AALBORG UNIVERSITET

MASTERS PROJECT

Forecasting Danish stock prices in 2022 under macroeconomic distress using macro indicators - a deep learning approach.

Supervisor: Thibault Laurentjoye

Melani Lærke Pedersen og Mikkel Bak Lyngø

Aalborg Universitet - Danmark

23-10-2023

BUSINESS SCHOOL
AALBORG UNIVERSITY

Titelblad

Education:	Cand.Oecon
University:	Aalborg Universitet
Group number:	1
Group members:	Melani Lærke Pedersen & Mikkel Bak Lyngø
Supervisor:	Thibault Laurentjoye
Title:	Forecasting Danish stock prices in 2022 under macroeconomic
.	distress using macro indicators - a deep learning approach.
Numb. of keystrokes:	186,333 characters
Numb. of pages:	78 pages

Authors

Melani Lærke Pedersen
Study number. 20211191

Mikkel Bak Lyngø
Study number. 20184438

Abstract

This study delves into the unpredictable nature of the stock market, acknowledging the challenges posed by its volatility and instability. Numerous theories exist on forecasting stock prices, yet no universal explanation fits all scenarios due to the diverse nature of stocks. This research takes a novel approach by employing LSTM models, to predict stock prices. It examines whether incorporating external macroeconomic factors into LSTM could improve forecasting accuracy. This analysis specifically focuses on a period marked by macroeconomic distress in autumn 2022. The study explores the optimization of LSTM and LSTM-X models to effectively handle volatility in the data being forecasted. Despite the findings indicating that both LSTM and LSTM-X models did not fully meet their potential, valuable insights were gained. Discussions and arguments presented in this study propose potential adjustments for future forecast models, aiming to simplify the handling of volatility and optimize forecasting accuracy.

Table of Contents

List of Figures

List of Tables

1 Introduction

Renowned economist John Kenneth Galbraith described stock markets and their relationship to the macroeconomy in his book from 1955, The Great Crash 1929, which dealt with the aftermath of the infamous 1929 financial meltdown in the United States. In this publication, Galbraith stated the following:

> *"The stock market is but a mirror, which provides an image of the underlying or fundamental economic situation" (Galbraith, 2009, p. 88).*

This indicates that there should be a relation between the stock market and macroeconomic conditions for a given economy. The relationship between the development of the stock market and macroeconomic variables is something empirical research has not yet fully described. Economists disagree on whether changes in the economy causes stock markets to change, or changes in the stock market causes economic changes, or the two simply follow the same patterns independently of each other. What economists do agree on, however is the strong correlation between the stock market and the economy as a whole (Bosworth et al., 1975). The empirical evidence is not unanimous in its explanation of the stock market, as there is not yet a consensus on which macroeconomic factors influence the development of the stock market. For this reason, there is a natural curiosity to further investigate the factors that influence the development of the stock market in Denmark. At the same time, the development of the Danish stock market has been characterized by more volatility than usual in the year 2022, which is why identifying and studying the driving factors for this development is particularly interesting.

After a long period of low interest rates and economic growth, 2022 demonstrated signs of slowing growth, both globally and in Denmark. The reason for this is that at the start of

2022, the global economy was characterized by significant disruptions in both production and consumption patterns. The strong recovery in economic activity following the Corona pandemic, together with accommodative economic policies, gave rise to strong demand. Meanwhile supply in parts of the economy was hampered by logistics and production constraints derived from the pandemic. Furthermore, Russia's invasion of Ukraine at the end of February 2022 exacerbated these imbalances in the global economy. This was particularly evident in global commodity markets, where supply difficulties led to very high increases in energy and food prices (Callesen et al., 2023). Together, these events have naturally had an impact on economic activity in Denmark, which is reflected in the Danish OMXC25 index, see figure 1.

Figure 1: OMXC25 stock prices in 2022.

The OMXC25 index shows that 2022 has been relatively more volatile compared to previous years. In the first half of 2022, a sharp drop in the stock market is observed, while the market has been more volatile in the second half of the year. Thus, the market in 2022 has been very uncertain, and it is clear that the OMXC25 reflects the economic disruptions that have affected both domestic and international markets.

For many, the stock market is unpredictable and incomprehensible, which has prompted a great deal of research to investigate the correlation between different macroeconomic variables and the volatility of stock prices. Several studies have investigated how stock forecasting through systematic patterns can provide investors with competitive advantages and the ability to earn above-normal profits. This contradicts the so-called 'efficient market hypothesis', which states that it is impossible to predict the future price of an asset based on the information contained in the historical prices of that asset (Thune, K. (2021)). This is supported by the 'random walk hypothesis' which states that it is not possible to forecast financial assest due to the randomness of stock fluctuations. However, several studies contradict these hypotheses about financial markets, which is assumed for this thesis.

In recent times the interest in social data science has increased significantly, which is an ever-growing field within technological science, and is continuously evolving as a predominant aspect of future forecasting methods. As technology advances, the forecasting discipline is becoming increasingly more comprehensive. The reason for this is that deep learning allows for far more sophisticated algorithms and therefore enables more nuanced and complex forecasting. It has been found interesting to further investigate how the development of deep learning models opens up new and better performing models for forecasting. A number of studies in the literature finds that deep learning models provides more accurate predictions of the stock markets relative to more naive statistical models. In particular, long short-term memory (LSTM) models are cited as having the ability to predict fluctuations in stock prices more accurately. It is additionally important to emphasize that among forecasting models, there is no consensus on one particular model that performs best in all markets in the literature.

Therefore, this thesis will contribute to the application of deep learning models in forecasting stock prices within the OMXC25. The aim of this thesis is to develop stock forecasting models utilizing the advantages of LSTM networks to predict the stock price of tomorrow as accurately possible. Furthermore, this thesis will examine the inclusion of external variables to investigate whether information from macroeconomic indicators can improve the accuracy of forecasting. This information is utilized to improve the forecast of the highly volatile period of autumn 2022 in the Danish economy.

1.1 Problem definition

There exists extensive literature on the forecasting of the stock market, and recent publications seem to agree that systematic patterns are present in stock prices, which can be forecasted. However, there is a lack of consensus of whether external features may improve the accuracy of these. Incorporating external variables for a period characterized by high volatility, could potentially feed information to forecasting models, which is why this thesis aims to investigate the following:

Does inclusion of macroeconomic features increase the performance of LSTM models in forecasting Danish stock prices? How is this applied to a period of macroeconomic distress during 2022?

Given the problem statement, this thesis will therefore investigate whether the LSTM models for the individual stocks perform better with or without these external features. The reasoning behind exploring the idea of adding external variables for a very volatile period, is that simple models have evidently already shown great results in academia on forecasting stock prices in more stable periods. However, when large volatility in the underlying structure of the data appears, external features could potentially feed information to the forecasting models optimizing their accuracy. Therefore the forecasting period is a period characterized by macroeconomic distress which is much reflected in autumn 2022. In this context, there is only a limited amount of literature that relates to this in a Danish perspective, which is why this project endeavors to make a contribution within this area.

1.2 Scope definition

In order to answer the questions in the problem statement and to achieve an adequate and in-depth thesis, it has been necessary to make some restrictive choices. The literature review in this thesis provides an overview of the academic literature that concerns the use of deep learning, and particularly LSTM models, for forecasting stock prices. Including reviews on the impact of macroeconomic variables on forecasting outcomes. The selected stocks are from the OMXC25 index which reflects the development of the most prominent Danish stock market. In addition, the project considers a time span from primo 2018 to ultimo 2022.The reason for this particular amount of data is mainly based on the literature review, as evidence suggest that five years of daily data provides enough information for sufficient forecasting of future values (Kumbure et al., 2022).

In this thesis, the assumption is that investors want information about the closing price tomorrow, i.e. day-ahead forecasts. This assumption is made as this allows the investor to make adjustments to his positions in the market on a daily basis. Thus, the forecasting models in this project are not intended a role for more advanced high-frequency day-trading, as these models will require a much larger amount of data and higher frequency data.

It is essential that the stock market in the year 2022 is the focal point in this thesis, as especially the fall of 2022 contains high volatility in stock prices. Therefore, autumn 2022 will be the period for forecasting, as this study emphasizes the ability of forecasting models to forecast during periods of large fluctuations. Thus, the models in this project include outliers for a more realistic approach to forecasting. This should enable the models to capture large future fluctuations by learning from previous outliers in stock prices.

As this thesis aims to optimize forecasting of Danish stock prices using macroeconomic variables, the data also contains information about selected macroeconomic features. The reason for including external macro variables in the study is that macroeconomics and financial markets are closely linked. With major turmoil in the macroeconomy in 2022, it is therefore considered interesting to investigate how macro factors can help forecast stock prices. Among the selected macro variables, not all are equally relevant to the different stocks. Therefore, feature selection is performed to identify the most significant macro factors for each of the dependent variables. Not all possible macro variables are included in the study, with only the most relevant being included in the feature selection process. At the same time, only variables that occur in daily frequency are included, as it is considered that interpolation is not ideal in this context. In this thesis, a delimitation occurs in terms of not including all possible types of variables, such as company-specific factors, technical indicators or others to the analysis. These factors are not discussed, as this is outside the scope of the project. However, it is acknowledged that these factors have an impact on stock fluctuations. In relation to this, training time of the forecasting models is a limitation in terms of processing power of the available computers.

The accuracy of the forecast are considered the most essential in this thesis, which is why there is a distinction from commenting on an investor's financial gain by using this forecast. The reason for this is that financial gains are also tied to risk level, transaction costs and various other parameters an investor considers in his investment strategy. Indirectly, this project also deals with the efficient market hypothesis. The weak form of the efficient market hypothesis states that it is impossible to forecast the future price of an asset based on the information contained in the historical prices of an asset. This means that the market behaves as a random walk and as a result makes forecasting impossible (Thune, K., 2021). This thesis assumes that these hypotheses does not hold, hence the models presented should be able for forecast the future prices.

Restrictions also occur in the form of methods, where a lot of different tests and models are used in the existing literature, each of which has its strengths and weaknesses. In this thesis, the choice of methods is to some extent based on availability, where the primary knowledge about LSTM models is acquired through the academic content from the courses from the Master's program in Economics at Aalborg University. The forecasting models are characterized as being one-step ahead models. The choice of this method is due to the fact that this method appears more accurate, relative to multi-step ahead, in terms of forecasting changes in daily stock prices. The reason for this is that multi-step ahead forecasts often converge towards a long-term equilibrium and thereby forecast a long-term trend rather than daily fluctuations. In addition to the academic content of the Master's program, the LSTM models have been expanded to include external variables.

1.3 Project structure

The introduction, problem statement and the scope of the project creates a foundation for the following project structure, consisting of 10 sections.

Section 2 is a statement of the Danish stock market and the development in commodity prices in 2022. In addition, a description of how different sectors in the Danish economy have been affected differently in 2022.

Section 3 is a detailed review of the theoretical foundation of the forecasting models applied in this project, including an introduction to the evaluation metrics used in this thesis.

Section 4 is a literature review, that clarifies the existing literature on the subject of financial stock forecasting. This gives an insight into the state-of-art methods within stock prediction, and provides guidance to the choices made in this thesis.

Section 5 describes the data used in this thesis, in the form of descriptive statistics. Furthermore, the properties of the selected stocks are investigated individually and the selected external variables are introduced.

Section 6 presents the methods, approaches and decisions made for this thesis. Describing the general workflow, and provides insights into the architectural choice's and limitations made for the deep learning models. Furthermore, the handling of varied training data lengths and external variables are described.

Section 7 presents the selection and results of the ARIMA, LSTM and LSTM-X forecasting models. Additionally, analysing the impact of different training data lengths on forecasts and presenting the combinations of external variables included in the LSTM-X models.

Section 8 compares the different forecasting models performance using evaluation metrics, Diebold-Mariano test and Pesaran-Timmermann test.

Section 9 discusses some of the results from the analysis and the application of such forecasts. It also provides a perspective for further study.

Section 10 summarizes the project's conclusions, as a final answer to the problem statement.

