) where p is the order of the autoregressive model, d is the degree of differentiation and q is the order of the moving average model. When d is equal to 0, the ARIMA model is equivalent of an ARMA model.

A common problem with ARIMA is that this model is linear. This implicates that a particular forecasting of a certain company will probably not work well with another company forecasting (Hiransha, et al., 2018). Since the forecast of ARIMA is linear, some important opportunities can be missed in the stock market. To avoid these issues, another approach can be used called Artificial Neural Network.

2.6 Artificial Neural Network

Because of the slowdown in growth of performance of transistors that are used to generate computer calculations and because the number of data accumulated grow over time, it is important to find new ways to achieve major improvements in the processor area to keep progress going (Jouppi, et al., 2018). An area that is getting a lot of attention is the Artificial Neural Network (ANN) since the power of ANN, compared to other methods, only grows with large datasets (Jouppi, et al., 2018).

ANN comes from the idea of Artificial Intelligence (AI) which can be described as "tasks that requires human-like intelligence" (Garbuio & Lin, 2019). Under the AI domain, exists the Machine Learning (ML) subdomain which is the use of certain algorithms to perform tasks through analysis from a set of examples. When a set of new data is fed to the ML system, the algorithm can perform the same task as before without the need to retrain the algorithm. ANN is a subdomain of ML, and can be described as algorithms that are composed of an interconnected group of nodes where each node is considered a neuron (see Figure 1).
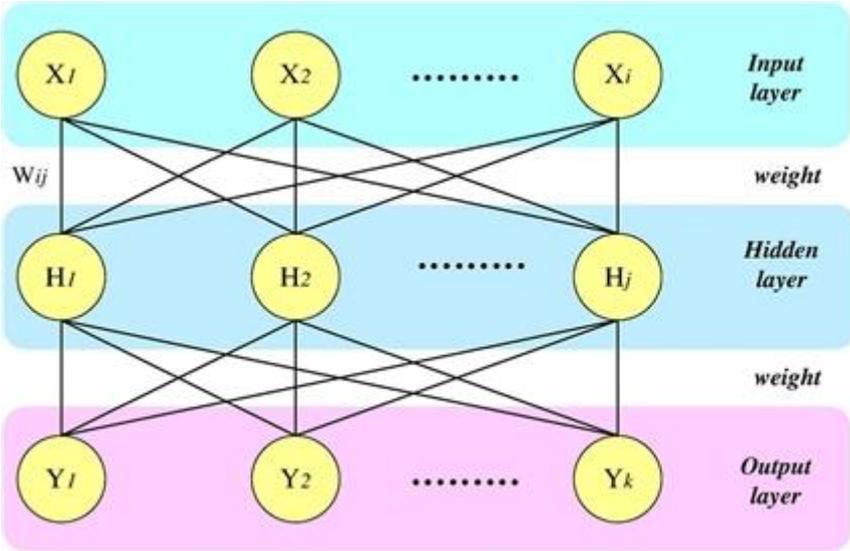
ANN is usually simply represented as neurons, which are usually represented as layers that hold some numbers (usually a number between 0 and 1 or between -1 and 1 depending on the function used on the input layer, called activation). The ANN is also divided into an input layer (represented as Y in Figure 1), a hidden layer (represented as H in Figure 1), an output layer (represented as X in Figure 1) and weights between these layers, which are numbers that multiplied by a neuron, can give a certain output to the next layer (represented as Wij in Figure 1).

The input layer is the data that it is inserted into the ANN algorithm, which for example can be represented as the stock market price of a certain company or the number of sales for a day of some company. As mentioned before, the raw data is processed through an activation function and the data is transformed to a number between 0 and 1 or between -1 and 1, depending on the activation function. The output layers are probability values that the researcher expects to predict and is the last layer after the hidden layers.

To compute what value to send to the next neuron, a weight value is given to each neuron before sending to the next neuron and a well-trained ANN will have weights that represents the outputs to each layer.

Mathematically, the performance of each neuron (neuron P) can be described as:

Equation 6- Formula for an ANN neuron

$$u_p = \sum_{i=1}^{n} w_{pi} x_i$$

$$y_p = \varphi(u_p + b_p)$$

Where $x_1,\ldots,x_n$ are the input parameters; $w_{p1}, \ldots , w_{pn}$ are the connection weights of neuron P; u_p is the input combiner; b_p is the bias; φ is the activation function; and y_p is the output of the neuron (Moghaddam, et al., 2016). The bias is a number chosen by the researcher to filter a value, meaning that the next neuron will only activate if u_p has a value higher than the bias. Each neuron from a layer will be connected to all neurons of the next layer.

The hidden layers are layers that hopefully will obtain patterns from the original data and it can be decided through experimentation to obtain the best result for the output.

In resume, ANN is a computational result of neurons multiplied by weights that it is divided in layers, resulting in an output layer. To be able to achieve the desired result, the ANN need to first have a training data with some feedback of the result, so that the right weights can be calculated to the neural network. This feedback can be calculated through a difference of the output and the future value in the case of time-series. The number of times that the program runs the loop to get the best weights are called Epochs. When these weights are adjusted, the prediction data can be inserted into the ANN with the trained weights and neurons for prediction and since the weights are already calculated, the speed to output a prediction is faster than during the training phase.

Currently there are many types of well-known ANN, but the major classifications are: supervised learning network, unsupervised learning network, associative supervised learning and optimization application network (Tsai & Hung, 2016).

The supervised learning network are usually associated with a "teacher-given, real-valued labels or targets yielding errors" (Schmidhuber, 2015), meaning that with the help of some specific signals created by the researcher, the ANN will try to reproduce this signals with a good accuracy. In a time-series data, this can be done by giving the lag result (the next day value, for example) to check how well the weights are doing.

The unsupervised neural network on other hand does not need any label or classification from the beginning, but this essay will focus on the supervised neural network.

A study done in 2009 showed that the most used techniques in ANN were the feed forward neural networks (FFNN) and recurrent networks (Atsalakis & Valavanis, 2009). The main difference between those techniques is that recurrent networks includes a loop connection where the weights of the previous layers can be calculated again, while FFNN only calculates from one direction.

ANN have been widely used to predict the stock market (Tsai & Hung, 2016) because of how well ANN can predict non-linearity of the data, the complexity of prediction the high degree of tolerance to faulty data (Kattan, et al., 2011).

The disadvantage of the ANN is that the models developed are still a trial and error basis and that the result cannot be explained, usually mentioned as a "black box" on how the ANN give its output (Palma & Sartoris, 2016).

2.6.1 Metrics of ANN

A metric is a function that is used to judge the performance of the ANN model (Keras Documentation, 2019). When the ANN is during the training phase, the model needs some metrics to define if the weights of the model needs to optimize. There are many ways to measure the metrics, but this study used the adaptive moment estimation (Adam) as optimization and the loss as measurement of the fitness of the model.

The Adam optimizer is an efficient way to rescale the weights of the ANN which is an appropriate approach for problems that are noisy or with sparse gradients (Kingma & Lei Ba, 2015) and the loss is the value that represents how far the output is from the prediction value during the training phase. Another metric used for a better interpretation of the result of the ANN is the accuracy. Accuracy is a binary result, which it yields a result if the output is equal or not equal to the actual data.

