**Semra Kafa**

# Artificial Intelligence in EPS forecasting

**Master's thesis spring 2020**
**Oslo Business School**
**Oslo Metropolitan University**
**Master's program in business administration**

# Preface

This thesis concludes my Master's in Business Administration with a major in Financial Analysis at Oslo Metropolitan University. My objective with this research is to draw a picture of the application of Artificial Intelligence in EPS forecasting. My passion for Finance and a big interest in Artificial Intelligence, have been my main motivation to choose this topic. The research process has been challenging and very exciting. During this time, I have studied different aspects of financial forecasting and gained valuable knowledge of how Artificial Intelligence is used in the financial world.

I would like to thank my supervisor, Associate Professor Einar Belsom at the Department of Oslo Business School, for his support and valuable input during this process. His committed guidance and constructive feedback have been imperative for the realization of this study. I would also like to thank all other parties contributing to this thesis with inputs, ideas, and technical support.

Oslo, 31.05.2020

Semra Kafa

# Abstract

In this paper, I present a study where the main focus is on how Artificial Intelligence is used in EPS forecasting. Unique to this study is that I compare the performances of one Neural Network model and one Time Series model for EPS forecasting. A sample of 10 international companies was chosen for this research paper. The artificial model was developed in MatLab, while the Time Series analysis was developed in Excel. This study compares these two approaches while emphasizing their strengths and weaknesses. When comparing MAPE, MSE, and MAD the results are in favor of the statistical approach. However, the performance analysis concludes that there are many factors included in a forecasting model development and the choice of approach is case and complexity dependent.

# Sammendrag

Denne masteroppgaven presenterer en studie om bruken av kunstig intelligens for EPS prediksjon. Unikt for denne oppgaven er at jeg sammenligner en kunstig intelligens modell og en tidsserieanalyse modell for prediksjon. Et utvalg av 10 internasjonale selskaper var valgt for oppgaven. Den kunstig intelligens modellen ble utviklet i MatLab og den tidsserieanalysen ble utviklet i Excel. I denne oppgaven sammenligner jeg disse to tilnærmingene og fokuserer hovedsakelig på deres styrker og svakheter. Ved sammenligning av MAPE, MSE og MAD indikerer resultatene at den statistiske modellen gjør det bedre. Likevel konkluderer jeg i resultat analysen at det er mange faktorer ved utformingen av en prediksjonsmodell og valget av metode avhenger av omfanget og kompleksiteten av enkelte case.

# Table of Contents

# 1. Introduction

With the rapid changes in technology, the world around us changes a lot and so does the financial market and the way we understand finance too. The methods we used to understand and analyse finance 20 years ago are getting old and inapplicable to use in the new dynamic environment. The increased globalization of the economy has brought with it many opportunities, but many challenges too.

Finance is one of the most important aspects of the business. When we say finance, we usually talk about money management, investing, borrowing, lending, budgeting, saving, and future forecasting. It is particularly the last one I am going to work on in this paper. A financial forecast is an estimate of the future. Using historical data and key economic indicators, economists within finance try to paint a picture of how the world will look like in one, five, or sometimes ten years from now.

In finance, we like predictions, even though we know that in practice the future is unpredictable. Historically, the best prediction machines have been human brains. According to what Professor Shing (2018) explains in Neuroscience news, the brain is a "prediction machine" that is constantly working on getting information from the environment and predicting the future using internal models. Rather than passively waiting to get information, our brain is working full time to gather data from the world around us and from the daily experiences we have.

Our brains are equipped with the "tools" to extract minimally analysed information and to use that information to make analogies and map it to similar representations in memory. This gives us what we also call "experience", from all the information gathered our brain maps and models the world as we see it. While our existing memories are used to make predictions, they are also constantly being updated. But forecasting in our brain does not happen by simply activating all the memory that we have gathered. Instead, our brain activates only the most important information depending on the context we are in. Finally, when all the aspects are considered our brain provides us with predictions about what might be expected in the current situation. According to Bar (2007), our accumulated experience creates an ever-evolving platform for predictions, and the influence of these predictions can be observed on many levels.

Financial forecasting means estimating the future performance of a business using historical data about the company and the industry of concern. Different statistical methods have been developed, so that one may predict the future. The most widely used are regressions with time series like Time Series Analysis, Autoregression (AR), Autoregressive Integrated Moving Average (ARIMA), and Vector Autoregression (VAR). Many factors in financial forecasting need to be considered. The complexity of the financial environment and macroeconomics have been increasing together with the increased connectivity and globalization. Today in economics we talk more about the global than the local. When we are doing business, we think about all of our competitors and not only the ones who are nearby. The increased complexity and number of factors affecting companies are challenging the classical linear forecasting techniques. Models based on Artificial Intelligence that can give us results calculated from relations we cannot see are becoming more and more used.

In recent years we hear a lot about the use of Artificial Intelligence in finance, but the truth is that Artificial Intelligence has been here for a while and has already proven its position in the financial world. When analysing the applications of neural networks in Finance Fadlalla and Lin (2001) found out that financial analysts and big companies were using neural network applications already before we all started using the internet. Troy Buckner from Hyman Beck and Company became a successful portfolio manager with his neural network-based trading system in 1997. He generated a 32% return while managing the short-term portfolio of the company. 55% of the companies that Fadlalla and Lin (2001) analyzed were already using Artificial Neural Networks (ANN) for stock market forecasts.

The objective of the research presented in this paper is to make a comparison between using Time Series analysis and Supervised Machine learning for EPS forecasting. Many studies have focused on the comparison between ANN and statistical models in forecasting. However, researchers in most of these studies have worked on other genres like share price and bankruptcy predictions for example. This study is meant as a contribution to the area of EPS forecasting. The final goal is to provide new and additional evidence on which forecasting methods are most accurate and applicable in different cases. Earnings per share (EPS) forecasts are of high value as they provide useful information about the firms' prospects and are seen as a significant factor in investments and stock price evaluation. Managers inside the company are also using EPS forecasts when making important decisions and working on budgeting and capital investments. Because EPS is seen as a very valuable parameter both inside and outside the company and since there is not enough research focusing on this topic, I chose to work on EPS forecasting.

In this study, I work on analyzing two types of prediction approaches. The neural network model I chose is a Nonlinear Autoregressive model with External (Exogenous) Input (NARX) and the statistical model is a Time Series analysis. Based on quarterly data for 10 international companies I aim to predict the future quarterly EPS values for these companies. From the forecasting results, I further calculate mean absolute percentage error (MAPE), mean squared error (MSE), and mean absolute deviation (MAD). Eventually, I analyze and discuss the interpretation of these metrics.

This paper is organized as follows. In the next section, I provide a literature review of Artificial Neural Networks, the classic time series models, and how these are used in EPS forecasting. After discussing the theoretical perspective of this paper, I present the data used in this study, the research methodology, and the results of my work. I close with a performance analysis of the models and suggest implications for further research.

# 2. Literature review

In this section, I present previous research in EPS forecasting. I review the concept and construction of both the artificial and the statistical approach. The literature review focuses on drawing a picture of currently available knowledge for how artificial intelligence and time series have been used for EPS predictions.

## 2.1 Artificial neural network

### *2.1.1 Brief review*

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience, analogy, and example without being explicitly programmed. The main goal with machine learning is developing programs that given a data set can learn for themselves. The script improves its performance with the time as the samples available for learning increase. The main objective of using Machine learning is that the algorithm will find natural patterns from the data and will generate insight, helping the analysts make better decisions. According to The MathWorks Inc.,(2016), there are two main types of Machine learning – supervised learning and unsupervised learning. Supervised learning trains the model on known input and output so that it can predict the future, unsupervised models group, and interpret the input data.

Supervised learning is divided into two types of models: classification and regression. Classification models' only goal is to distinguish between different categories in data. Regression models are the ones that are mostly used in finance. Instead of categorizing, these models give continuous responses to the data. Applications of regression techniques include forecasting stock prices, handwriting recognition, and acoustic signal processing. Using supervised learning we can create a model that can produce future predictions based on historical data. The model receives both a set of input data and a set of output data. Based on these two datasets the algorithm is trained to generate reasonable predictions to new input data. As explained in Introducing Machine Learning by The MathWorks Inc., (2016) these models have been used for optimizing energy usage in large buildings and detecting low-speed car crashes for example.

Artificial neural network (ANN) is a computer model in which construction is based on the way the biological human brain works. It consists also of neurons that receive a number of inputs signals and produces an output signal. The neurons are connected with links where every link has a numerical weight. Numerical weights represent the long-term memory of ANN. The information processing of a neuron starts with a signal that is received from the input links, the signals are computed and then sent to the output links. The output can consist of either a solution or further input to other neurons. One of the main differences between the linear models for prediction and the models using ANN is the ability to learn. According to Negnevitsky (2011), learning is the essential and fundamental characteristic of the neural network, where the machine can learn from errors just like human beings.

The great interest for the neurocomputing (ANN-based computing) models comes from their great capacity for information processing, learning, and modeling complex real-world problems. What makes artificial problem solving similar to our biological problem solving is their information processing characteristics such as robustness, fault and failure tolerance, and ability to handle imprecise and fuzzy information to name a few. According to Basheer and Hajmeer (2000), the main goal when creating artificial models is to develop a mathematical framework that will make possible for the ANN to learn by copying the way human brain process and use knowledge, but much faster.

### 2.1.2 Technical construction

An ANN is constructed from several very simple, but highly interconnected processors called neurons, which are analogous to the biological neurons in the human brain. When creating an artificial model, we start by deciding how many neurons the model should have and how they will be connected. This process is referred to as network architecture.
The neurons and their connections are established by links, and each link has a numerical weight. The long-term memory of ANN is represented by these weights. As Negnevitsky (2011) emphasizes what is referred to as "learning" in ANN is repeatedly adjusting the weights between the neurons.

The idea that is still on the basis of how ANN works, was presented for the first time in 1943 by Warren McCulloch and Walter Pitts. Based on their suggestion when the neuron gets the input layer, it is calculating the weighted sum of the signals and comparing it to a threshold value, ϴ. If the weighted sum is greater or equal to the threshold the neuron becomes activated and the output is +1, if it is lower than the threshold the output is – 1.  This function is called the sign function.