2 The stock market and macroeconomy in 2022

2.1 The Danish stock market in 2022

After years of soaring stock markets, investors around the world have faced significant turbulence in 2022. Since the beginning of 2020, the global economy has been hit by two main collective shocks that have caused major disruptions to the international economic system. First in the form of the Corona pandemic and then in the form of Russia's invasion of Ukraine. Both events can be characterized as negative supply shocks and therefore, in isolation, pull in the direction of higher prices and lower economic activity. Economic activity and growth are an interaction between a wide range of factors, and in an increasingly globalized world, what happens in the major economies or countries that have access to important raw materials, such as oil and natural gas, is of great significance. Commodity prices affect supply chains in large parts of the global economy and can therefore be important indicators of price increases, and thus the outlook for economic activity for Danish companies. These supply shocks resulted in significant bottlenecks in international supply chains, and the mismatch between supply and demand was also exacerbated by accommodative monetary and fiscal policies across countries, which supported demand despite supply constraints. In particular, Russia's unexpected invasion of Ukraine in February 2022 led to significant price increases, especially in European energy prices and in food prices where Russia and Ukraine are important exporters, such as natural gas and grains. The price increases were also exacerbated by the political reactions from both Western countries and Russia, which led to a significant reduction in the import of Russian energy to Europe, creating a risk that European countries would be short of energy for the winter of 2022 (DORS, 2022, p. 48).

It is essential that economic growth is the foundation of stock markets, which is why changes in growth expectations can create volatility in stock prices. The economic imbalances in the global economy in 2022 caused Danish stock prices measured by the OMXC25 index to decrease by a total of 13.5 percent for the entire year, where large fluctuations in the index were observed throughout the year, (Callesen et al., 2023, s. 9). Due to these fluctuations, the stock market in 2022 can therefore be characterized as a so-called Bear Market, which is a market condition that occurs when the market falls by 20 percent or more in a given period of the year, which was observed four times during 2022. Especially in the first half of 2022, stock markets fell sharply, while in the second half of the year, the market has moved up and down without a clear trend. Derived from these conditions, a phenomenon called Bear Market Rally was also observed, described as a short-term counter-reaction to the general market condition (Jyske Bank, 2022). Here, investors no longer believe in a market downturn and start buying again, causing the price to rise briefly. As inflation and interest rates continued to rise, this discouraged investor expectations, causing the price to decrease again. This naturally creates a state of volatility in the stock markets, where the market, especially in the autumn, was very uncertain and subject to significant price fluctuations in both directions.

For investors, decisions to buy or sell stocks can be described as being driven by market psychology, as expectations for the future are driven by this (BankInvest, 2022). This is where the complexity and unpredictability of the market originates. This aspect of stock trading is about people and how investors assess the risk of investing. Here, optimism and confidence in the markets have a significant impact on how willing investors are to trade stock prices at higher levels and at a more expensive valuation relative to expected earnings and interest rates. Conversely, pessimism and risk aversion create more pressure on stock prices. Investors' worsening expectations for growth prospects, especially in the autumn of 2022, reflected that downgraded expectations for international growth would further impact the Danish economy. This was in response to household purchasing power being eroded over the course of the year as a result of higher inflation, which increased by a total of 8.5 percent in 2022. In isolation, consumer prices in October 2022 increased by 11.4 percent year-on-year, being the highest rate of increase observed in 40 years (source, NB2023, page 10). This caused further monetary tightening during the autumn to put a damper on the economy. Together with a drop in real wages for 2022 as a whole, this meant that the consumer confidence reading in the autumn of 2022 decreased to its lowest level in 48 years (Callesen et al., 2023, p. 10). Overall, these aspects driving the expectations naturally put negative pressure on domestic demand, which for investors downgrade expectations. These downgraded expectations was reflected in the Danish stock market, especially in more cyclical stocks.

The high inflation, and the monetary policy counter-reactions, led to large fluctuations in the financial markets in 2022, with this being particularly pronounced in the bond markets. The 10-year Danish government bond increased by 2.7 percentage points. The rate on 30-year fixed-rate convertible mortgage bonds increased from around 1.8 percent in January to around 4.8 percent at the end of December. Similar increase could be observed on short-term mortgage bonds (Callesen et al., 2023, p. 9). This reflected the increase in government bond yields and growing risk premiums. As a result, financing conditions for Danish households and businesses tightened significantly during 2022, both in nominal and real terms. The Danish central bank's monetary policy interest rates has an effect on the financial markets and the real economy. This is a crucial relationship, which makes it possible to use the interest rate on government bonds as a proxy variable for the development in inflation and monetary policy interest rates. This is a central assumption, as this thesis aims to investigate the effect of macroeconomic variables, and thus also inflation and interest rates, when forecasting Danish stock prices.

2.2 The impact of commodity prices in 2022

A central part of the research in this thesis discusses the effect of developments in macroeconomic variables for forecasting Danish stock prices. As most of the macroeconomic uncertainty during 2022 primarily originates from commodity price shocks and monetary policy counter-reactions, it is essential to investigate these factors in more detail. Empirical studies discussing the impact on commodity prices, in the form of energy price increases on consumer prices, indicates that the direct effect, i.e. the impact on consumer energy prices, is generally high and rapid (DORS, 2022, p. 58). This is especially true for changes in the oil price, where the price change is typically passed on to the consumer price within about a month. However, changes in the natural gas price have a slightly slower impact and typically take three to nine months. The reason why changes in the price of natural gas are slower to affect consumer prices may be that the natural gas market has different supply and demand dynamics compared to the oil market. Here, natural gas supply chains, especially in countries dependent on imports, such as Denmark, can involve multiple intermediaries and infrastructure. Thus, these aspects can cause delays in price adjustments. The indirect effect on non-energy consumer prices, such as producer prices, is typically more gradual (DORS, 2022, p. 58), as energy is only a limited part of the overall value chain for companies. This is consistent with the fact that the fluctuations in consumer prices and OMXC25 only became apparent in the autumn of 2022, several months after Russia's invasion of Ukraine. For these reasons, forecasting in the autumn of 2022 is of particular interest, as it will be possible to examine the ability of the forecasting models to predict the fluctuations in the stock market, mainly caused by the occurrence of these macroeconomic disturbances.

2.3 The development in relevant sectors

The macroeconomic imbalances in 2022 have affected Danish companies in different sectors of the economy have differently. Therefore, this thesis will investigate how adding external macroeconomic factors in forecasting models can optimize predictions about the development of four different Danish stocks from different sectors of the economy. All companies in the world stock market belong to one of a total of 11 sectors. This division of sectors is based on the Global Industry Classification Standard, GICS (Voergaard, P., 2020). It is often the case that some sectors do well while others do less well. Spreading the stocks across several sectors ensures that the survey is not overly exposed to the development in one sector, even though the economic barometers indicate that the weakening of growth in the fourth quarter of 2022 was noticeable across sectors (ECB, 2023).

The first of the selected stocks is Royal UNIBREW, which belongs to the stable consumer goods sector. Royal UNIBREW produces and sells alcoholic and non-alcoholic beverages, including beer, malt beverages, cider, soft drinks, juices and water (Euroinvestor, 2023). In 2022 the economy experienced negative supply shocks, with commodity price increases impacting the prices that consumers face both directly and indirectly, as companies pass on increasing prices to consumers. In addition, there has been a decline in real wages in 2022, where evidence shows that households reduce their consumption quota when they are pessimistic about the development of economic activity (DORS, 2022, p. 98). Since the company primarily sells stable consumer goods, which consumers are less likely to do without even in times of crisis, it can be argued that the development of the stock price is thus less affected by cyclical fluctuations. Conversely, it can be argued that the price of beer and soft drinks is not perfectly inelastic, which is why increased production costs have had an impact on the company's dividends. Thus, it is to be expected that the addition of macroeconomic factors that drive the development in consumer prices and production costs can to some extent be optimizing in forecasting the stock price of Royal UNIBREW, as this sector is relatively less cyclical.

The second of the selected stocks is Novo Nordisk, which belongs to the healthcare sector. With around a third of the global branded diabetes care market, Novo Nordisk is the leading supplier of diabetes products in the world. The company manufactures and markets a range of human and modern insulin, injectable diabetes treatments, oral antidiabetic agents and obesity treatments (Euroinvestor, 2023). In 2022, Danish exports benefited from the recovery in export markets, which supported growth in Denmark as the economic turnaround began to appear domestically. The increase in goods exports was particularly supported by exports of medicines, which are less cyclical (Callesen et al., 2023, p. 10-11). In addition, the price of medicine can generally be argued to be very elastic, as demand will remain unchanged to some extent. It is therefore to be expected that the addition of macroeconomic factors, that drive the development of prices for the production of medicines, may be less optimizing in forecasting the stock price of Novo Nordisk, as this sector is less cyclical.

The third of the selected stocks is Jyske Bank, which belongs to the financial sector. Jyske Bank is engaged in banking, mortgage and leasing activities. Its subsidiaries provide other financial or ancillary services (Euroinvestor, 2023). Higher interest rates in 2022 further boosted net interest income for banks, increasing the institutions' dividends and resilience against loan losses. Conversely, high inflation and the prospect of slowing growth worsened the ability of some bank customers to service their debt, which poses a risk of write-downs for institutions. In this context, it can be observed that bank lending to corporate customers has increased significantly in 2022, which to some extent reflects the fact that corporate customers have needed ongoing funding (Callesen et al., 2023, p. 15). This is partly due to challenges in supply chains as well as high commodity and energy prices, which could indicate that corporate customers have been burdened by the macroeconomic imbalances.

It is therefore to be expected that adding macroeconomic factors reflecting interest rates and the customers' ability to raise and service their debt can be optimizing in forecasting the share price of Jyske Bank.

The fourth of the selected stocks is Rockwool, which belongs to the industrial sector. Rockwool primarily manufactures and sells building materials, including insulation and roofing systems (Euroinvestor, 2023). The development of industrial production, which is closely linked to exports, performed well in 2022. However, industrial production declined at the end of 2022, mainly in the energy-sensitive industries, in which Rockwool is included, despite the reduction of supply bottlenecks (ECB, 2023). In addition, the sharp increases in interest rates during 2022 made it significantly more expensive to buy and borrow for construction and renovation of real estate. It is therefore to be expected that the addition of macroeconomic factors reflecting price increases in industrial production and interest rates, in relation to real estate investments, can be optimizing in forecasting the stock price of Rockwool.

Therefore, it is interesting to investigate how macroeconomic factors can help improve forecasts of the development of Danish companies' stock prices, linked to different sectors of the economy, in a period of large fluctuations in the macroeconomy.

3 Theoretical foundation

The following section aims to clarify the theoretical aspects of both ARIMA and LSTM models. First, a brief introductory description of ARIMA models is given, followed by a description of the Engle-Granger test for cointegration and the Cross-Correlation function for multicollinearity used for external variable selection. Further, the concepts and the dynamics of deep learning models are described, with a special attention on describing Long-short-term-memory networks. Concluding with the theoretical explanations of the used metrics and test for evaluation.

3.1 Econometric Foundation

In this project, the Autoregressive Integrated Moving Average (ARIMA) models serves as the foundational predictive tool (or baseline model) for stock price forecasts. This model relies on autoregressive processes (AR), reflecting a series' dependence on its prior values, and moving average processes (MA), denoting its dependency on past error terms. The 'I' denotes the number of differences needed to make the time series stationary, which involves subtracting the previous value from the current value to achieve stationarity, called integration. The integrated ARMA(p,q) model is thus called ARIMA(p, d, q) (Prabhakaran, 2019). It is essential for ARIMA models that time series data is stationary for accurate forecasting and reliable results. To determine if a time series is stationary, the Augmented Dickey-Fuller (ADF) test is applied to determine whether a unit root is present in the dataset. If such a unit root is present the time series is non-stationary and the time series can exhibit trends or seasonality, which is not ideal for model predictions. Furthermore, the ARIMA models are estimated based on the Box Jenkins methodology (Newbold, 1975).

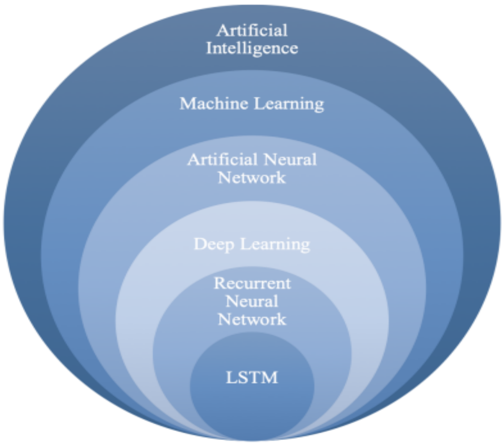
This ADF test is also used in the initial part of the Engle-Granger (EG) test, as this test assumes that all variables must be integrated of the same order to be included, in this case I(1) for all variables. If this is not the case, OLS estimates for such a model will converge faster or they might follow the same trend (Enders, 2017). The EG test is used to analyze the long-term relationship between the development of stock prices and selected macroeconomic variables, later to be applied in the LSTM-X models. This test implies cointegration between the two non-stationary time series, which means that a stable long-run relation exists between them, even in the case of short-term fluctuations. This thesis values this relation as it allows for understanding the underlying economic or financial dynamics that exists between stock prices and the external variables. However, it is acknowledged that cointegration only suggests a statistical long-run association and does not provide insights about the causality between these variables.