2.7 Previous research

Most of the previous research conducted on artificial neural network has generated positive results for prediction of the stock market. The stock market index of the articles used on this essay were NASDAQ, DOW30, Tehran Stock Exchange, Exchange rates of coins, Bucharest Stock Exchange, National Exchange of India and the New York Stock Exchange. The dates vary from 1996 to 2017, but some studies used 6 months of data while others used 21 years of data. The type of ANN used were most back propagation (LSTM for example) and some used other types of ANN as a simpler feed forward or a more complex hybrid neural network. The training data also varies in numbers where some used almost 96% of the data for training, while others used as low as 50% of the data. All the studies also had great results on accuracy varying

from 75% to 98.25% which suggests that ANN is a good method for forecasting the stock market (see table 1).

Table 1 - Overview of previous research on Artificial Neural Network

Authors and Year	Forecasted Index and Predicted Time Interval	Time Interval	Type of ANN	Input Variables	Result
(Moghaddam, et al., 2016)	NASDAQ index	28 Jan 2015 – 18 June 2015	Feed Forward	Training: 70 days Testing: 29 days	Accuracy: 0,9622
(Rababaah & Sharma, 2015)	DOW30 and NASDAQ100	3 Jan 2000 – 13 Jan 2012	Back Propagation	Training: 85% Evaluation: 15%	Accuracy: 0,9825
(Safa & Panahian, 2018)	87 companies from Tehran Stock Exchange	8 Dec 1999 – 30 Jan 2017	Harmony Search Algorithm	75% training data, 25% evaluation data	Accuracy: 0,965
(Galeshchuk & Mukherjee, 2017)	Exchange rates EUR/USD, GBP/USD, JPY/USD	2010 – 2015	Back Propagation	50% training data, 50% evaluation data	Accuracy: 0,750075
(Cocianu & Grigoryan, 2015)	Bucharest Stock Exchange	1 Mars 2009 – 30 November 2014	Back Propagation	200 samples for training, 100 forecasting	Forecasting error: 0,0012
(Hiransha, et al., 2018)	National Stock Exchange of India (NSE) and New York Stock	1 Jan 1996 – 30 June 2017	Back Propagation (LSTM), Recurrent Neural	4861 samples for training, 200 samples forecasting	MAPE: Between 3,85 and 11,64

	Exchange (NYSE)		Networks and Convolutional Neural Network		
(Garcia, et al., 2018)	German DAX-30 stock index	8 Dec 99 – 30 Jan 2017	HyFIS	87% training data, 13% test sample	76.24% mean hit ratio

2.8 Analysis model

To be able to answer the purpose of this study, the following Artificial Neural Network and ARIMA models for data collection, procedure and analyzing was made:

2.8.1 Artificial Neural Network model for analysis

The ANN model created in this essay has its foundation from previous researches and was prepared as following: The first step done for the ANN analysis was to collect the relevant raw data and to remove all the unnecessary data. When only the relevant data was left, the prediction target was included as how the values of the next day went down or up. The next step was to normalize and scale the data, to yield a better result, followed by a separation of the training data and the evaluation data. The last part of the ANN was to decide what method to use for calculation of the ANN, how many layers and neurons, how many epochs and then run the ANN to see the results (see Figure 2).
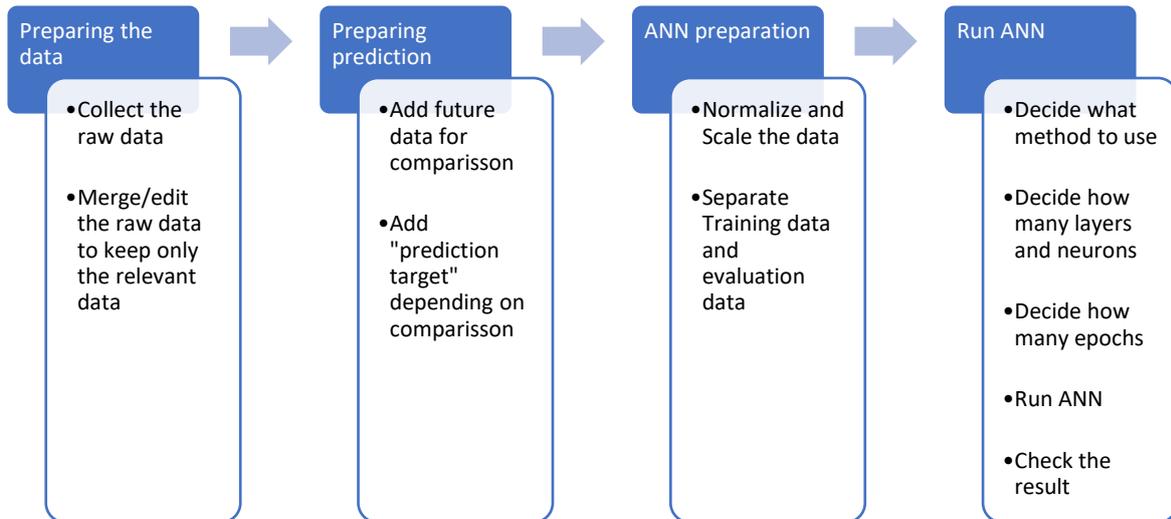
Figure 2- Overview of the steps to run ANN

2.8.2 ARIMA model for analysis

The ARIMA model, similar to the ANN was also prepared according to previous studies. First the data was collected, and all the non-relevant data was excluded. The data was separated into two parts, the training data and the prediction data. The order was decided according to previous studies and lastly the program was executed for prediction of the stock market.
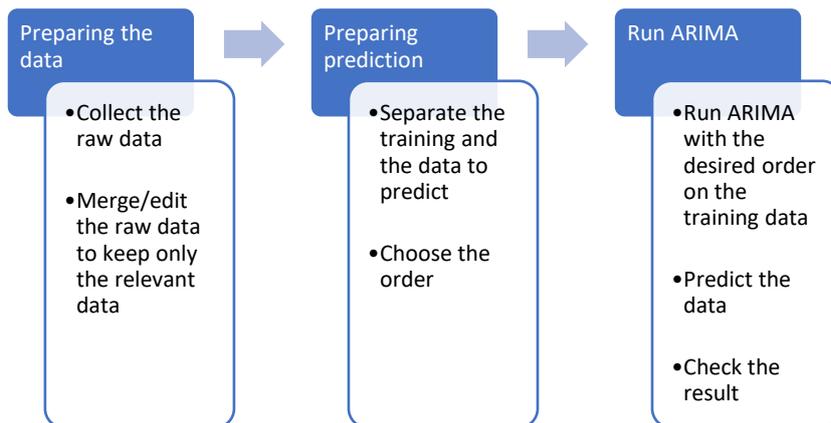
Figure 3- Overview of the steps to run ARIMA

2.9 Hypothesis

According to previous studies mentioned in this essay, ARIMA and ANN have had good results in accuracy for forecasting the future prices of various stock markets. To be able to answer the purpose of this essay the following hypotheses was generated:

H0 (null hypothesis): There is no relationship between the real value and the predicted value using ARIMA on the Swedish Stock Market (OMXS30).

H1: There is a relationship between the real value and the predicted value using ARIMA on the Swedish Stock Market (OMXS30).

H2 (null hypothesis): There is no relationship between the real value and the predicted value using ANN on the Swedish Stock Market (OMXS30).

H3: There is a relationship between the real value and the predicted value using ANN on the Swedish Stock Market (OMXS30).

3. Methodology

This chapter presents the research paradigm concerning choice of method, how the database search was conducted, which data was elected and how it was preprocessed, the architecture of the ARIMA model and the ANN model, the software used and lastly the ethical considerations used in the study.

