Norwegian School of Economics

Bergen, Fall, 2020

Something Old, Something New

A Hybrid Approach with ARIMA and LSTM to Increase Portfolio Stability

Kristian Senneset and Mats Gultvedt

Supervisor: Ivan Belik

Master thesis, Economics and Business Administration,

Business Analytics

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

In this thesis we seek to examine how modern forecasting approaches can improve estimations of stock pair correlations, and derived from this, contribute to making portfolios more stable. Volatility of financial markets have experienced increases due to the ongoing global pandemic. This amplifies the issues that investors face when assessing the risk related to their investments. We construct a hybrid model consisting of an ARIMA component to explain the linear tendencies of correlation, and a Long Short-Term Memory component to explain the non-linear tendencies. Our approach is populated by data from constituents of Oslo Stock Exchange ranging a time span from 2006 through the third quarter of 2020. Our results indicate that modern approaches to forecasting accrue stronger predictive performances than the conventional methods. Across all test periods our proposed hybrid model achieves an RMSE of 0.186 compared to an average benchmark RMSE of 0.237. However, the implications of these findings are ambiguous as the increase in predictive performance cannot be said to definitively outweigh the increase in cost of implementation. Our thesis contributes to the existing literature by exhibiting the untapped potential of how modern approaches to forecasting can improve accuracy of quantitative inputs for decision making.

Contents

List of Figures

List of Tables

1. Introduction

Refining the accuracy of inputs that are used as decision basis is a continuous issue across all business industries. The conventional theories base their approaches to estimations and calculations of inputs on simplistic statistical methods. In line with technological developments and availability of data, modern frameworks for forecasting has been established. Many researchers have found such modern forecasting approaches to outshine the conventional methods when applied on a variety of data sets.

However, modern approaches to forecasting have not been widely adopted for the issue of estimating inputs regarding investment risk. Research has been heavily focused on forecasting prices and returns on investment objects, while the equally important decision factor, risk, has not been covered to the same degree. The purpose of this thesis is to investigate quantitative methods for approaching risk in investment objects. The thesis relies on well-established concepts of portfolio theory, as well as modern approaches to making estimations for use in financial applications. It should be noted however that this thesis is not predominantly a thesis on the research field of finance. It is rather an exploration of how data analysis can support business decisions, here applied on a decision problem from the field of finance.

Examining this research area is of importance because dealing with levels of risk subject to dynamic conditions is something that most decision makers must deal with incessantly. To investigate methods for approaching risk in a meaningful manner, we must first delimit the topic to an appropriate scope. In the following section we will provide the thematic boundaries and an outline of the contents of this thesis.

1.1 Problem Definition

Risk is omnipresent in the world of business, but to provide a meaningful contribution to the literature we must delimit the topic sufficiently. An element of risk that is quantifiable and abundantly recorded is the price movements, and thereby derived risk, of financial instruments. A possible approach to improve risk assessments could be investigating how advanced methods of making estimations can contribute to more robust and stable investment portfolios. Furthermore, an interesting aspect of risk assessment using advanced methods, is reviewing their ability to contribute over a time span that is affected by unlikely, but highly impactful circumstances, also known as *Black Swan Events*. This specific element of the narrative is motivated by the ongoing global pandemic, Covid-19, which is forcing decision makers to prioritize risk assessments. For these reasons, the objective of this thesis can be delimited to the following research question:

How can modern approaches to forecasting contribute to more stable portfolios?

The research question is substantiated by two central elements of analysis: A comparative design with assessment of predictive performance across methods, materialized through our set of benchmark models (1), and a critical assessment of the method contribution's sensitivity to financial black swans (2).

Based on the background information hereunder, we lay the foundation for examining how estimations that investors rely on, can be improved. The succeeding literature review provides an overview of how risk has been estimated historically by practitioners, as well as emerging methods that can be utilized in this regard. The remainder of the thesis is structured as follows. Firstly, we define a proposed model inspired by existing literature and present an experimental approach to demonstrate how modern techniques, such as machine learning, can improve financial estimations. This experiment must be regarded only as a display of one possible application of modern forecasting approaches, meant to pose as a basis of analytical discussion. Secondly, we describe the data selection and the preprocessing required for it to populate our suggested methods. Thereafter, an explanation of how we decide to evaluate our model is included. The results from the model are then presented and evaluated before we ultimately discuss our findings with respect to our research question and related limitations.

2. Background

The year 2020 has involved substantially increased levels of uncertainty worldwide. As the spread of Covid-19 continues, national measures such as social distancing and quarantining go hand in hand with fears of contagion and increasing layoffs. The International Monetary Fund has developed a measure for tracking uncertainty related to social, political and economic circumstances across the globe, constructed by performing textual analysis on reports for each country (World Uncertainty Index, 2020). This measure, called the World Uncertainty Index, has in 2020 reached heights that are unprecedented for as long as uncertainty has been tracked by the IMF.

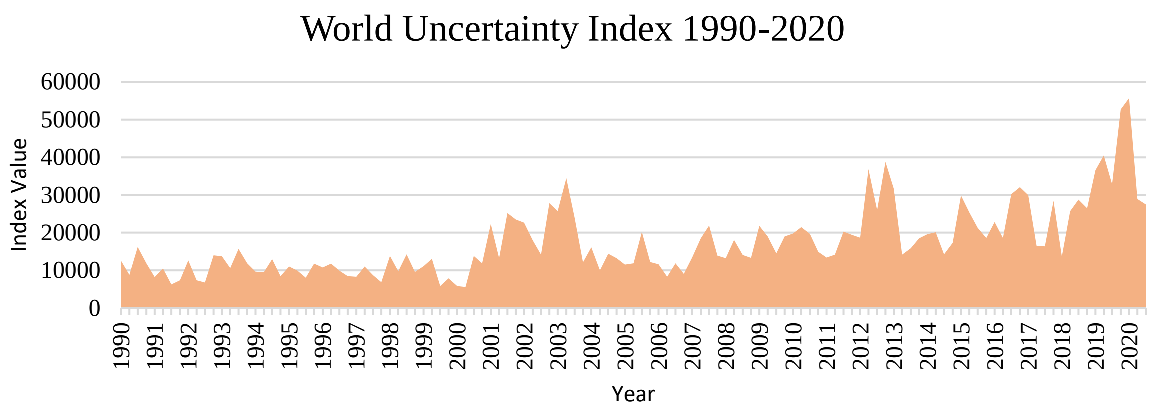

Figure 2.1 World Uncertainty Index. Data: (World Uncertainty Index, 2020)

New heights of global uncertainty naturally have impacts on the global financial markets. As Hites Ahir, the senior officer responsible for the World Uncertainty Index, described it in a recent index update; increasing levels of uncertainty historically coincides with periods of low economic growth and tighter financial conditions (Ahir, Bloom, & Furceri, 2020). In June 2020, The World Bank published a report with the title *Global Economic Prospects*. They claim that the global pandemic has enkindled the deepest global recession in decades and include baseline forecasts which projects a 5.2% contraction in global GDP during 2020 (World Bank, 2020). This global increase in uncertainty has provided motivation for our research question, as the implications derived from deviations in data driven decisions, will bear substantial impact.

2.1 Financial Black Swans

Dealing with uncertainties is an everlasting challenge for all participants of the global market. Covid-19 is not the first pandemic that exceeds expectations and leads to unforeseen financial impacts, and it will surely not be the last. The idea that improbable events collectively are very likely to occur, has among others, been discussed by mathematician David J. Hand who has written a book on the subject called *The Improbability Principle*. In essence, he argues that improbable events in reality occur quite regularly (Hand, 2014). Complementing Hands literature, Nassim Nicholas Taleb coined the term Black Swan Events in 2001 when he published his book *Fooled By Randomness,* and further in *The Black Swan: The Impact of the Highly Improbable* released in 2007. The latter book discusses the extreme impacts of rare and unpredictable events (Taleb, 2007). It has been an area of discussion whether the current pandemic can be defined as a black swan event or not, and the author himself has weighed in arguing that it should not be (Avishai, 2020). However, the virus and its impacts fit the broader definition of an unlikely event with extreme consequences, and the key takeaway from Taleb's contribution still stands regardless of the validity of definition. That is, humans should not seek to explain unlikely events by simplistic explanations in hindsight. Rather than attempting to predict unlikely events, one should build robustness for their adverse effects.

This thesis will base on the assumption that Covid-19 and its impacts on financial markets are representing a financial black swan. The event is virtually impossible to predict and has tremendous effects on the returns and risk related to financial investments. This background information helps address our research question appropriately. Based on Taleb's literature we seek to analyze forecasting contributions with consideration to financial black swans. According to his perspective we will assess the contributions from modern forecasting techniques with consideration to their robustness to a financial black swan. This leads us to investigate what constitutes risk in the financial markets and how it can be mitigated. In the next segment we will therefore explain how risk is quantified in decision making tools.

2.2 Investment portfolios

Investors are always looking for ways to obtain returns while mitigating the risk they are taking. Therefore, in traditional portfolio theory, the performance of the investments is usually considered a combination of the main components; expected returns and the risks related to these investments. One of the most prominent influencers of portfolio theory is Harry Markowitz, who defined the *Modern Portfolio Theory*. His dissertation on portfolio selection is still highly relevant to this day, even though it was published as early as 1952. His theory was based on the idea that every investor seeks to maximize their returns for any given level of risk (Markowitz, 1952). Some investors are risk averse, while some seek the thrill of higher-risk investments. Regardless of the risk aversion level, the investor is interested in finding the portfolio within their risk desirability, likely to yield the highest returns. This can also be considered such that investors prefer portfolios with less risk for any level of return. The set of optimal portfolios for any desired level of risk, or alternatively level of return, is called the efficient frontier (Markowitz, 1952). Furthermore, the theory is based on the concept that the risk level of a portfolio can be reduced by diversifying through unrelated securities. Therefore, the overall risk related to a portfolio can be calculated as a function of the variances of portfolio assets, along with the correlation between each pair of assets. Alas, the correlation between investment objects can be considered a proxy for the risk involved with investments.

There have been countless attempts at trying to predict future stock prices employing any thinkable method available. Being able to predict the expected stock prices accurately would mean that one of the two components practitioners assess when constructing portfolios are known entities. However, the same can be said about the risk component derived from correlations. Better predictions of future correlation, which employ modern methods, could potentially lead to better foundations for building effective investment portfolios. Going back to Nassim Taleb's petition to build robustness for unlikely and extreme events, this could be addressed by improving estimations of the future correlation between stock pairs. Our research question relates to how portfolios can become more stable through applying forecasting methods for constructing inputs. In this sense the stability of a portfolio relates to the actual variance on returns achieved from portfolio compositions. Regardless of the risk preference of an investor, more accurate inputs will aid in attaining the desired risk profiles of investments.

In the next chapter we will therefore seek guidance from the literature as to how such improvements can be made with the support of modern techniques. We will present an overview of how risk has been quantified historically followed by literature on the broader field of estimating future values of financial time series. The former provides reference, or a starting point for analysis, while the latter provides inspiration regarding favorable methodology for estimating values that can be used by decision makers to construct portfolio strategies.

3. Literature Review

In this section, we begin with describing how practitioners historically have approached the problem of quantifying risk. These traditional methods are often based on naïve projections or simply assumptions of constant correlation. In earlier years of the Modern Portfolio Theory, it was subject to criticism because of the assumptions it relied on for measuring risk through correlation coefficients (Low, Faff, & Aas, 2016). The simplest method used by practitioners of Modern Portfolio Theory, the Full Historical Model, assumed that correlation for any combination of assets in the investment horizon would be equal to the preceding observation. This is equivalent to producing naïve forecasts and are optimal when data follow a random walk, which is the case for many financial time series (Hyndman & Athanasopoulos, 2018). However, random walk forecasts were not deemed accurate enough, which culminated in an alternative approach for estimating future correlation in portfolios. This model, called the Constant Correlation model, was built on the assumption that any deviation from the market mean correlation coefficient was due to random fluctuations (Elton, Gruber, & Urich, 1978). Hence, correlation coefficients were according to this method estimated by projecting the mean correlation coefficient of all constituent pairs for the investment horizon. A third approach attempting to find better estimations of correlation coefficients also culminated, called the Single-Index Model (Elton, Gruber, & Urich, 1978). The Single-Index Model employs the market return to partly explain a pair of financial instruments' price movement in relation to each other. However, none of these statistical methods for projecting correlation coefficients have been satisfactory when it comes to estimation performance.

The aforementioned models employed in the Modern Portfolio Theory assumed that correlation coefficients are constant and fixed. Reflection of correlation is vital as it provides stability in portfolios through encouraging diversification. However, findings discussed by Preis et al. (2012) show that the average correlation among stocks scales linearly with market stress. Thus, naïve estimations on correlation coefficients are subject to large errors as uncertainty changes. The diversification effect responsible for protecting portfolios is diminished in times of market losses which, inconveniently, is when it is needed the most. Chesnay & Jondeau (2001) also provides an empirical study which points out that periods with high levels of financial turbulence and uncertainty, tend to generate positive correlations between stock prices, as contractions in the economy affect most companies. These studies imply that correlation coefficients are likely to deviate from historical quantities, which provides further support towards the criticism of assuming fixed correlation coefficients. Following this, diversification derived from analysis of correlation coefficients is useless if it only works when market conditions remain unchanged. Alas, diversification as a stability measure need to account for changes in correlation of price movements and cannot rely on assumptions of fixed entities. Markowitz himself also addressed this criticism stating that his assumed task was to develop a framework for outputting efficient risk-return combinations, given inputs such as means and variances of individual securities and the correlation between them (Markowitz, 2002). He further assumed that it was not his task to provide these inputs and ensure their accuracy, but rather the task of security analysts. The field of forecasting has evolved tremendously since the time of Markowitz and we are therefore interested in investigating modern approaches to forecasting applicable to this problem, such as automated forecasting frameworks, machine learning, neural networks and the combination of such methods.

The remaining research presented revolves around forecasting financial time series, and some highly favored frameworks for this research field. The literature review is an essential segment of the thesis process, as there is a multitude of available methods in the field of financial time series analysis. All these methods come with their own benefits and detriments. The following sections seek to review literature on time series forecasting with long-established methods such as ARIMA, more modern methods in deep learning techniques such as neural networks, and lastly, several hybrid models employing a combination of methods.