In addition, Cross-Correlation function (CCF) has been performed among all the external variables, in order to account for possible multicollinearities (Enders, 2017). This refers to the situation of high correlation between two or more predictor variables, thus one variable can be linearly predicted from lagged values of the others. By addressing multicollinearity, it is possible to streamline the models to only include the most significant external variables. This multicollinearity test states, that as the correlation coefficient value between the variables approach 1 (or -1), there is an increasing linear relation. The values 0.8 or 0.9 are often associated with "high" correlation, however, these are not strict boundaries derived from a theoretical rationale, rather rule-of-thumb based values used in practice. In this thesis, the boundary of 0.8 is applied, as correlations above this value can provide similar predictive information in the LSTM-X forecasting models. Once again, high correlation does not mean that a causal relation exists.

3.2 AI and deep learning models (introduction to this field)

The following subsection aims to clarify the theoretical aspects for deep learning models, including Long Short Term Memory models, LSTM. This is done to give the reader a better starting point for understanding the forecasting model used in this project. The LSTM-model (Long-short term memory) was invented by Sepp Hochreiter and Jürgen Schmidhuber (Hochreiter & Schmidhuber, 1997). Figure 2 provides an overall illustration of the structure of artificial intelligence, AI, showing where LSTM originates. This also shows the aspects that will be explained in more detail in this section.

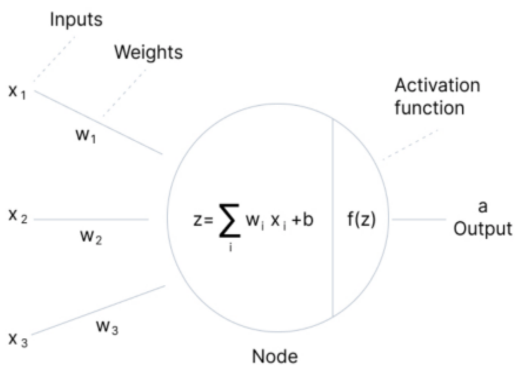
Figure 2: Subcategories of Artificial Intelligence.



It is argued that deep learning techniques are particularly suitable for stock market prediction because many factors are considered to influence stock prices and often in complex and non-linear contexts. The reason for this is that these models use artificial neural networks to capture correlations and learn from large amounts of data. AI refers to the simulation of human intelligence in machines, which are programmed to mimic complex functions associated with the human mind, including rationalization, learning, and prediction (Su&Yang, 2022). Within AI there is the subcategory of machine learning, ML, which is characterized by the ability of algorithms to learn patterns and make decisions from data without being explicitly programmed. Basically, ML models aim to find statistical regularities to enhance the model's capabilities through experience (Ayodele, 2010). One of the more widely used methods in ML for classifying or constructing regression models for predicting numerical values is supervised machine learning, SML. Here, the algorithm generates a function that maps input to the desired output (Delua, 2021).

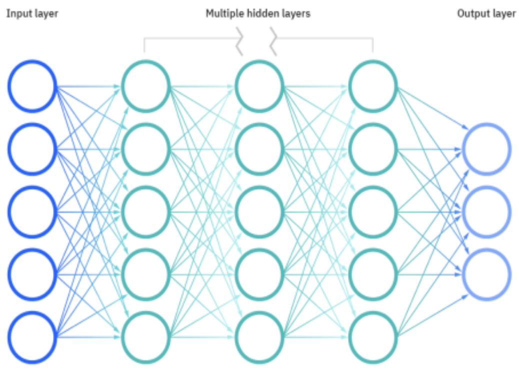
By moving closer to the center in figure 2, the subcategory artificial neural networks, ANN, is found. A special property of a neural network (ANN) is that such an estimation model is able to automatically learn and recognize from the input data. This is done by the network being able to change its parameters through training to achieve more accurate predictions. For feed-forward networks, this training process occurs through the use of backpropagation, which will be described in more detail later. ANN models are fundamentally inspired by the way biological neurons in humans signal to each other (Chen, 2021). The transformations of the models take place in hidden layers and is considered non-transparent, which makes it more difficult to map the process and dynamics from input to output, revealing complex and non-linear relationships. In this case, an ANN model can create many different weights and combinations between layers in the effort to achieve a forecast output, as illustrated in figure 3.

Figure 3: Nodes of a neural network (Baheti, 2022).



Here, the weights assigned to each input by the model determine how much that input contributes to the output. These weighted inputs are then summed and a bias term, b, is added, as shown in the equation in figure 3. The node also displays an activation function, which introduces nonlinearities into the neural network, allowing the model to learn complex patterns. Thus, in ANN models, it is the model's adjustment of these input weights that is the key aspect (IBM Cloud Education, 2020). In this thesis, deep learning is used to classify patterns in data using neural networks with three or more hidden layers, as illustrated in figure 4.

Figure 4: An example of a deep network (Kavlakoglu, 2020).



The circles illustrates the different layers and nodes, and how these can appear more complex in deep learning. It appears that each hidden layer, representing a number of neurons which is associated with weights and inputs (IBM Cloud Education, 2020).

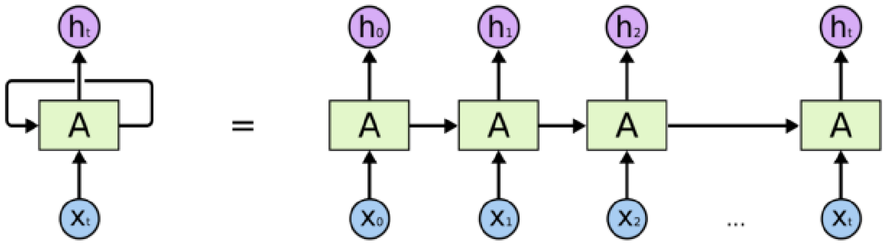
3.3 Model architecture

In this subsection, the architecture of the LSTM model is specified, which relates to the specific connectivity of the cells within the model, which are the actual building blocks of the networks of this thesis.

3.3.1 Recurrent Neural Network

Within deep learning, there is a subcategory of recurrent neural networks, RNN, which attempt to incorporate the way in which the human brain processes sequential data. RNN models are more dynamic and have the ability to remember information from previous observations. The structure of RNN is illustrated in figure 5. Here, input, x goes into a cell denoted as A, which subsequently creates an output h.

Figure 5: XX Recurrent Neural Network (Olah, 2015).

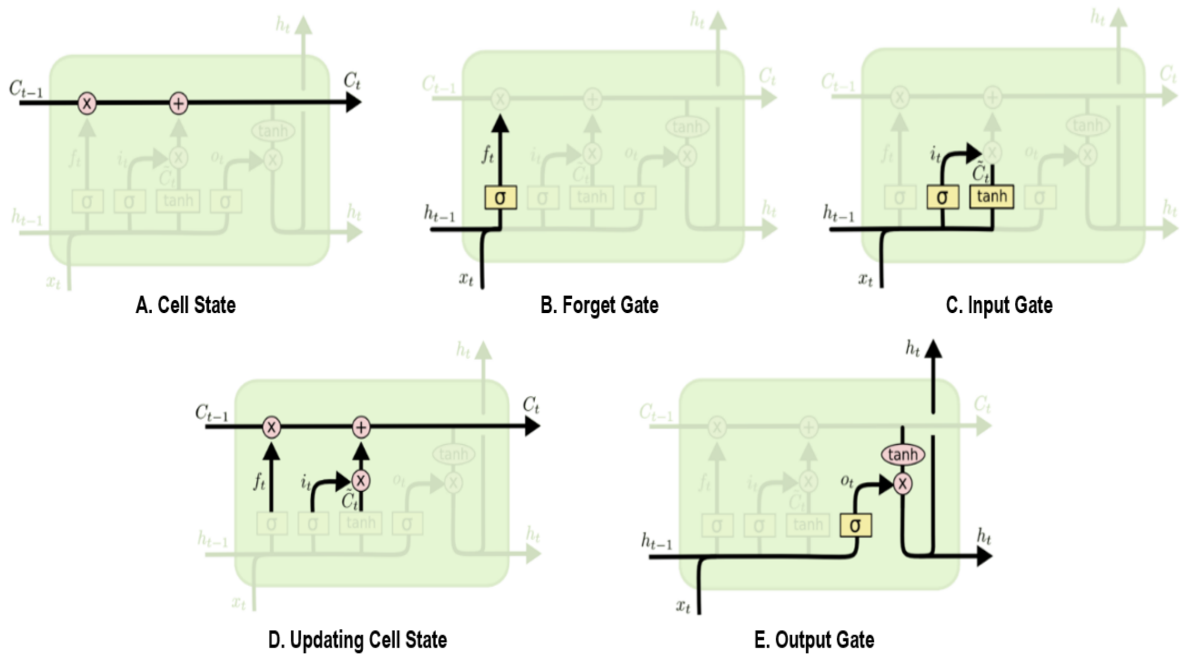


In figure 5, the left side illustrates a recurrent cell with a loop representing the information being passed trough the cell multiple times before converted to h_t. On the right side this recurrent cell are unrolled to present the longer informational processing through one single cell. This creates a possibility for the cells to reevaluate past information, giving it a memory to better learn complexities within data. Hence, past sequences will have a decisive impact on future output (Olah, 2015). Thus, the essence of RNN models is that these can incorporate information over time. On the other hand, there are some difficulties with the simple RNN model when this time interval becomes sufficiently large. This is called long-term dependency difficulties and is an important reason why this project uses LSTM models for forecasting, which are described in more detail in the next subsection.

3.3.2 Long Short Term Memory

The core essence of LSTM (Long Short-Term Memory) is encapsulated by the cell state, which enables the LSTM to learn from both short-term and long-term information, thus designing the model to avoid long-term dependency problems. The cell state, denoted as C, is illustrated in figure 6.A and can be described as a conveyor belt running through the entire chain, engaging in some minor linear interactions (Olah, 2015).

Figure 6: The information processing within LSTM (Olah, 2015).



Through this conveyor belt, LSTM has the ability to remove or add information along the way, where most gates for this consist of a so-called sigmoid layer, $sigma$. In addition, there is a gate in the form of a point multiplication operation, $times$. A sigmoid layer can take a value in the range [0;1] and indicates how much information should pass through this specific gate. Here, a sigmoid value of zero indicates that no information is passed through, whereas the value one means that all information is relevant and therefore sent through the gate.

The first step in LSTM is to decide which information is selectively removed from C_t. This initial process takes place in the so-called *forget gate*, which is illustrated in figure 6.B. Each hidden layer receives information from the previous layer, C_{t-1} previous output, h_{t-1}, and the new input, x_t. This information is channeled through various gates, each processing the information in its unique way. The information added to C_t from the forget gate, f_t is given by the following:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$$

Here, W_f denotes a weight matrix, to which b_f denotes a bias vector, which is continuously updated according to new information. These both act as an additional set of weights (DeepAI, n.d.). Before sending the processed information to C_t, this is multiplied by $sigma$, which indicates the proportion of still relevant information after this first interaction.

The next interaction in the LSTM network describes which of the new information should be sent to C_t, called the *input gate*. Two actions are taken, illustrated in figure 6.C. The first action determines which part of the information should be updated i_t. The second action, called the *than layer* constructs a vector consisting of new possible values denoted $\tilde{C}_t$. These two interactions are described by the following formulas:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = tanh(W_c[h_{t-1}, x_t] + b_c)$$

From the above, it is evident that the construction of the formulas is identical to that of the forget gate, where the respective weight matrices, are multiplied, and the respective bias vectors are added for each of these processes. The degree of information that is carried forward is calculated by multiplying with $sigma$ for the input gate layer and the $tanh$ function in the tanh layer. Finally, the input gate layer and tanh layer are combined as illustrated in figure 6.D, thereby updating C_t by the following formula:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t$$

The formula indicates how the new possible values, which are scaled according to how much the model updates each value based on the total information. The last interaction in the LSTM network determines what the layers output will be, based on the cell state. In this, the information is filtered so that the output contains only the selected information from the cell state, as illustrated in figure 6.E. Here, a sigmoid layer, σ, is used to determine the proportion of information to be used in the output, which is given by the following:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_0)$$

Then the information from the current cell state, C_t, is passed through the tanh function, $tanh$. Thus, the information is converted to values in the interval [-1;1], which can then be multiplied by the information from the sigmoid gate, o_i, which is given by the following:

$$h_t = o_t tanh(C_t)$$

Finally, the final output of h_t for each sequence is given. In addition, C_t indicates the processed information that is subsequently transferred to the hidden layer in the next sequence, after which the whole process repeats. This clarifies all operations associated with a time step in the hidden neuron in the LSTM network is a sequence-to-sequence model. Thus, there will only be one output after the last layer, and by constantly filtering out unnecessary information, the LSTM model avoids difficulties with vanishing or exploding gradients to a greater extent. This allows weights to be changed correctly when training this network in the model.