3.1 Overview of methodology and approach

In this essay, a positive paradigm was used as research approach. The positive paradigm assumes that all observations obtained are a result of measurable phenomena, meaning that the result obtained from a research should be logical and usually associated with data in numeric form, most commonly called quantitative research data (Collis & Hussey, 2014, pp. 42-57). This study was also analytical in nature since it explored the association between the models, through quantifying and testing the hypothesis. A deductive research was conducted meaning that it was tested by empirical observations using secondary data obtained from Yahoo Finance (Collis & Hussey, 2014, pp. 4-8).

3.2 Literature search and source criticism

To be able to conduct a deductive research and produce credible theories and methods, a search on scientific databases was used. The main database used in this research was Business Source Complete (EBSCOhost) which is an international database with articles concerning business administration and related topics.

The first explorative search was "artificial neural network" limited to peer reviewed journals and sorted by relevance. Since this search had a sizable result, other searches were used to find relevant articles (see table 2).

Table 2- Search words, filters and number of results to find relevant articles

Search words	Filters	Number of results
Artificial Neural Network	Peer Reviewed, English	13849
Artificial Neural Network	Peer Reviewed, English, Full Text	2358
Stock Market	Peer Reviewed, English, Full Text	42576

Trading	Peer Reviewed, English, Full Text	19886
Artificial Neural Network & Stock Market	Peer Reviewed, English, Full Text	132
Artificial Neural Network & Trading	Peer Reviewed, English, Full Text	66
Artificial Neural Network & Trading & Stock Market	Peer Reviewed, English, Full Text	30

The articles used in this essay include some articles from these searches and similar articles obtained through the references of the searched articles.

The criticism of this method to obtain articles is that all studies found and used in this essay had a significant positive result for Artificial Neural Network in the stock market. This can possibly be a distortion of reality since articles with no significant results are more prone to remain unpublished (Kicinski, 2013).

3.3 Selection of data, data collection and data processing

The stock market history data of the 30 most traded stocks in the Swedish Stock market (see Appendix A) was obtained from Yahoo Finance. The data obtained for the analysis in the programs ranged from 1991-07-09 to 2019-04-26 (see Appendix B). The OMX data file was divided into columns containing the date, Open, High, Low, Close, Adjusted Close and Volume. Since most of the values from the Volume is 0 and the values of Close and Adjusted Close is equal, the Volume and the Adjusted Close values were removed from the data. The sum of the valid values after removing these columns is 7128 values for each column. The values used for the ANN model were all remaining columns, or 28512 values totals while the values used for the ARIMA model was only the close data, or 7128 values.

Table 3 - Number of values used for each model

Model	Columns used	Number of values
ARIMA	Close	7128
ANN	Open, High, Low, Close	28512

3.3.1 ANN data preprocessing

The data for the ANN model was separated into the training data, the validation data and the prediction data. According to previous research (Moghaddam, et al., 2016; Rababaah & Sharma, 2015; Safa & Panahian, 2018) the first 80% of the total data can be used for training while the last 20% can be used as validation data.

All the data was scaled to a range between -1 and 1 before running through the ANN for better analysis by the MinMaxScaler function build into the Python program.

3.4 Statistical processing and analysis

3.4.1 Architecture of the ARIMA model

The Arima model is quite straight forward since ARIMA only needs four variables to be run: the previous data to predict future data, the order of the autoregressive model (p), the degree of differentiation (d) and the order of the moving average model (q). In this essay the (p,d,q) used was (1,1,1), (1,2,1) and (2,2,1), the training data used was between 1991-07-09 to 2018-12-28 and the result was compared to a forecast data that range from 2019-01-02 to 2019-04-26. The result was compared by the mean absolute percentage error (mape), the mean error (me), the mean absolute error (mae), the mean percentage error (mpe) and the root-mean-square error (rmse), Standard Error (SE) and the probability to the result be only noise (P).

3.4.2 Architecture of the Artificial Neural Network

The ANN model used had chunks of 20 consecutive days as input layer, followed by 3 Long Short-Term Memory (LSTM) hidden layers with 128 neurons each, followed by a hidden layer with 32 neurons of an activation function called Rectified Linear Unit (ReLU). The last layer was the output layer, which was composed of two neurons with the softmax activation. One of the neurons is a probability for the next day prediction going up and the other one is a probability for prediction of the next day to go down.
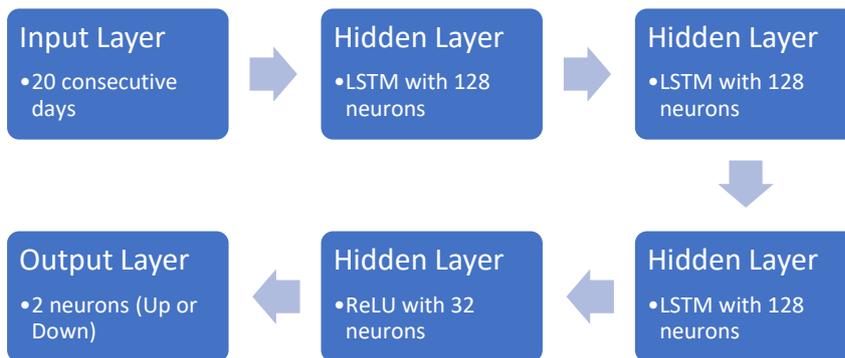
Figure 4- ANN architecture to predict the next day value of OMXS30.

3.4.2.1 Long Short-Term Memory (LSTM)

One of the methods used in the hidden layers in this essay is the Long Short-Term Memory (LSTM). A basic explanation of LSTM is that LSTM is a neural network with an additional "memory cell" as hidden layers (see figure 5).
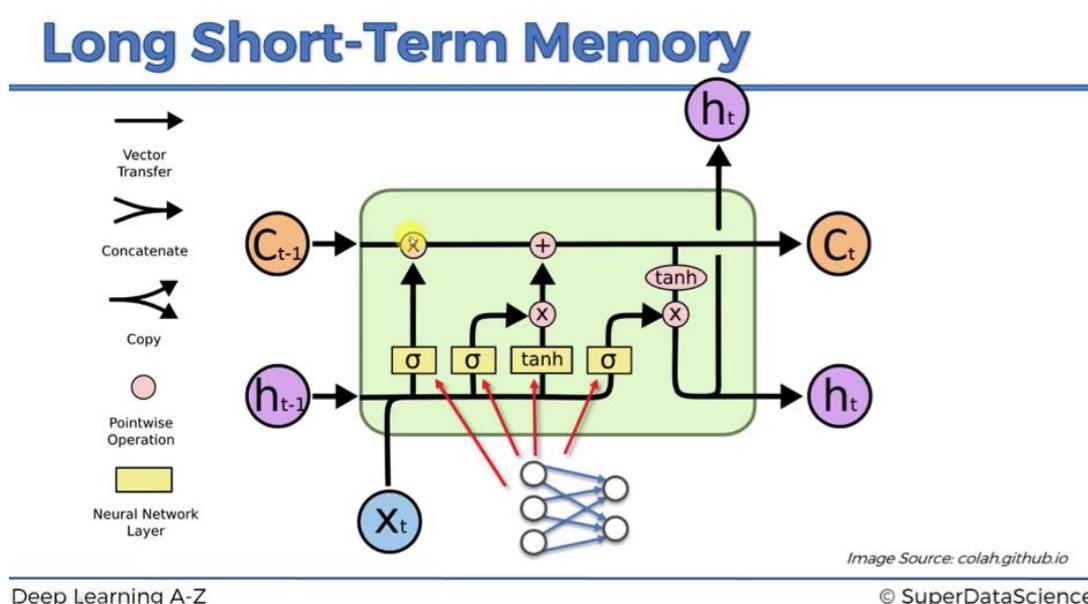
Figure 5- Representation of LSTM (SuperDataScience.com, 2019)

The data flows from h_{t-1} to a sigmoid function called "forget gate layer". This function receives the data from h_{t-1} and X_t and output a value between 0 and 1, where 1 represents "remember the data" and 0 represents "forget the data". The next step is to analyze the data through "tahn",

which has a mathematical representation as C_t = tanh(Wc.[h_{t-1}, x_t]+b_C) which gives a result between -1 and 1. In the last step the model outputs an information depending on the information from previous data which can be influenced by old data depending on the results of the previous steps.