The neuron uses the following transfer function:

$$X = \sum_{i=1}^{n} x_i w_i$$

(2.1-1)

$$Y = \begin{cases} +1 \ if \ X \geq \theta \\ -1 \ if \ X < \theta \end{cases}$$

(2.1-2)

The output function is represented by the following formula:

$$Y = sign\left[\sum_{i=1}^{n} x_i w_i - \theta\right]$$

*(2.1-3)*

Based on the same principle but changing the output values to 1 and 0, we get the step function. The step and sign activation functions together form the group of hard-limit functions. These are often used in decision making for classification and pattern recognition tasks. 15 years later in 1958 Rosenblatt introduced the mechanics of the single artificial neuron and the procedure for training a simple ANN – a perceptron. The perceptron is the simplest form of neural network. According to Negnevitsky (2011), the function of it is based on a single neuron, adjustable synaptic weights, and a hard limiter.

The perceptron can be trained on a set of examples using a special learning rule. However, this rule has an accurate performance only for linearly separable classes. To make the perceptron work for nonlinear classes as well, an additional layer of neurons is added between the input and the output layer. The neurons in this layer do not interact with the external environment and this layer is called the hidden layer. As Basheer and Hajmeer (2000) further explain, neural networks with one or more hidden layers are called multilayered.

The non-linear classification work of ANN is processed in the hidden layer. The input layer accepts signals from the outside world, while the output layer accepts signals for the outside world. The hidden layer establishes the output pattern of the entire network, it detects patterns, features, and connections hidden in the input data. Figure 2.1 shows the illustration presented in the book of Negnevitsky (2011)

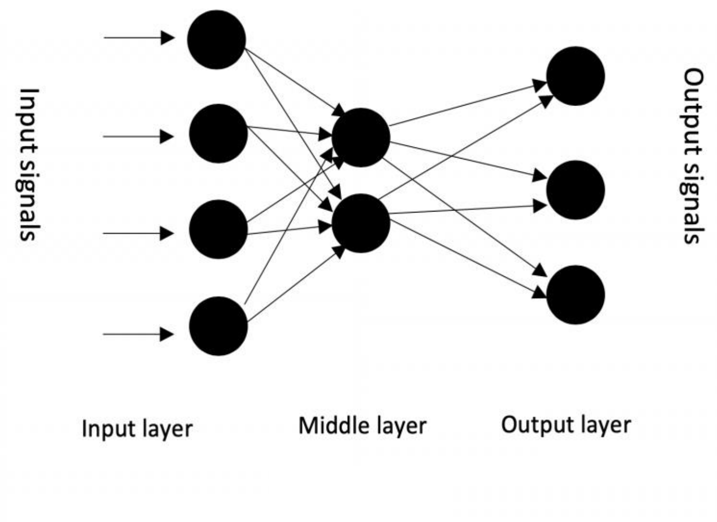


*Figure 2.1*

It is usual to use ANN with one or two hidden layers, where every layer consists of from 10 to 1000 neurons. There is also possible to include more than two hidden layers, but this will increase the computational demand for processing the script. However, according to Negnevitsky (2011), most practical implications use only one hidden layer.

The feedforward error-backpropagation learning algorithm is the most famous procedure to train artificial neural networks according to Basheer and Hajmeer (2000). The term backpropagation refers to the way error computed at the output side is propagated backward from the output layer into the hidden layer and finally to the input layer. The errors between the neural network outputs and the target outputs are used to adjust the interconnection weights between the neurons. The algorithm corrects itself backward until the neural network outputs fit in the target values within a prespecified tolerance. These networks are the most widely used types of networks.

### *2.1.3 ANN in financial forecasting*

The application of ANN models involves the interaction of many diverse variables that are highly correlated. These are frequently assumed to be nonlinear, unclearly related, and too complex to be described by a mathematical model according to Huang, Lai Keung, Nakamori, Wang, and Lean (2007). There are mainly three unique features of ANN compared to classic statistical models regarding Huang, et al. (2007) that makes it increasingly attractive. The Neural Networks are nonlinear data-driven. These models are universal functional approximators, so they can approximate any continuous functional form to any desired accuracy. Finally, the ANN model can generalize. Once learning from the data presented to it, the algorithm can give correct output to new data even if there is noisy information. When forecasting we start with historical data and try to predict future situations, this is an ideal application for the artificial network. We give the algorithm enough data to learn from the past, so giving it the information of today it can forecast tomorrow.

In their paper Fadlalla and Lin (2001) address the use of Neural Networks in Finance. The major areas where ANN applications are used are characterized by data intensity, unstructured nature, a high degree of uncertainty, and hidden relations. They found that in most cases the artificial applications outperformed the traditional statistical models. Applications used in T-bills forecasting, asset management, and portfolio selection have also shown significant success. In their study, the researchers chose a sample of 40 studies to analyze and see how and in which areas in finance the neural networks are mostly applied. In 11 of the sample papers, the authors compared the performance of neural network models with the performance of statistical models in forecasting. In 10 of 11, the authors concluded that the neural networks perform better than the statistical models. These conclusions support also the suggestion, that the complexity of financial forecasting may be more suitable for ANN models than classical statistical models.

Earlier research on EPS forecasting by Zhang, Cao and Schneiderjans J. (2004) including 283 firms spanning 41 industries concludes also that the neural network approach incorporating fundamental accounting variables results in forecasts that are more accurate than linear forecasting models. By fundamental accounting variables, the authors refer to accounts receivable, inventory, and capital expenditures. Abarbanell S. and Bushee J. (1997) examined how firm-specific earnings news, industry context, and macroeconomic trends affect the future earnings of the firm. According to their study, there are some accounting variables that can give us information about the future earnings performance like account receivables, capital expenditures, inventory, and gross margin of the company. They concluded also that macroeconomic variables do not have a direct effect on company earnings but can form the relationship between the fundamental signals and the futures earnings indirectly. The usefulness of fundamental analysis for prediction earnings is also tested and confirmed in a recent study made by Hancock and Seng (2012). Their research and regression analysis confirmed the explanatory power of inventory and capital expenditures for EPS, as well as other variables like labor force and effective tax rate.

In their study Zhang, Cao and Schneiderjans J. (2004) compared both univariate models and multivariate models. When testing the prediction accuracy of univariate ARIMA and univariate NN model they concluded with better performance for the NN model. However, as they also emphasize their evidence of the superiority of the univariate NN model was weak. Their research shows that specifically including a collection of valuable fundamental variables improves the forecast accuracy of NN models compared to linear models. The importance of the selection of highly informative accounting variables is shown also in the study of Falas, Charitou and Charalambous (1994). Results obtained in their comparison study show only 2% overall improvement of forecasting accuracy when using a NN model. Their conclusion addresses the importance of investigation and selection of the right accounting variables when creating a model based on the neural network, as this is crucial for the overall learning and success of the algorithm.

However, it is important to emphasize that there are not only studies with positive results. In their research Callen L., et al. (1996) show that the resulting errors using NN are significantly larger than those generated by a linear time series model ARIMA. The main difference between these two studies is that Zhang, Cao and Schneiderjans J. (2004) included fundamental accounting information in their NN model. Abarbanell S. and Bushee J. (1997) address the nonlinear relationship between fundamental financial signals and future EPS, which makes the linear models not applicable at least in theory.

## 2.2 Statistical models

The history of research on finance and economic modeling is long. Time series analysis is one of the most widely used methods for forecasting in econometrics. A time series is a collection of observations made periodically through time. Time series are used in many economic areas like predictions of GDP, unemployment rate, interest rates, and stock prices. When we talk about time series analysis, we talk about a group of methods used in econometrics. According to Chatfield (2000), forecasting methods can be classified into three categories: judgmental forecasts, univariate methods, and multivariate methods. Judgmental forecasts are based on subjective judgment, intuition – typically financial analysts can make these types of forecasts. Univariate forecasts where the predictions depend on past and present values of a single dependent variable. Multivariate methods where forecasts on a given dependent variable depend on its own past values and past values of at least one more explanatory variable. Stock and Watson (2015) explain that the main idea behind time series econometrics is that we can predict the future analyzing the past values of the variable we want to predict. This means that using linear regression models with time series, we assume that the future will be like the past, the concept also called stationarity. Stationarity requires the future to be like the past, at least in a probabilistic sense.

The autoregressive integrated moving average (ARIMA) method developed by Box and Jenkins is one of the most important and widely used methods in forecasting as stated also in the research paper of P. G. Zhang (2003). ARIMA models are quite flexible and they can represent several different types of time series like pure autoregressive (AR), pure moving average (MA) and combined AR and MA (ARMA) series, their major limitation is their linear form. The assumption of linear correlation in these models makes them irrelevant to use when working with real-world non-linear patterns. According to Singh Vaisla and Kumar Bhatt (2010), statistical models have proven their effectiveness in forecasting over the years. However, as stated by the researchers, the prediction ability of these models is reduced as the data series become more complex.

In their book Introduction to Econometrics Stock and Watson (2015) address the different types of models and how they are used. Autoregression (AR) is the most basic one of the time series analysis. In this model, the forecasts are made by a regression model that relates a time series variable to its past values. The mathematical expression of the population for the time series $Y_t$ is:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

*(2.2-1)*

If there is historical data on Y and I want to forecast its future value $Y_{t+1}$ based on equation (2.2-1) the equation of the forecast is:

$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T$$

*(2.2-2)*

Where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated using the historical data. Forecasts are an estimation of the future and there are also wrong estimations, the forecast error is the mistake made by the forecast model. It is the difference between the actual value and the estimated value.

Forecast error = $Y_{T+1} - \hat{Y}_{T+1|T}$

*(2.2-3)*

There are different ways of measuring and analyzing the forecast errors. These are further explained and discussed under the performance analysis.

The AR model is an important forecasting tool and is used as a basis of many fundamental ideas in time series forecasting and the models that have been developed later on (Chatfield C. , 2000). Yet, as Stock and Watson (2015) point out as well, Economic theory often suggests using other variables that could help increase the forecasting accuracy of the interest variable. Including one or more explanatory variables than the dependent variable is another group of time series analysis called Regression with Multiple Predictors.