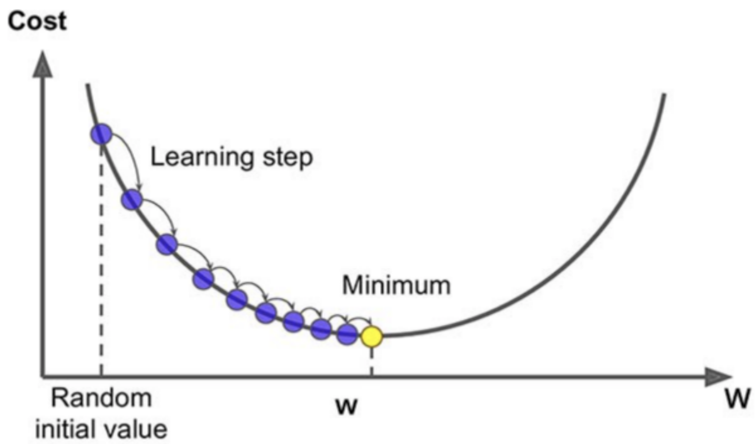
3.4 How does deep learning work?

It is essential that neural networks undergo a training process to update the model's parameters. This is done to increase the accuracy of the model's predictions. The training is mainly done by using an interplay between two algorithms in the form of a stochastic gradient descent algorithm and a backpropagation algorithm.

3.4.1 Backpropagation

Backpropagation, which is short for backward propagation of errors, is used in the LSTM models to train the models to better learn patterns in the sequential data. This is a feature that allows the model to make adjustments to its internal parameters, such as weights, by minimizing the difference between its predicted output and the actual values in the data. Basically, backpropagation works by processing inputs at each time step, thus updating internal states of the network. It then generates an output sequence consisting of predictions that are similar to the input sequence. Here, a loss function (such as RMSE or MAE) is used to quantify the difference between these predicted outputs and the actual values, providing a so-called error value to represent the distance between these sequences. This error value sequence is sent back through the network one time step at a time (IBM Cloud Education, 2020). This latter operation is called backpropagation through time (BPTT). In addition, gradients are generated at each of these time steps, showing how much the error value changes by changing the LSTM parameters. These gradients are accumulated for the entire sequence and subsequently used in an optimization algorithm, which in this project is a stochastic gradient descent algorithm. This aims to adapt the LSTM parameters so that the error value is reduced for the entire sequence. This action of reducing the error value is repeated until the model converges, or the number of iterations given the selected number of epochs. This backpropagation continues until the LSTM model is able to produce accurate predictions, as illustrated in figure 7.

Figure 7: XX Rectified Linear Unit funktionen (Amidi & Amidi, 2019).



Through backpropagation, the initial value of the loss function approximates the global minimum. At this point, the stochastic gradient descent algorithm has found internal parameters in the model that minimize the loss function (Gudimalla, 2021).

3.4.2 Stochastic gradient

It is essential that the application of neural networks is based on the assumption of being able to construct models that are able to produce predictions that approximate the actual values in the test data. To achieve such an approximation of the actual values, the optimization algorithm plays an essential role. This aims to minimize a given loss function $C(w, b)$ as a function of weights, w, and bias, b (Nielsen, 2019). In order to make use of this stochastic gradient descent algorithm, it is necessary to first apply backpropagation, as described earlier, to calculate the gradient of the loss function. This thesis uses a stochastic gradient descent algorithm for this subsequent minimization. In general, compared to a traditional gradient descent algorithm, this algorithm adds a stochastic or random aspect to the training of the network. Here, each training of this network will be assigned a gradient of the loss function to update the internal parameters of the model. These are therefore not updated with an average gradient for the entire training process. This action continues until convergence is reached or the selected number of epochs is reached. An important aspect of achieving the global minimum is the chosen hyperparameters, see section 6, where an excessive learning rate can cause the algorithm to overshoot this minimum. The advantages of this algorithm are that the network is better able to avoid local minima where the gradient is zero. Since this may not be the global minimum, which is the actual goal of the optimization.

3.4.3 Vanishing gradient problem

When backpropagation occurs over many stages during training, the gradients, i.e. derivatives of the loss function with respect to the model parameters, tend to vanish to very small values. These vanishing gradients have the effect of slowing and reducing the network's learning, causing the loss function weights to be inadequately updated and therefore cannot be minimized. This is particularly relevant for the network in LSTM models, as these are constructed from gates that enable the model to handle and learn long-term relationships in the data. Specifically, this happens for activation functions such as sigmoid or tanh, where the gradients are small for large input sequences. As these gradients multiply during back propagation, they tend to decay exponentially through the layers of the network. To accommodate this vanishing learning of potentially relevant information, a activation function such as ReLu [1], which does not suffer from vanishing gradient since the gradients are

[1] A description of activation functions and ReLU are found in Appendix A 12.1.1.

held positive, can be used for this purpose.

3.5 Evaluation methods

In this subsection, evaluation methods are presented, which are used to compare, evaluate and discuss the selected LSTM models. Comparing the results obtained with these proposed models is the final part of the analysis. To be able to do this, it is essential to choose a metric that is convenient in time-series data to obtain a fair comparison between the different models. Therefore, the most prominent evaluation methods from the literature are used.

3.5.1 Mean absolute error & Root mean squared error

There are different valid metrics for comparing the performance of different forecasting models. The most essential would be the root mean square error (RMSE), where larger errors have a disproportionately large effect on RMSE. The reason for this is that squaring causes larger errors to be given a greater weight. This means that RMSE is very sensitive and able to handle potential outliers, and the general comparison stipulates that a lower RMSE is better than a higher one (Hami & Pougnet, 2015).

The root mean squared error, RMSE indicates the difference between estimated, $\hat{y}_t$, and actual values, y_t, for the models in order to measure the forecast models accuracy. This evaluation metric can be expressed as the following:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}$$

The mean absolute error (MAE) is another metric that can be used to evaluate, which also indicate the difference between estimated, $\hat{y}_t$, and actual values, y_t. However, this averages the absolute differences between the actual and predicted values, and can thus be expressed as the following:

$$MAE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mid y_t - \hat{y}_t \mid}$$

In this case, the errors are scaled linearly, where all errors are weighted equally (Kampakis, 2020). As with the RMSE, a larger MAE is not preferred, as it indicates larger variance in the errors.

3.5.2 Diebold-Mariano test

The Diebold-Mariano test is used to compare the forecast accuracy of two or more forecasting models, and it tests whether a given forecast is significantly more accurate relative to another in terms of making accurate predictions (Zaiontz, 2021). The test uses a loss function to calculate forecast errors, which is given by the following:

$$\bar{d} = \frac{1}{H}\sum_{i=1}^{H}[g(e_{1i}) - g(e_{2i})]$$

The Diebold-Mariano statistic is given below, in which the method uses j-step ahead of forecasts $g(e_{1i})$ and $g(e_{2i})$:

$$DM = \frac{\bar{d}}{(\gamma_0 + 2\gamma_1 +, ..., +2\gamma_t)/(H + 1 - 2_j + H^{-1}j(j-1))}$$

The test is performed for the null hypothesis given by $\bar{d} = 0$, which states that the models have equal predictive accuracy and thus there is no significant difference between the forecast errors of the different models (Mcleman-Hasselgaard & Nielsen, 2017).

The alternative hypothesis states that one model is superior to the others. In this case, several alternative hypotheses are stated. The first states that forecasts are significantly different from each other, the second states that one forecast appears significantly more accurate relative to another, whereas the last states that one forecast appears significantly less accurate relative to another. The test thus provides a way to identify the model among several models, in terms of forecasting performance.

3.5.3 Pesaran-Timmermann test

The Pesaran-Timmermann test is a statistical evaluation method specifically suited for assessing the ability of forecasting models to capture directions and trends in stock prices, i.e. increases or decreases in prices. Thus, the test does not assess the models' ability to predict the actual size of these price changes or the size of forecast errors (Pesaran & Timmermann, 1992). The test for a given time series, y_t, and a forecast of this time series given by, x_t, is defined as follows:

$$S_n = \frac{\hat{p} - \hat{p_*}}{[v\hat{a}r(\hat{p} - v\hat{a}r(\hat{p_*})]^{1/2}}$$

Here $\hat{p_*}$ indicates the probability that x_t and y_t are both greater than zero summed with the probability that they are not. In addition, $\hat{p}$ indicates the proportion of times the actual and predicted values have the same direction. Here $\hat{p}$ is in the interval [0;1], where $\hat{p} = 1$ implies that all forecasts are of the correct direction. In addition, the null hypothesis states that x_t and y_t are independently distributed, i.e. x_t is not able to forecast y_t (Pesaran et al., 1992).

4 Literature review

This section carries out a review of existing literature on the methodologies and outcomes related to stock price prediction, and the role of LSTM networks within this field. This section aims to outline the academic context, discussing key contributions and findings of prior works, uncovering trends and findings in stock price forecasting. This sets the foundation for the analysis and discussions in this thesis. Given the implications of successful forecasting models on companies' competitive advantage, many researchers have shared methodologies and results, contributing to the academic and practical discussion on stock forecasting. This section is divided into two subsections; the first subsection reviews comparative studies between LSTM and other forecasting methods in predictive accuracy, especially a comparison with ARIMA and other deep learning methods. The second subsection review the literature about the impact of including external features into LSTM models performance.

4.1 A review on LSTM performance in stock forecasting

In many research studies concerning time series forecasting of financial data, LSTM models seems to outperform the traditional ARIMA models. Namini et al. (2018) compares the performance of LSTM and ARIMA focusing on the reduction of error rates in predictions achieved by each model. They predict on historical monthly financial data, primarily indices, and historical monthly economic data which includes indexes for medical care commodities, housing, etc., all for different periods of 33+ years ending in 2018. The results shows that LSTM clearly outperform the ARIMA models. The average RMSE of the financial time series for LSTM is 64.213 compared with 511.481 for ARIMA, meaning an average of 87.447 reductions in error rates achieved by LSTM. For the economic data the study shows an average reduction in RMSE of 84.445 achieved by LSTM. In 2019 Namini et al. (2019) an expansion of the experiment was conducted, incorporating daily data and introducing a Bidirectional LSTM [2] to their study. By exclusively focusing on the daily frequency addition to this research, LSTM outperformed ARIMA by an average of 43.23 reductions in error rates [3], confirming LSTM's superiority over the ARIMA model. However, they conclude, that the Bidirectional LSTM demonstrated marginally better performance compared to the standard LSTM. Hence, indicating a preference for more advanced models in forecasting financial time series.

Xiao et al. (2022) achieved similar results when comparing LSTM and ARIMA models for forecasting daily stock prices on a dataset of 50 stocks spanning from 2010 to 2018. Their results emphasizes that while both ARIMA and LSTM are capable of predicting daily stock prices, the LSTM model outperforms the ARIMA model. Their underlying rationale for this improved performance is the deep-rooted nature of stock price dynamics, which are influenced by various historical factors. Each change in stock price at a particular moment is influenced by previous changes, showing a continuous pattern in its evolution. Given the LSTM's ability to retain memory of past events, it is capable of understanding these intricate and complex patterns, thereby making it a more effective tool for future stock price predictions. Interestingly, Xiao et al. (2022) also discovered that the efficacy of LSTM-based predictions varied across different stocks. Utilizing the exact same LSTM model across all analyzed stocks, they reported that the model performed notably better in forecasting the prices of Google and Netflix as opposed to Apple and Amazon. This indicates that the success of predictive algorithms can be stock-specific, warranting further investigation into the nuances of how different stocks respond to LSTM-based forecasting.

[2] Unlike traditional LSTM models that process data sequentially from the beginning to the end, bidirectional LSTM models process sequences in both forward and backward directions simultaneously

[3] It is noteworthy that RMSE values for higher frequency data are generally smaller.