3.4.2.2 Activation Function

The other method used on the last hidden layer and on the output was the activation function. The activation function is a neuron which transforms the input data to a better descriptive data. The activation function used on the hidden layer was the Rectified Linear Unit (ReLU) which is a simple and effective activation widely used across the deep learning community (Ramachandran, et al., 2017). The mathematical representation of ReLU is:

Equation 7- ReLU formula

$$b(x_1, x_2) = max(x_1, x_2) \text{ and g(x)} = 0$$

This activation function gives an output of 0 if x is less than 0 and x if x is bigger than 0 (Ramachandran, et al., 2017).

Another activation function used in this essay was the softmax function which is a function that normalizes the input value to a probability distribution, so that the sum of the values used in the output layer of the neural network is 1. The mathematical representation of the softmax function with discrete observation y = (y_1, …, y_n) $\in$ {1, …, K}^{n} is:

Equation 8 - Softmax Formula

$$p(y = k|x; \beta) = \frac{e^{\beta_k^t x}}{\Sigma_{k'=1}^{K} e^{\beta_{k'}^T x}} \text{ (Bouchard, 2007)}$$

3.4.3 Software

The software used for this research was Python 3.6 with the following libraries: Keras for ANN, Statsmodels for ARIMA, Matplotlib for graphics, Sklearn for statistics and Pandas for data manipulation.

3.5 Methods' validity and reliability

To achieve as high generalizability as possible and to minimize the occurrence of errors and bias in the result, a few considerations were made during the writing of this essay. A way to achieve a high degree of validity is to create a result that reflects what the researcher want measure (Collis & Hussey, 2014, pp. 52-54). To be able to do that, this essay only used peer reviewed articles and data from well-known databases to avoid errors and bias.

It is also important for a scientific study to try to achieve a high degree of reliability, which is to achieve high accuracy and precision of the measurements and results so that when the study is duplicated, the result should be the same (Collis & Hussey, 2014, pp. 52-54). In this study, the architecture of each model has been described in detail and tested multiple times to be sure that the result is reliable.

3.6 Ethical considerations

According to Collis & Hussey (2014, pp. 30-35) there are a few principles that a researcher should follow when writing an essay, such as the wellbeing of research participants, dignity, informed consent, privacy, confidentiality, anonymity, deception, affiliation, honesty and transparency, reciprocity and misrepresentation. Since all data obtained in this study are public, meaning that they are not considered sensitive data and no confidential data from organizations or persons have been used in this essay, no ethical principles were broken.

4. Results and analysis

This chapter presents the descriptive result from the different ARIMA models, a close look on the best performing ARIMA model with the respective forecast slope vs. the actual data and the descriptive result of the ANN model.

4.1 Result from ARIMA(p,d,q)

In the ARIMA model all the training data was fed to the program, which was the data from 1991-07-09 to 2018-12-28 of the stock market price for the Swedish Stock Market (OMXS30). The data used to forecast was the data for stock market prices from the Swedish Stock Market (OMXS30) between 2019-01-02 to 2019-04-26. The training data of ARIMA results in a slope that describe the training data and compares to the evaluation date and returns the MAPE, ME, MAE, MPE, RMSE, SE and P. To be able to acquire the best result for ARIMA, this essay evaluated three different values for p,d,q to identify the best order. The p,d,q analyzed in this essay was: (1,1,1), (1,2,1) and (2,2,1).

4.1.1 Result of AR and MA standard error and P

The first ARIMA evaluated was ARIMA(1,1,1). The standard error and P value were as follow:

Table 4 - Result of the standard error and p-value of ARIMA(1,1,1)

ARIMA (1,1,1)	SE	P
AR1	0,079	0,000
MA	0,072	0,000

This values shows that the result of ARIMA(1,1,1) is significant, with $P<0,05$. The standard error is also low, meaning that the sample represents the population.

The second ARIMA evaluated was ARIMA(1,2,1) and the result was as follow:

Table 5 - Result of the standard error and p-value of ARIMA(1,2,1)

ARIMA (1,2,1)	SE	P
AR1	0,014	0,001
MA	0,001	0,000

As before, the P-value and SE is also low, meaning that ARIMA(1,2,1) is also significant and representative.

The same occurs with ARIMA(2,2,1), even though the P-value of AR2 is closer to the 0,05 limit which can be seen below:

Table 6 - Result of the standard error and p-value of ARIMA(2,2,1)

ARIMA (2,2,1)	SE	P
AR1	0,014	0,001
AR2	0,014	0,046
MA	0,001	0,000

The standard error and p values shows that the best model is the ARIMA(1,2,1) since the standard error and the P-values were the lowest of all the models.

4.1.2 Result of the Error values of ARIMA(p,d,q)

For each of the ARIMA model used in this essay, some error values were calculated to check how well the model was doing against the predicted data. The result can be seen in the table below:

Table 7 - Result of error values of ARIMA(1,1,1), ARIMA(1,2,1) and ARIMA(2,2,1)

ARIMA	mape	me	mae	mpe	rmse
(1,1,1)	0,092	-144,886	145,613	-0,091	159,229
(1,2,1)	0,088	-138,345	139,395	-0,087	153,484
(2,2,1)	0,088	-138,374	139,423	-0,087	153,508

The error values of all ARIMA tested in this study shows that the model with the most accurate results was the ARIMA(1,2,1). This is due to that the estimated values were the lowest compared to the other models used.