AutoRegressive Integrated Moving Average, or ARIMA, is a forecasting framework developed by Box and Jenkins (1970), and is one of the most widely utilized methods of forecasting economic and financial time series (Hyndman & Athanasopoulos, 2018). Studies have been conducted on financial time series such as electricity prices, housing prices, and stock prices. Weiss (2000) employed the ARIMA framework to construct models that predicted electricity prices of mainland Spain with good results. The ARIMA model designed predicted prices with an average error of about 10%, both with explanatory variables and without. Raymond (1997) used an ARIMA model to identify trends in Hong Kong's real estate prices and concluded that ARIMA models are particularly good frameworks for forecasting on the short-term due to slow changes in the short-term factors. The autoregressive component was helpful in determining the trending effects of the housing prices while the moving average components contributed with determining turning points. These two components, which in addition to some level of data differencing, make up the ARIMA framework, were successful in tracking the direction of changes in the real-estate prices. Similarly, Adebiyi, Adewumi and Ayo (2014) found that ARIMA models have a strong potential for predicting for the short-term. They built an ARIMA model for stock price prediction on two constituents, Nokia and Zenith Bank. The model predictions were satisfactory, and they concluded that ARIMA models can compete reasonably well with emerging forecasting techniques such as artificial neural networks in short-term prediction.

Among machine learning applications in the field of stock market predictions, Galler, Kryzanowski and Wright performed a pioneering study in 1993. They developed a classifier model using deep learning and proceeded to correctly classify 72 % of directional movements on one-year-ahead stock returns (Kryzanowski, Galler, & Wright, 1993). In addition to being able to classify directional movements, Olson and Mossman (2003) showcased the potential for machine learning to be used in regression models. They forecasted one-year-ahead point predictions on the Canadian Securities Exchange. Both studies could report that their deep learning model could outperform the existing regression models using traditional techniques. Among the newest and most popular techniques within machine learning for forecasting time series is the application of neural networks. In particular, Long Short-Term Memory networks, or LSTM networks, have been employed diligently in recent times.

Literature on utilizing LSTM in predictive modeling of financial markets is historically scarce, despite being suitable for financial time series predictions. There are several reasons why such literature might be lacking, which can be broken down into two main reasons. Firstly, challenges related to backtesting financial strategies deteriorates the value of findings. Alas, struggles with backtesting mean that separating what are successful trading strategies for only a specific place in time, and those applicable for the future, is severely challenging (Lopez de Prado, 2018). Secondly, there are predominant incentives for keeping significant findings unpublished as that will more likely lead to financial benefits. However, due to the growth in computational efficiency and the availability and popularization of machine learning in the last few years, the activity in this field has increased. Huck, Anh and Krauss published a paper in 2017 where they compared different machine learning techniques for stock price prediction. Interestingly, they did not outperform traditional techniques but performed well in periods with high volatility and market decline, such as the dot-com bubble in the late 90s and the 2008 financial crisis (Krauss, Anh, & Huck, 2017).

Usually, LSTM networks are employed when working with vast amounts of data, but there are examples of successful application on training with fewer data points in the literature. Siami-Namini, Tavakoli and Namin (2018) built an LSTM network to predict time series of financial data and managed to obtain forecasts with average errors of between 13 and 16 %. In the same year, Fischer and Krauss (2018) built LSTM networks to model S&P 500 constituents' directional movements. They found LSTM networks to outperform other alternatives within machine learning that are considered *memory-free*, such as Random Forest and a logistic regression classifier. In the next segment we will complement the literature review with some studies that delve into combining the methods mentioned above, so-called *hybrid models*.

Hybrid models have the fundamental advantage that it combines two or more individual models, which means the models have the potential of complementing each other. This leads to being able to exploit the advantages of each model’s characteristics. In 2003 Peter G. Zhang published a study on the combination of the ARIMA model and a neural network. He proposed that since ARIMA models and neural networks often were subject to comparisons of predictive strength for time series, with varying conclusions, it should be investigated whether a hybrid model taking advantage of both models' strengths was beneficial. In the study, he investigated different time series, including sunspot data, Canadian lynx data and exchange rates. He displayed that neither ARIMA, nor neural networks individually, were suitable for a wide range of time series. Most time series include both linear and non-linear relationships between observations, and a hybrid model consisting of methods favorable for each type of relationship is advised according to his findings (Zhang, 2003). This pioneering study, establishing a framework for a hybrid between ARIMA and neural networks, has inspired several studies in recent times.

A study conducted by Temür, Temür and Akgün (2019) employed a hybrid model made up of an ARIMA component and an LSTM network to forecast housing prices in Turkey. They found results that corresponded with Zhang's (2003) literature. The best accuracy was achieved with the mentioned hybrid model, and the difference in predictive power between the hybrid and the individual models was significant. Furthermore, Zhang’s (2003) literature has also provided methodological inspiration for a study by Choi (2018) where the effectiveness of an ARIMA and LSTM network hybrid model on predicting S&P 500 constituents correlation coefficients were investigated. Choi found that the hybrid model produced forecasts on correlation coefficients for stock pairs, which improved significantly upon traditional correlation projection methods. During the work on this thesis, we have let us inspire by these methodological frameworks and wish to build a similar hybrid model for Oslo Stock Exchange constituents to demonstrate the potential usefulness of neural networks for financial time series forecasting.

Without having touched upon the specific approach of this thesis, it should still be pointed out how this thesis contribute to the literature. As far as we know there is no existing literature on making predictions of correlation coefficients employing the methods included in the literature reviewed for Oslo Stock Exchange constituents. We will come back to the specifics of selected approach and data in later chapters. Furthermore, the time span investigated in this thesis involves both the financial crisis of 2008 and the Covid-19 pandemic of 2020. We find no existing literature discussing the impact of black swans on estimates of correlation coefficients. The literature review contributes to explaining why our research question should be addressed by presenting a problem that traditionally has been addressed by simple statistical methods, despite the emergence of methods for forecasting that is applicable to the problem. All this considered, this thesis should complement the existing literature in a meaningful way.

Substantiated by background information and the literature review above, we will in the next chapter propose our approach to explaining how modern forecasting techniques can aid decision makers in constructing stable portfolios. The approach chapter consists of our preferred method of addressing the research question but is naturally only one way of doing just that. We will however emphasize the reasons for our selection of approach.

4. Approach

In this chapter we will introduce our proposed model in the first section. The second section consists of the benchmark models we include in our approach which addresses the research question by providing a comparative design of analysis. The third section introduces an additional evaluation approach based on a portfolio sampling. Ultimately, the last section of this chapter describes the data which will populate our proposed model and benchmark models.

4.1 Hybrid Model

Inspired by the literature reviewed, we present a hybrid method, using an ARIMA model combined with an LSTM model to predict the correlation coefficients between each pair of stock. The method rests on the assumption that the time series data is composed of both linear and non-linear tendencies (Zhang P. , 2003), expressed in the following equation.

$$x_t = L_t + N_t + \epsilon_t \tag{4-1}$$

Where the notation L_t represent the linearity in the data x_t at time step t, N_t represent the non-linearity and ϵ_t is the error term. As discovered through the literature review, hybrid models have emerged in recent years as a method of improving forecasts from individual models through combination. We are encouraged by the literature on this research and aspire to answer our research question with the help of these techniques. Dependent on the predictive performance derived from such methodology, this can aid decision makers by exhibiting the potential contribution of forecasting techniques in supplying inputs to frameworks for strategizing portfolios. There are a multitude of methods that are applicable for forecasting both the linear and non-linear component, and there are benefits and detriments to every method. In the following segments we will provide a rationale for the elected hybrid components, ARIMA and LSTM, an explanation of how they are implemented, and a description of the data selection process.

4.1.1 Hybrid section I - ARIMA

ARIMA models have been a popular method of choice for researchers attempting to predict future values of financial time series (Hyndman & Athanasopoulos, 2018). Studies have shown that ARIMA models excel in forecasting several different types of econometric time series and is often able to outperform more complex and extensive methods (Levenbach, 2017). As discovered in the literature review, ARIMA models have proven to be particularly good frameworks for forecasting the short-term linear tendencies of financial time series. The ARIMA model uses linear functions of past data to forecast future values and has been favored by researchers due to its simplicity in both comprehension and application (Fattah, Ezzine, Aman, Moussami, & Lachhab, 2018). In addition, financial time series are generally likely to inherit some seasonal effect, which ARIMA is well suited for handling (Hyndman & Athanasopoulos, 2018). The relative simplicity of ARIMA makes it enticing in a business sense as it eases implementation due to less requirements in preprocessing of data, computational efforts, and its wide applicability. In summation, ARIMA is an easy-to-implement framework that is applicable for forecasting financial timeseries at a low computational cost. Naturally, a wide range of methods could account for explaining the linear tendencies of financial time series data but based on the aforementioned reasons we will employ ARIMA.

The ARIMA framework combines *autoregressive processes* and *moving average processes,* aiming to describe the autocorrelations in the data (Box & Jenkins, 1970). The additional *integrate* component involves applying differencing on the time series to convert non-stationary time series into stationary (Box & Jenkins, 1970). In short, the ARIMA method involves a selection process to identify the number of lags to be used for the autoregressive and moving average parts that best fit the observed time series, as well as a level of differencing. The term autoregression refers to the procedure of regressing the variable against itself, using the previous p values. Similarly, moving average uses the past q forecast errors in a regression-like model (Hyndman & Athanasopoulos, 2018). Additionally, it is often necessary to apply a level of differencing d, to obtain a stationary time series. This process results in a ARIMA model of order (p, d, q). A detailed description of the ARIMA method can be found in Appendix A2.

Our ARIMA approach is based on a stepwise automatic model selection algorithm developed by Hyndman and Khandakar (Hyndman & Khandakar, 2008), and implemented using the function *auto.arima* from the R-package *forecast* (Hyndman R. , et al., 2020). In short, the algorithm applies different model orders and calculates the relative goodness of fit with the Akaikes Information Criteria (AIC). The algorithm returns the model with the lowest AIC. We do not wish to force any model order on the time series input, as we seek to keep this section of the hybrid model as automated as possible.

After fitting a model on all the correlation time series, the residuals from the ARIMA predictions are stored. As the ARIMA model predictions are assumed to have explained a substantial amount of the linear relationships in the data, the residuals are thought to contain the non-linear relationship and are used as input in the second section of the hybrid method.

4.1.2 Hybrid section II – Neural Network

Neural networks have surged in application the last decade and is recognized to handle and model a multitude of complex non-linear problems (Haykin, 2008). A neural network consists of nodes, organized in layers, that are connected with weights. In general, data is presented to the network in the input layer, passed through nodes in one or more hidden layers, before calculating an output in the output layer. Figure 4.1 displays these layers for a Feed-Forward Neural Network (FNN).

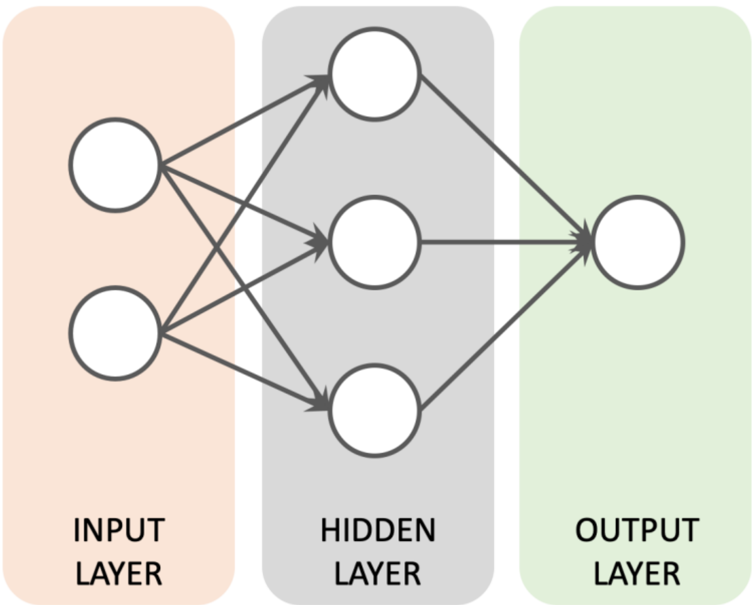

Figure 4.1 Feed-Forward Neural Network structure (Bouvet Norge, 2020).

As the name suggests, the information in a FNN is passed forward through the layers in a single direction. The arrows that connect the nodes each has a weight that regulate the information passed through each connection. The network aims to optimize these weight parameters w, as well as bias parameters b, in order to predict values $\hat{y}$ that minimize a loss function L. Thus, the predicted values are a function of the input x and the network parameters θ so that $\hat{y} = f(x, \theta)$. The loss function expresses the accuracy of the predictions $L(\hat{y}, y) = L(f(x, \theta), y)$. The network learns by updating the loss function iteratively with an optimization algorithm that adjust the parameters θ in a direction that reduces distance between the predicted values and the true values. This optimization process is called back-propagation and uses *gradient descent*, which is an iterative optimization for identifying a local minimum, to find the optimal values for the parameters (Lecun, Bottou, Orr, & Müller, 2012).

Recurrent Neural Network (RNN) is a subdivision of neural networks, which has a structural feature allows the network to contain information from sequential input across time steps (Dupond, 2019). The nodes in the hidden layers in the RNN is looped, allowing the sequential input to be interpreted iteratively. Information from the input is stored in each iteration as a *hidden state* and the hidden layers inherits these states from previous iterations. Thus, the hidden state can be described as the *working memory* of the network. A representation of this concept is displayed in figure 4.2.