According to Gorr (1994), univariate time series models for forecasting are simple and well established. In his research, he emphasizes that there have been done a lot of improvements in this statistical area and that it is unlikely for neural networks to bring any major advancement for forecasting. In their study Callen L., Kwan C.Y., Yip C.Y., & Yuan (1996) also argue that the simple linear time series models are well equipped for great forecasting accuracy. In addition to that they point out that, simple time series analysis has through the years shown better forecasting performance than more complex models like ARIMA. In their study, Callen et al. (1996) concluded that even though quarterly earning are seen as seasonal and non-linear, the linear time series analysis yields better performance than the neural network model.

While time series forecasting methods assume a correlation between the past and the future, there are also studies concluding that time series of predicting EPS based only on annual EPS data have done no better than simple random walk models (Conroy & Harris, 1987). In their research Hopwood, Mckeown and Newbold (1982) studied if the prediction rate of earnings using univariate time series is improved if instead of using annual data, they use quarterly data. They found that while annual models generally are close to random walks, there seems to be a relationship between past performance and current performance when using quarterly data. Therefore Hopwood, Mckeown and Newbols (1982) also concluded that using quarterly earnings increases the prediction power substantially. There have been other studies as well supporting the results that quarterly earnings cannot be described as random walk and they are not independent (Griffin A. (1977), Brown and Rozeff S. (1979)).

## 2.3 Literature discussion

As discussed above, studies are concluding that univariate statistical models perform well, when based on past EPS performance. There are however also studies showing that NN models based on past EPS values and fundamental variables have great prediction accuracy. There is particularly one question that arises from this literature review. Does the complex Neural Network model outperform the standard Time Series analysis which has been used in forecasting for many years? Does the increased complexity decrease the errors in forecasting?

Whether statistical or ANN method will give the best results in forecasting EPS is difficult to say based on the literature review only. On one hand, we have articles showing that univariate time series predictions on earnings/EPS have been very successful like Brown and Rozeff S. (1979), Griffin A. (1977), Hopwood, Mckeown and Newbold (1982). While on the other hand, we have articles supporting the idea that the complexity of EPS forecasting is more suitable for Neural Networks like the work of Zhang, Cao and Schneiderjans J. (2004).

Context sensitivity is the third possibility that has been discussed over the years. In their study Hill, Marques, O'Connor, and Remus (1994) did a comprehensive literature review analyzing many studies that compare artificial neural network models and different statistical models. They concluded that the artificial approach is just as good as and “occasionally better” than the regression models. However, they addressed their belief that there are only in certain circumstances the artificial approach outperforms the statistical approach. Therefore they further suggested that studies on analyzing and describing these conditions need to be done.

The possibility of neural network models being superior in only certain types of situations is discussed also in the editorial publication of Chatfield C. (1993). He suggests that there is a need for more systematic evidence where the two approaches are compared in a statistically justified way. Whether artificial models or statistical models are better in prediction is still a question that we need to work on with our data and research, so we can hope that one day economists will all agree on which of these methodologies is best to use in different types of forecasting and cases.

# 3. Data and Methodology

One of the main goals of this research paper was to have generalizable results. Therefore, there were three main criteria when choosing the companies for the study. The first criterion was that the company has to operate on an international level as in this way I will minimize the geographical economic effects on the company. The second criterion was that the company does not have any missing values in its quarterly reports for the past 10 years, as this is crucial for the models. The last criterion was that the companies will be from different industries, so the industry factor is minimized in the same way as the geographical.

Based on the literature review I decided to use quarterly EPS values rather than annual EPS values. The quarterly data for 10 companies were collected from www.stockrow.com. Stockrow is an online database directly connected to the US Securities and Exchange Commission (SEC). All the companies with more than $10 million in assets whose securities are held by more than 500 owners must file annual and other periodical reports according to SEC. Data like quarterly income sheet, balance sheet, share price, and key financial indicators for 10 big international companies with at least 10 years of history were collected. The next step of the research was to define the variables that are going to be used in the analyses. After defining the variables for the models that are going to be used the data was formatted according to the requirements for the different models.
The companies are NIKE, Walmart, Exxon Mobile, Amazon, Apple, Chevron Corporation, General Electric, Microsoft, Verizone Communication, and IBM.

It is also important to mention that different data may need a different type of modeling. What I have seen from previous studies as well is that to be able to get great forecasting results, the researchers have to think in terms of the data-model match. While working with the data I collected I tested different models with different complexity to see how the data fits the models and vice versa. To test the different hypotheses from the literature review I decided to make a comparison between the performance of two models. Specifically, I constructed one neural network time series NARX model and one univariate Time Series model. The models and their constructions are further explained in the sections below.

## 3.1 Neural Network model

The choice of fundamental variables is based on previous studies by Abarbanell S. and Bushee J. (1997) and Hancock and Seng (2012). Since both studies found a highly significant relationship between inventory and EPS these are the variables I chose to include in my NN model. Past EPS values were included as they are considered to involve important information for the future development of the company. EPS and Inventory are variables that can easily be taken from the available yearly reports for the companies.

After calculating and formatting the data in the right configuration, it was divided into three categories *input, target, and test.* In total 40 observations of quarterly inventory and quarterly EPS were collected for all the companies. *Inputs* included Inventory and EPS for the period of the first quarter of 2010 to the last quarter of 2018 included (36 observations). The model is supposed to use the past EPS values and inventory values to predict one step ahead of performance for the company. As Hyndman J. and Athanasopoulos (2018) emphasize in their book the only way to determine the accuracy of a model is to test it with new data. Therefore, the model was trained with data from 36 observations, and the last 4 observations were used for testing the model.

To develop the Neural Network model, I used the Neural Network application in MatLab. Since MatLab has already developed neural network-scripts within its applications, this reduced both the time for creating the model and the probability of making mistakes. The data I had for the companies was calculated, formatted, and divided into three groups. After testing different models that can be used, I decided to use a Nonlinear Autoregressive model with

External (Exogenous) Input (NARX). This is a model with backpropagation that gives the opportunity to predict series *y(t)* given *d* past values of *y(t)* and other series *x(t)*. The function of the model is

$$y(t) = f(x(t-1), \ldots x(t-d), y(t-1), \ldots, y(t-d)$$

*(3.1-1)*

Figure 3.1 represents a visualization of the function

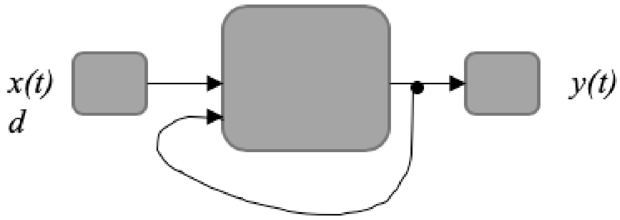


*Figure 3.1*

The NARX model is a dynamic artificial model used for modeling of nonlinear relations. According to P. Menezes Jr and Barreto A. (2008), this model has been successfully used for predictions including financial time series, biomedical time series modeling, and communication network traffic prediction for example. In their research paper, the authors emphasize also that in particular when there are noisy time series the NARX models outperform the standard linear methods. Their study compared the prediction abilities of different models and in particular the feedback loops to improve the prediction performance. Their results show that NARX is a prediction tool that consistently outperforms standard neural network-based models. It is important to emphasize that when studying and working on my research on Matlab, NARX was also the main model recommended for working on forecasting. This is also why I concluded that this will be the most appropriate model to use in my research.

The process of relations recognition is divided into three timesteps within the model. This means that the given dataset of *inputs* is further divided into three groups. Step one is *training*, there 70% of the data is presented to the network during training, and the network is adjusted according to its errors to create better connections. The second step is *validation*. During this process, 15% of the data is used to measure the generalization of the model and improve it. The last step includes *testing*, where 15% of the data is tested to provide an independent measure of the performance of the model after the training. It is important to mention that these percentages can be changed based on what the researcher wants to focus on. Some would prefer to have more data for training the model while within big data sets it can be preferable to use more than 15% for testing or validation.

When creating the model, I was also required to choose the number of neurons and the number of time delays within the network. As the inputs for training were given when working with the further development of the accuracy I tried to experiment with the numbers of neurons and time delays to test how they will affect the training and performance of the model. After experimenting for different companies, I came to the conclusion that 10 neurons and 4-8-time delays within the network seem to give more appropriate results and these were the parameters I used further for the other companies as well.

The final step of training the network is choosing the algorithm that fits best the available data. In this research, I chose the Levenberg-Marquardt algorithm. This algorithm requires more memory to train but less time. Training stops automatically when the model stops improving in good generalization. The neural network model for each company was trained until *training, valuation, and testing* of the model gave an R-squared over 0.90. Every time the model is trained, a report of its performance is created. This report summarizes the errors and includes a regression line showing how good does the information from the inputs fits the given outputs. Figures 3.2 and 3.3 show the results of the training for NIKE.

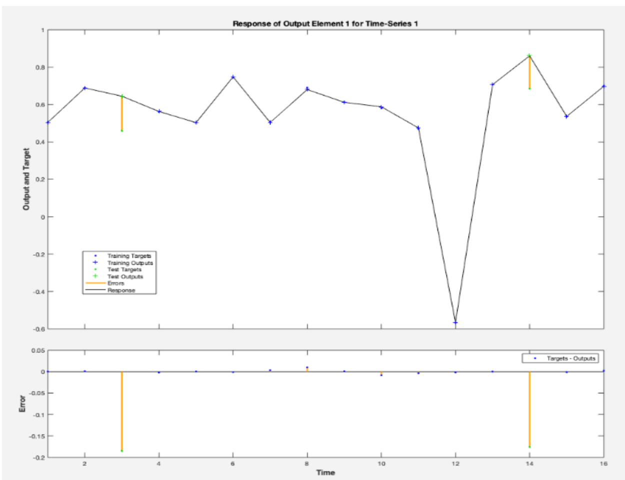


*Figure 3.2*

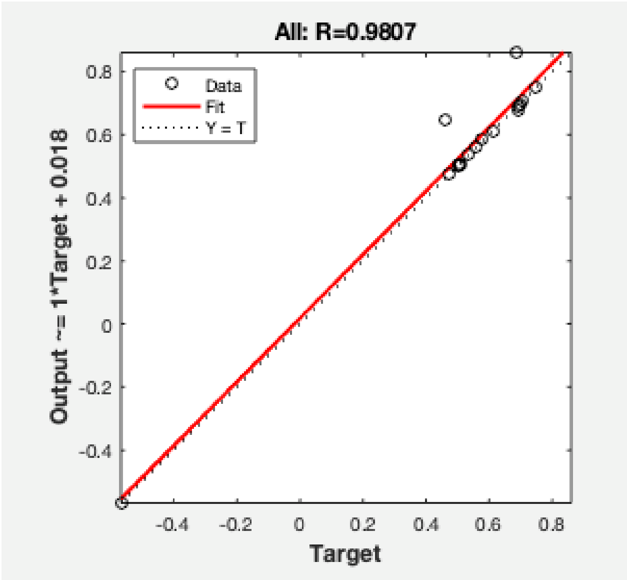


*Figure 3.3*

The regression line is used as a measure of how good the model is in creating relationships between inventory and past EPS values to predict future EPS values. R-squared over 0.90 indicates a good fit between the inputs and the outputs in the training set. However, as it is emphasized by Hyndman J and Athanasopoulos (2018) the good training fit will not necessarily mean a good forecast.