4.1.3 A closer look on ARIMA(1,2,1)

The ARIMA(1,2,1) gave the best results on both p-values and error values compared to ARIMA(1,1,1) and ARIMA(2,2,1). The forecast from ARIMA(1,2,1) can be seen in Figure 6 and 7 below:
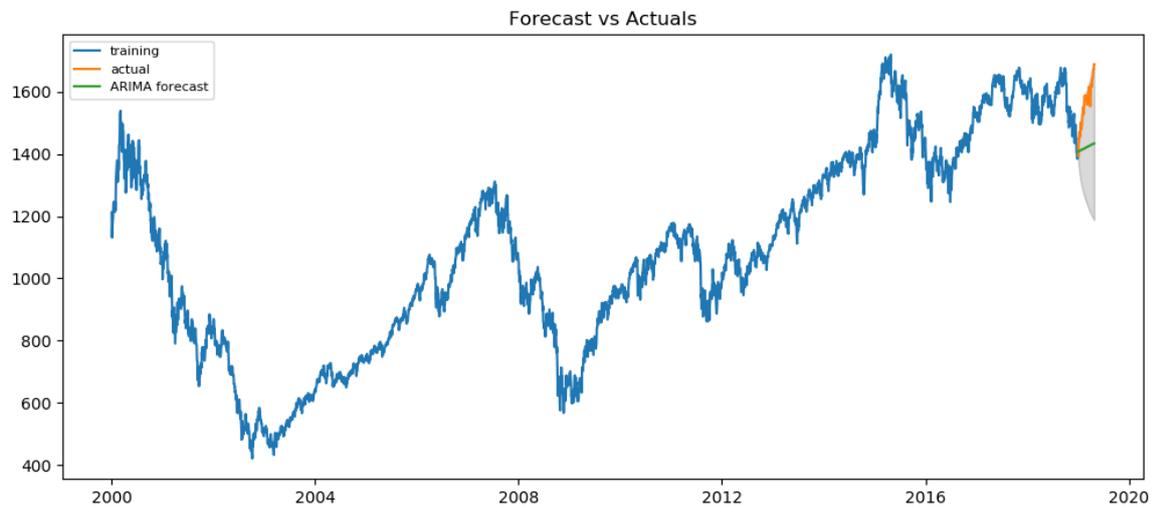
Figure 6 - ARIMA(1,2,1) training data, forecast data and the ARIMA forecast
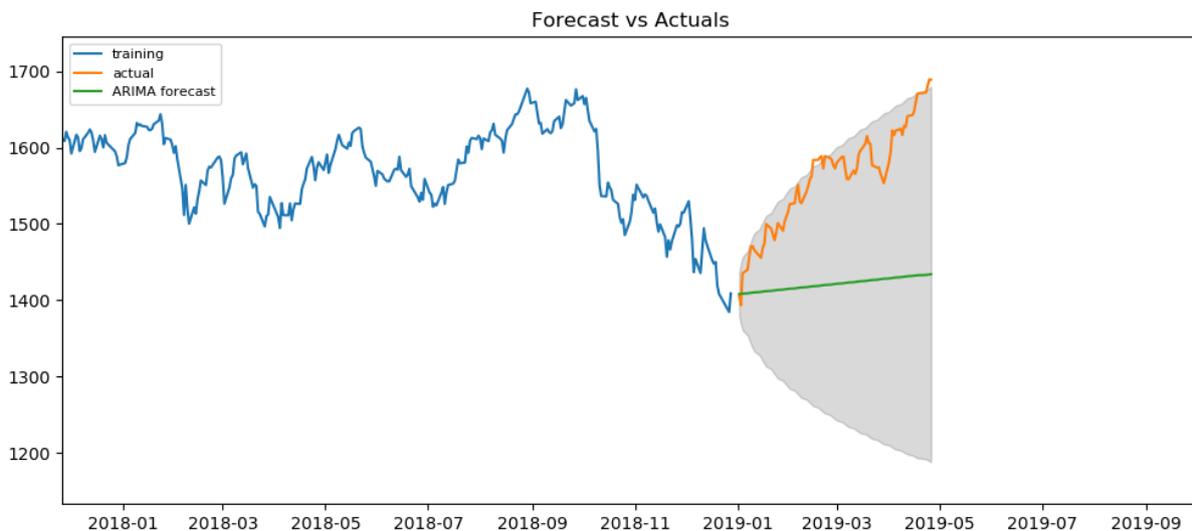
Figure 7 - The training data from 2008, the actual data and the ARIMA forecast slope

As can be seen in the pictures above, the ARIMA(1,2,1) forecasted a slope that would go up in the next days, which accords with the actual data.

4.2 Result of the Artificial Neural Network model

The ANN used in this study had one input layer, four hidden layers and one output layer. The input layer was chunks of 20 consecutive days data, the hidden layers were composed of three

hidden layers with the LSTM neuron, one with ReLU neuron and an output with 2 softmax neurons.

The ANN is usually evaluated through the accuracy, as other studies have proved before (Galeshchuk & Mukherjee, 2017; Safa & Panahian, 2018; Rababaah & Sharma, 2015; Moghaddam, et al., 2016). The result shows the loss of each EPOCH to check how far from the real data the predicted value is. The accuracy is a comparison between the next day value of the OMXS30 index and the prediction of the ANN model. As expected, the best model created by ANN showed better results for each EPOCH (see Figure 8 and 9) during the training phase, since the weights got adjusted to a better value for each interaction with the ANN yielding the right output. The accuracy had its accuracy peak at EPOCH 14 with the value of 0,9742 (see Figure 8) and the lowest lost also at EPOCH 14 with the value of 0,05936 (see Figure 9)
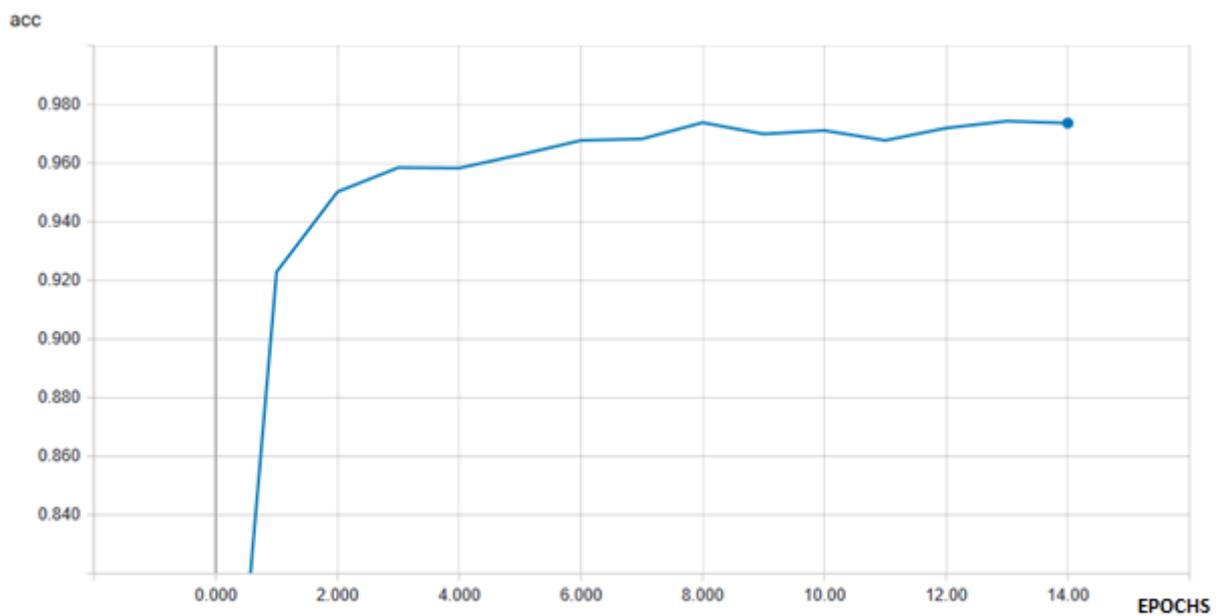
Figure 8 - Accuracy of the ANN during the training phase
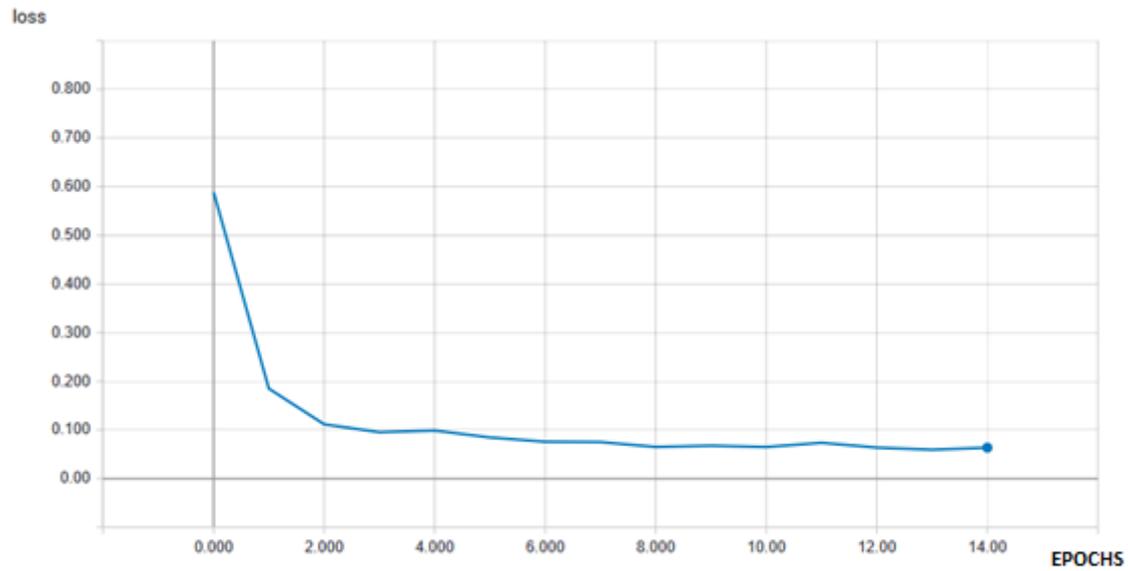
Figure 9 - Loss of the ANN during the training phase