Xio et al. (2022) and the studies mentioned above have highlighted the superior performance of LSTM models compared to ARIMA models. While these results establish the efficacy of LSTM models, it is crucial to shed light on the strengths and weaknesses of LSTM in the broader landscape of deep learning techniques in the existing literature. Hence, a nuanced understanding of how LSTM's performance measures up against other deep learning algorithms in the task of stock price prediction, is provided.

Niu et al. (2020) compared the performance of an LSTM model with other deep learning and hybrid models for predicting daily closing prices of various indices - Hang Seng Index, S&P500, London FTSE, and Nasdaq - over the period of 2010-2019. The results indicated that the LSTM model generally outperformed other standalone deep learning models such as Extreme Learning Machine (ELM), Convolutional Neural Networks (CNN), and Back-Propagation Neural Network (BPNN) for most indices, except for London FTSE, where its performance was only marginal over CNN and BPNN. Specifically, for the S&P500, LSTM recorded the lowest RMSE at 26.129, outperformed ELM's 27.407, CNN's 29.194, and BPNN's 35.726. The most striking difference was found in the Nasdaq index, where LSTM's RMSE of 77.565 was significantly lower than ELM's 103.792, CNN's 95.438, and BPNN's 91.649. The research paper also explored the addition of Variational Mode Decomposition (VMD) and Empirical Mode Decomposition (EMD) to these models. VMD is a technique that decomposes signals through a unique optimization process, while EMD simplifies complex stock price movements into more understandable patterns. The researchers concluded that EMD and VMD were effective in revealing the underlying trends and intricate patterns in stock prices, which are often inconsistent or complex. This improvement was evident as the combined models achieved better forecasting metrics (RMSE, MAE, MAPE etc.) than the standalone models, including LSTM.

Fischer and Krauss (2018) applied LSTM networks to a large-scale financial market prediction task on S&P 500 stocks from 1992 to 2015. They developed a comprehensive analysis through a sophisticated LSTM model, defining a binary classification problem based on whether the daily return was above or below the cross-sectional median return of all stocks during each period. Their objective was to examine the performance of memory-based against memory-free deep learning models. Initially, their findings indicated that LSTM networks outperformed memory-free classification methods such as a Deep Neural Net (DNN), Logistic Regression Classifier (LOG), and a Random Forest (RAF) - with the exception during the financial crisis - in various sizes of long-short portfolios. This relative advantage also holds true with regard to prediction accuracy where a Diebold–Mariano test confirms superior forecasts of the LSTM networks compared to the applied benchmarks. Hence, this emphasizes the application of memory-based models such as LSTM, due to their ability to learn and recognize temporal patterns, remember long term dependencies and their adaptability to new trends, hold significant advantages over memory-free models in predicting stock prices.

In addition, they evaluated the financial performance of 100,000 randomly sampled portfolios in the sense of Malkiel's monkey throwing darts at Wall Street Journal's stock page (Malkiel, 2007), and concluded that even the "best" performing monkey does not nearly match the outcomes produced by the models implemented in this study.

Additionally, they investigated the financial profitability of these models. The LSTM showed superior performance from 1992 to 2000. Between 2001 and 2009, a period of market moderation, the LSTM continued to secure positive returns after transaction costs every year, maintaining its dominance up to the financial crisis. However, during the financial crisis, the Random Forest experienced high returns due to its robustness to noise and outliers. From 2010 to 2015, the performance of the Random Forest declined, whereas the LSTM preserved capital after transaction costs, consistently achieving higher accuracy scores in almost all years. They further unveiled common patterns in the LSTM selected stocks, noting that the portfolio often included stocks with below-average momentum [4], high volatility, and strong short-term reversal traits, aligning with recognized market anomalies. They suggested that future research should explore more refined patterns detected by LSTM networks and the "black-box" of deep learning from financial data as well as validate the profitability of these patterns with advanced, rule-based trading strategies. In conclusion, Fischer and Krauss have effectively shown that LSTM network can efficiently extract valuable insights from noisy financial time series data. Hence, LSTM networks represents a significant advancement in this field as well.

4.2 A review on the inclusion of external variables in performance of stock forecasting

It is of great importance for this thesis to additionally review the literature that explores, how incorporating external features can influence and potentially enhance the predictive capabilities of stock forecasting models. The literature presented in this subsection will provide insights into the benefits, challenges, and nuances of including various external data points into LSTM-based stock prediction frameworks.

Du et al. (2019), formulated a research focusing on the application of LSTM for predicting the closing prices of the Apple stock for a ten year range from November 2008 to November 2018. They examined the LSTM model's predictive performance, when using single-feature inputs compared to multi-feature inputs. The included features consisting of open price, higher price, lower price, adjusted closing price, and volume. The results showed that the LSTM model with multiple features was superior, recording a MAE of

[4] !!!

0.033 compared to the 0.155 of the single-featured model. This emphasizes the impact of incorporating external features to improve the accuracy of stock predictions. However, it is essential to acknowledge that these results are solely dependent to Apples stock, limiting the generalizability of the findings to the field of stock market forecasting.

Nti et al. (2019) research different types of macroeconomic variables effect on predictive power of stocks for different sectors in Ghana. They develop a ARIMA and a LSTMRNN model to predict the 30-day ahead stock prices of stocks in five sectors; Petroleum, Banking, Technology, Pharmaceuticals, and Telecommunication industries. The dataset contains historical stock prices and 42 initially selected macroeconomic indicators in the period of January 2002, through December 2018. Their results showed that the proposed LSTM-RNN model with external factors outperformed the baseline ARIMA with external factors in all five sectors, indicating the superiority of deep learning models compared to the more restricted statistical models. In their research, they discovered that the significance of different macroeconomic factors varied across each of the sectors, indicating that the impact of each macroeconomic variable is different and varies among the different sectors. Focusing on sectors relevant to this thesis, in the banking sector, the most influential macroeconomic features are Year-on-Year, Danish Kroner, Minimum Daily Wages, Swedish Kroner, and US dollars, indicating a clear relation to the exchange rates. Furthermore, for the pharmaceutical sector, the most significant features are Total Government Liabilities, Savings and Time Deposits, Daily Minimum Wages, Currency Outside Banks, and the Treasury Bill rate for 91 days, reflecting a clear relation to government and bank-related variables. However, it is crucial to acknowledge that these findings are specific to Ghanaian stocks and sectors and may not necessarily reflect the dynamics observed in Danish stocks and sectors.

These findings confirm that variations in macroeconomic variables do affect stock volatility on the Ghana Stock Exchange. However, the degree to which each macroeconomic variable impacts stock price movements varies across different sectors of the Ghanaian economy, with each variable having a unique influence. They also contend that similar studies suffers from two initial challenges; the lack of causative direction and multicollinearity among predictors. Consequently, they argue, that this leads to uncertainty on the validity of the results and conclusions drawn from such studies, since the choice of macroeconomic variables seems combined or grouped together in one model in the multivariate analysis, without considering their individual distinct impacts or characteristic.

Lu et al. (2020) developed a study predicting daily stock prices from 1991-2020 focusing on the Shanghai Composite Index. The study developed a variety of models including Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), CNN-RNN, and CNN-LSTM to compare their forecasting performances. All these models included external features derived from trade data. The dataset consists of eight features: opening price, highest price, lowest price, closing price, volume, turnover, ups and downs, and change. The goal was to examine the efficiency of each model in incorporating these external features for improved accuracy in stock price predictions. When comparing the Root Mean Square Error (RMSE) values of each model, the LSTM model showcased superior performance with a value of 41.003, followed by RNN at 42.957, CNN at 42.967, and MLP trailing at 49.799. The results indicate that the LSTM model including external variables, outperformed the other deep learning models also including external variables in terms of predictive accuracy, with the exception of the hybrid models CNN-RNN and CNN-LSTM. The incorporation of external features in LSTM models, as demonstrated by this study, holds significant promise for improving the precision of stock price forecasting.

Kumbure et al. (2022) provides an in-depth literature review, in which they investigate machine learning techniques that are applied for stock market prediction, with a focus on the types of variables used as features. They analyzed 138 journal articles, published between 2000 and 2019, uncovered 2173 unique variables including technical indicators, macroeconomic variables, fundamental indicators, and others. In this literature review, variables such as technical indicators and macroeconomic variables have been considered the most influential ones affecting stock price movements. However, their sources utilized varying sets of features, whereas there is no general consensus about, which variables are the most relevant for stock market forecasting. When considering the technical indicator variables, the most commonly used variable is the closing price with lag 1, since it represents the most recent information. For the macroeconomic variables, exchange rates, commodities, and treasury stand out as most predominant variables. They identified 279 unique macroeconomic variables, 33 of which is concerning commodities, where precious metals such as gold, silver and copper, and energy-related commodities such as crude oil and gas, with crude oil being the most frequently used commodities. The largest group of macroeconomic variables are treasury variables, containing governmental debt obligations, such as treasury bills (up to 1 year of maturity), treasury notes (2 - 10 years), and treasury bonds (over 10 years). It is notable that in the category of 'others' variables, contains stock and market variables, in the form of several major global equity market indices, used as input for a stock market prediction model. These indices include FTSE 100, DAX, NASDAQ, S&P 500 etc. which are also frequently used by researchers.

From the review of the most recent studies, they also unveiled the rapid rise in the use of deep learning techniques, especially after 2015 and has continued even in 2020 and 2021. More than 50 percent of the source articles concerning deep learning techniques were published between 2015 and 2019. Another interesting fact is that LSTM networks have witnessed a growing interest in stock market predictions among the variations of deep learning models. While their source results are not directly comparable, an in-depth analysis by Kumbure et al. highlighted the highest performing for each machine learning approach. In regards to LSTM, they singled out the work of Shen and Shafiq (2020), who proposed a deep learning approach incorporating LSTM, feature expansion, recursive feature elimination (RFE), and PCA to predict the Chinese stock market. Their results indicate that the proposed method performs better than other forecasting models, such as SVM, MLP, RF, LR, and ARIMA, achieving the highest accuracy of 93%. Furthermore, they mentioned Liu and Long (2020) that explored the performance of LSTM networks in predicting the CSI 300 stock index price with sentiment analysis using text data from internet stock message boards. They achieved an 80.2% accuracy, exceeding the performance of all baseline models referenced. Additionally, they highlighted that specific deep learning techniques, such as LSTM and CNN, have gained prominence in recent studies. However, it has been demonstrated that a combination of these techniques proves more effective than utilizing each of them independently.

The reviewed literature clarifies the general trends in the results and methodologies of empirical studies, highlighting a common superiority of LSTM over ARIMA. This unveils a noticeable gap in literature regarding the forecasting of the Danish stock market, which this thesis tries to accommodate. Additionally, it is acknowledged that macroeconomic variables may have varying impacts on different stocks. It would be beneficial to explore, which macroeconomic factors influence specific Danish stocks and whether this information could improve the accuracy of forecasting.

5 Datamanagement and descriptive statistics

This section will undertake a descriptive analysis of the data, including a statistical overview relevant to the data's characteristics. Initially, data used in this thesis and relevant information will be presented. The section will also describe the preprocessing techniques and data management strategies implemented for the dataset utilized in this study. Subsequently, the dependent variables will be presented, examining the individual stocks and their statistical properties, as well as the reasoning behind their selection. Concluding this section, with a presentation of the external variables chosen for this thesis.

5.1 About the data and preprocessing

The purpose of the forecasting models in this project is for investors to be able to use the generated information to optimize the trading of stocks from day to day. This is why daily stock price data has been applied for these models. The time period of the data is from primo 2018 to ultimo 2022. In addition, the forecasting period is from 09-01-2022 to and including 12-29-2022. This exact forecasting period is chosen due to the macroeconomic turmoil in 2022, which had a visible impact on the financial markets in the autumn of 2022. This thesis uses daily closing prices because relative to average prices, open prices, etc., it reflects the final trading consensus of the value of the stock in question for the day. This makes closing prices a reliable data starting point for the analysis.

The datasets for this thesis are collected from different data sources. The stock prices are collected from NASDAQ Nordic (Nasdaq, 2023), bond rates are collected from MarketWatch (MarketWatch, 2023), exchange rates are collected from Danmarks Nationalbank (Nationalbanken, 2023), and additional data for the external variables are collected from Market Insider (Markets Insider, 2023). Data for GDP is collected from Statistics Denmark

(DST, 2023). In addition, the descriptive analysis in this thesis has primarily been carried out in R, see code in appendix B.