After the models were created and their test results gave the required match between the inputs and the outputs, the next step was to test the prediction ability of the network with real data. Figures 3.4, 3.5 and 3.6 show the variation between the predicted and real EPS values for the first, second and third predicted quarter.

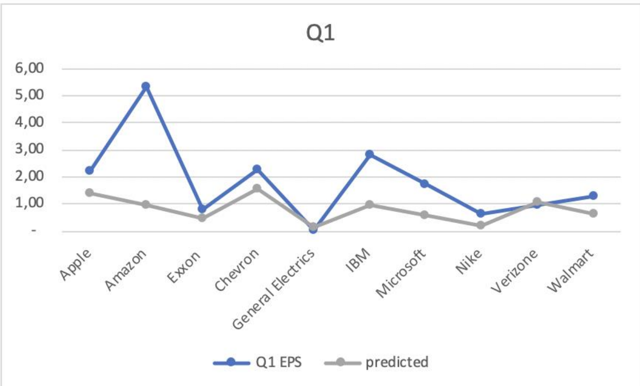


*Figure 3.4*

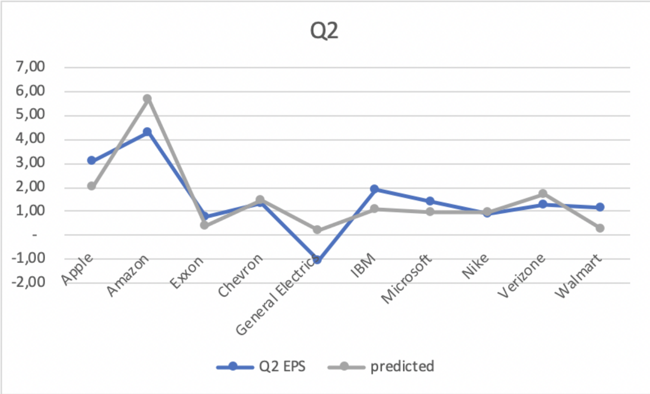


*Figure 3.5*

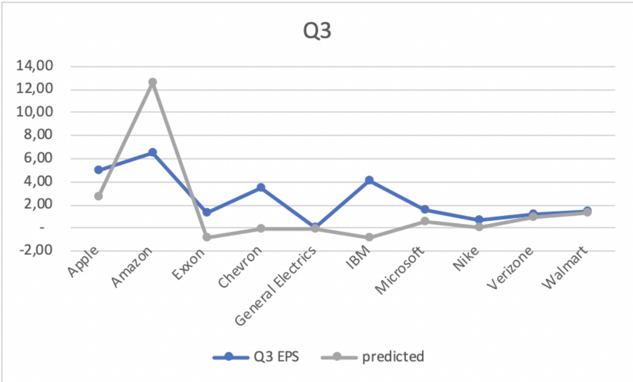


*Figure 3.6*

## 3.2 Statistical model

The first step of the statistical forecasting process is the same as the first step while working with NN models, to decide which model to use. Given the time scope of my study and the data available I had to choose a model that will be appropriate both regarding data and time. Therefore, I decided to use a simple univariate Time Series Regression model with moving average, seasonal effect, and trend modeling. As in the NN model, the first 36 points were used to create the model and the last 4 data points were used for forecasting and testing. In time series analysis we assume that there is linearity and based on analysis of the past observations, we calculate the future.

The Time Series analysis started with calculating the moving average (MA) for the 36 data points. A moving average can be calculated by the following formula:

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j}$$

(3.2-1)

Here $m = 2k + 1$, by averaging the values within $k$ periods $t$ the trend-cycle at time $t$ is calculated. As explained in the book Forecasting: Principles and Practice by Hyndman J & Athanasopoulos (2018) the purpose of this method is to eliminate some of the randomnesses in the data, by leaving smooth trend-cycle components. After calculating the MA with $k=4$ for the data I had, the next step was to conduct the centered moving average (CMA) of the moving average. The reasons for calculating CMA were to smooth the moving average and estimate the trend-cycle from seasonal data. CMA was calculated for 2 periods of MA.

The next step was to extract the seasonality $S_t$ and irregularity $I_t$ components of the data based on CMA and the EPS observations. This was done by dividing the observation value $Y_t$ on the CMA for the same period. Based on the EPS value relative to the average I was able to say how much of the difference is based on seasonality and irregularity. Since these two components were now conducted in one value ($S_t + I_t$) I had to extract the seasonality of this value. Assuming that there is a quarterly seasonal component I averaged the quarterly values and found the seasonality component $S_t$. After conducting the seasonality and irregularity components I had to discover the trend for the data. This was done by deseasonalizing the data by dividing $Y_t$ on $S_t$.

To find out what the trend in the data is, I conducted a regression analysis for the deseasonalized data. The trend was calculated using the following formula:

$$T_t = \beta_0 + \beta_1 * t$$

(3.2-2)

The final step of the analysis was to calculate the forecasted value $\hat{Y}_t$. This was done by using the formula:

$$\hat{Y}_t = T_t * S_t$$

(3.2-3)

Table 3.1 shows the calculations explained above for NIKE.

| | | | Quarterly EPS for Nike | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Yt/CMA | Quart average St,It | Yt/St | Tt | |
| t | | NIKE | EPS $Y_t$ | MA(4) | CMA(4) | St,It | St | Deseasonalize | Trend | Forecast |
| 1 | Year 1 | Q1-2010 | | | | | 0,70 | 0 | 0,27 | 0,19 |
| 2 | | Q2-2010 | 0,27 | | | | 1,16 | 0,23 | 0,28 | 0,33 |
| 3 | | Q3-2010 | 0,29 | | | | 1,19 | 0,24 | 0,29 | 0,34 |
| 4 | | Q4-2010 | 0,24 | 0,27 | 0,27 | 0,89 | 0,85 | 0,28 | 0,30 | 0,25 |
| 5 | Year 2 | Q1-2011 | 0,28 | 0,27 | 0,27 | 1,00 | 0,70 | 0,39 | 0,30 | 0,21 |
| 6 | | Q2-2011 | 0,32 | 0,28 | 0,29 | 1,10 | 1,16 | 0,27 | 0,31 | 0,36 |
| 7 | | Q3-2011 | 0,35 | 0,29 | 0,30 | 1,17 | 1,19 | 0,29 | 0,32 | 0,38 |
| 8 | | Q4-2011 | 0,26 | 0,30 | 0,30 | 0,85 | 0,85 | 0,30 | 0,33 | 0,28 |
| 9 | Year 3 | Q1-2012 | 0,31 | 0,31 | 0,30 | 1,01 | 0,70 | 0,44 | 0,34 | 0,24 |
| 10 | | Q2-2012 | 0,29 | 0,30 | 0,30 | 0,99 | 1,16 | 0,25 | 0,35 | 0,40 |
| 11 | | Q3-2012 | 0,31 | 0,29 | 0,29 | 1,09 | 1,19 | 0,26 | 0,35 | 0,42 |
| 12 | | Q4-2012 | 0,21 | 0,28 | 0,30 | 0,70 | 0,85 | 0,25 | 0,36 | 0,31 |
| 13 | Year 4 | Q1-2013 | 0,48 | 0,33 | 0,34 | 1,44 | 0,70 | 0,69 | 0,37 | 0,26 |
| 14 | | Q2-2013 | 0,37 | 0,34 | 0,36 | 1,02 | 1,16 | 0,31 | 0,38 | 0,44 |
| 15 | | Q3-2013 | 0,44 | 0,38 | 0,39 | 1,13 | 1,19 | 0,37 | 0,39 | 0,46 |
| 16 | | Q4-2013 | 0,30 | 0,40 | 0,39 | 0,78 | 0,85 | 0,36 | 0,40 | 0,33 |
| 17 | Year 5 | Q1-2014 | 0,39 | 0,37 | 0,38 | 1,03 | 0,70 | 0,55 | 0,40 | 0,28 |
| 18 | | Q2-2014 | 0,40 | 0,38 | 0,40 | 1,01 | 1,16 | 0,34 | 0,41 | 0,48 |
| 19 | | Q3-2014 | 0,56 | 0,41 | 0,42 | 1,32 | 1,19 | 0,47 | 0,42 | 0,50 |
| 20 | | Q4-2014 | 0,38 | 0,43 | 0,44 | 0,86 | 0,85 | 0,45 | 0,43 | 0,36 |
| 21 | Year 6 | Q1-2015 | 0,46 | 0,45 | 0,46 | 0,99 | 0,70 | 0,65 | 0,44 | 0,31 |
| 22 | | Q2-2015 | 0,50 | 0,47 | 0,49 | 1,03 | 1,16 | 0,43 | 0,45 | 0,52 |
| 23 | | Q3-2015 | 0,69 | 0,51 | 0,52 | 1,33 | 1,19 | 0,58 | 0,45 | 0,54 |
| 24 | | Q4-2015 | 0,46 | 0,53 | 0,54 | 0,85 | 0,85 | 0,54 | 0,46 | 0,39 |
| 25 | Year 7 | Q1-2016 | 0,56 | 0,55 | 0,55 | 1,01 | 0,70 | 0,80 | 0,47 | 0,33 |
| 26 | | Q2-2016 | 0,50 | 0,55 | 0,56 | 0,90 | 1,16 | 0,43 | 0,48 | 0,56 |
| 27 | | Q3-2016 | 0,75 | 0,57 | 0,57 | 1,30 | 1,19 | 0,63 | 0,49 | 0,58 |
| 28 | | Q4-2016 | 0,51 | 0,58 | 0,60 | 0,85 | 0,85 | 0,60 | 0,50 | 0,42 |
| 29 | Year 8 | Q1-2017 | 0,69 | 0,61 | 0,63 | 1,10 | 0,70 | 0,98 | 0,50 | 0,35 |
| 30 | | Q2-2017 | 0,61 | 0,64 | 0,62 | 0,99 | 1,16 | 0,53 | 0,51 | 0,60 |
| 31 | | Q3-2017 | 0,58 | 0,60 | 0,59 | 0,98 | 1,19 | 0,49 | 0,52 | 0,62 |
| 32 | | Q4-2017 | 0,47 | 0,59 | 0,43 | 1,09 | 0,85 | 0,56 | 0,53 | 0,45 |
| 33 | Year 9 | Q1-2018 | - 0,57 | 0,27 | 0,29 | -1,98 | 0,70 | - 0,81 | 0,54 | 0,38 |
| 34 | | Q2-2018 | 0,71 | 0,30 | 0,31 | 2,27 | 1,16 | 0,61 | 0,55 | 0,63 |
| 35 | | Q3-2018 | 0,69 | 0,32 | | | 1,19 | 0,58 | 0,55 | 0,66 |
| 36 | | Q4-2018 | 0,54 | | | | 0,85 | 0,63 | 0,56 | 0,48 |
| **37** | **Year 10** | | | | | | **0,70** | | **0,57** | **0,40** |
| **38** | | | | | | | **1,16** | | **0,58** | **0,67** |
| **39** | | | | | | | **1,19** | | **0,59** | **0,70** |
| **40** | | | | | | | **0,85** | | **0,59** | **0,50** |