During the validation phase, the most precise values were also at EPOCH 14 with the values of 0,9892 for accuracy and 0,0600 for loss (see Figure 10 and 11).
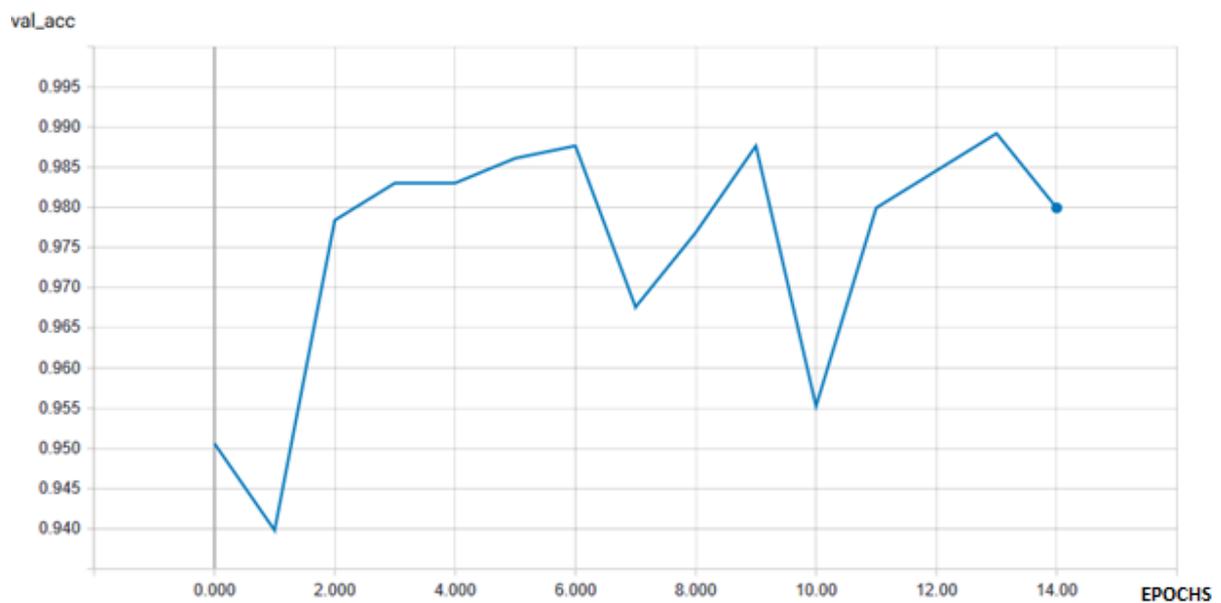
Figure 10 - Accuracy of the ANN during the validation phase
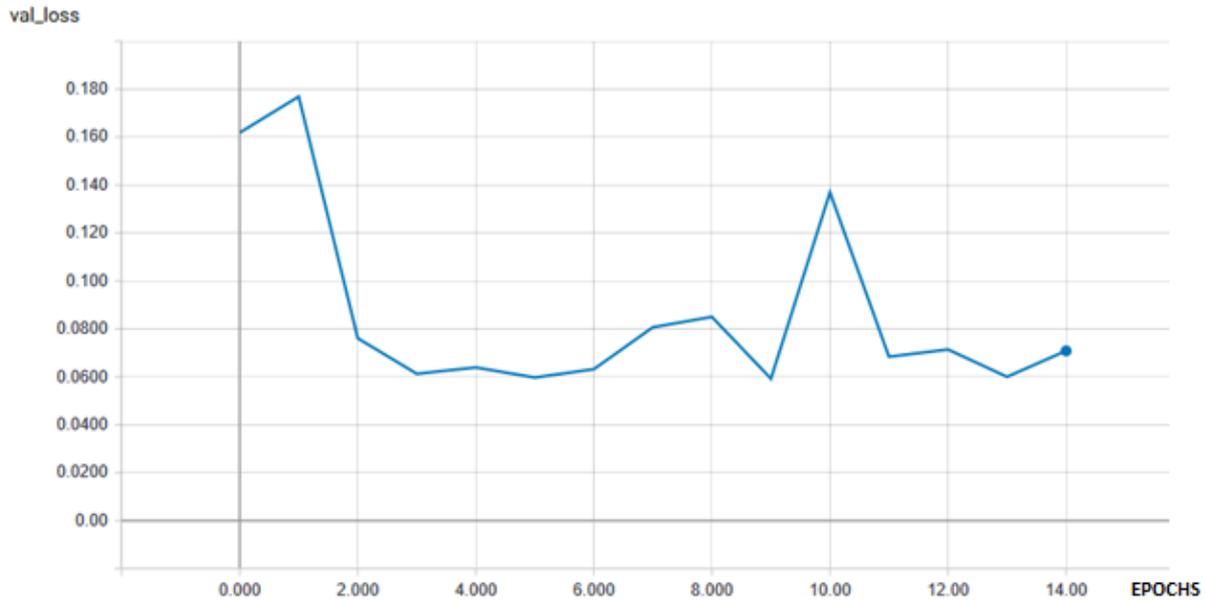
Figure 11 - Loss of the ANN during the validation phase

The result of the ANN during the training phase and the validation phase showed high accuracy and loss (see table 8). The biggest difference between the training phase and the evaluation phase was that the validation phase showed less consistency, which is expected since the weights were not being fitted to this new data.

Table 8 - Best values for Accuracy and Loss for the training phase and validation phase

Best value	Accuracy	Loss
Training phase (EPOCH 14)	0,9742	0,05936
Validation phase (EPOCH 14)	0,9892	0,0600

To be able to confirm or reject the hypothesis, 30 rounds of the ANN was ran producing the lowest accuracy as 0,9324 and the highest accuracy as 0,9892. The mean of all the results was 0,9579 with the confidence level (p-value 0,95%) as 0,0061, meaning that the null hypothesis 3 was rejected.