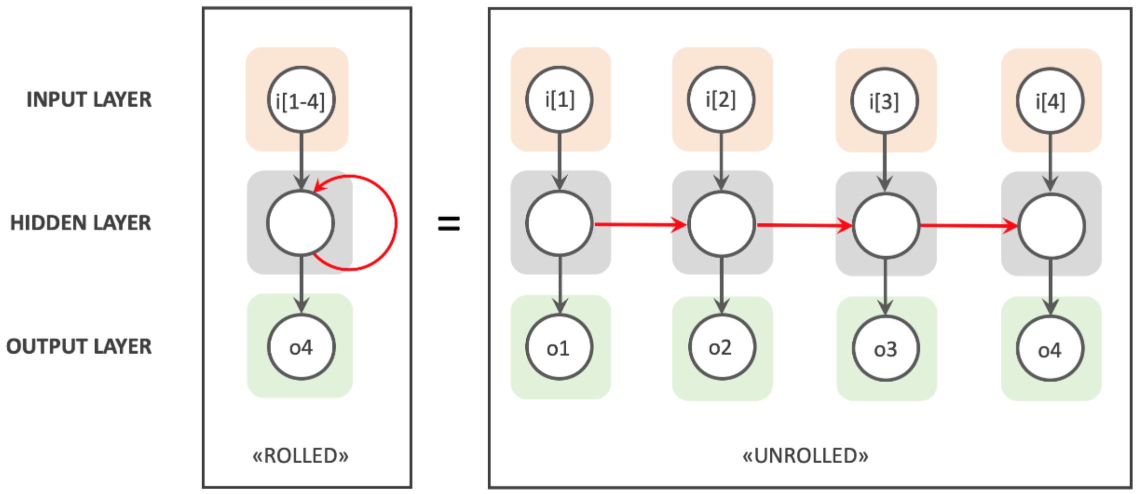

Figure 4.2 General structure of an RNN. An input sequence with four timesteps will create four identical copies of the network structure and the hidden state is passed onto the next time step. Source: (Bouvet Norge, 2020).

As depicted, RNN can be described as a chain of identical neural networks, one for each time step in the sequential input, looped together. When optimizing the loss function in an RNN, all time steps in the sequential input is passed through the loop before each update. One iteration of this procedure is called an *epoch*. As the neural networks in the unrolled RNN are identical, they also share the same adjustable weights and biases that the function looks to optimize.

The passing of the hidden states in an RNN, as shown by the red arrows in the figure, also comes with some limitations, as it often struggles to control the information over long sequences. The resulting effect of these hidden states on the network outputs either decays rapidly or explodes exponentially over time (Hochreiter, Bengio, Frasconi, & Schmidhuber, 2001), and a graphical representation is depicted in figure 4.3. This problem is often referred to as the *vanishing gradient problem* and introduces a problem when attempting to model dependencies in long sequences (Bengio, Simard, & Frasconi, 1994).

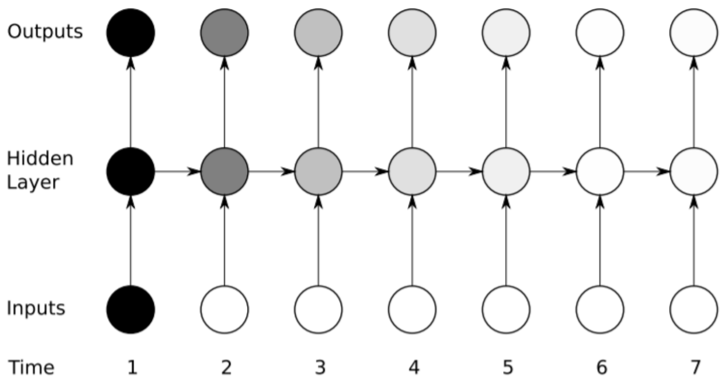

Figure 4.3 The vanishing gradient problem. Source*:* (Graves, 2012)

There have been several attempts to create a modified RNN architecture to deal with the aforementioned problem, and we have selected the Long Short-Term Memory (LSTM) approach in this thesis.

In addition to the working memory through the hidden states, the LSTM has a *cell state*, that serves the function of a long-term memory. This allows it to persist and contain information over longer time periods and sequences. The cell state is regulated by *gates* that control what information to remove from the previous time step and what information to add from the input in the current time step. The cell state and the gates are the mechanisms of the LSTM that tackles the vanishing gradient problem. A more detailed description of RNN and LSTM is found in Appendix A3 and A4.

For our modelling task, the LSTM has desirable features, as we want the model to have the ability to use information from sequences in an early time step for forecasting current time steps. In theory, this facilitates the possibility for the model to extract information and learn from previous data such as from the financial crisis of 2008 and apply this when forecasting periods with similar circumstances. However, complex LSTM models are computationally heavy and time consuming to train. Additionally, it can be challenging to design and tune a network to obtain a model that does not just fit the observed data well, but also learn the true relationship in the data and forecasts well out-of-sample. For this reason, we focus the construction of the LSTM model to a simple and generalized structure to reduce the time, computational power and the size of the dataset required to train and use such a model. This entails a probable decrease in performance accuracy but increases usability and allow decision makers and portfolio managers to refit the model on a variety of time series to support the forecasting task of their interest.

The input used for the LSTM model consisted of the residual values derived from the forecasts of the ARIMA model. The residual data is divided so that the last time step is treated as a target value Y and the model is trained on the remainder of the previous observations X. Furthermore, the LSTM requires the data to be three-dimensional, on the following form, $[Samples, Time\ Steps, Features]$.

With the selection of ARIMA and LSTM as the components of the hybrid model, we can present the following flowchart of the hybrid model:

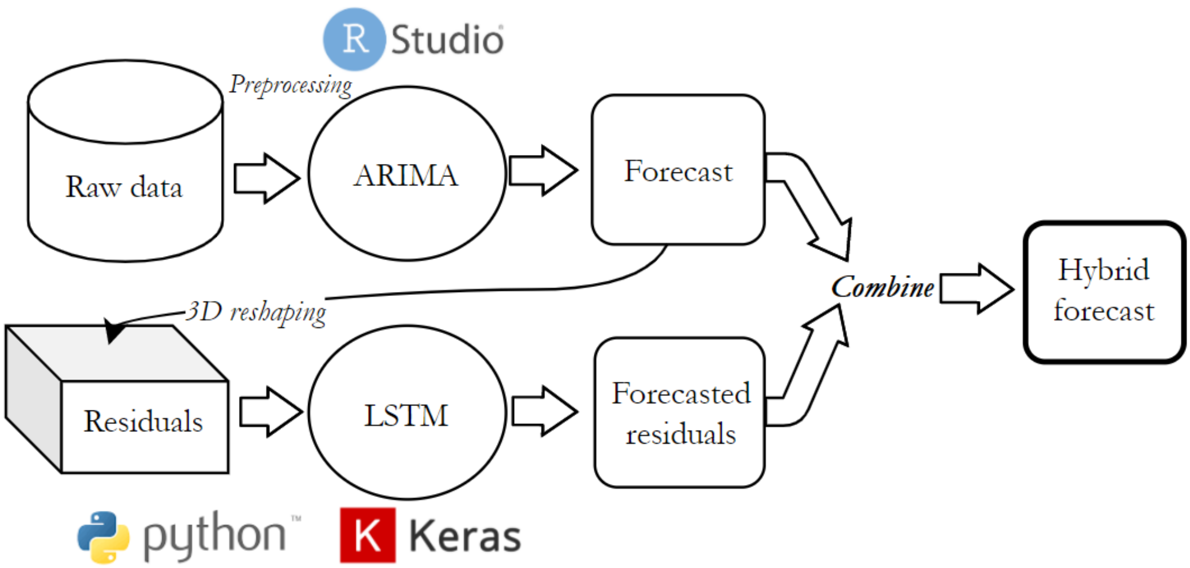

Figure 4.4 Illustration of the hybrid model. The residuals are contrived from the forecast of the ARIMA model, reshaped, and used as input in the LSTM model. The LSTM forecasts the residual which is combined with the ARIMA forecast to produce the final hybrid forecast.

For the LSTM we need to address some hyperparameters and design choices for model optimization. Furthermore, we have also performed some measures to reduce the problem of overfitting. We will in the following segments elaborate on these aspects.

LSTM Model architecture

Additionally, there are several hyperparameters and design choices to be selected when building the architecture for the LSTM model. There has been extensive research in exploring methods to optimize the selection. However, these methods entail a tedious and computationally demanding task (Hutter, Hoos, & Leyton-Brown, 2011). For simplicity and due to computational limitations, some of the model choices are selected and assumed to be fixed throughout the development of the final model, and some have been found through trial and error. A short description of how the model design and hyperparameters are selected will follow.

The complexity of the network can be controlled with the selection of number of hidden layers and number of nodes in each layer. As we want a simple structure, we only employ a single hidden layer, and limit the number of nodes in the hidden layer by searching between the interval [5, 20].

For the calculations in each cell, the *ADAM* optimizer function is used as it is regarded as a robust selection to the choice of the remaining hyperparameters (Goodfellow, Bengio, & Courville, 2016). In order to merge the output from all the cells into a single value, the output layer employs a doubled-hyperbolic tangent function. Multiplying the hyperbolic tangent function by two will ensure that the final predictions are transformed into the range $[-2,2]$, which encompasses the minimum and maximum value that the residuals of the correlations can take. To determine the learning rate, Greff et. al. suggests a procedure of starting with a high value (e.g. 1.0) and divide by 10 until performance stops increasing (Greff, Srivastava, Koutník, Steunebrink, & Schmidhuber, 2017). Through the design and selection of hyperparameters, a main weakness related to neural networks can be addressed, namely overfitting. In the next segment we will therefore describe how our approach is designed with respect to this problem area.

Overfitting

Neural networks have a tendency to fit a model too closely to the training data provided (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). This is known as overfitting and causes problems as it leads researchers to believe they have found a good model for their problem. However, as the models are used to produce real forecasts, they realize the predictive performances is not coherent with the assumed predictive strength. Alas, neural networks are often subject to developing models that correspond too closely with the specific dataset, and therefore fails to predict future observations reliably.

When building a generalized model, it is also a well-known practice to incorporate a validation set in the development of the model (Kohavi & Provost, 1998). This way, the data can be separated into train, validation and test data and use the validation set, hereby referred to as the development set, to prevent overfitting on the observations in the training set. The data split is further outlined in section 4.4.3. We will implement the development set in the model development through an *early stopping* process. When training the model, a performance measure for the development set is calculated and registered every epoch. Whenever the model has not improved the performance on the development set for 10 epochs, the training ends, and the weight and bias parameters from the epoch with the best performance is saved and used as the final model. Additionally, another common measure to reduce issues of overfitting is through regularization. Regularization is the act of making modifications to the learning algorithm which seeks to reduce the out-of-sample error, but not the in-sample error (Goodfellow, Bengio, & Courville, 2016). Out-of-sample error refers to the ability of predicting observations that is previously unseen to the algorithm, while in-sample error relates to predictions on the data which the algorithm is based upon.

One method of regularization is carried out through the inclusion of dropout layers. Dropout regularization is a way to debias the layer, by turning off any given node during training of the model with a probability p (Zhang, Lipton, Li, & Smola, 2020). This is contributing to reducing risk of nodes becoming interdependent which is a prevalent source of overfitting (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014). We investigate the effect of the dropout rate on the accuracy of the model on the train and development set incrementally. Additional regularization steps can be performed by conducting weight regularization, of which we separate between two main types. These are known as the Lasso regularization (L1) and the Ridge regularization (L2) (Martins, 2019). Weight regularization aims to penalize certain weights in the loss function, and their values are found by investigating the effect of different combinations of model hyperparameters on predictive performance. In summary, overfit has been addressed through employment of a development set and tuning of hyperparameters.

In parallel with reviewing literature and defining a proposed model, we have examined different models applicable for estimating correlations on the investment horizon. As previously mentioned, these models are not solely meant to provide inspiration for our proposed model, but also to provide reference for examining the performance. To ensure a comparative design in the analysis of the performance of our proposed model we have therefore include a range of models as benchmarks. To evaluate the performance of the hybrid model and the benchmarks, we have used the Root Mean Squared Error (RMSE) and the Mean Average Error (MAE). The justification and details of these evaluation metrics are presented in Appendix A5. In the next section we will briefly elaborate on our selection of benchmark models.

4.2 Benchmark models

The predictive performance of our hybrid model is compared to a total of seven benchmark models, whereas four are referred to as conventional approaches of projecting correlation coefficients for portfolio optimization based on historical coefficients. The remaining three benchmarks consists of the two methods in the hybrid model, evaluated individually, as well as an alternative hybrid model, which are referred to as forecasting methods.

4.2.1 Historical Model

The simplest method of projecting correlation coefficients for use in portfolio optimization presupposes that correlation for any pair of stock constituents will be persistent (Elton, Gruber, & Urich, 1978). Correlation coefficients used in the Historical Model will thus always be equal to the corresponding coefficient according to the most recent observation.

$$r_{ij}^{t} = r_{ij}^{t-1} \tag{4-2}$$

i, j: stock constituent index in the correlation matrix

4.2.2 Constant Correlation Model

The next method we use as benchmark employs the mean correlation coefficient for all stock constituents for projecting future correlations. The Constant Correlation model presupposes that any discrepancy from the mean are random deviations (Elton, Gruber, & Urich, 1978). Hence, the estimation of future correlations for each pair should be equal to the most recent observation of the average correlation.

$$r_{ij}^{t} = \frac{\sum r_{ij}^{t-1}}{n(n-1)/2} \tag{4-3}$$

i, j: stock constituent index in the correlation matrix

n: number of stock constituents

4.2.3 Single Index Model

The Single Index Model presupposes that the movement of the market return can be employed to make better estimates for future correlation coefficients (Elton, Gruber, & Urich, 1978). A key assumption in the Single Index Model is that stocks most often have positive covariance as they respond to the same macroeconomic factors. Nonetheless, companies are affected diversely by different economic factors. Following this reasoning the Single Index Model assumes that covariances of each stock pair are calculated by multiplying the respective betas and the market variance. The estimation of future correlation coefficients in the Single Index Model is expressed as

$$r_{ij}^{t} = \frac{\beta_i \beta_j \sigma_m^2}{\sigma_i \sigma_j} \tag{4-4}$$

i, j: stock constituent index in the correlation matrix
m: market index

4.2.4 Overall Mean

Elton, Gruber and Urich (1978) conducted a study comparing a wide range of statistical methods for estimating correlation coefficients including the models described above. Among all the statistical methods compared they found the Overall Mean to achieve the best predictive performance. The Overall Mean assumes that correlation coefficients for a given pair of investment objects are estimated as their mean relationship of price movements over time. The estimation of future correlation coefficients employing Overall Mean is expressed as,

$$r_{ij}^{t} = \frac{\sum_{t-1}^{1} r_{ij}^{t}}{n} \tag{4-5}$$

i, j: stock constituent index in the correlation matrix
n: number of observations for each pair

4.2.5 ARIMA

The ARIMA method is also included as a benchmark. The *auto.arima* models previously selected in the methodology section are used to create out-of-sample predictions for the development and test sets. This enables us to interpret to which degree the ARIMA by itself can explain the variation in the data, and thus provide insight about how each hybrid component is contributing to its performance. These predictions are compared to the actual values for the sake of calculating accuracy metrics.