The use of daily prices means that the models have enough data to train properly. In addition, the ARIMA model require consistency in terms of the size of the input sequences, thus every year must consist of the same number of observations. Following this reasoning, a small number of random observations have been removed from some of the years to achieve a frequency of 248 daily observations a year. Given five years of data, this gives a total of 1240 observations in the data set for every variable. Furthermore, the data for the external variables has been adjusted to the number of data observations for the stock prices. The external variables are matched with the dates of the dependant variables, meaning observations outside these dates are removed. If it was encountered that external variable contained too few observations, so-called “last value carried forward” was used, i.e. missing values in the variables were replaced with the lagged value. The reason for replacing with the lagged value is that it is considered better to repeat values, relative to generating arbitrary values.

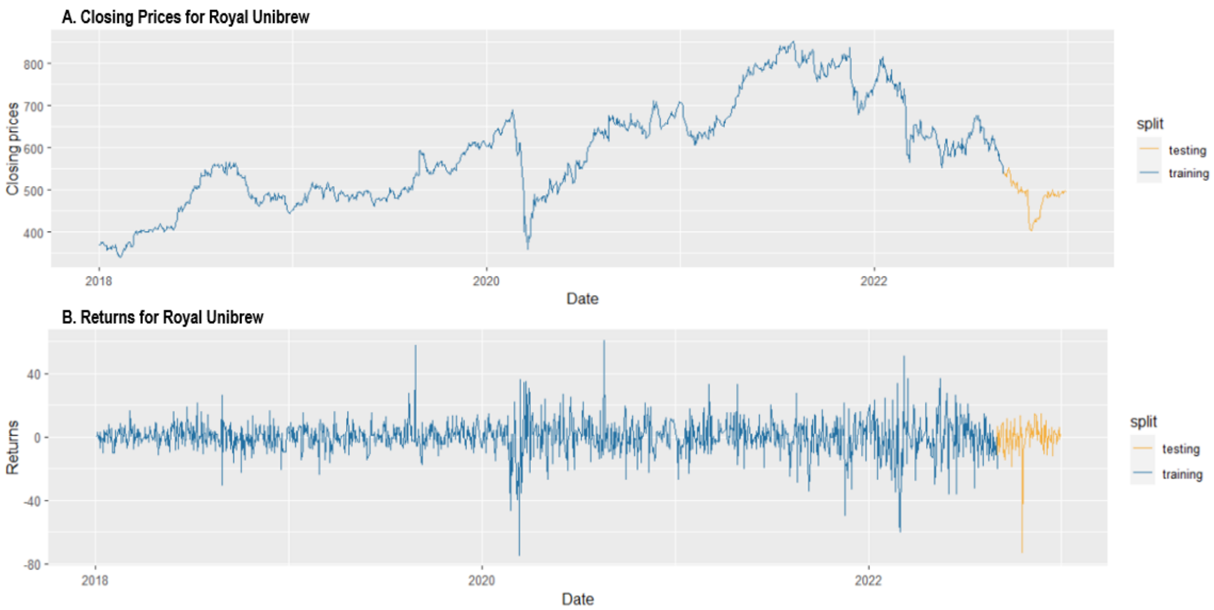
5.2 Dependent variables

The thesis is centered on the Danish stock market, which is why the dependent variables are selected stocks form the Danish OMXC25 index. This index contains the 25 most traded stocks on the Danish stock market and is thus the primary index to measure economic activity in Denmark. Additionally, for this project, an examination of various types of stocks is desired, specifically stocks from companies representing different sectors of the economy. Given section 2, it is expected that stocks from different sectors will react differently to economic fluctuations, making it relevant to investigate their varying impact on forecasting. The chosen stocks have been specifically selected to exhibit different cyclical patterns, whereas cyclical assets are advantageous in periods of growth and defensive assets are preferable in periods of recession (Voergaard, 2020). Thus, cyclical assets can often be described as high-risk, whereas defensive assets are characterized as low-risk. The selected stocks are thus Royal Unibrew, Novo Nordisk, Jyske Bank and Rockwool.

5.2.1 Royal Unibrew

The development in the stock prices (a) and differentiated stock prices (b) of Royal Unibrew from primo 2018 to ultimo 2022 is illustrated in Figure 8. From (a) Royal Unibrew seems to have a general upward going trend until medio 2021, despite the large amount of volatility. From medio 2021 it adapts a downward trend that really seems to take over in 2022, whereas the volatility equivalently seems to increase. This is confirmed in (b) where it is possible to identify changes in volatility over time. It can be observed that volatility appears to be increasing over the entire data period. Here, in particular, volatility seems to have increased from the start of 2020 onwards. The hit of the COVID-19 pandemic is clear in 2020, which had a significant impact on global stock markets, and the beverage industry was no exception. Comparing Royal Unibrew's stock price development in 2022 with the OMXC25, see section 1, it can be seen that fluctuations are very similar to the general market development, described in section 2.

Figure 8: Royal Unibrew closing prices (a) and returns (b).



The forecasting period, marked with yellow, differs in trend from the rest of the data, with the exception of primo and medio 2022, as observed in (a). Additionally, the volatility seen in (b) appears to have stabilized slightly in the test period, with the exception of significant outliers in September 2022. Understanding these changes in volatility can be crucial for forecasting, as the model will perform better if the training data is similar to the data to be forecasted. In order to compare the standard deviation across the different data periods in more detail, an annualized standard deviation [5] is calculated, which represents the volatility of the two periods. Table 1 provides an overview of the annualized volatility for these periods.

Table 1: Annualized volatility for Royal Unibrew

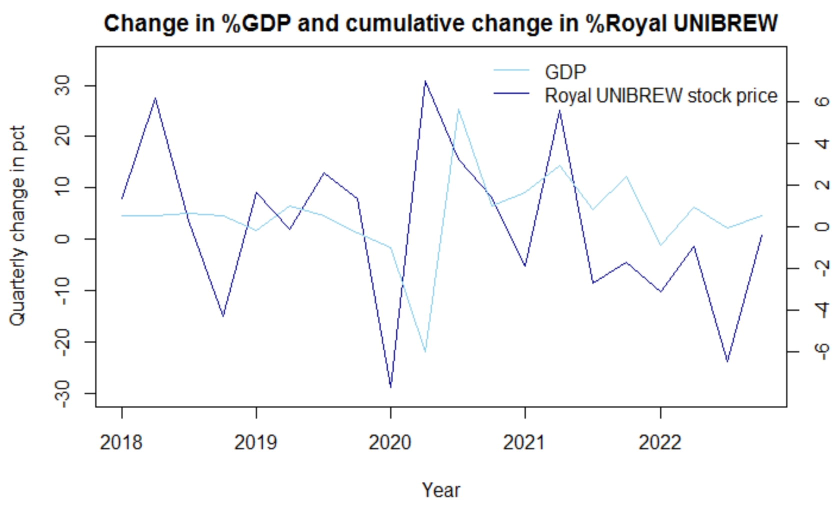
Training period	Forecasting period
0.303	0.344

[5] Annualized volatility are calculated as the standard deviation of daily returns times the square root of yearly observations (248)

It is essential to investigate how the relationship between the volatility of the test data differs from the volatility of the training data, as this can have a significant impact on forecasting models. It can be observed that the test period for Royal UNIBREW contains more volatility relative to its training period. If both training and test data have similar volatility, all other things being equal, the models will have a better starting point to provide more robust and accurate forecasts on unseen data. This is described in more detail in section 6.

Being a brewing and beverage company in a defensive sector, this stock is expected to be less sensitive to changes in the economy. To examine the sensitivity of stock prices to economic fluctuations, the cumulative percentage changes in stock prices and GDP each quarter are illustrated in Figure 9.

Figure 9: % change in GDP and % change in Royal Unibrew per. quarter.



In figure 9, the left y-axis indicates the percentage change in stock prices, and the right y-axis indicates the percentage change in GDP. From figure XX, it can be observed that Royal Unibrew, seems to be a cyclical stock, since it appears to be relatively sensitive to economic fluctuations. The stock price fluctuations appear to move in the same directions as fluctuations in GDP. This is somewhat in contrast to initial expectations given its sector being categorized as defensive, see section XX. In order to support the above observations, the Spearman correlation [6] between stock prices and GDP is calculated. The Spearman correlation is used to capture nonlinear correlations, where the sign of the correlation indicates whether the stock is pro-cyclical or counter-cyclical. The correlation coefficient is 0.136, meaning that Royal Unibrew has a positive correlation with GDP, which confirms that Royal Unibrew is considered pro-cyclical. Thus, an increase in GDP will lead to an increase in the stock price and vice versa. This is in line with the observations from figure XX. Furthermore, the size of this correlation coefficient alone does not indicate whether the stock is either defensive or cyclical. This will require a much more comprehensive study of company-specific and sector-specific factors, which is considered to be outside the scope of this project.

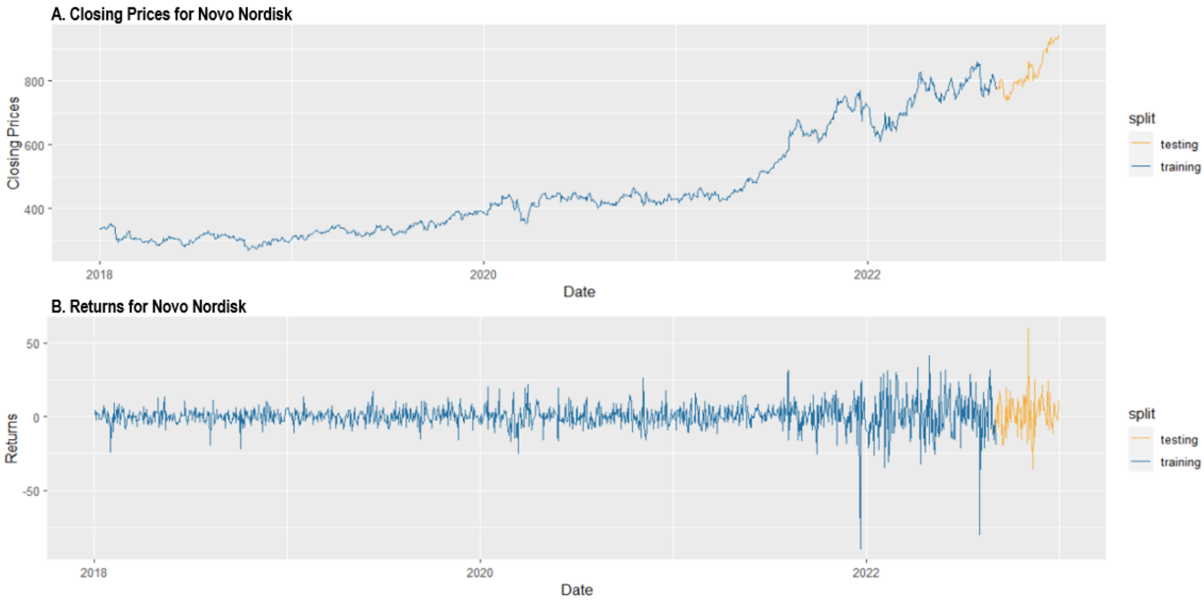
5.2.2 Novo Nordisk

The development in the stock prices (a) and differentiated stock prices (b) of Novo Nordisk from primo 2018 to ultimo 2022 is illustrated in Figure 10. From (a) Novo Nordisk seems to have a general upward going trend for the entire data period. From medio 2021 this increasing trend seems to accelerate, whereas the volatility equivalently seems to increase, which is confirmed in (b), from which it is possible to identify changes in volatility over time. It can be observed that volatility appears to be increasing over the entire data period. However, from medio 2021, and in particular from primo 2022, onwards volatility seems to have increased significantly. Additionally, the impact of the COVID-19 pandemic can be observed to have decreased the stock price for a moment in 2020. Comparing Novo Nordisk's stock price development in 2022 with the OMXC25, see section 1, it can be seen that fluctuations are not similar to the general market development, described in section 2. This, however, confirms the sector-specific characteristics, see section (statement). Here,

[6] Note that correlation does not imply causation.

it was argued that the increase in exports of goods in 2022 was particularly supported by exports of pharmaceuticals, which are less cyclical. Therefore, it is not surprising that Novo Nordisk does not follow the development of the OMXC25 index.

Figure 10: Novo Nordisk closing prices (a) and returns (b).