One of the shortcomings of this study is that the time provided to do this study did not allow further tests to get better results in accuracy and loss of the ANN. Another shortcoming of this essay is that the nature of ANN is still a "trial and error" basis, meaning that the method used in this essay has a possibility to not be the most efficient.

4.3 Hypothesis result

The result of each hypothesis can be seen below:

Hypothesis	
H0 (null hypothesis): There is no relationship between the real value and the predicted value using ARIMA on the Swedish Stock Market (OMXS30).	REJECTED
H1: There is a relationship between the real value and the predicted value using ARIMA on the Swedish Stock Market (OMXS30).	CONFIRMED
H2 (null hypothesis): There is no relationship between the real value and the predicted value using ANN on the Swedish Stock Market (OMXS30).	REJECTED
H3: There is a relationship between the real value and the predicted value using ANN on the Swedish Stock Market (OMXS30).	CONFIRMED

Since both Hypotheses 1 and 3 was confirmed, the result indicated that the Artificial Neural Network method can be used instead of ARIMA to predict the price of the Swedish Stock Market (OMXS30).

5. Conclusions and future research

This chapter summarizes the contribution this essay generated to the research area, it also discusses the implications of the result and concludes with a discussion about possible future researches related to this field of study.

5.1 Conclusions

Many studies have used the stock market as a benchmark for prediction performance due to the complexity, uncertainty, noise and nonlinearity of the time series data (Rababaah & Sharma, 2015; Safa & Panahian, 2018). One of the most used models for such prediction is the ARIMA model (Aras, et al., 2017).

This essay examined if a newer method called Artificial Neural Network could be used instead of ARIMA, as a valid substitute method for prediction of the stock market. This suggestion is due to that ANN has been proven to get good results with complex and nonlinear data (Jouppi, et al., 2018). Most studies in this area has not been focused on the European stock market, therefore this essay aimed to use ANN to predict the Swedish stock market and compare with results from ARIMA.

The result showed that both ARIMA and ANN can be used as methods to analyze a future value of the stock market and the result of the ANN also coincides with previous studies made on the stock market (see table 9), which suggests that ANN is a valid method to use in the European stock market as well.

Table 9 - Result of previous study compared to this essays result

Authors and Year	Result (Accuracy)
(Moghaddam, et al., 2016)	0,9622
(Rababaah & Sharma, 2015)	0,9825
(Safa & Panahian, 2018)	0,965
(Galeshchuk & Mukherjee, 2017)	0,750075
Melo, Alberto, 2019 (this study)	0,9892

The contribution of this essay is to show that the computational method called ANN can be used instead of the more statistical approach called ARIMA using the Swedish stock market

data, which is a data that has not been found on previous researches. Investors, banks, insurance companies, pension funds and hedge funds can use this method for analysis of the stock market.

5.2 Future research

This study has created a broad base for future research. A suggestion for a new approach could be to try different ANN architectures to see how it would affect the result. Other types of architectures could be to change the number of neurons or the number of layers of the ANN. Another example is to add more inputs to the architecture, for example to add sentiment analysis as input, which is a quantification of emotions, evaluations and attitudes toward a specific subject (Mostafa, 2019). One example of such sentiment analysis is to retrieve data from well-known finance analysts from Twitter or from well-known finance websites.

It could also be of interest to study the effect of changing the neurons to a different method. Instead of LSTM, convolutional neural network could be used. It has been considerable better than previous state-of-the-art methods for image and video classification and prediction (Krizhevsky, et al., 2017).

It is important to note that even though this study generated high accuracy, it does not imply that this ANN is profitable in practice, thus creating a demand for future studies regarding the possible profitability of Artificial Neural Network.

References

Aras, S., Deveci Kocakoc, I. & Polat, C., 2017. Comparative study on retail sales forecasting between single and combination methods. *Journal of Business Economics and Management,* 18(5), pp. 803-832.

Atack, J. & Neals, L., 2009. *he Origins and Development of Financial Markets and Institutions: From the Seventeenth Century to the Present.* Cambridge: Cambridge University Press.

Atsalakis, G. & Valavanis, K., 2009. Surveying stock market forecasting techniques - part II: Soft computing methods. *Expert Systems with Applications,* Volume 36, pp. 5932-5941.

Bin, L., Chen, J., Puclik, M. & Su, Y., 2017. Predicting Extreme Returns in Chinese Stock Market: An Application of Contextual Fundamental Analysis. *Journal of Accounting & Finance,* 17(3), pp. 2158-3625.

Bolla, L., 2017. Fundamental Indexing in Global Bond Markets: The Risk Exposure Explains It All. *Financial Analysts Journal,* 73(1), pp. 101-120.

Bouchard, G., 2007. Efficient Bounds for the Softmax Function andApplications to Approximate Inference in Hybrid models. *workshop for approximate Bayesian inference in continuous/hybrid systems,* pp. 1-9.

Cocianu, C.-L. & Grigoryan, H., 2015. An Artificial Neural Network for Data Forecasting Purposes. *Informatica Economica,* 19(2), pp. 34-45.

Collis, J. & Hussey, R., 2014. *Business Research: a practical guide for undergraduate and postgraduate students.* 4 ed. s.l.:Palgrave.

Contreras, I. e. a., 2012. A GA Combining Technical and Fundamental Analysis for Trading the Stock Market. *Applications of Evolutionary Computation: EvoApplications,* Volume 7248, pp. 174-183.

Corporate Finance Institute, 2019. *What is the Capitalization-Weighted Index?*. [Online] Available at: https://corporatefinanceinstitute.com/resources/knowledge/trading-investing/capitalization-weighted-index/

Dzikevicius, A., Saranda, S. & Kravcionok, A., 2010. The accuracy of simple trading rules in stock markets. *Economics and Management,* Volume 15, pp. 910-916.

Enders, A. & Brandt, Z., 2007. Using Geographic Information System Technology to Improve Emergency Management and Disaster Response for People With Disabilities. *Journal of Disability Policy Studies,* 17(4), pp. 223-229.

Fama, E. F., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance,* 25(2), pp. 383-417.

Galeshchuk, S., 2017. Technological bias at the exchange rate market. *Intelligent Systems in Accounting, Finance & Management,* 24(2), pp. 80-86.

Galeshchuk, S. & Mukherjee, S., 2017. Deep networks for predicting direction of change in foreign exchange rates. *Wiley,* pp. 100-110.

Garbuio, M. & Lin, N., 2019. Artificial Intelligence as a Growth Engine for Health Care Startups: Emerging Business Models. *California Management Review,* 61(2), pp. 59-83.

Garcia, F., Guijarro, F., Oliver, J. & Tamosiuniene, R., 2018. HYBRID FUZZY NEURAL NETWORK TO PREDICT PRICE DIRECTION IN THE GERMAN DAX-30 INDEX. *Technological and Economic Development of Economy,* 24(6), pp. 2161-2178.

Hill, R., Griffiths, W. & Lim, G., 2018. *Principles of Econometrics.* s.l.:Wiley Custom.

Hiransha, M., Gopalakrishnan, E. A., Vijay, M. & Soman, K., 2018. NSE Stock Market Prediction Using Deep-Learning Models. *Procedia Computer Science,* Volume 132, pp. 1351-1362.

Hochreiter, S. & Schmidhuber, J., 1996. LSTM CAN SOLVE HARD LONG TIME LAG PROBLEMS. *NIPS'96 Proceedings of the 9th International Conference on Neural Information Processing Systems,* pp. 473-479.