4.2.6 LSTM

For the same reason as adding a stand-alone ARIMA model for predicting future correlations we also add one for the LSTM method. This time LSTM are given past correlations as input instead of residuals from ARIMA. Parameter tuning through trial and error quickly revealed to have little impact on the accuracy of the stand-alone LSTM. Hence, we resolved to keeping pre-defined model parameters identical to those identified for the hybrid model.

4.2.7 Hybrid: ARIMA-Random Forest

Neural networks have been a widely popular method in the realm of Machine Learning in the recent years. We wanted to make sure that the perceived usefulness of neural networks among researchers is not inflated. As an assurance, we elected to make predictions using an alternative machine learning method as a replacement for the LSTM within the same hybrid methodology. Similarly to the LSTM, a Random Forest (RF) model requires restructuring of the data. Each quarterly correlation coefficient is treated as the outcome variable and is supplied with lagged values of the time series as predictors.

Random Forest is a popular and effective machine learning algorithm which utilizes ensemble learning, an algorithm which combines multiple learning models to improve the overall performance. Random Forest constructs a multitude of decision trees which individually produces a prediction, either in the form of a class in classification problems or point predictions for regression problems (Breiman, 2001). For each tree, a random subset of the training data is drawn and used to calculate its output. The output of a Random Forest model is either the mode of the classes predicted in classification, or the mean prediction across the decision trees in a regression problem. One key advantage of using Random Forests models is that the generalization error converges to a limit as the number of trees in the forest increases.

In other words, in accordance with the Strong Law of Large Numbers, overfitting is seldom a problem for Random Forest models (Breiman, 2001).

As the Random Forest model solely constitute a component of one of our benchmark models, we limit the optimization of hyperparameters to initial trial and error. Furthermore, there are in practice only two user-specified hyperparameters: the number of trees in the forest and the number of variables in the random subset at each node. In general, the model is most often not overly sensitive to these parameters (Liaw & Wiener, 2001). Nevertheless, the hybrid model combining ARIMA and Random Forest is not meant to represent an optimized regression on time series employing Random Forest, but rather provide a reference point for assessing the predictive power of our proposed model.

In addition to the comparative analysis provided by the benchmark models described in this section, we also want to address the research question in a practitioner’s sense. Therefore, we will in the next section describe an additional method of evaluation which incorporates the portfolio variance of returns that can be derived from our results.

4.3 Portfolio Selection

The portfolio-based evaluation described hereunder constitutes an expansion on the already established evaluation approach. Our intention is to provide an insight into how estimations on correlation impacts the variance of returns for individual portfolios of constituents.

The portfolio selection will be based on a random sampling from the population of investment objects. The random sample will be performed 10 times, each including five investment objects. We can then compare the total portfolio variance derived from estimated correlation matrices, as well as the correlation matrices based on actual data. The number of samples is selected as a compromise between time consumption and the evaluation value attained. This will represent a display of how estimation errors impact the actual variance of returns on investors’ portfolios. This is useful because it portrays the quantitative results from the investors perspective.

As this evaluation method is time consuming, we have elected to compare the best performing conventional method and forecasting method from the comparative evaluation. The overall variance of a portfolio is a product of each investment objects individual variance as well as the covariance between all portfolio constituents. For simplicity we resort to equal weighting between portfolio constituents. Our methods will provide correlation matrices and we can employ these to calculate total portfolio variance through the following equation:

$$Portfolio\ variance = [w_1\sigma_1 \cdots w_n\sigma_n] \times \begin{bmatrix} 1 & r_{12} & \cdots & r_{1n} \\ r_{21} & 1 & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & \cdots & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} w_1\sigma_1 \\ \vdots \\ w_n\sigma_n \end{bmatrix} \quad (4\text{-}6)$$

Where the correlation matrix for each portfolio is multiplied with a vector of weighted standard deviations of asset returns and a transpose of the same vector.

The proposed methods, benchmarks, and portfolio evaluation described above need to be populated by data. Based on the research question it is clear that the data should consist of time series data on some sort of financial assets. This could include properties, commodity, stocks, currency, and a range of other tradeable assets. Because we are particularly interested in the relationship of price movements between market constituents, and a considerable number of them, we find it favorable to populate our methods with stock data. The following section describe the data gathering and preprocessing steps made to the data.

4.4 Data

In this section we will describe the data that we have selected for populating our methodological approach. As briefly mentioned, we have decided to employ stock data, which have an obvious advantage when it comes to availability. Furthermore, we have selected to focus on constituents of the Oslo Stock Exchange, as it will represent a set of financial investment objects that are not widely investigated in our field of research. This thesis relies on obtaining stock prices for the constituents of OSEBX. We want to focus our work on constituents of OSEBX as it consists of a representative sample of all listed shares on Oslo Stock Exchange (Oslo Børs, 2020). In addition, the list of shares on OSEBX are routinely revised to, among other things, ensure ample liquidity. Before the raw data we collected can populate our methodological approach, it requires some preprocessing, which will be described in the following segment.

4.4.1 Preprocessing

In this thesis we have decided to investigate the period between 2006 through the third quarter of 2020. This starting point provides a sizeable sequence of data, as well as it includes the financial crisis occurring in 2008. This time span corresponds with 3 700 trading days. Among the original list of OSEBX tickers there are 69 different tickers, however many of these have not been listed on Oslo Stock Exchange for the entire period. We want to ensure that our methods are populated by long series of data that span a multitude of market cycles. Therefore, the initial filtering of companies consists of only keeping stocks that have been registered on Oslo Stock Exchange for the entirety of the 3 700 days. This leaves us with a dataset of 38 companies and their adjusted closing prices, presented in table 4-1.

Table 4-1 Companies included in our dataset

Companies included

Company Name	Ticker	Company Name	Ticker
ABG Sundal Collier Holding	ASC	Medistim	MEDI
AF Gruppen	AFG	NEL	NEL
Aker	AKER	Nordic Semiconductor	NOD
Aker Solutions	AKSO	Norsk Hydro	NHY
American Shipping Company	AMSC	Norwegian Air Shuttle	NAS
Atea	ATEA	Olav Thon Eiendomsselskap	OLT
Axactor	AXA	Orkla	ORK
Bonheur	BON	PGS	PGS
DNB	DNB	Photocure	PHO
DNO	DNO	Schibsted ser. A	SCHA
Equinor	EQNR	SpareBank 1 SR-Bank	SRBANK
Frontline	FRO	Stolt-Nielsen	SNI
Gaming Innovation Group	GIG	Storebrand	STB
Golden Ocean Group	GOGL	Subsea 7	SUBC
Hexagon Composites	HEX	Telenor	TEL
Kitron	KIT	TGS-NOPEC Geophysical Company	TGS
Kongsberg Automotive	KOA	Tomra Systems	TOM
Kongsberg Gruppen	KOG	Veidekke	VEI
Lerøy Seafood Group	LSG	Yara International	YAR

Even though all companies have been registered on the stock exchange for the entire time span, there are still a few occurrences of NAs in the dataset. This is due to stocks not being traded on certain days which could indicate trading halts, or simply the stock being so illiquid that it has not been traded for a day. Since the models we will work with require complete data for all rows, we decide to impute these NAs by replacing them with the previous observed value. This ensures that we can calculate correlation coefficients for every stock pair and days in the dataset. Also, we register that table 4-1 includes the major companies from the Oslo Stock Exchange, and is diversified on a multitude of different industries, displayed in table A-

2 in the appendix. Thus, we view our selection of companies to be sufficiently representative for the OSEBX.

Furthermore, because we want to measure all variables in a comparable metric, and price levels vary substantially among the companies include, we decide to transform our adjusted prices to returns. This enables evaluation of relationships among variables despite originating from price series of unequal values. For decision-makers employing a framework like the one we present, returns in favor of prices better summarize the investment opportunity in a complete and scale-free manner. Correspondingly, we have calculated one-period simple returns as expressed in equation 4-7.

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \qquad (4\text{-}7)$$

R_t: return at time t
P_t: asset price at time t

4.4.2 Model Inputs

Our initial dataset consists of daily observations; however, we are interested in producing quarterly forecasts. This is because quarterly data points can encompass more information about which phase of the market cycle they belong to. Quarterly data allows for market fluctuations, for example in the form of financial black swans, to be more visible because the time periods extend over a considerable part of the market cycles. From the dataset of 3 700 daily observations, we will employ all of them, corresponding to 59 quarters of stock observations ranging from Q1 2006 through Q3 2020. Correlation coefficients are calculated based on daily data from each quarter. The correlation coefficient for the stock pairs, or the sample Pearson correlation coefficient, which indicates the strength of the relationship between two stocks (CFI, 2020), are calculated employing the equation:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{4-8}$$

x_i, y_i: return for stock x and y

$\bar{x}$, $\bar{y}$: mean of return for stock x and y

The final correlation matrix consists of each stock pair and their quarterly correlation coefficients. Having 38 companies to choose from, the number of unique possible pairs is 703[1]. The correlation matrix that we will use in the constructed models consequently consists of 41 477 data points. An interesting point when describing our data is that due to our methodological approach, output data also constitutes input data. Residuals derived from the ARIMA model is used as input data in the LSTM model.

This selection of data will populate our selected models, but we also need to select an approach for interpreting and validating derived results. Therefore, we are dependent on defining a strategy for quantitative evaluation of the results, which will be presented in the following segment.

4.4.3 Data Split

The characteristics of the data we examine in this thesis as described in the previous section imposes certain constraints on the design of validation and evaluation approach. However, we will begin by describing the usefulness of splitting the data for the purpose of evaluating forecast performance appropriately. In order to train the proposed models, we are dependent on creating a data split which allows for evaluating how our models perform when predicting correlations that were not used in fitting the models. This approach is commonly referred to as a train-test-split where the data is separated into two splits, namely a train portion and a test portion (Hyndman & Athanasopoulos, 2018). The training data is used for estimating the forecasting model parameters and optimizing these based on the desired evaluation metric. The test portion of the data is then employed to evaluate the accuracy of forecasts produced from the model. This split of the data reliably gives indications of the model's true forecasting power.

---

[1] Total number of unique pairs $= \frac{n(n-1)}{2} = \frac{38*37}{2} = 703$

When it comes to the imposed constraints derived from the characteristics of our data, this is essentially due to the dimension of time. When working with time series forecasting, the usual methods of cross-validation are not possible, as the order of the data is essential. The alternative to cross-validation often used for time series validation requires splitting the data into several train and test splits with a rolling time window (Hyndman & Athanasopoulos, 2018). This involves either a sliding window with a fixed window size or an expanding window as observations are added for every time step. Usually, when employing a walk forward methodology like this, the model is retrained for every observation added to the window (Kirkpatrick & Dahlquist, 2010). This is referred to as walk forward optimization, where the model parameters are continuously optimized at each time step. This method of reoptimizing each time step's parameters leads to a trade-off between improved estimates at a computational cost for creating many models. Therefore, we opt for employing a walk-forward evaluation rather than optimizing at each time step. Training the model for one split of data, validating it on the development set, and then evaluating using the walk forward principle is a tolerable compromise.

We choose to create a time series cross-validation split with a constant training set size. A constant training set size implies that for each observation we add at the end of the series we remove one from the beginning. Since the LSTM model we have decided to employ requires a substantial input of data, we have opted to keep the number of data splits low, in order for the training set size to remain large. As previously mentioned, we have also included a development set in order to prevent overfitting to the initial training set. Lastly, we use the walk forward concept explained above to include three test sets for measuring the performance of true forecasts. The data split is visualized in figure 4.5.

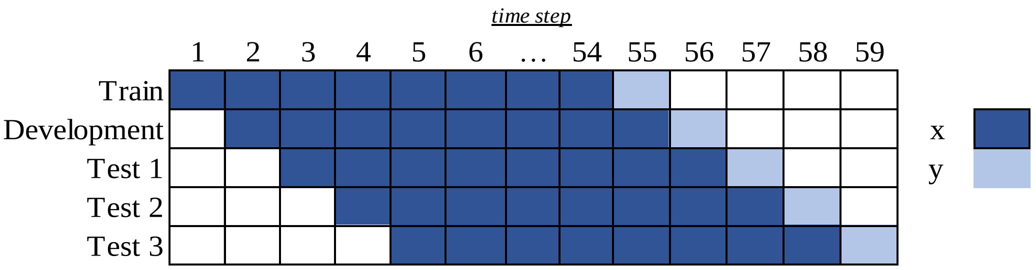

Figure 4.5 Illustration of the data split. The light blue squares depict the forecasted time step as target value *y* for each data set.

When interpreting results, it is advisable to understand the inherent difference within our set of time steps. In the background information we touched upon how Covid-19 has had tremendous effects on global economy. This also includes Oslo Stock Exchange. Figure 4.6 clearly display the negative effects of Covid-19 during the start of 2020. In particular, the first period used for out-of-sample forecasts, Test 1, involves significant negative returns for the OSEBX. From the literature review it is established that periods of market loss coincide with positive correlation coefficients across pairs. Hence, an initial expectation would be that estimating correlation coefficients for Test 1 is particularly difficult because correlation differs from the most occurrent situations. A summary of statistics for correlation in the test periods included can be viewed in Appendix A1. This summary clearly show that our data coincides with the literature and that the periods of market loss, generally involve higher positive correlations. Mean correlation among all stock pairs in Test 1 are more than three times higher than in the preceding quarter. It should also be noted that figure 4.6 display that these market movements are not unprecedented in the investigated time span. An ideal model would be able to learn from these previous time sequences to understand that in periods of market loss, correlation coefficients move collectively in the positive direction.