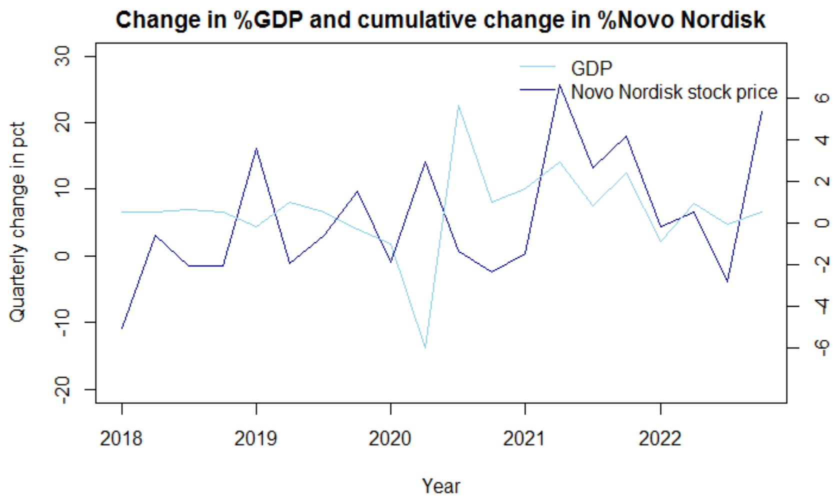
The forecasting period, marked with yellow, do not differ in trend from the rest of the data, with the exception of primo 2022, as observed in (a). Additionally, the volatility seen in (b) appears to have stabilized slightly in the test period, with the exception of significant outliers in October 2022. Understanding these changes in volatility can be crucial for forecasting, as the model will perform better if the training data is similar to the data to be forecasted. In order to compare the standard deviation across the different data periods in more detail, an annualized standard deviation is calculated, see Table 2. It can be observed that the test period for Novo Nordisk contains less volatility relative to its training period.

Table 2: Annualized volatility for Novo Nordisk

Training period	Forecasting period
0.258	0.244

As described earlier, Novo Nordisk's stock price is expected to be less sensitive to changes in the economy. To examine the sensitivity of stock prices to economic fluctuations, the cumulative percentage changes in stock prices and GDP each quarter are illustrated in Figure 11.

Figure 11: % change in GDP and % change in Novo Nordisk per. quarter.

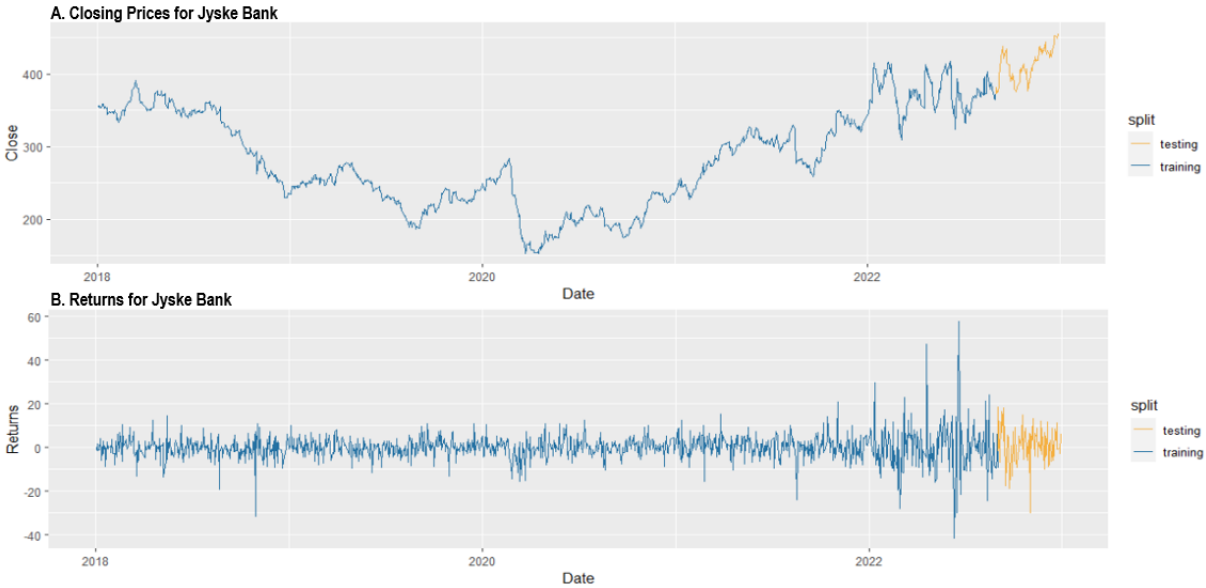


In figure 11, the left y-axis indicates the percentage change in stock prices, and the right y-axis indicates the percentage change in GDP. From figure 11, it can be observed that Novo Nordisk, seems to be a defensive stock, since it appears to be relatively insensitive to economic fluctuations. The stock price fluctuations appear to not always move in the same directions as fluctuations in GDP. This is in line with initial expectations given its sector being categorized as defensive, see section 2. In order to support the above observations, the Spearman correlation between stock prices and GDP is calculated. The correlation coefficient is 0.023, meaning that Novo Nordisk has a positive correlation with GDP, which confirms that Novo Nordisk is considered pro-cyclical. Thus, an increase in GDP will most often lead to an increase in the stock price and vice versa. This is in line with the observations from figure 11.

5.2.3 Jyske Bank

The development in the stock prices (a) and differentiated stock prices (b) of Jyske Bank from primo 2018 to ultimo 2022 is illustrated in Figure 12. From (a) Jyske Bank seems to have a general upward going trend for the data as a whole, despite the large amount of volatility. Initially, a decreasing trend is observed until medio 2019. From medio 2019 it adapts a upward trend until the start of 2020, due to the impact of the COVID-19 pandemic. From medio 2020 an upward trend really seems to take over, whereas the volatility equivalently seems to increase. In (b) the changes in volatility over time confirms these observations. It can be observed that volatility appears to increase significantly from primo 2022 onwards. Comparing Jyske Bank's stock price development in 2022 with the OMXC25, see section 1, it can be seen that fluctuations are not very similar to the general market development, described in section 2. It was argued that higher interest rates in 2022 further boosted net interest income for banks, increasing institutions' dividends and resilience to loan losses. Therefore, it is not surprising that Jyske Bank does not follow the development in the OMXC25 index.

Figure 12: Jyske Bank closing prices (a) and returns (b).



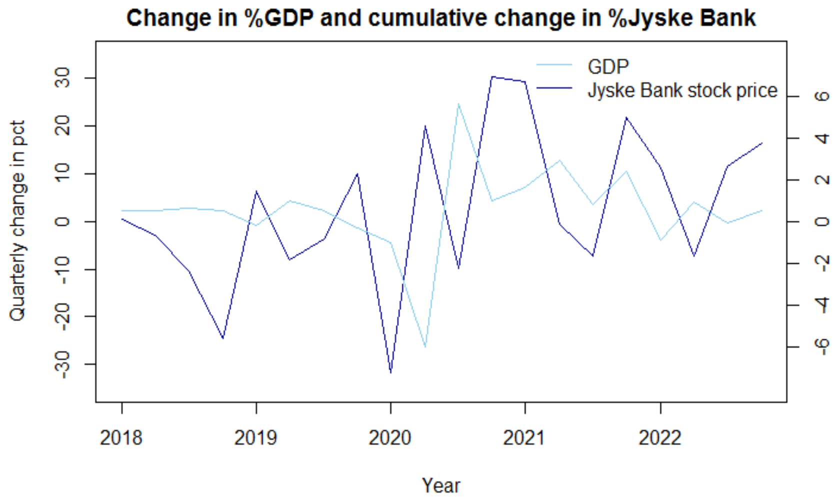
The forecasting period, marked with yellow, do not differ in trend from the rest of the data in general, with the exception of primo 2018 until medio 2020, as observed in (a). Additionally, the volatility seen in (b) appears to have stabilized slightly in the test period, with the exception of some significant outliers. In order to compare the standard deviation across the different data periods in more detail, an annualized standard deviation is calculated. Table 3 provides an overview of the annualized volatility for these periods. It can be observed that the test period for Jyske Bank contains less volatility relative to its training period.

Table 3: Annualized volatility for Jyske Bank

Training period	Forecasting period
0.322	0.314

As described earlier, Jyske Bank's stock price is expected to be less sensitive to changes in the economy. To examine the sensitivity of stock prices to economic fluctuations, the cumulative percentage changes in stock prices and GDP each quarter are illustrated in fgure 13.

Figure 13: % change in GDP and % change in Jyske Bank per. quarter.

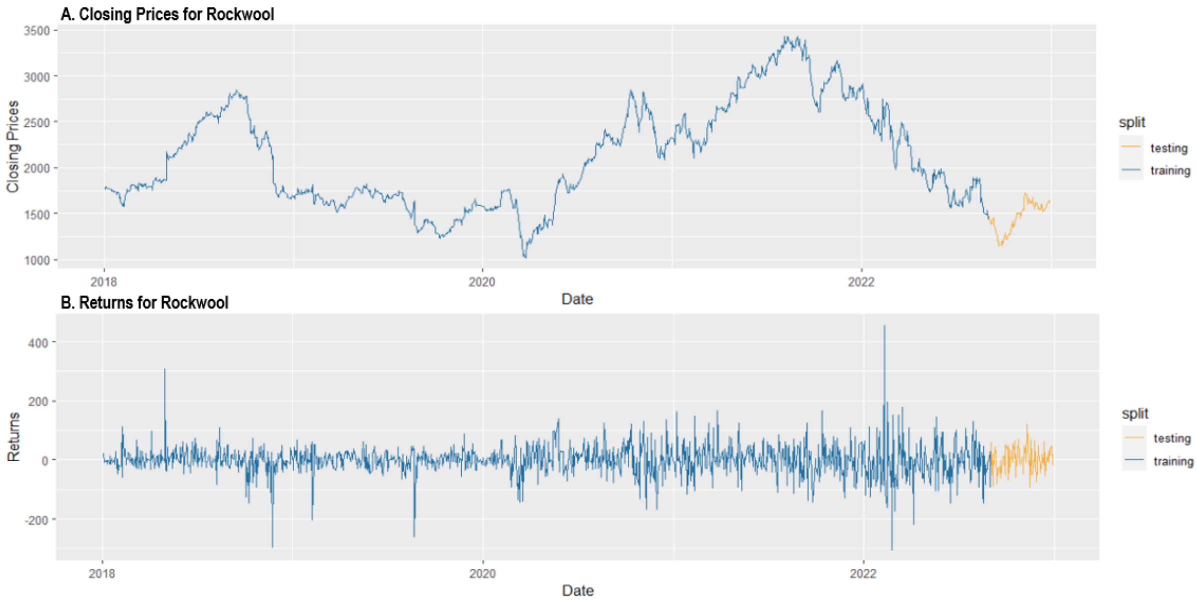


In figure 13, the left y-axis indicates the percentage change in stock prices, and the right y-axis indicates the percentage change in GDP. From figure 13, it can be observed that Jyske Bank, seems to be a defensive stock, since it appears to be relatively insensitive to economic fluctuations. The stock price fluctuations appear to not always move in the same directions as fluctuations in GDP. This is somewhat in contrast to initial expectations given its sector being categorized as cyclical, see section 2. In order to support the above observations, the Spearman correlation between stock prices and GDP is calculated. The correlation coefficient is -0.019, meaning that Jyske Bank has a negative correlation with GDP, which confirms that Jyske Bank is considered counter-cyclical. Thus, an increase in GDP will most often lead to a decrease in the stock price and vice versa. This is roughly in line with the observations from figure 13, where stock price fluctuations sometimes seem to move in the opposite direction as fluctuations in GDP. In this case, however, there is only a very slight tendency for stock prices and GDP to move in the opposite direction due to the relatively weak negative correlation. In addition, the size of this correlation coefficient alone does not indicate whether the stock is either defensive or cyclical.

5.2.4 Rockwool

The development in the stock prices (a) and differentiated stock prices (b) of Rockwool from primo 2018 to ultimo 2022 is illustrated in Figure 14. From (a) Rockwool seems to have a general downward going trend for the data as a whole, despite the large amount of volatility. Initially, an increasing trend is observed until medio 2018. From medio 2018 it adapts a downward trend until medio 2020. From medio 2020 an upward trend is observed until medio 2021. From medio 2021 until the end of 2022 a downward trend is seen, with a slight upward trend at the very end of 2022. Additionally, the impact of the COVID-19 pandemic can be observed to have decreased the stock price for a moment in 2020, which is confirmed from observations in (b). It can be observed that volatility appears to increase significantly from medio 2020 onwards. Comparing Rockwool's stock price development in 2022 with the OMXC25, see introduction, it can be seen that fluctuations are very similar to the general market development, described in section 2.