*Table 3.1*

Figures 3.7, 3.8, 3.9 show the forecasted and the real values for all companies for the three forecasted periods.

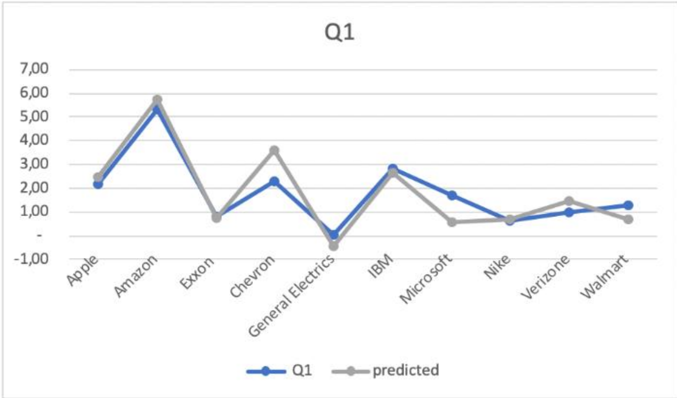


*Figure 3.7*

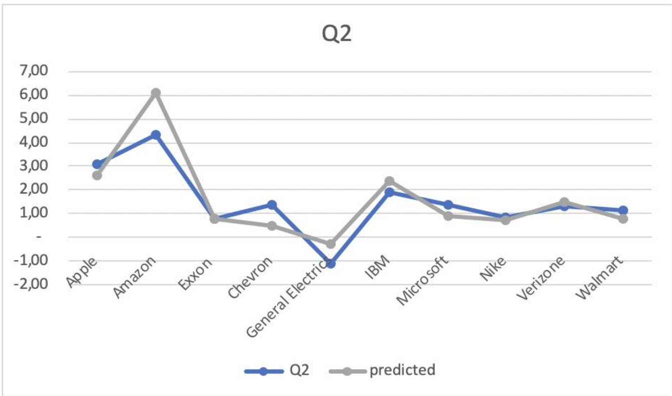


*Figure 3.8*

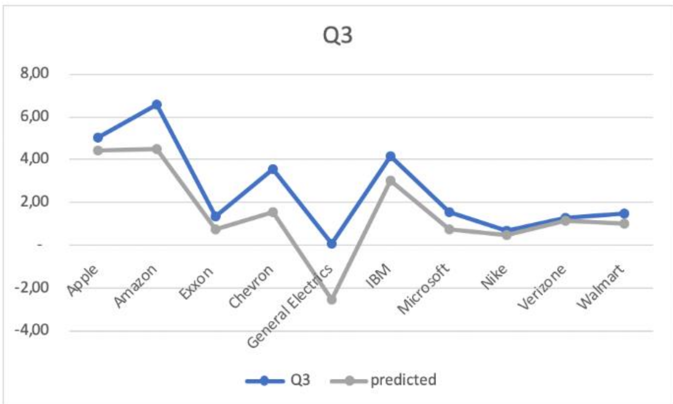


*Figure 3.9*

# 4. Performance analysis

The evaluation of the forecast models is one of the most important parts of comparative analysis. To measure and analyze the performances of the models I decided to use Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and Mean Absolute Deviation (MAD). MAPE, MSE, and MAD were used to classify and compare model performances also in the research of Zhang, Cao and Schneiderjans J. (2004), Kumar and Pradhan P. (2010), Jarett E. (1990) and many other studies. Before analyzing the results, it is of course useful to see them first. Figures 4.1, 4.2, and 4.3 show the real values and the forecasted values for both the Neural Network and the Time Series models for the three forecasted periods.

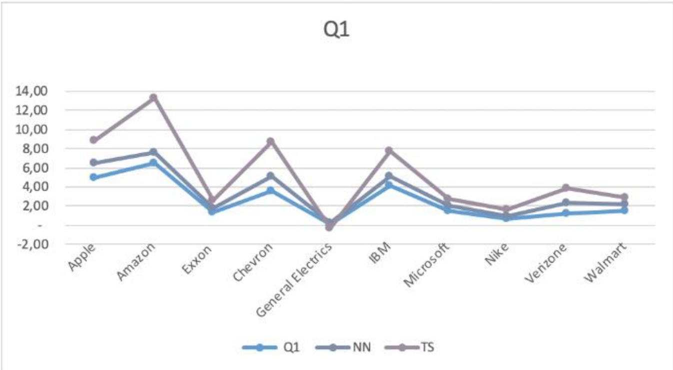


*Figure 4.1*

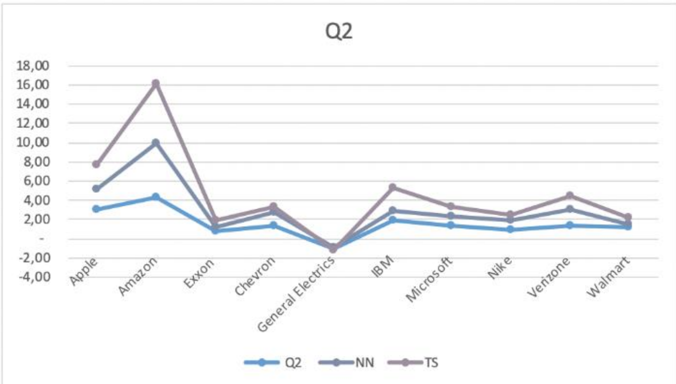


*Figure 4.2*

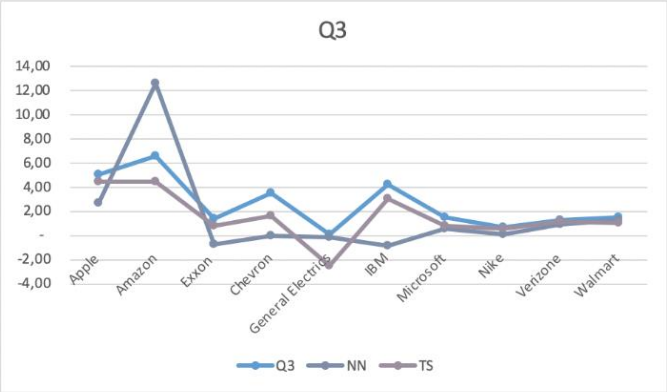


*Figure 4.3*

Taking a look at figures 4.1 and 4.2 it is easy to see that the NN model seems to be overall better in its performance than the TS model. However when we go on to figure 4.3 the differences in forecasting performance are in favor of the statistical model. There seems to be a negative relationship between the time period and the forecastability of the Neural Network model. Performance analysis is a complex process and it is difficult to make conclusions based on forecast fitting only. We can measure forecast accuracy by summarizing and examining forecast errors in different ways.

Mean absolute percentage error = MAPE is frequently used to compare two forecasting models, where MAPE is calculated by using the following formula:

$$MAPE = mean\left(\left|\frac{100e_t}{y_t}\right|\right)$$

*(4-1)*

Table 4.1 summarizes the results for MAPE.

| MAPE | NN | TS |
|---|---|---|
| Apple | 39 % | 13 % |
| Amazon | 68 % | 27 % |
| Exxon | 81 % | 17 % |
| Chevron | 46 % | 60 % |
| General Electrics | 409 % | 2333 % |
| IBM | 77 % | 19 % |
| Microsoft | 54 % | 51 % |
| Nike | 56 % | 19 % |
| Verizone | 23 % | 25 % |
| Walmart | 47 % | 38 % |
| *Average* | *90 %* | *260 %* |

*Table 4.1*

The average MAPE for the NN is 90% while for the TS is 260%. However, when we take a closer look it is obvious that the results are highly influenced by the extreme values for General Electrics. As explained by Hyndman J. and Athanasopoulos (2018) even though MAPE is a good error measure it can take extreme values whenever $Y_t$ is close to zero. This is the case with General Electrics. Since the values of EPS for Q1 and Q3 are 0,01 and 0,08 the MAPEs for both models are extremely high. After removing the outliers the average MAPE for the Neural Network model is 55%, while it is 30% for the Time Series model. For seven of nine observations, the MAPE is lower using the linear model. Based on MAPE the Time Series analysis seems to have smaller error percentages and better accuracy after removing the outliers.