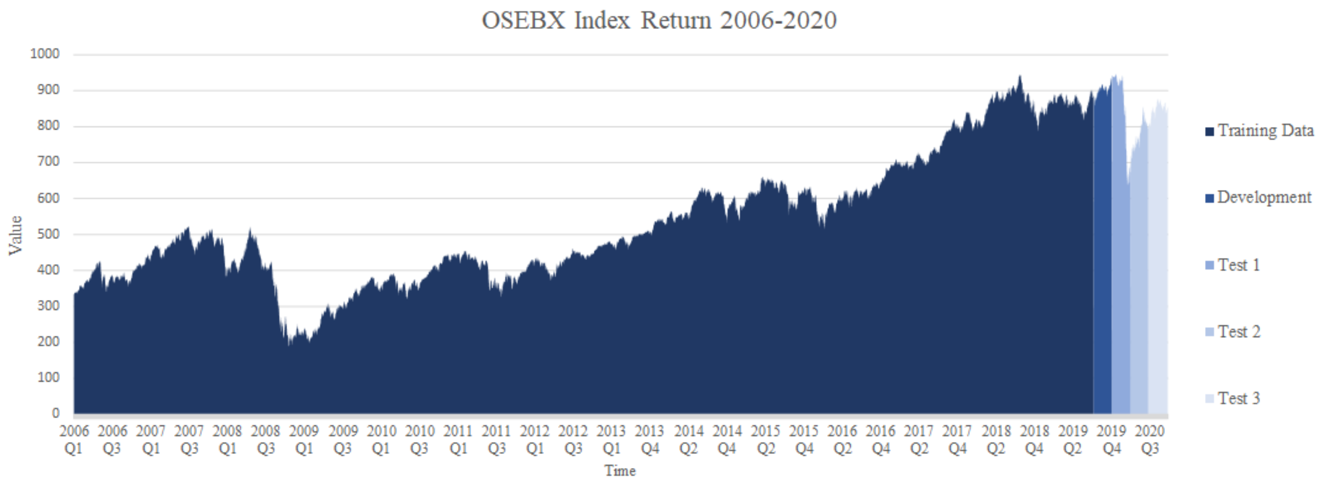

Figure 4.6 Visualization OSEBX index return from Q1 2006 to Q3 2020.

In this chapter we have described the elected approach for estimations, evaluation, and data population. The final hyperparameters for the LSTM model along with a general overview of R and Python implementation can be reviewed respectively in Appendix A6 and Appendix A7. In the next chapter we will present the results derived from the approach described in this chapter. The findings will address the research question directly and form the basis for further discussion of the research question and the subsidiary question related to financial black swans.

5. Results

The following chapter will exhibit our results which will serve as a basis for answering our research question of how modern approaches to forecasting can improve stability in portfolios. Subsequently, an assessment of the variation of results across the development and test sets is presented. Additionally, we display the effects of these predictions on sample portfolio variances. This addresses the research question by allowing us to closer examine the stability of models over time. Ultimately, we compare our findings to the existing literature.

5.1 Predictive Performance

In this section we present the predictive performance of the hybrid model in comparison with the benchmark models outlined in segment 4.2. For each data partition, the models produce out-of-sample forecasts for one quarter ahead and the accuracy is assessed with the performance measures RMSE and MAE.

Table 5-1 Performance for all models and benchmarks, measured in RMSE and MAE. Our proposed ARIMA-LSTM hybrid model is denoted as 'HYBRID'. The lowest RMSE and MAE for each test set is highlighted in bold face.

		Root Mean Squared Error					*Mean Absolute Error*				
		Dev	**Test 1**	**Test 2**	**Test 3**	**Avg.**	**Dev**	**Test 1**	**Test 2**	**Test 3**	**Avg.**
Conv.	**Full Hist.**	0.197	0.389	0.283	0.209	0.269	0.159	0.344	0.239	0.169	0.228
	Constant Corr.	0.215	0.508	0.331	0.237	0.323	0.170	0.483	0.283	0.196	0.283
	Single Index	0.161	0.391	0.242	0.329	0.281	0.128	0.352	0.203	0.280	0.241
	Overall Mean	0.152	0.349	0.199	0.150	0.212	0.122	0.313	0.166	0.120	0.180
Forecast	**ARIMA**	0.151	0.361	0.194	0.150	0.214	0.121	0.328	0.160	0.119	0.182
	LSTM	0.180	0.300	0.203	0.160	0.211	0.149	0.266	0.163	0.128	0.176
	ARIMA-RF	**0.135**	0.360	0.172	**0.134**	0.200	**0.110**	0.330	0.143	**0.106**	0.172
	HYBRID	0.149	**0.292**	**0.155**	0.147	**0.186**	0.120	**0.259**	**0.126**	0.118	**0.156**
	Avg.	0.168	0.369	0.222	0.189	0.237	0.135	0.334	0.185	0.154	0.202

In general, the forecasting methods outperform the conventional method in almost all test periods. However, the Overall Mean model stands out among the conventional methods and has both RMSE and MAE values close to the forecasting methods in most periods. Furthermore, the performance of our ARIMA-LSTM hybrid model stands out in Test 1 and Test 2 and is only outperformed by the ARIMA-RF on the first and last evaluation period, however marginally. However, the ARIMA-RF performs significantly worse on Test set 1 and 2, resulting in a higher average RMSE, depicted in the *Avg.* column in table 5-1. The components of the hybrid model have a worse average predictive performance when applied individually and have a similar mean RMSE. As the individual LSTM model perform significantly better in Test 1 than the individual ARIMA, it has a slightly lower average RMSE.

The results are similar when reviewing the MAE. The ARIMA-LSTM hybrid performs somewhat better than the other forecasting methods and the best performing conventional model Overall Mean. The predictive power of the hybrid model is significantly better than the remaining conventional methods.

5.2 Performance Stability

The main strength of the ARIMA-LSTM hybrid is its performance stability through the test periods. Whereas the other forecasting models perform similarly well to the hybrid in the first and last test periods, they have a significant drop in accuracy in the second and third period tested. This tendency can be extracted from the average RMSE across all models, which is shown in the bottom row of table 5-1. From the first to the second test period, the average RMSE increases from 0.168 to 0.369. The ARIMA-LSTM hybrid also experiences a drop in RMSE, but clearly outperforms the other models in the second test period, excluding the individual LSTM, with an RMSE of 0.292. This trend continues in the third testing period, however with somewhat less distinction.

Across the four sets tested, the hybrid had a standard deviation of 0.061 for the RMSE, displayed in table 5-2. Comparing this to the closest performing conventional and forecast model according to RMSE, the Overall Mean and the ARIMA-RF, that had a standard deviation of 0.081 and 0.094 respectively, the hybrid appear to achieve high stability in its predictions.

Table 5-2 Standard deviation for RMSE in test sets for all models

Standard Deviation of RMSE for each model

Model	**St.dev**
Full Historical Model	0.076
Constant Correlation Model	0.115
Single Index Model	0.087
Overall Mean Model	0.081
ARIMA	0.087
LSTM	0.054
ARIMA-RF	0.094
ARIMA-LSTM	0.061

5.3 Portfolio Variance

Table 5-3 displays the portfolio variance calculated using Overall Mean method and the hybrid model for the 10 sample portfolios each made up of five randomly selected stocks from our dataset. The variance for each of the 10 portfolios is summed and compared to the actual variance. A more detailed view of the portfolio results can be found in Appendix A8.

Table 5-3 Portfolio variance for Overall Mean method and hybrid model, compared to the actual value. The most accurate variance for each test set is highlighted in bold face.

Sum Variance for all Portfolios

Set	Overall Mean	Hybrid	Actual
Dev	**0.713**	0.753	0.732
Test 1	0.365	**0.395**	0.601
Test 2	0.380	**0.407**	0.448
Test 3	**0.398**	0.471	0.369

In summation, the predictions from the hybrid model used to calculate the portfolio is closer to the actual variance than the estimations from the Overall Mean method in Test 1 and 2. In these periods, both the methods estimate a lower variance than what was actually observed, but since the hybrid predicts slightly higher, its performance is better. However, for the development set and Test 3, the hybrid overestimates the variances, and the Overall Mean method is able to produce estimations closer to the observed variances. Table 5-4 displays the accumulated absolute deviation from the methods' estimation and the actual variance for each portfolio.

Table 5-4 Absolute deviation between actual portfolio variance and the estimations from the method in all test sets. The lowest deviation for each test set is highlighted in bold face.

Absolute Deviation Accumulated

Set	Overall Mean	Hybrid
Dev	0.107	**0.073**
Test 1	0.237	**0.207**
Test 2	0.076	**0.057**
Test 3	**0.046**	0.103
Sum	0.465	**0.441**

The same information can be extracted from this table; the hybrid model's estimations for the portfolio variance are better in Test 1 and Test 2, whereas the Overall Mean is better in the development set and test 3. Accumulated over the four test sets, the deviation between the estimations of the hybrid model and the actual values is slightly lower than for the estimations from Overall Mean method, shown in the bottom row of the table.

5.4 Findings in Relation to Previous Literature

As displayed in the previous sections, our proposed hybrid model had the lowest average RMSE and MAE, and the highest stability in prediction accuracy over the four time periods tested. This is mainly in line with the findings made by previous literature employing a similar framework of methodology. Choi’s (2018) implementation of a similar hybrid model to predict stock correlations for S&P500 constituents showed that the ARIMA-LSTM model had a significantly lower RMSE compared to an equivalent set of financial methods such as Constant Correlation, Full Historical and the Single-Index Model. Our results display a similar improvement compared to these methods, but neither the best performing traditional method in our research, the Overall Mean model, nor other forecasting benchmarks was included in Choi’s experiment. The performance of the Overall Mean model is in line with Elton, Gruber, and Urich’s paper (1978) where it was the best performing model among a similar set of correlation forecasting methods in a comparative experiment. However, their paper and other research also identifies the Constant Correlation as one of the best performing methods. In contrast, the Constant Correlation method clearly yielded the highest average RMSE, and was the worst performing method in three out of the four periods tested.

In summation, our ARIMA-LSTM hybrid model was able to achieve significantly higher accuracy in predicting correlation coefficients than most of the conventional methods, measured in both RMSE and MAE. In comparison with the other forecasting models, as well as the Overall Mean model, the hybrid achieved a somewhat lower average RMSE and MAE. The hybrid also showed the lowest variation in prediction accuracy across the test periods, as well as notably lower RMSE in Test 1 in which most of the benchmarks exhibit a large decrease in performance accuracy. When comparing the hybrid model and the Overall Mean model’s ability to predict the correlations between a randomly selected set of stocks in a portfolio, the differences in predictions diminished somewhat, but the hybrid still performed slightly better. The results presented in this chapter provides a basis for answering the research question through a discussion in the next chapter.

6. Discussion

In this chapter we will discuss the results taking the previously stated research question into consideration. Firstly, we will discuss the implications of our results in relation to the research question. Secondly, we will discuss some barriers for adoption of modern forecasting approaches, before we acknowledge some limitations of the thesis. These limitations build up to our proposal for further research, which will constitute the last segment of this chapter.

6.1 Implications for the Research Question

The research question relates to how modern approaches to forecasting can contribute to making portfolios more stable. The research question is also substantiated by a supplementary question of to which degree these contributions are sensitive to financial black swans. To discuss these questions meaningfully it is desirable to first understand the dynamic tendencies of our testing periods. The five data splits used for training, development, and evaluation, stem from fundamentally different time periods. This is clearly illustrated by examining the OSEBX Index Chart in figure 6.1, segmented into our testing periods. Test 1 incorporates most of the financial impact from Covid-19, Test 2 incorporates the recovery, while Test 3 shows market tendencies similar to the development set in addition to the most recent years of training data.

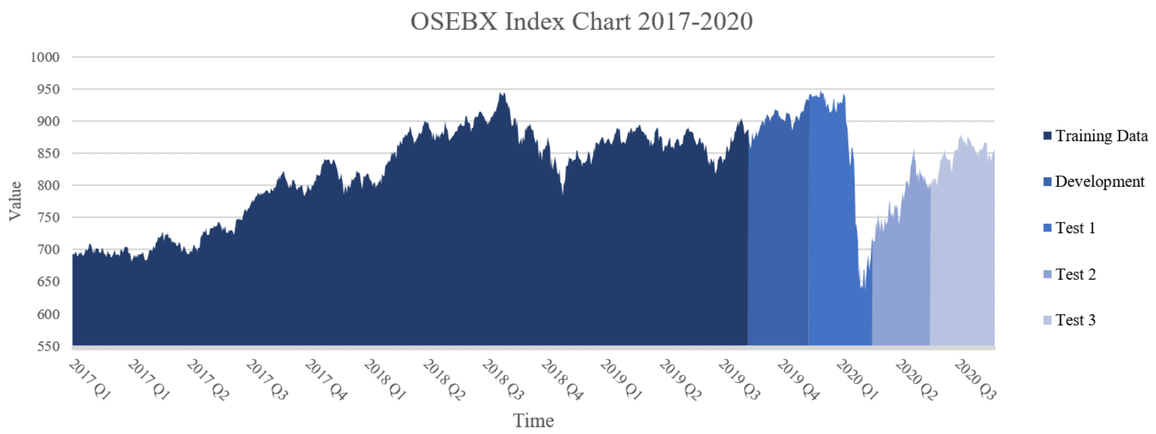

Figure 6.1 Visualization of OSEBX index return from Q1 2017 to Q3 2020.

To begin with, the employment of more accurate forecasts of correlation coefficients will provide better basis for perceiving the risk related to a potential portfolio constituent. The closer the correlation coefficients used for determining constituents of a portfolio are to the real correlation among constituents in the investment horizon, the closer investors will come to the true efficient frontier of portfolios. The portfolio selection segment of the evaluation contributes to displaying how the predictive performance has impact on smaller subsets of the data, and thus leads to better assessments of portfolio variance.

Reviewing the results presented in the preceding section, the proposed hybrid model employing a neural network achieve the best predictive performance. Hence, employing neural networks can provide more accurate forecasts as an alternative to other means of defining correlation in the portfolio optimization problem. A stable portfolio will exist on the low-risk end of the efficient frontier, yielding moderate levels of return with a low portfolio variance. Defining correlation coefficients as a proxy for risk, is essential in this trade-off. With this in mind, the results presented show tendencies that the potential contribution of modern forecasting approaches to the stability of portfolios is significant.