Jouppi, N., Young Cliff, Patil, N. & Patterson, D., 2018. A Domain-Specific Architecture for Deep Neural Networks.. *Communications of the ACM 61,* 61(9), pp. 50-59.

Kattan, A., Badullah, R. & Geem, Z. W., 2011. *Artificial Neural Network Training and Software Implementation Techniques.* Hauppauge: Nova Science Publishers, Incorporated.

Keras Documentation, 2019. *Usage of metrics.* [Online] Available at: https://keras.io/metrics/ [Acesso em 19 May 2019].

Kicinski, M., 2013. Publication bias in recent meta-analyses. *PLOS one,* 8(11), pp. 1-10.

Kingma, D. & Lei Ba, J., 2015. ADAM: A Method for Stochastic Optimization. *ICLR,* pp. 1-15.

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

Krizhevsky, A., Sutskever, I. & Hinton, G. E., 2017. ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM,* 60(6), pp. 84-90.

Law, J., 2014. stock exchange. *A Dictionary of Finance and Banking.*

Law, J., 2019. *Fundamental analysis.* [Online] Available at: http://www.oxfordreference.com.proxybib.miun.se/view/10.1093/acref/9780198789741.001.0 001/acref-9780198789741-e-1549

Lee, Y.-C., Wu, C.-H. & Tsai, S.-B., 2014. Grey system theory and fuzzy time series forecasting for the growth of green electronic materials. *International Journal of Production Research,* 52(10), pp. 2931-2945.

McCleary, R., McDowall, D. & Bartos, B., 2017. *Design and Analysis of Time Series Experiments.* Oxford: Oxford University Press.

Merh, N., Saxena, V. & Pardasani, K., 2011. Next Day Stock Market Forecasting: An Application of ANN and ARIMA. *IUP Journal of Applied Finance,* 17(1), pp. 70-84.

Moanta, C.-P. & Ioana, G.-M., 2018. CHARACTERISTICS OF FOREX MARKET AND TRADING STRATEGIES BASED ON TECHNICAL ANALYSIS. *Revista Tinerilor Economisti,* 15(30), pp. 27-43.

Moghaddam, A. H., Moghaddam, M. H. & Esfandyari, M., 2016. Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science,* Volume 21, pp. 89-93.

Mohanram, P., Saiy, S. & Vyas, D., 2018. Fundamental analysis of banks: the use of financial statement information to screen winners from losers. *Review of Accounting Studies,* 23(1), pp. 200-233.

Mostafa, M. M., 2019. Clustering halal food consumers: A Twitter sentiment analysis. *International Journal of Market Research,* 61(3), pp. 320-337.

Nasdaq, 2019. *Vad är OMX Stockholm 30 index?.* [Online] Available at: http://www.nasdaqomxnordic.com/utbildning/optionerochterminer/vadaromxstockholm30ind ex

Nigam, S., 2018. A Soft Computing Technique for Stock Price Prediction. *Annual International Conference on Operations Research & Statistics,* Volume 1, pp. 103-107.

Palma, A. A. & Sartoris, A., 2016. Weak-Form Market Efficiency of the Brazilian Exchange Rate: Evidence from an Artificial Neural Newtork Model. *Latin American Business Review,* 17(2), pp. 163-176.

Rababaah, A. & Sharma, D., 2015. Integration of two different signal processing techniques with artificial neural network for stock market forecasting. *Journal of Management Information and Decision Sciences,* 18(2), pp. 63-80.

Ramachandran, P., Barret, Z. & Quoc, V. L., 2017. Searching for Activation Functions. *arXiv*, pp. 1-13.

Rösch, D. M., Subrahmanyam, A. & Van Dijk, M. A., 2017. The Dynamics of Market Efficiency. *Review of Financial Studies,* 30(4), pp. 1151-1187.

Safa, M. & Panahian, H., 2018. P/E Modeling and Prediction of Firms Listed on the Tehran Stock Exchange; a New Approach to Harmony Search Algorithm and Neural Network Hybridization. *Iranian Journal of Management Studies (IJMS),* 11(4), pp. 765-786.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural Networks,* Volume 61, pp. 85-117.

Schreder, H., 1962. The Stock Market. *The Journal of Finance,* 17(2), pp. 245-258.

Scikit Learn, 2019. *4.3. Preprocessing data.* [Online] Available at: https://scikit-learn.org/stable/modules/preprocessing.html#preprocessing-scaler
Sharma, D. K. & Rababaah, A. R., 2014. Stock Market Predictive Model Based on Integration of Signal Processing and Artificial Neural Network. *Academy of Information and Management Sciences Journal,* 17(1), pp. 51-70.

Smith, D., Wang, N., Wang, Y. & Zychowicz, E., 2016. Sentiment and the Effectiveness of Technical Analysis: Evidence from the Hedge Fund Industry. *Journal of financial and quantitative analysis,* 51(6), pp. 1991-2013.

SuperDataScience.com, 2019. *Recurrent Neural Networks (RNN) - Long Short Term Memory (LSTM).* [Online] Available at: https://www.superdatascience.com/blogs/recurrent-neural-networks-rnn-long-short-term-memory-lstm/

Tsai, J.-M. & Hung, S.-W., 2016. Supply Chain Relationship Quality and Performance in Technological Turbulence: An Artificial Neural Network Approach. *International Journal of Production Research,* 54(9), pp. 2757-70.

Xuemin, Y. & Lingling, Z., 2017. Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach. *Review of Financial Studies,* 30(4), pp. 1382-1423.

Yahoo! Finance, 2019. *OMX Stockholm 30 Index (^OMX).* [Online] Available at: https://finance.yahoo.com/quote/%5EOMX/chart?p=%5EOMX#eyJpbnRlcnZhbCI6Im1vbnRoIiwicGVyaW9kaWNpdHkiOjEsImNhbmRsZVdpZHRoIjoxLCJ2b2x1bWVVbmRlcmxheSI6dHJ1ZSwiYWRqIjp0cnVlLCJjcm9zc2hhaXIiOnRydWUsImNoYXJ0VHlwZSI6ImxpbmUiLCJleHRlbmRlZCI6ZmFsc2UsIm1hcmtldFNlc3Npb2b

Yan, X. & Zheng, L., 2017. Fundamental Analysis and the Cross-Section of Stock Returns: A Data-Mining Approach. *Review of Financial Studies,* 30(4), pp. 1382-1423.

Appendix A – List of the companies on OMXS30

- Nordea Bank Abp

- Kinnevik AB

- AstraZeneca PLC

- Telia Company AB (publ)

- Telefonaktiebolaget LM Ericsson (publ)

- Tele2 AB (publ)

- Swedbank AB (publ)

- SSAB AB (publ)

- Skanska AB (publ)

- Alfa Laval AB (publ)

- ABB Ltd

- Skandinaviska Enskilda Banken AB (publ)

- Investor AB (publ)

- Hexagon AB (publ)

- Svenska Cellulosa Aktiebolaget SCA (publ)

- Essity AB (publ)

- AB Electrolux (publ)

- Svenska Handelsbanken AB (publ)

- H & M Hennes & Mauritz AB (publ)

- Getinge AB

- AB Volvo (publ)

- Atlas Copco AB

- Swedish Match AB (publ)

- Securitas AB

- Autoliv, Inc.

- AB SKF (publ)

- ASSA ABLOY AB (publ)

- Sandvik AB

Appendix B – Plot from OMX Stockholm 30 Index (Yahoo! Finance, 2019)