It would also be interesting to see what MAPE is relative to the 10-year average EPS value. This would give a better understanding of how far the predictions are from the average EPS values for the companies. Table 4.2 summarizes these results.

| MAPE | NN | 10y avrg. EPS | MAPE/EPS | TS | 10y avrg. EPS | MAPE/EPS |
|---|---|---|---|---|---|---|
| **Apple** | 39 % | 2,00 | 19 % | 13 % | 2,00 | 7 % |
| **Amazon** | 68 % | 1,52 | 45 % | 27 % | 1,52 | 18 % |
| **Exxon** | 81 % | 1,50 | 54 % | 17 % | 1,50 | 11 % |
| **Chevron** | 46 % | 2,03 | 23 % | 60 % | 2,03 | 30 % |
| **General Electrics** | 409 % | 0,07 | 5845 % | 2333 % | 0,07 | 33329 % |
| **IBM** | 77 % | 2,99 | 26 % | 19 % | 2,99 | 7 % |
| **Microsoft** | 54 % | 0,71 | 76 % | 51 % | 0,71 | 72 % |
| **Nike** | 56 % | 0,45 | 124 % | 19 % | 0,45 | 42 % |
| **Verizone** | 23 % | 1,15 | 20 % | 25 % | 1,15 | 22 % |
| **Walmart** | 47 % | 1,14 | 42 % | 38 % | 1,14 | 33 % |

*Table 4.2*

As the overall results show that the mean absolute percentage error is lower relative to the 10-year EPS average, the extreme values get even more extreme. Before removing the General Electrics, the average MAPE/10-year EPS average was 627% for the artificial model and 3357% for the statistical model. After removing General Electrics from the table the average MAPE/10-year EPS average for the artificial model is 48% and for the statistical model 27%. A comparison of MAPEs both for the predicted time period and for a larger time period yields strong evidence in favor of the TS model when excluding the outliers. There is seemingly a conflict between the graphs shown in figures 4.1 and 4.2, and the results MAPE numbers show. Figures 4.1 and 4.2 are obviously showing a smaller gap between the real values and the predicted NN values, while the mean absolute percentage errors are smaller for the TS model.

In order to determine if these results apply only to these samples or the population as a whole, I decided to conduct t-tests for all the metrics I have calculated. The objective is to see if these results are statistically significant. The t-test I am calculating in this paper is a t-test for two samples with assumed unequal variances. The null hypothesis in this test is that the means in these populations are the same but the variances may be different. In order to reject the null hypothesis, I need a p-value that is lower than 0,05, which is the significance level I chose. If the p-value is larger, it means that there is no evidence that the means of this metric for the NN and the TS models are different for the population. This test is also called Welch`s test as it was developed by B. L. Welch in 1938.

The t-test for all NN and TS MAPEs shows a *p-value* of 0,24. Consequently, the differences in MAPEs are not statistically significant to reject the null hypothesis. However, the t-test for

MAPEs without General Electrics values gives a *p-value* of 0,004. It is very fascinating to see how much the outliers actually affect the results of the test. On one hand, we cannot reject the null hypothesis, while on the other hand removing two outliers results in a totally contradicting conclusion. I further examined also how the t-test performance was for the MAPEs relative to10-year EPS average. These tests were more univocal, the p-value with the outliers was 0,22 and without the outliers 0,07. Based on the *p-value* without the outliers, the zero hypothesis can be rejected. This means it is statistically justified to say that the means of MAPEs may be different. It is important to emphasize that the fact that the *p-value* is low enough to reject the null hypothesis, does not mean that the alternative hypothesis is true.

Results indicate that on average the statistical model also outperforms the artificial model when MSEs are compared. MSE is calculated by using the following formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} E_t^2$$

*(4-2)*

The summary of MSE results on company and prediction period level are presented in tables 4.3 and 4.4.

| | **Neural Network** *MSE* | **Time Series** *MSE* |
|---|---|---|
| **Apple** | 2,39 | 0,21 |
| **Amazon** | 18,98 | 2,59 |
| **Exxon** | 1,60 | 0,12 |
| **Chevron** | 4,37 | 2,15 |
| **General Electrics** | 0,56 | 2,60 |
| **IBM** | 9,64 | 0,49 |
| **Microsoft** | 0,81 | 0,73 |
| **Nike** | 0,20 | 0,03 |
| **Verizone** | 0,10 | 0,10 |
| **Walmart** | 0,43 | 0,25 |
| **Average** | **3,91** | **0,93** |

*Table 4.3*

| | Neural Network | Time series |
|---|---|---|
| *Prediction period* | *MSE* | *MSE* |
| Q1 | 2,54 | 0,42 |
| Q2 | 0,65 | 0,56 |
| Q3 | 8,53 | 1,81 |
| Average | 3,91 | 0,93 |

*Table 4.4*

Characteristic for MSE is that this metric "punishes" large errors relative to small errors. This can lead to misclassification and even wrong conclusions when the data consist outliers. In their book, Hyndman and Athanasopoulos (2018) emphasize the importance of not only detecting but also analyzing outliers. According to the authors, it is reasonable to remove the outliers only when we know that the data leading to these outliers is incorrect. However, if this is not the case it is important to examine and analyze the outliers. Therefore in this study, it would be appropriate to see overall average MSE both with and without the outliers. From table 4.3 the outliers can be identified. Amazon, IBM, and Chevron seem to have extreme MSE values compared to the majority of the data. The average NN MSE for all companies is 3,91 and for TS is 0,93. After removing the outliers the average NN MSE is 0,87 and for TS it is 0,58. This experiment shows that extreme values may be behind the big differences between these two models.

To be able to determine if these results are statistically justified I conducted a Welch`s test for MSEs in table 4.3 as well. The null hypothesis is that the means in these populations are the same but the variances may differ. The empirical results clearly indicate that the null hypothesis cannot be rejected. The *p-value* for a t-test with all companies is 0,08, while it is 0,27 after removing the outliers (IBM, Amazon, and Chevron). Based on the t-test, there seems to be no difference between the MSE mean of the NN and the TS models.

Randomness is very important in statistics because it reduces the probability of bias. Even though there may be a good reason to remove the outliers, it will be also useful to see how the average MSE changes when removing random companies from the sample. Without taking a look at results, randomly I took out two companies – Nike and Walmart. The average NN MSE with the outliers is still very high – 4,81. For the TS mode,l it is 1,12. After removing the outliers only 5 companies were left in the sample. Even though the difference got smaller, the results did not change. The statistical model had still better performance where the MSE average was 0,75. The artificial model had an MSE average of 1,09. This consistency suggests that based on the MSE metric only, the statistical model outperforms the artificial model. The t-test after removing Nike and Walmart had a *p-value* of 0,08, which is the same as the *p-value* for all MSEs. After removing the outliers and Nike and Walmart the *p-value* was 0,30. This experiment also confirms that the null hypothesis cannot be rejected. This consistency indicates that there is no empirical reason to claim that the MSE means for the NN and TS models may differ. Even though there are differences in this sample, there is no statistical reason to believe that these differences are existing in the population.

However, it is important for the analysis to try to understand what is happening in these samples. Why are the results as they are? Why does the statistical approach perform better? Why does the artificial model give extreme values? In order to evaluate the overall MSE differences between these two models, it is crucial to analyze the outliers. In the next section, I take a deeper look at the data for four companies based on their MSE results. I chose two companies with extreme results (Amazon and IBM) and two companies with normal results (Apple and Verizone). The objective is to analyze the input variables and discover potential reasons for the outlier results. In this way, I can see if there is a pattern within the data that may cause extreme values for the artificial model. The main reason for the MSE outliers is the high prediction errors for the NN model. Therefore it would be interesting to examine the differences between Apple, Verizone, IBM, and Amazon. MSEs for Apple and Veriozne are 2,39 and 0,10 using the artificial model. MSEs for IBM and Amazon are 18,98 and 9,64. Also, the MAPEs for Apple and Verizone are 39% and 23%, while the same metrics for IBM and Amazon are 68% and 77%. The main question is what may be the reason for the obvious confusion of the artificial model. Starting with Amazon one obvious thing is that there are continuous negative EPS values meaning that there are periods when EPS is negative for more than one quarter. This is not an issue neither with Apple nor with Verizone. IBM does not have this problem either. The development of the EPS and relative EPS changes do not show any meaningful differences either.

However, there is one thing that is common for IBM and Amazon and not for the other two companies. The trend of the Inventory seems to be inconsistent with the movement of EPS for IBM and Amazon. Figures 4.4 and 4.5 show the trend of the EPS and the trend of inventory for the four companies.