The supplementary question regarding financial black swans is more intricate to answer. Based on our evaluation, strict conclusions should not be drawn as our testing periods only include two quarters which are significantly affected of a financial black swan. However, the predictive performance in the black swan quarters, Test 1 and Test 2, can provide suggestions as to whether the contribution is resistant to black swan events. The proposed hybrid model employing ARIMA and LSTM has the best predictive power in the periods affected by a financial black swan.

A common issue with forecasting in general discussed previously in this thesis is overfitting to the training data. Such overfitting means that seemingly accurate models will not retain the predictive performance when used for true forecasts over the investment horizon. Contributions from a model that are impeded when a black swan event occurs can indicate that the model overfits. As presented in section 5.2 the proposed hybrid model with ARIMA and LSTM varies notably less in RMSE and MAE compared to the various models in general. This suggests that overfitting issues has been reduced as this model has lower fluctuations in predictive performance across multiple time steps.

As the research question presented in this thesis is substantiated by a consideration of resistance to black swans, it requires assessing predictive performance on two ends. A model that fulfills the desired outcome of the posed research question must have strong predictive performance across all testing periods, while not letting the negative impacts of the atypical Test 1 and Test 2 periods deteriorate the overall predictive performance. From the results table

5-1 we can deduce that our proposed hybrid model both has a strong predictive performance and that this is not a result of overfitting data which would lead to relatively large errors in Test 1 and Test 2.

In summary, the findings support the notion that modern approaches to forecasting can contribute to portfolio stability, and that these contributions to a sufficient degree are resistant to black swans. On the contrary, the results are ambiguous as to whether the LSTM in particular is able to improve the forecasts of ARIMA, as the ARIMA-RF model also performs well. However, both the individual LSTM and the hybrid employing an LSTM model performs significantly better than the remaining models in the periods affected by high market fluctuations. This could potentially be attributed to its long-term memory capacity and ability to store information about sequences over long training horizons. Nonetheless, the findings are unambiguous when it comes to the stronger predictive power of modern forecasting approaches, compared to the conventional methods. Better estimations on the back of modern forecasting approaches can help to provide accuracy in the implied risk decision makers are facing when investing. However, these findings rely on methodology with limitations and should not be followed by sentiments of undividedness when drawing conclusions. These limitations will be discussed in the section 6.3, but first we will provide a line of reasoning which might explain why this thesis, or similar studies, fail to bridge the gap between researchers and practitioners. It is important to understand why the modern and advanced methods, which might appeal to researchers due to their predictive power, not always achieves the same acceptance among enterprises or private investors.

6.2 Adoption Barriers

Our thesis provides findings with managerial implications that modern approaches can be employed to make better foundations for strategic decisions in portfolios. Further, this can be extrapolated to learning that expands on our research question. If our specific choice of investigation, portfolios, can be improved with modern approaches, then there is likely to exist other decision areas which have untapped potential in terms of exploiting machine learning and other emerging methods. This inference is however frivolous if businesses chooses not to adopt such modern approaches. Therefore, in the following discussion we will go into barriers of adopting new methods and the implications this have for further research. Admittedly, this is a digression from our specific research question that deals with the usefulness of modern forecasting methods in portfolios. However, it is a valuable discussion for understanding to which degree managerial implications can and should be drawn from our thesis.

One can draw parallels to the philosophical principle *Occam's razor.* This principle states that if two different explanations exist for the same phenomenon, then the simpler explanation should be preferred (Duignan, 2020). The philosophical principle has laid the foundation for a principle in computational learning theory, *Occam* learning, which states that given all other things being equal, a shorter explanation for observed data should be favored over a lengthier explanation (Blumer, Ehrenfeucht, Haussler, & Warmuth, 1987). For our thesis this can be conveyed as; if simpler methods provide decision makers with the same value in terms of predictive power and implicit learning about relationships as advanced models, decision makers should prefer the simple method. This is because simpler methods can be associated with lower costs, while more advanced methods are consequently more expensive. Hence, lack of adoption of modern approaches to forecasting, and data analysis in general, is an issue that can be reduced to a cost-benefit analysis of possible methods. The benefit is the predictive power of models and the implicit learning of the model, which can be derived from the degree of interpretability. The cost on the other hand is related to requirements such as competence, preprocessing, computational power, availability, time, and data volumes. Given a problem related to a set of data, it is likely that there is an extremely complex model in a sea of infinitely many different modifications of models and parameters, which is optimal for the problem. However, it is unfeasible to try to find this one ideal model and decision makers must therefore always appraise the possible models with consideration to the associated cost of identification and implementation. Alas, businesses should, and most likely will, determine their problem approaches based on a cost-benefit assessment.

If we accept notions that there are emerging and more advanced methods that possibly can provide better predictive power than the status quo, it is highly relevant to discuss why these are not adopted more widely among businesses. For such methods to be adopted and implemented in favor of the traditional methods, the perceived value added must outweigh the increased cost. This can happen in one of two ways. Either researchers must improve emerging methods to a point where their added values are so superior that businesses are forced to adopt them in spite of increased costs, or they must develop effective frameworks which dramatically lessens the burden of implementation. This thesis has relied on the latter to some degree. The relative simple method of ARIMA for explaining linear tendencies are specialized for each of the 703 time series in our data set, however implemented with the help of an automatic framework, while the LSTM network is generalized to reduce the computational cost. Our approach is therefore an example of how researchers can adhere to the cost-benefit consideration, but it is only one of many possible approaches to do just that. This leads us to the limitations of our thesis and our encouragement to future research.

6.3 Limitations

The findings of the methodology reported in the last section should be considered in light of some limitations. Although our results suggest promising potential with the application of modern approaches for forecasting the correlation coefficient between stocks, it is important to highlight some of the limitations the approach does not cover in its current state. This section includes identifying said limitations, in addition to discussing some of these limitations in detail. Lastly, alternative approaches and direction for future research is proposed.

First and foremost, the most fundamental constraint when performing research in general should be referred to. This thesis is subject to constraints regarding to time and the time required for different processes. In particular, the computational burden of problems increases in parallel with time necessities. Naturally, such constraints lead to a need for simplifications across a multitude of thesis elements.

One of these simplifications relates to the data gathering. Only included companies' historical price movements were collected as data foundation, limiting our model and benchmarks to univariate time series architecture. A multivariate time series model with additional explanatory variables would have entailed a stronger foundation to draw empirical conclusions from. Therefore, the univariate time series structure limits the validity of the generalization of the hybrid model. This could mean that application on different time periods, or on other stock pair correlations, would have resulted in less accurate predictions.

Another ramification of the all-embracing time constraint is the simplifications applied to the size of our data. Neural networks have a fundamental advantage of being able to handle vast amounts of data, which means that this thesis is limited in its review of this learning algorithm. Including additional points of data through covering a wider timespan, or supplementary variables as mentioned in the preceding paragraph, would increase the utility value of the neural network. Exhibiting awareness of this limitation could be viewed as paradoxical when considering the transformation from daily to quarterly data. This transformation undeniably reduces the number of data points fed to the models. However, this has been considered a trade-off between time span covered and computational feasibility. Employing quarterly data allows for covering a substantially wider time span, without making the computational burden unmanageable.

A desire to provide sufficient amounts of data to the neural network led to this thesis relying on simplifications regarding the validation split of data. Normally cross-validation in time series involves including numerous data splits to validate on many time steps. However, the data hungriness of neural networks led to keeping the window size of the time series cross-validation large, which consequently reduces the number of possible evaluation steps. Having said that, it should be regurgitated that each time step in our data split involves predicting and calculating performance metrics that are composed of 703 correlation coefficients. On account of this, making a generalized model on the initial training data and evaluating it on a limited amount of testing sets is determined an acceptable compromise.

The constraints related to time and computational efforts also induce need of limitation in model applications. In this thesis automated frameworks have been used to reduce the time consumption of certain methodological steps. These automated frameworks are great for this reason; however, they can also bring about suboptimal solutions or potentially hampered learning as it reduces researcher involvement. In other cases, for instance in the inclusion of benchmark models, parameter tuning is limited which will have depreciating effects on the validity of findings. Limited time also led to this thesis not making efforts to look inside the black box of the LSTM (Beizer, 1995). Traditionally, applications of advanced neural networks for decision making have received criticism for being used in favor of interpretable models. This criticism has triggered a response where researchers have developed methods for backpropagating through neural networks, which allows for learning causality mechanisms. On the other hand, this thesis and the research question presented does not make efforts to explain the causalities of changes in correlation coefficients, but rather discuss the potential value of modern learning applications in financial forecasting. These presented limitations are the starting point for a discussion on how this specific method can be embroidered in further research, and thus contribute to better predictive power and improve implicit learning through increased interpretability.

6.4 Future Research

Further research in the field with regards to these limitations can be motivated by the reasoning that increased value added from modern approaches to forecasting will lead to a higher probability of real-life business adoption. Improvements in predictive power can be achieved by exploring different learning algorithms, or alternative data foundations. Parameter optimization is costly both in time and computational efforts, which is significantly limiting the scope of this thesis. Furthermore, inclusion of explanatory time series is a field of further research that is likely to yield interesting findings, which further strengthens the predictive power that modern forecasting approaches can display to attract practitioners. As mentioned earlier, however, we would also like to point out that further research that improves the predictive power of these new methods should be done simultaneously with attempts to automate and simplify their implementation on real issues in order to best reduce barriers to business adoption. We will therefore attribute the same emphasize to developing frameworks for implementation of modern approaches, as to constantly expanding the complexity of models.

7. Conclusion

This thesis aimed to investigate how modern approaches to forecasting can contribute to more stable portfolios. The background and literature review provided a reasoning for why this research question needs to be addressed. The background information explained a long-lasting problem of constructing inputs for portfolio strategy. This problem is also surrounded by dynamic conditions represented in this thesis by the inclusion of Covid-19 as an example of a financial black swan. Furthermore, the literature review revealed a decision area which traditionally is solved with simpler statistical methods, despite developments of modern forecasting methods that are applicable to the problem.

The aforementioned elements led us to our proposed model consisting of an ARIMA component and an LSTM component, which was responsible for explaining the linear and non-linear tendencies in the data, respectively. Our experimental approach was populated by data on Oslo Stock Exchange returns, including a time span that encompassed several peaks and throughs. This coincided with our attempt to substantiate our research question with an element of sensitivity to financial black swans. The approach also included a range of benchmarks consisting of conventional methods for estimating correlation, the individual components of the hybrid model and an alternative machine learning method for the non-linear tendencies of a hybrid model. This ensured a comparative design of the experiment which aimed to provide findings related to our research question.

Our approach, populated by the elected data, provided findings which illustrated an untapped potential of modern approaches to forecasting in providing input accuracy in portfolio strategy. The elected forecasting methods of our thesis accrued a predictive performance that overall was stronger than the conventional methods across all test sets. In addition, the dynamic conditions represented in a financial black swan encompassed by Test 1 and Test 2 did not deteriorate the predictive performance enough for these contributions to lose its value. This implies that practitioners equipped with modern forecasting approaches can achieve more accuracy in their inputs and thus achieve their desired level of portfolio stability.

However, the discussion has also addressed why common practice may deviate from scientific findings, such as the ones presented in this thesis. The Overall Mean, which is a relatively simple statistical method for estimating future correlation, had a predictive performance comparable to the forecasting methods. Considering that this method is substantially less costly, in reference to computational cost and time demand, it is difficult to strictly determine that forecasting methods with their predictive power displayed in this thesis is worthwhile for business adoption. We therefore want to encourage researchers to focus their attention to efforts on reducing the cost of modern forecasting approaches, in order to bridge the gap between researchers and practitioners. In addition, our findings were affected by the main limitation which is related to the scarcity of time. Provided more time this thesis could include additional, data, methods and optimization of model parameters, which would be expected to increase the predictive performance achieved.

This thesis has contributed to the literature by displaying how previously unused modern approaches to forecasting can be utilized for estimation of inputs required for decision making. Alas, we contribute to the literature by providing an example of how modern forecasting approaches can provide more stability in portfolios by increasing the accuracy of correlation coefficient estimations. Forecasting of correlation coefficients is naturally only one specific area of decision-making inputs, and we believe that there are a multitude of potential areas to investigate in further research.

8. References

Adebiyi, A., Adewumi, A., & Ayo, C. (2014, March). Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *J. Appl. Math*. doi:614342:1-614342:7.

Ahir, H., Bloom, N., & Furceri, D. (2020, April 4). *IMF Blog*. Retrieved December 8, 2020, from Global Uncertainty Related to Coronavirus at Record High: https://blogs.imf.org/2020/04/04/global-uncertainty-related-to-coronavirus-at-record-high/

Armstrong, J. (2001). Evaluating forecasting methods. *Chapter 14 in Principles of forecasting: a handbook for researchers and practitioners*.

Avishai, B. (2020, April 21). *The Pandemic Isn't a Black Swan but a Portent of a More Fragile Global System*. Retrieved from The New Yorker: https://www.newyorker.com/news/daily-comment/the-pandemic-isnt-a-black-swan-but-a-portent-of-a-more-fragile-global-system

Beizer, B. (1995). *Black-Box Testing: Techniques for Functional Testing of Software and Systems.*

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning Long-Term Dependencies with Gradient Descent is Difficult. *Transactions On Neural Networks, 5*(2), pp. 157-166.

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1987). Occam's razor. *Information processing letters, 24*(6), pp. 377-380.

Bouvet Norge. (2020, June 11). *Convolutional Neural Networks : The Theory*. Retrieved October 10, 2020, from Bouvet Deler: https://www.bouvet.no/bouvet-deler/understanding-convolutional-neural-networks-part-1

Box, G., & Jenkins, G. (1970). *Time Series Analysis: Forecasting and Control.* San Fransisco: Holden Day.

Breiman, L. (2001). Random Forests. *Machine Learning, 45*, pp. 5-32. doi:10.1023/A:1010933404324

CFI. (2020). *Correlation - A statistical measure of the relationship between two variables*. Retrieved October 6, 2020, from Corporate Finance Institute: https://corporatefinanceinstitute.com/resources/knowledge/finance/correlation/

Chesnay, F., & Jondeau, E. (2001). Does Correlation Between Stocks Really Increase During Turbulent Periods? *Economic Notes*(30), pp. 53-80.

Choi, H. K. (2018). Stock price correlation coefficient prediction with ARIMA-LSTM hybrid model. Seoul, Korea: Korea University. Retrieved from Retrieved from: https://arxiv.org/pdf/1808.01560v5.pdf

Chollet, F., & others. (2015). Keras. Retrieved from https://github.com/fchollet/keras

Duignan, B. (2020, August 5). *Topic: Occam's razor*. Retrieved November 2, 2020, from Brittanica: https://www.britannica.com/topic/Occams-razor

Dupond, S. (2019). A thorough review on the current advance of neural network structures. *Annual Reviews in Control, 14*, pp. 200-230.

Elton, E., Gruber, M., & Urich, T. (1978, December). Are Betas Best? *The Journal of Finance, 33*(5), pp. 1375-1384. doi:10.2307/2327272

Fathi, O. (2019). Time series forecasting using a hybrid ARIMA and LSTM model. France: Velvet Consulting.

Fattah, J., Ezzine, L., Aman, Z., Moussami, H., & Lachhab, A. (2018). Forecasting of demand using ARIMA model. *International Journal of Engineering Business Management, 10*(2). doi:10.1177/1847979018808673.

Fischer, T., & Kraus, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270*(2), pp. 654-669. doi:10.1016/j.ejor.2017.11.054

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning.* Cambridge, MA.: MIT Press.

Graves, A. (2012). *Supervised Sequence Labelling with Recurrent Neural Networks* (Springer. utg.).

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017, October). LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems, 28*(10), pp. 2222-2232. doi:10.1109/TNNLS.2016.2582924

Grolemund, G., & Wickham, H. (2011). Dates and Times Made Easy with lubridate. *Journal of Statistical Software, 40*(3), pp. 1-25. Retrieved from https://www.jstatsoft.org/v40/i03/

Hand, D. J. (2014). *The Improbability Principle: Why Coincidences, Miracles, and Rare Events Happen Every Day.* New York: Scientific American / Farrar, Straus and Giroux.

Haykin, S. (2008). *Neural Networks and Learning Machines: Third Edition.* Upper Saddle River, New Jersey: Pearson Education, Inc.

Hochreiter, S., & Schmidhuber, J. (1997). LONG SHORT-TERM MEMORY. *Neural Computation, 9*(8), s. :1735{1780. doi:10.1162/neco.1997.9.8.1735

Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J. (2001). Gradient Flow in Recurrent Nets: the Difficulty of Learning Long-Term Dependencies. *A Field Guide to Dynamical Recurrent Neural Networks*.

Hutter, F., Hoos, H. H., & Leyton-Brown, K. (2011). Sequential Model-Based Optimization for General Algorithm Configuration. *Lecture Notes in Computer Science, vol 6683*. doi:10.1007/978-3-642-25566-3_40

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software, 27*(1), pp. 1-22. doi:10.18637/jss.v027.i03

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting, 22*(4), ss. 679-688.

Hyndman, R., & Athanasopoulos, G. (2018). 3.1 Some simple forecasting methods. *Forecasting: principles and practice*. Australia: Otexts. Retrieved December 8, 2020, from Forecasting: principles and practice: https://otexts.com/fpp2/simple-methods.html

Hyndman, R., & Athanasopoulos, G. (2018). Forecasting: principles and practice. Melbourne, Australia. Retrieved December 8, 2020, from 8 - ARIMA models: https://OTexts.com/fpp2

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., . . . Yasmeen, F. (2020). forecast: Forecasting functions for time series and linear models. Retrieved from https://pkg.robjhyndman.com/forecast/

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., . . . Yasmeen, F. (2020). forecast: Forecasting functions for time series and linear models. R package version 8.13. Hentet fra https://pkg.robjhyndman.com/forecast/

Kirkpatrick, C. D., & Dahlquist, J. R. (2010). Technical Analysis: The Complete Resource for Financial Market Technicians. *FT Press.*, p. 548.

Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning, 30*, pp. 271-274. doi:10.1023/A:1007411609915

Krauss, C., Anh, X., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research, 259*(2), pp. 689-702. doi:10.1016/j.ejor.2016.10.031

Kryzanowski, L., Galler, M., & Wright, D. W. (1993, July). Using Artificial Neural Networks to Pick Stocks. *Financial Analysts Journal, 49*(4), pp. 21-27. doi:10.2469/faj.v49.n4.21

Lecun, Y., Bottou, L., Orr, G., & Müller, K.-R. (2012). Efficient BackProp. I Y. Lecun, L. Bottou, G. Orr, & K.-R. Müller, *Neural Networks: Tricks of the Trade.* (ss. 9-48). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-35289-8_3

Levenbach, H. (2017). *Change & Chance Embraced: Achieving Agility with Smarter Forecasting in the Supply Chain.* Delphus Publishing.

Liaw, A., & Wiener, M. (2001). Classification and Regression by RandomForest. *R News, 2*, pp. 18-22.

Lopez de Prado, M. (2018). *Advances in Financial Machine Learning.* John Wiley & Sons Inc.

Low, R., Faff, R., & Aas, K. (2016). Enhancing mean–variance portfolio selection by modeling distributional asymmetries. *Journal of Economics and Business, 85*, ss. 49–72. doi:10.1016/j.jeconbus.2016.01.003

Markowitz, H. (1952, March). Portfolio Selection. *The Journal of Finance*(7(1)), pp. 77-91. doi:10.2307/2975974

Markowitz, H. (2002). Efficient Portfolios, Sparse Matrices, and Entities: A. *Operations Research, 50*(1), pp. 154-160. doi:10.1287/opre.50.1.154.17774

Martins, T. G. (2019). Deep Learning Lecture 1 - Prevent overfitting in Keras. Department of Mathematical Sciences, NTNU. Retrieved from https://www.math.ntnu.no/emner/MA8701/2019v/DeepLearning/7-prevent_overfitting.html

McElreath, R. (2016). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* CRC Press.

Olson, D., & Mossman, C. (2003, January). Neural network forecasts of Canadian stock returns using. *International Journal of Forecasting, 19*, pp. 453-465. doi:10.1016/S0169-2070(02)00058-4

Oslo Børs. (2020). *Hovedindeksen OSEBX*. Retrieved December 8, 2020, from https://archive.is/oeAvZ#selection-791.0-803.12

Preis, T., Dror, Y., Kenett, H., Stanley, E., Helbing, D., & Ben-Jacob, E. (2012, October 18). Quantifying the Behavior of Stock Correlations Under Market Stress. *Scientific Report*(2). doi:10.1038/srep00752

Raymond, Y. T. (1997). An application of the ARIMA model to real-estate prices in Hong Kong. *Journal of Property Finance, 8*(2), pp. 152-163. doi:0958-868X

Siami-Namini, S., Tavakoli, N., & Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. *IEEE International Conference on Machine Learning and Applications, 17*, pp. 1394-1401. doi:10.1109/ICMLA.2018.00227

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research, 15,* pp. 1929-1958.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable.* New York: Random House.

Temür, A., Akgün, M., & Temür, G. (2019). Predicting housing sales in Turkey using ARIMA, LSTM and hybrid models. *Journal of Business Economics and Management, 20,* pp. 920-938. doi:10.3846/jbem.2019.10190.

Weiss, E. (2000). Forecasting commodity prices using ARIMA. *Technical Analysis of Stocks & Commodities, 18*(1), pp. 18-19.

Wickham, H., François, R., Henry, L., & Müller, K. (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.2. Retrieved from https://CRAN.R-project.org/package=dplyr

World Bank. (2020). *Global Economic Prospects, June 2020.* Washington, DC: World Bank. doi:10.1596/978-1-4648-1553-9

World Uncertainty Index. (2020, November 20). *Home: World Uncertainty Index (WUI): Global*. Retrieved November 8, 2020, from https://worlduncertaintyindex.com/

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2020). Dive into Deep Learning. Retrieved December 8, 2020, from https://d2l.ai

Zhang, P. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, *50,* pp. 159-175. doi:10.1016/S0925-2312(01)00702-0.

9. Appendix

The appendix is arranged according to the chronological references throughout the text. Firstly, we present a descriptive summary of the collected data. Secondly, we present theory regarding ARIMA, RNN and LSTM. A brief explanation of the elected evaluation metrics is then included. Thereafter, we provide a short description of the approach implementation in R and Python. Lastly, disaggregated results from the portfolio evaluation are presented.

A1: Data Description

Table A-1 Descriptive test data summary

Summary Statistics for Correlation in Test Periods

Data Set	Quarter	Return OSEBX	Median	Mean	S.d.
Dev	2019 Q4	5 %	0.133	0.144	0.159
Test 1	2020 Q1	-20 %	0.501	0.483	0.157
Test 2	2020 Q2	4 %	0.278	0.265	0.200
Test 3	2020 Q3	8 %	0.153	0.182	0.154

Table A-2 Companies included in the dataset, with ticker and industry

Company Name	Ticker	Industry
Af Gruppen ASA	AFG	Construction & Engineering
Aker ASA	AKER	Oil & Gas Related Equipment and Services
Aker Solutions ASA	AKSO	Oil & Gas Related Equipment and Services
American Shipping Company ASA	AMSC	Freight & Logistics Services
ABG Sundal Collier Holding ASA	ABG	Investment Banking & Investment Services
Atea ASA	ATEA	Software & IT Services
Axactor SE	AXA	Banking Services
Bonheur ASA	BONHR	Electrical Utilities & IPPs
Dnb ASA	DNB	Banking Services
Dno ASA	DNO	Oil & Gas
Equinor ASA	EQNR	Oil & Gas
FRONTLINE LTD	FRO	Oil & Gas Related Equipment and Services
GAMING INNOVATION GROUP INC	GIG	Hotels & Entertainment Services
Golden Ocean Group Limited	GOGL	Freight & Logistics Services
Hexagon Composites ASA	HEX	Containers & Packaging
Kitron ASA	KIT	Electronic Equipment & Parts
Kongsberg Automotive ASA	KOA	Automobiles & Auto Parts
Kongsberg Gruppen ASA	KOG	Aerospace & Defense
Leroy Seafood Group ASA	LSG	Food & Tobacco
Medistim ASA	MEDI	Healthcare Equipment & Supplies
Norwegian Air Shuttle ASA	NAS	Passenger Transportation Services
Nel ASA	NEL	Renewable Energy
Norsk Hydro ASA	NHY	Metals & Mining
Nordic Semiconductor ASA	NOD	Semiconductors & Semiconductor Equipment
Olav Thon Eiendomsselskap ASA	OLT	Real Estate Operations
Orkla ASA	ORK	Food & Tobacco
PGS ASA	PGS	Oil & Gas Related Equipment and Services
Photocure ASA	PHO	Pharmaceuticals
Schibsted ASA	SCHA	Media & Publishing
STOLT-NIELSEN LIMITED	SNI	Freight & Logistics Services
Sparebank 1 SR Bank ASA	SRBNK	Banking Services
Storebrand ASA	STB	Investment Banking & Investment Services
Subsea 7 SA	SUBC	Oil & Gas Related Equipment and Services
Telenor ASA	TEL	Telecommunications Services
TGS NOPEC Geophysical Company ASA	TGS	Oil & Gas Related Equipment and Services
Tomra Systems ASA	TOM	Professional & Commercial Services
Veidekke ASA	VEI	Construction & Engineering
Yara International ASA	YAR	Chemicals

A2: ARIMA

A univariate ARIMA model attempts to predict a value in a response time series by utilizing linear combinations of its past values and errors. This requires stationarity in the time series. A stationary time series is a time series whose properties, such as mean, variance and autocorrelation, do not depend on the time at which the series are observed (Hyndman & Athanasopoulos, 2018).

Consider the general ARIMA model of order (p, d, q)

$$y_t^d = c + \sum_{i=1}^{p} \varphi_i \, y_{t-1}^d + \varepsilon_t + \sum_{j=1}^{q} \theta_j \, \varepsilon_{t-j} \tag{9-1}$$

Where: y_t^d is stationary at time t with d levels of differencing, c is a constant intercept, φ_i is a parameter denoting the coefficient related to the previous p values of y_t, and ε_t is an error term $\sim\mathcal{N}(0, \sigma_\varepsilon^2)$, θ_j is a parameter denoting the coefficient related to the past q values of the error term.

Box and Jenkins suggest an iterative three-stage process for estimating an ARIMA model (Box & Jenkins, 1970). Firstly, the order (p, d, q) of the model is selected based on the time series' observed characteristics. Typically, the time series is visually inspected to identify how many differencing levels must be applied to obtain stationarity. One level of differencing is equal to computing the difference between consecutive observations, expressed as $z_t = x_t - x_{t-1}$. Additional computations like logarithmic or Box-Cox transformations can also be applied to stabilize the variance. Then, the autocorrelation function, regularly referred to as the ACF, can be used to measure the linear dependence between observations separated by a time lag p. Further, the partial autocorrelation function, referred to as PACF, can be used to determine how many autoregressive terms q are necessary (Hyndman & Athanasopoulos, 2018). Secondly, the parameters $\varphi_i \; and \; \theta_j$ for the selected model (p, d, q) are estimated. These coefficients are typically computed with maximum likelihood estimation to best fit the selected model.

The goodness of fit of the calculated model is often measured by Akaikes Information Criteria (AIC) (McElreath, 2016).

The AIC can be written as

$$AIC = -2\log(L) + 2(p + q + k + 1) \quad (9\text{-}2)$$

Where L is the likelihood estimate of the data, p and q are the number of past values and past error terms included in the model as parameters, and k is an indicator where $k = 1$ if the intercept coefficients $c > 0$ and 0 otherwise.

Lastly, the model fitted is evaluated and the autocorrelations from its residuals are checked to satisfy certain assumptions. The residuals are expected to resemble white noise and show low levels of autocorrelation. If the autocorrelations still contain some large values, the values for p and q can be adjusted and the three-stage process is repeated.