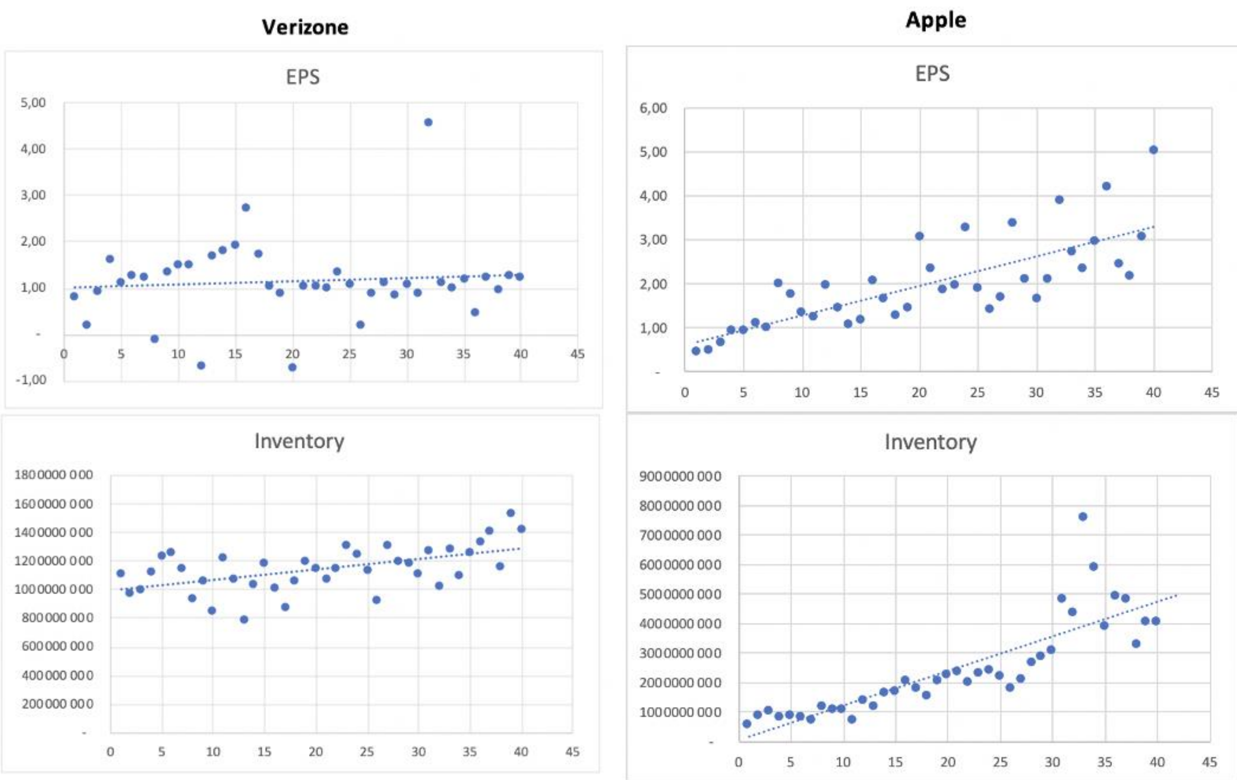


*Figure 4.4*

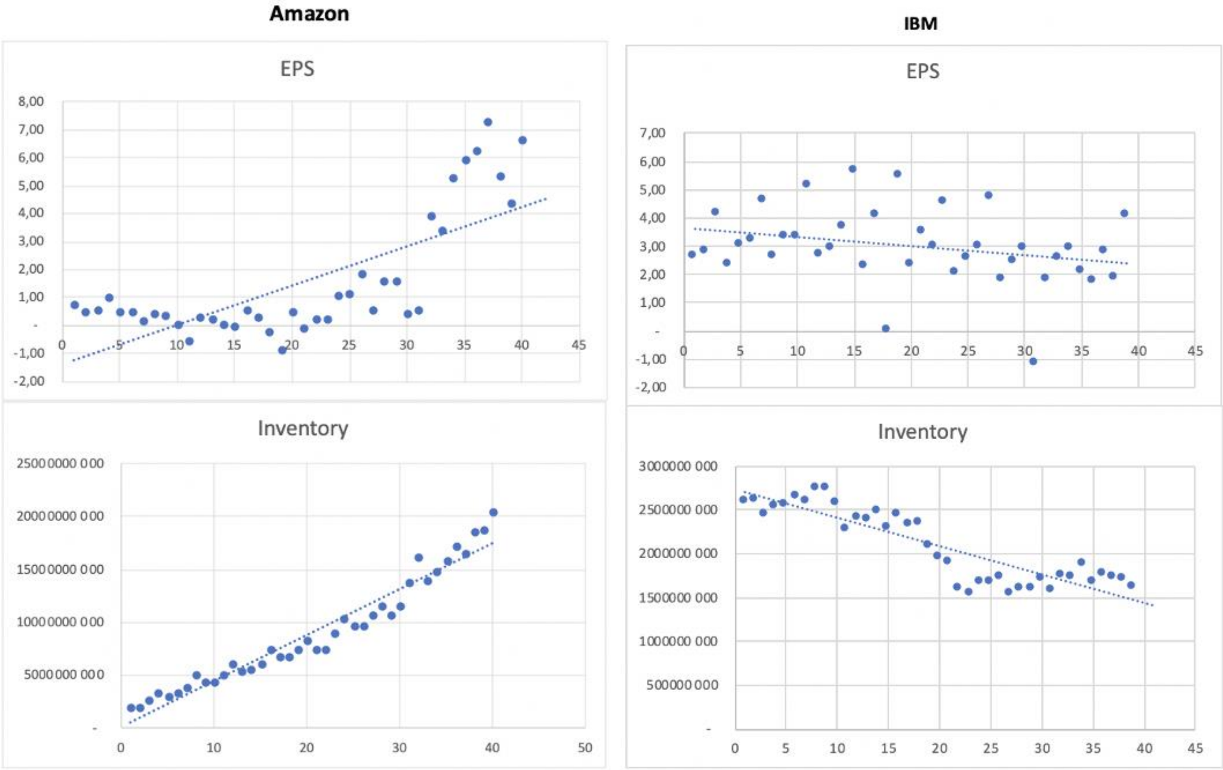


*Figure 4.5*

This situation is most obvious for IBM. To see more concrete the differences in how EPS changed relative to inventory I decided to calculate the average EPS percentage change relative to the percentage change of inventory. The formula for this calculation is:

$$\frac{EPS_{\Delta}}{Inventory_{\Delta}} = \frac{1}{n}\sum_{t}^{n} \frac{\frac{(EPS_t - EPS_{t-1})}{EPS_{t-1}}}{\frac{(Inventory_t - Inventory_{t-1})}{Inventory_{t-1}}}$$

(4-3)

For Apple and Verizone the average relative change was 1,11 and 4,03. While the average relative change of EPS to inventory for Amazon and IBM was 17,91 and – 76,29. The big differences between how inventory and EPS move within time maybe also one of the reasons why the NN was not able to create stronger relationships within the training data set. Since inventory is the second variable to explain and forecast EPS it is realistic to think that if their movements do not seem to be related, it will be difficult for the artificial model to create a good estimation of the information that the inventory variable brings. From previous studies discussed under the literature review, it has been revealed that inventory is one of the variables contributing to better accuracy in forecasting. However, this may not be the case for all types of companies. The way inventory moves relative to EPS seem to have an effect on the weights between the neurons in the artificial model. It is however uncertain if this contribution is significant.

The last metric to be analyzed is MAD. This metric shows how far the distance between each data point and the mean is. MAD is calculated by using the following formula:

$$MAD = \frac{1}{n}\sum_{i=n}^{n} E_t$$

*(4-4)*

The summary of MAD results on company and forecast period level are presented in tables 4.5 and 4.6.

| | Neural Network *MAD* | Time Series *MAD* |
|---|---|---|
| **Apple** | 1,39 | 0,44 |
| **Amazon** | 3,91 | 1,44 |
| **Exxon** | 0,93 | 0,22 |
| **Chevron** | 1,44 | 1,40 |
| **General Electrics** | 0,53 | 1,29 |
| **IBM** | 2,55 | 0,58 |
| **Microsoft** | 0,85 | 0,81 |
| **Nike** | 0,39 | 0,14 |
| **Verizone** | 0,29 | 0,27 |
| **Walmart** | 0,59 | 0,50 |
| **Average** | **1,29** | **0,71** |

*Table 4.5*

| | Neural network | Time Series |
|---|---|---|
| *Prediction period* | *MAD* | *MAD* |
| Q1 | 1,04 | 0,50 |
| Q2 | 0,68 | 0,58 |
| Q3 | 2,14 | 1,06 |
| Average | 1,29 | 0,71 |

*Table 4.6*

Table 4.5 shows that MAD for the artificial model is 1,29 and 0,71 for the statistical model. These results however are not surprising, since the same companies leading to extreme MSE values are increasing the MAD as well. Amazon and IBM can be detected as outliers in table 4.5. Removing the outliers decreases MAD of the artificial model to 0,80 and MAD of the statistical model to 0,63. Before removing the outliers the performance of the statistical model is 82% better in MAD, while without the outliers this performance is reduced by 55%. This change addresses once again the importance of detecting and analyzing outliers.

There is always a possibility that these results are related only to this sample only. This is why I conducted a t-test for MAD as well. The null hypothesis is that there is no difference between the mean MAD of the artificial model and the statistical model, while variances may differ. The t-test with the outliers gave a *p-value* of 0,08, while the t-test without the outliers gave a *p-value* of 0,24. The empirical results for MAD clearly indicate that the null hypothesis cannot be rejected. Therefore the conclusion is that there is no statistical evidence that the mean absolute errors of the two models are different. Obviously, the evidence from the t-tests shows that for MSE and MAD there is no proof that the results of this reasearch can be generalized. However, the t-test of MAPE without outliers indicated that the means between the two models may be different and the null hypothesis can be rejected. Then again, when using the statistical tests, there is always the possibility of making type I or type II error.

As emphasized earlier, the main question is not which of the models is doing greater, but rather why is it so. As Kaastra and Boyd (1996) also highlight in their study “the black box nature” of artificial models makes them more challenging to analyze. This has been also a reason for these models to be criticized. When working with time series it is much easier for economists to discover and explain errors. This is due to the fact that we are included in the whole process and we know what is happening. However, this is not the case with neural networks. There are many things in this “tedious software” as Kaastra and Boyd (1996) call it, that we cannot see or explain. We do not exactly know what is happening within the network and how the network discovers trends or seasonality for example. While in the linear model we know exactly how to take the trend component out of the sample if there is such a component in the neural network is not certain.

Obviously, the evidence is in favor of the statistical approach. Then again, there may be many reasons behind these results. One of the most important factors for the neural network’s success is data elements. The software needs many data points to be able to create connections and then based on these connections make forecasting. However, the number of data points that are “enough” is uncertain. The importance of the optimal amount of data is also addressed in the research of Foster, Collopy and Ungar (1992). In their research, they worked with short and noisy time series as I did in this research. Their results were also in favor of the statistical approach. Even though the flexibility of the artificial approach makes it very attractive, this flexibility can also lead to overfitting. Overfitting is one of the main issues with artificial models. The model creates relations that does not exist because of the noise in the data.

There are also numerous variations in how a neural network can be constructed. The design procedure includes variable selection, data collection, data preprocessing, training, testing and validation, network adjustments, evaluation, training and implementation. The architecture of the network defines the number of neurons, layers, and the type of interconnections. Even though the flexibility of the artificial models provides us with a powerful forecasting tool, as Kaastra and Boyd (1996) also emphasize the design of the model can be complicated and challenging. In their study, they also address the importance of experimenting and testing the neural network applications as they are relatively new in financial economics.

# 5. Conclusion

The objective of this study is to make a comparative analysis of two forecasting approaches. The main question is if the artificial neural network model is superior in forecasting than the time series analysis. I develop two models and document that the statistical approach shows better performance in forecast accuracy when MAPE, MSE, and MAD are compared. However, these results are not statistically significant. The evidence further indicates that even though in my study the statistical model performed better, there is no empirical reason to say that these results can be generalized.

To claim that in general statistical approach is better than the artificial approach is not justified nor on the basis of the research results or the basis of the literature review.
On one hand, there are many studies concluding with the superiority of the artificial approach, while on the other hand there are also many studies concluding with the contrary. Even though Zhang et al. (2004) have done a very similar study to the one I conducted, their results are totally different.

In their study, Hill et al. (1994) and also Chatfield (1993) addressed the possibility of the artificial model being context-sensitive. Based on previous research and the results I conducted, there seem to be boundaries for when the artificial models are superior to the statistical models. The complexity of these models, however, make these boundaries difficult to identify. There is definitely not much of research focusing on this topic. Therefore, it will be interesting to see future studies work on identifying these limitations.

Finally, I also examine the relationship between inventory and EPS trend and find that whenever these trends do not move in the same way, there seem to be challenging for the artificial model to make good predictions. There is empirical evidence from previous research showing that inventory is one of those fundamental variables increasing prediction accuracy when included in models. Yet, most of these conclusions come from studies where statistical models were used. It will be very interesting to analyze this issue further with more companies and data and to see if there is a reasonable explanation or this is just a coincidence. I leave this issue to be addressed by further research.

# References


Abarbanell S., J., & Bushee J., B. (1997). Fundamental Analysis, Future Earnings and Stock Price. *Journal of Accounting Research*, pp. 1-24.

Bar, M. (2007, June 4). The proactive brain: using analogies and associations to generate predictions. *TRENDS in Cognitive Science*, pp. 280-289.

Basheer, A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological methods*, pp. 3-31.

Brown, L., & Rozeff S., M. (1979). Univariate Time-Series Models of Quarterly Accounting Earnings per share: A Proposed Model. *Journal of Accounting Research*, pp. 179-189.

Callen L., J., Kwan C.Y., C., Yip C.Y., P., & Yuan, Y. (1996, 12). Neural network forecasting of quarterly accounting earnings. *International Journal of Forecasting*, pp. 475-482.

Chatfield, C. (1993). Neural networks: Forecasting breakthrough or passing fad? *International Journal of Forecasting*, pp. 1-3.

Chatfield, C. (2000). *Time-series forecasting.* London: Chapman & Hall/CRC.

Conroy, R., & Harris, R. (1987, June 6). Consensus Forecasts of Corporate Earnings: Analysts' forecasts and time series methods. *Managements Science*, pp. 725-738.

Fadlalla, A., & Lin, C.-h. (2001, July-August 4). An Analysis of the Application of Neural networks in Finance. *INTERFACES*, pp. 112-122.

Falas, T., Charitou, A., & Charalambous, C. (1994). The Application of Artificial Neural Networks in the Prediction of Earnings. pp. 3629-3633.

Foster, W., Collopy, F., & Ungar, L. (1992). Neural Network Forecasting Of Short Noisy Time Series. *Computers chem. Engng*, pp. 293-297.

Gorr, L. W. (1994). Research prospective on neural network forecasting. *International Journal of Forecasting*, pp. 1-4.

Griffin A., P. (1977). The Time-Series Behavior of Quarterly Earnings: Preliminary Evidence. *Journal of Accounting Research*, pp. 71-83.

Hancock, J. R., & Seng, D. (2012, February 1). Fundamental Analysis and the Prediction og Earnings. *International Journal of Business and Management*, pp. 32-46.

Hill, T., Marques, L., O'Connor, M., & Remus, W. (1994). Artificial neural network models for forecasting and decision making. *International Journal of Forecasting*, pp. 5-15.

Hill, T., O'Conor, M., & Remus, W. (1996, July). Neural Network Models for Time Series Forecasts. *Management Science*, pp. 1082-1092.

Hopwood, S. W., Mckeown, C. J., & Newbold, P. (1982). The Additional Information Content of Quarterly Earnings Reports: Intertemporal Disaggregation. *Journal of Accounting Research*, pp. 343-349.

Huang, W., Lai Keung, K., Nakamori, Y., Wang, S., & Lean, Y. (2007). Neural Networks in Finance and Economics Forecasting. *International Journal of information Technology & Decision Making*, pp. 113-140.

Hyndman J, R., & Athanasopoulos, G. (2018). *Forecasting Principles and Practice, 2nd edition.* April: OTexts: Melbourne, Australia.

Hyndman, R. (2006, June). Another look at forecast-accuracy metrics for intermittent demand. *FORESIGHT*, pp. 43-46.

Jarett E., J. (1990). Evaluating Methods for Forecasting Earnings Per Share. *Managerial Finance*, pp. 30-35.

Kaastra, I., & Boyd, M. (1996, March 23). Designing neural network for forecasting financial and economic time series. *Neurocomputing*, pp. 215-236.

Kumar, R., & Pradhan P., R. (2010, November). Forecasting Exchange Rate in India: An Application of Artificial Neural Network Model. *Journal of Mathematics Research*, pp. 111-117.

Negnevitsky, M. (2011). *Artificial Intelligence: A Guide to Intelligent Systems.* Edinburgh Gate Harlow Essex CM20 2JE England: Pearson Educationn Limited.

Neuroscience news. (2018, March 2). *Neurosciencenews.com*. Retrieved February 27, 2020, from https://neurosciencenews.com/brain-predictions-8563/

P. Menezes Jr, M. J., & Barreto A., G. (2008). Long-term time series prediction with NARX network: an empirical evaluation. *Neurocomputing*, pp. 3335-3343.

Singh Vaisla, K., & Kumar Bhatt, A. (2010, 06). An analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting. *International Journal on Computer Science and Engineering*, pp. 2104-2109.

Stock, H. J., & Watson, W. M. (2015). *Introduction to Econometrics.* Edingburgh, England: Pearson Education Limited.

The MathWorks Inc. (2016). Introducing Machine Learning. *Introducing Machine Learning*. The MathWorks Inc.

Welch, B. (1938, February). The Significance of the Difference Between Two Means when the Population Variances are Unequel. *Biometrika*, pp. 350-362.

*www.stockrow.com*. (n.d.).

Zhang, P. G. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, pp. 159-175.

Zhang, W., Cao, Q., & Schneiderjans J., M. (2004). Neural Network Earnings per Share Forecasting Models: A Comparative Analysis of Alternative Methods. *Decision Science*, pp. 205-237.

# Appendix

## Table 1: Forecasting results of the artificial model

| ***Neural network*** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Company** | **Q1** | **predicted** | **error** | **Q2** | **predicted** | **error** | **Q3** | **predicted** | **error** |
| **Apple** | 2,20 | 1,41 | 0,79 | 3,08 | 2,04 | 1,04 | 5,04 | 2,70 | 2,34 |
| **Amazon** | 5,32 | 0,98 | 4,34 | 4,31 | 5,67 | -1,36 | 6,57 | 12,59 | -6,02 |
| **Exxon** | 0,79 | 0,48 | 0,31 | 0,76 | 0,42 | 0,34 | 1,34 | -0,80 | 2,14 |
| **Chevron** | 2,27 | 1,57 | 0,70 | 1,37 | 1,45 | -0,08 | 3,53 | -0,02 | 3,55 |
| **General Electrics** | 0,01 | 0,12 | -0,10 | -1,07 | 0,20 | -1,28 | 0,08 | -0,12 | 0,20 |
| **IBM** | 2,82 | 0,97 | 1,85 | 1,89 | 1,06 | 0,83 | 4,14 | -0,84 | 4,98 |
| **Microsoft** | 1,72 | 0,59 | 1,13 | 1,40 | 0,97 | 0,43 | 1,53 | 0,54 | 0,99 |
| **Nike** | 0,63 | 0,22 | 0,41 | 0,87 | 0,97 | -0,10 | 0,71 | 0,06 | 0,65 |
| **Verizone** | 0,98 | 1,09 | -0,11 | 1,29 | 1,70 | -0,41 | 1,26 | 0,92 | 0,34 |
| **Walmart** | 1,29 | 0,63 | 0,66 | 1,17 | 0,28 | 0,89 | 1,51 | 1,29 | 0,22 |

## Table 2: Forecasting results of the time series model

| ***Time Series*** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Company** | **Q1** | **predicted** | **error** | **Q2** | **predicted** | **error** | **Q3** | **predicted** | **error** |
| Apple | 2,20 | 2,46 | -0,26 | 3,08 | 2,59 | 0,49 | 5,04 | 4,46 | 0,58 |
| Amazon | 5,32 | 5,76 | -0,44 | 4,31 | 6,09 | -1,78 | 6,57 | 4,47 | 2,10 |
| Exxon | 0,79 | 0,76 | 0,03 | 0,76 | 0,78 | -0,02 | 1,34 | 0,74 | 0,60 |
| Chevron | 2,27 | 3,62 | -1,35 | 1,37 | 0,48 | 0,89 | 3,53 | 1,57 | 1,96 |
| General Electrics | 0,01 | -0,44 | 0,45 | -1,07 | -0,30 | -0,77 | 0,08 | -2,56 | 2,64 |
| IBM | 2,82 | 2,67 | 0,15 | 1,89 | 2,39 | -0,50 | 4,14 | 3,04 | 1,10 |
| Microsoft | 1,72 | 0,59 | 1,13 | 1,40 | 0,89 | 0,51 | 1,53 | 0,73 | 0,80 |
| Nike | 0,63 | 0,67 | -0,04 | 0,87 | 0,70 | 0,17 | 0,71 | 0,50 | 0,21 |
| Verizone | 0,98 | 1,48 | -0,50 | 1,29 | 1,50 | -0,21 | 1,26 | 1,15 | 0,11 |
| Walmart | 1,29 | 0,69 | 0,60 | 1,17 | 0,76 | 0,41 | 1,51 | 1,03 | 0,48 |