A3: RNN

We utilized a Recurrent Neural Network, commonly referred to as an RNN, in order to make the final predictions for the correlation coefficients. The general structure of a neural network is that of a network of mathematical functions, known as neurons or nodes, that is joined by connection weights (Graves, 2012). RNNs are a type of sequential neural network that, in contrary to Feed-Forward Networks, can use its output data from a previous time step as input data in the next time step, through a feedback loop (Dupond, 2019). Thus, RNNs allow the model to capture dependencies, and store this information over time and sequences as *hidden states*, making such models suitable for time series forecasting. The recurrent structure of a RNN is displayed in figure A-1.

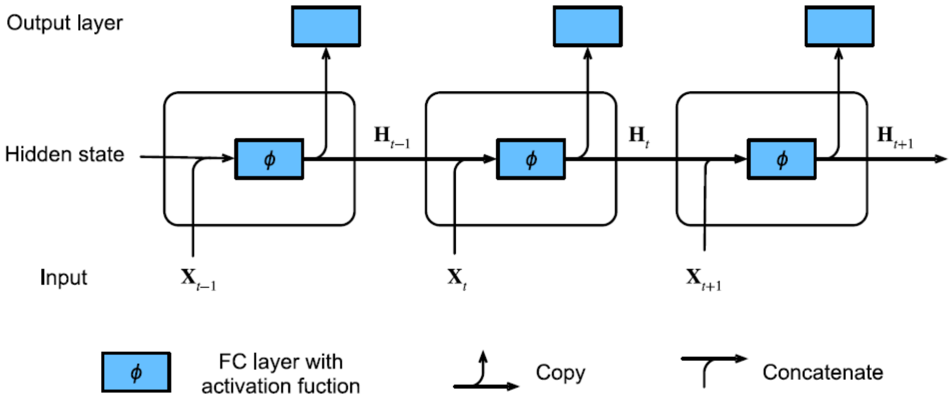

Figure A-1 RNN structure. Source: (Zhang, Lipton, Li, & Smola, 2020)

Generally, the RNN updates its hidden state H_t, given a sequence of input values $x = [x_1, x_2, \ldots x_t]$ and the hidden state of the previous time step H_t, as shown in the following equation

$$H_t = \varphi(WX_t + UH_{t-1} + b) \quad (9\text{-}3)$$

Where φ represents an *activation function*, which serves as a *gate* that transforms and maps the input values. The model aims to learn the parameters W and U, as well as the bias term b. Furthermore, the activation is passed forward to the next layer of nodes until it reaches the output node where it produces the final predictions. The network seeks to optimize the parameters by minimizing a *loss function* that computes the difference between the model predictions $\hat{y}$ on the training data and the true target value y. For regression tasks, the most common loss function is the squared error, $(y - \hat{y})^2$, but different functions can be selected depending on the specificity of the task. In practice, the network draws a randomly selected subset, referred to as a batch, of the training samples at fixed size and calculates the loss of the predictions. The parameters are then updated through a process called *backpropagation through time*, based on an optimization algorithm that improves the *loss function*. The network also uses a pre-defined *learning rate* when deciding how much to update the parameters each iteration. A single iteration of this process is called an *epoch* and the number of iterations is, in addition to the size of the selected subset, learning rate and optimization algorithm used, a hyperparameter that should be tuned in order to find an appropriate model.

However, due to the recurrent connections in these RNN structures, the resulting effect of these hidden states on the network outputs either decays rapidly or explodes exponentially over time (Hochreiter, Bengio, Frasconi, & Schmidhuber, 2001), depicted in figure A-2. This problem is often referred to as the vanishing gradient problem and poses a problem when attempting to model dependencies in long sequences (Bengio, Simard, & Frasconi, 1994). There have been several attempts to create a modified RNN architecture in order to deal with the aforementioned problem, and we have selected the Long Short-Term Memory (LSTM) approach in this thesis.

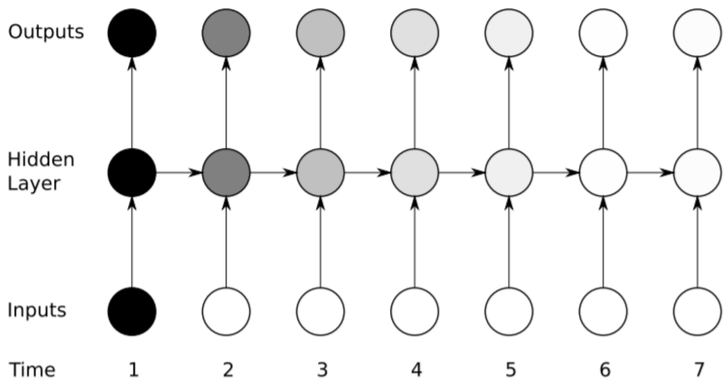

Figure A-2 Vanishing gradient problem. Source*:* (Graves, 2012)

A4: LSTM

In 1997, Hochreiter & Schmidhuber developed the LSTM network to address the problem of long-term information preservation without the risk of exploding or vanishing gradients. A LSTM model introduces four different gates*, the forget gate, the input gate, the input candidate gate and the output* gate, that gives the model the ability to decide when to remember and when to ignore inputs in the hidden state by using a specified algorithm (Zhang, Lipton, Li, & Smola, 2020). Additionally, a cell state C_t is calculated, stored, and passed on to the following time step, serving as the *long-term memory* in the model (Fathi, 2019). Figure A-3 depicts how the gates in a node interact with data passed from the previous node and how it calculates what output H_t to pass on to the next node.

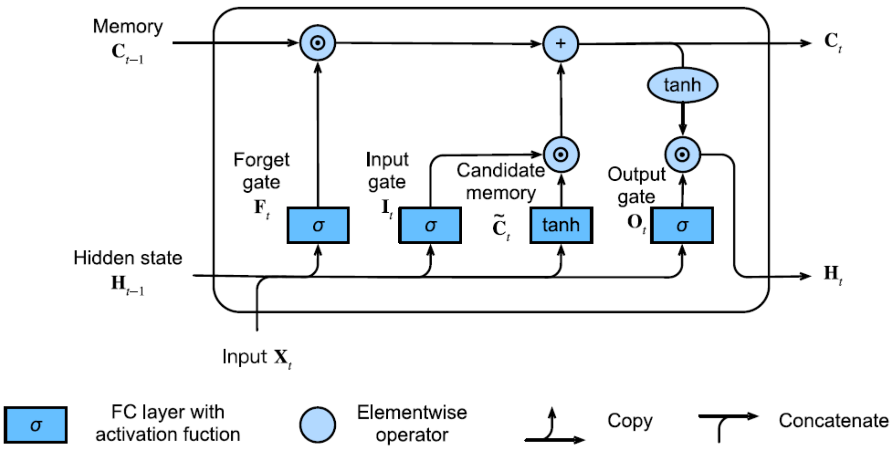

Figure A-3 Graphical illustration of the inner structure of an LSTM cell. The operations performed in each gate is explained below. Source: (Zhang, Lipton, Li, & Smola, 2020)

The forget gate F_t represents a forgetter that is pointwise multiplied to the previous cell state C_{t-1} to drop values that are deemed unnecessary, as well as keeping those who are necessary for the predictions. The calculations in the forget gate is expressed through the following equation

$$F_t = \sigma(W_f x_t + U_f H_{t-1} + b_f) \tag{9-4}$$

The input value x_t and the hidden state from the previous block H_{t-1} is weighted with the parameters W_f and U_f, where the subscript f refers to the forget gate. Additionally, the gate's bias parameter b_f is added, before a sigmoid function σ is applied, ensuring that the output is mapped between 0 and 1.

The input gate I_t decides how much information from the input that will be added to the cell state and follows the same structure as the forget gate.

$$I_t = \sigma(W_i x_t + U_i H_{t-1} + b_i) \tag{9-5}$$

$$\tilde{C}_t = tanh(W_c x_t + U_c H_{t-1} + b_c) \tag{9-6}$$

The input candidate gate $\tilde{C}_t$ uses a $tanh$ function to create a set of candidate values that is combined with I_t through pointwise multiplication. If the forget gate approximates 1 over time and the input gate approximates 0, most of the past cell states C_{t-1} are saved and used in current time steps. This enables the model to better identify long-term dependencies which reduces the effect of the vanishing gradient problem (Zhang, Lipton, Li, & Smola, 2020).The $tanh$ function is a hyperbolic tangent function which renders values between -1 and 1. The combination that updates the cell state uses pointwise multiplication, described in the following equation

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \tag{9-7}$$

In other words, the new cell state is stripped for information the model deemed unnecessary and will encompass information from the new input that it deems valuable.

Then, this final cell state C_t is stored and passed on to the next time step, and also used in the calculations for the output in the output gate in equation 9-9.

$$O_t = \sigma(W_o x_t + U_o H_{t-1} + b_o) \tag{9-8}$$

$$H_t = O_t \odot tanh\,(C_t) \tag{9-9}$$

Again, weights are multiplied with the inputs and the gate's bias parameter is added before a sigmoid function is applied to perform the output gate calculations. Lastly, the output gate calculates what values to use as output in the hidden state H_t by combining the $tanh$ applied cell state with O_t using pointwise multiplication.

A5: Performance Metrics

A5.1 MSE

Historically, the MSE is a popular pick as an accuracy measure for forecasting due to its theoretical relevance in modelling statistics (Hyndman & Koehler, 2006). The MSE is calculated as a sum of the squared errors for each observation, divided by the total number of observations, shown in equation 9-10. Due to the squaring of the errors, the MSE will penalize large deviations between the true values y and the forecasted values $\hat{y}$ more heavily. For the validity of our models, we wish to avoid obtaining a model that performs extremely well in some circumstances and very poorly in others. For this reason, MSE is our preferred metrics as it is better to display the stability and generalization of the models across all the time series. Additionally, the root of the MSE (RMSE) increases interpretability as it expresses the prediction error in the same units as the variable we are estimating. Alas, we will use the RMSE when presenting the results.

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{9-10}$$

A5.2: MAE

We also include the Mean Absolute Error (MAE) as a performance measure. Contrary to the MSE, the MAE is less sensitive to outliers and large prediction deviations which has caused some authors to favour the metric for forecast accuracy evaluation (Armstrong, 2001). In other words, the MAE penalises the errors for all the observations i equally, which captures the overall performance better, but is less suitable when a potential outlier has a great negative effect for the practical use of the model. Therefore, we include MAE as an additional performance metric, meant to supplement the MSE. The calculation of MAE is displayed in equation 9-11.

$$\frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{9-11}$$

A6: Hyperparameter Selection

Table A-3 Final hyperparameters used in the LSTM model

Hyperparameter	Value
Number of hidden layers	1
Number of nodes	10
Loss function	*MSE*
Activation function	*tanh*
Optimization algorithm	*adam*
Epochs	100
Batch size	64
Early Stopping	10
(L1, L2) regularization weights	(0.2, 0.0)
(L1, L2) regularization bias	(0.2, 0.0)
Learning rate	0.001
Dropout rate	0.1

A7: R and Python Implementation

All of our models besides from the LSTM are implemented in the R programming language. The LSTM network is implemented using keras version 2.3.1 (Chollet & others, 2015) with built in tensorflow version 1.13.1, in a Jupyter Notebook environment. We employed lubridate for manipulating dates (Grolemund & Wickham, 2011), dplyr for data manipulation and subsetting (Wickham, François, Henry, & Müller, 2020), forecast for all ARIMA-related tasks (Hyndman R. , et al., 2020) and randomForest for fitting benchmark RF model and forecasting (Liaw & Wiener, 2001).

A8: Sample Portfolios

Table A-4 Stock tickers in the 10 randomly sampled portfolios

	Sample Portfolios									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Ticker	BON	AFG	AMSC	AXA	DNO	AFG	LSG	AKSO	AKER	AMSC
	EQNR	KOA	DNO	EQNR	GOGL	AXA	SCHA	DNB	ASC	BON
	KOA	NOD	KIT	NAS	KIT	EQNR	SNI	MEDI	OLT	DNB
	NOD	PGS	KOA	SIN	OLT	KOA	SUBC	NHY	STB	KOG
	STB	YAR	NOD	STB	SCHA	MEDI	TGS	TOM	TEL	ORK

Table A-5 Portfolio variances. Derived from the correlation predictions with the Overall Mean model and the hybrid model, compared to the actual variances

	Overall Mean										
Set	1	2	3	4	5	6	7	8	9	10	Sum
Dev	0.085	0.032	0.044	0.089	0.034	0.079	0.115	0.045	0.019	0.171	0.713
Test 1	0.029	0.032	0.044	0.052	0.034	0.047	0.047	0.035	0.018	0.028	0.365
Test 2	0.031	0.034	0.046	0.054	0.035	0.049	0.048	0.037	0.019	0.029	0.380
Test 3	0.033	0.037	0.049	0.057	0.036	0.051	0.049	0.037	0.020	0.029	0.398
Sum	0.178	0.135	0.182	0.252	0.138	0.225	0.258	0.154	0.076	0.256	1.855
	Hybrid										
	1	2	3	4	5	6	7	8	9	10	Sum
Dev	0.089	0.035	0.050	0.100	0.038	0.091	0.121	0.050	0.020	0.159	0.753
Test 1	0.030	0.035	0.050	0.058	0.038	0.052	0.048	0.038	0.018	0.028	0.395
Test 2	0.032	0.037	0.051	0.060	0.038	0.053	0.049	0.039	0.019	0.030	0.407
Test 3	0.040	0.042	0.063	0.071	0.042	0.060	0.054	0.042	0.021	0.034	0.471
Sum	0.191	0.149	0.215	0.289	0.156	0.256	0.273	0.168	0.078	0.251	2.026
	Actual										
	1	2	3	4	5	6	7	8	9	10	Sum
Dev	0.101	0.028	0.058	0.104	0.037	0.094	0.104	0.044	0.015	0.147	0.732
Test 1	0.050	0.051	0.097	0.085	0.056	0.084	0.066	0.056	0.030	0.028	0.601
Test 2	0.029	0.032	0.054	0.068	0.041	0.059	0.052	0.049	0.023	0.042	0.448
Test 3	0.026	0.024	0.039	0.054	0.034	0.053	0.055	0.036	0.019	0.028	0.369
Sum	0.206	0.134	0.248	0.311	0.167	0.290	0.277	0.185	0.087	0.246	2.150