Figure 14: Rockwool closing prices (a) and returns (b).



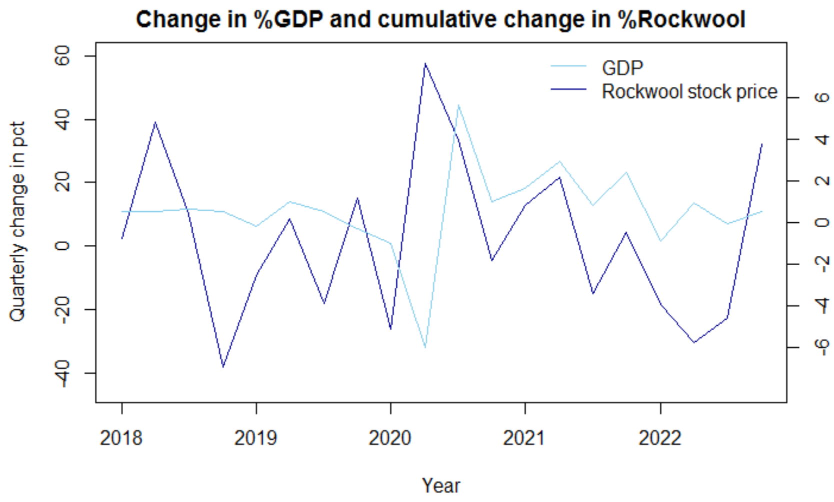
The forecasting period, marked with yellow, differs in trend from the rest of the data in general, with the exception of primo 2018 until medio 2018 and medio 2020 until medio 2021, as observed in (a). Additionally, the volatility seen in (b) appears to have stabilized slightly in the test period. In order to compare the standard deviation across the different data periods in more detail, an annualized standard deviation is calculated. Table 4 provides an overview of the annualized volatility for these periods. It can be observed that the test period for Rockwool contains a greater amount of volatility relative to its training period.

Table 4: Annualized volatility for Rockwool

Training period	Forecasting period
0.401	0.444

As described earlier, Rockwool's stock price is expected to be sensitive to changes in the economy. To examine the sensitivity of stock prices to economic fluctuations, the cumulative percentage changes in stock prices and GDP each quarter are illustrated in figure 15.

Figure 15: % change in GDP and % change in Rockwool per. quarter.



In figure 15, the left y-axis indicates the percentage change in stock prices, and the right y-axis indicates the percentage change in GDP. From figure 15, it can be observed that Rockwool, seems to be a cyclical stock, since it appears to be relatively sensitive to economic fluctuations. The stock price fluctuations appear to move in the same directions as fluctuations in GDP. This is in line with initial expectations given its sector being categorized as cyclical, see section 2. In order to support the above observations, the Spearman correlation between stock prices and GDP is calculated. The correlation coefficient is 0.208, meaning that Rockwool has a positive correlation with GDP, which confirms that Novo Nordisk is considered pro-cyclical. Thus, an increase in GDP will most often lead to an increase in the stock price and vice versa. This is in line with the observations from Figure xx, where stock price fluctuations seem to move in the same direction as fluctuations in GDP.

5.2.5 Summary

It can be seen from the above illustrations that stock prices move differently over time, partly due to the fact that the sectors are characterized by different structures. However, for all the stocks, a periodic decline in the beginning of 2020 can be observed, which denotes the decline in the stock market derived from the corona crisis, after which prices rose sharply. It was also illustrated how stock prices reacted to cyclical fluctuations in the economy. while acknowledging that stock price movements are also simultaneously influenced by the company's ability to innovate, as well as to be cost-effective, etc. These components, which exclusively affect the return of this specific company, are considered company-specific components, as these do not affect the return of other stocks. In general, it can be observed from the differentiated stock prices that the volatility of all four stocks, except for Rockwool, is increasing over time, with 2022 in particular appearing to be subject to great volatility in stock prices. It is therefore observed that the volatility is not constant over the entire data period, which is described in more detail below.

5.3 Statistical characteristics

Statistical tests' is carried out in order to clarify the characteristics and derive relevant information of the data, which is valuable information for the forecasting process. It is essential that the data for the four stocks is divided into a training period as well as a forecasting period. Therefore, it is possible to examine selected characteristics of each of the four stocks for the training period. Table 5 provides an overview of a number of different selected characteristics for the training period of the four stocks.

Table 5: Statistical characteristics of training data.

Stock	Jyske Bank	Novo Nordisk	Rockwool	Royal Unibrew
Mean	282.66	455.56	2125.08	591.72
Variance	4289.11	25687.55	314245.34	15113.15
Standard deviation	65.49	160.27	560.57	122.94
Skewness	0.08	0.97	0.40	0.19
Kurtosis	1.91	2.63	2.15	2.29
Maximum	417.30	860.00	3429.00	852.20
Minimum	152.70	267.30	1018.00	340.40

It can be observed that Jyske Bank clearly has the lowest mean, variance and standard deviation, whereas Rockwool has the highest. In addition, the scale of the minimum and maximum values is the highest for Rockwool. It is possible to examine whether the training periods for each stock appears normally distributed. For this, skewness and kurtosis are calculated, which are subsequently used in the Jarque-Bera test for normal distribution. Given the skewness test, which indicates the symmetry of the distribution around the mean, $S > 0$ for all the stocks. This indicates a right-skew distribution [7]. It can be observed that skewness for Novo Nordisk is close to 1, which is why this distribution has a high degree of skewness. Jyske Bank is observed to have skewness close to 0, which is why this distribution is roughly symmetrical. In most all cases, stocks exhibit a positive skewness, which means that the most extreme values are on the right side of the distribution and are thus positive values. In addition, positive skewness indicates the extent to which stock prices contain potential investment opportunities. In contrast, negative skewness may indicate higher risks, due to the higher negative values in such distribution. Thus, Novo Nordisk is observed to have the greatest investment potential cet par.

[7] skewness less than -1 or greater than 1 equals high degree of skewness

In addition, kurtosis indicates the degree to which the tails in a given distribution differ from the tails in a standard normal distribution, i.e. whether the tails contain extreme values (large outliers). The selected stocks in all cases take a value $K < 3$, which indicates flat tails and indicates a platycurtic distribution, which thus indicates minor deviations. Since most investors are risk-averse, they will prefer a low-kursosis distribution, as this implies lower fluctuations in prices and thus less risk. For Novo Nordisk it is observed that K is almost equal to 3, which indicates a standard normal distribution, also referred to as a mesocurtic distribution. that $K > 3$, which implies heavy tails and thus large outliers, referred to as a leptokurtic distribution.

Skewness and kurtosis can even be combined in a Jarque-Bera test for normality. With a p-value of almost zero for all the stocks, the null hypothesis is rejected. This may affect the results of the forecast models, although normality is not an exclusively necessary assumption for forecasting. These can also be made relatively more normalized through transformation, such as Box-Cox transformation, see section 7.

It is important to obtain initial knowledge about the behavior of the variables, and since autumn 2022 is pivotal for later forecasting, it is valid to present the structures of the data for this period. Table 6 therefore presents the same selected characteristics for the test period of the four stocks.

Table 6: Statistical characteristics of test data.

Stock	Jyske Bank	Novo Nordisk	Rockwool	Royal Unibrew
Mean	413.20	828.55	1455.15	485.72
Variance	486.12	3498.62	25383.65	1332.71
Standard	22.048	59.15	159.32	36.51
Skewness	-0.21	0.52	-0.38	-0.68
Kurtosis	2.17	2.07	1.95	2.86
Maximum	455.00	940.90	1725.00	552.60
Minimum	363.00	737.90	1150.00	403.30

One of the main differences between the training period and the test period can thus be observed to be skewness in most cases, which during the test period is negative for almost all the stocks, except Novo Nordisk. This may indicate higher risk for these stocks, given the higher negative values in this distribution, whereas Novo Nordisk is observed to have positive skewness. Now, essential characteristics have been clarified. The following section will provide a description of the external variables included for forecasting

5.4 External variables

For the purpose of answering the problem statement, external variables are included to examine whether additional information improve the performances of the forecasting models. From section 2, it is expected that these chosen variables will influence the fluctuations of stock prices. In the literature, many different types of external variables are found to influence stock prices, although this thesis will only test the effect of a subset of these. Thus, the initial selection of external variables is based on the existing literature. These variables are chosen due to being the most frequently mentioned factors and also shown to be the most significant in relation to stock price forecasting. A total 15 variables are included in the analysis in the form of stock indices, exchange rates, government bonds and commodities.

5.4.1 Stock indices

The stocks included in the indices of the FTSE100, OMXC25, S&P500 and DAX40 should provide a good indication of the stock market's general performance. These are included to provide information on the general performance of the British, Danish, American and German stock markets. Thus, they are used as proxies for the state of the economy in these countries. As Denmark is a small open economy, it is expected to be affected by changes in the larger and surrounding economies. In addition, the OMXC25 index is used as a simple-factor model to represent market volatility in Denmark. Furthermore, the S&P500 can be used as an indicator for the state of the U.S economy, as it includes 500 leading companies and covers approximately 80% of available market capitalization (S&P Global, 2023). The correlation between these indices and the dependent variables is expected to be positive, since a rise in the general stock market oftentimes would mean a rise in the individual stock prices as well.

5.4.2 Currency exchange rates

The currency exchange rates of the most significant Danish export markets are included, as the Danish economy is predominantly driven by export (Udenrigsministeriet, 2018). The included exchange rates are USD-DKK, GBP-DKK and EUR-DKK. The exchange rate is specified as the value of one foreign currency in Danish kroner. The EUR-DKK relation is included, although it is expected to have little impact on the dependent variables, since the monetary policy in Denmark is designed to ensure a fixed exchange rate against the euro. The Danish central exchange rate against the euro has been unchanged since 1987, and in 2022 the krone's exchange rate against the euro was stable on the strong side of the central exchange rate (Callesen et al., 2023, fig. 1). These exchange rates serve as a measure of the relative price level between Denmark and its most important trading partners. Thus, a rise in the value of a foreign currency relative to the Danish krone would make Danish goods cheaper relative to foreign goods cet. par. For this reason, it is expected that the correlation between the foreign exchange rates and the dependent variables would be positive. However, the correlation could be negative for the stocks of a Danish company that is reliant on imported production inputs.

5.4.3 Danish government bonds

The interest rates on 2-, 5- and 10-year Danish government bonds are included to provide information on the expectations of the financial market. The reasoning for including these Danish government bonds is that they are used as proxies for monetary interest rates and should therefore mimic market expectations. In addition, government bonds can be categorized as an alternative investment to the stock market. The interest rate of 2-year government bonds can be determined as a proxy for the short-term expectations in the market, the 5-year government bonds can be determined as a proxy for the medium-long-term expectations and the 10-year government bonds as a proxy for the long-term expectations. It is expected that a rise in the interest rate of any of these government bonds would be an indicator of negative expectations regarding stock performance. Thus, it is expected that the interest rate on government bonds would have a negative correlation with the dependent variables.

5.4.4 Commodities

The prices of different commodities are included to account for geo-political developments and the impact on general price levels in the economy. Gold and silver are known for being relatively safe investments during economic declines. Thus, the prices for these are expected to increase when the stock market is declining, as these can be seen as opposing alternative investments. Therefore, gold and silver should have a negative correlation with the dependent variables during economic declines and could have a positive correlation when stock markets are overheated. Copper is used across many sectors and can be seen as a leading indicator of economic health, as it is used practically everywhere – in homes and factories, in electronics and in power generation (Horowitz, 2022). Thus, the correlation between the copper price and the dependent variables is expected to be generally positive. Oil and gas prices are an essential part of most companies' production and transportation costs or supply chain. It is also a vital resource for heating and cooling. An increase in oil or gas prices might increase production costs, which would induce an increase in consumer prices cet. par. Thus, it is expected that an increase in oil or gas prices would have a negative correlation with the dependent variables.

6 Methodology

The main objective of this thesis is to evaluate and compare the performances of three types of forecasting models in predicting fluctuations in daily closing prices. The model is a day-ahead model, which traders can use to predict the closing price of tomorrow. This is a common resource in investment companies, banks etc. where traders can leverage stock price forecasting models in various ways to optimize their trading strategies, optimize their portfolios, manage risks and ultimately aim to improve their overall trading performance. The art of forecasting, where quantitative analysis intersects with human judgment and intuition, is a complex field of science. Developers of forecasting models constantly need to balance between model complexity and predictive accuracy along with considering elements of risk and reward due to their business strategies. This section clarifies the methods, techniques and choices made to carry out the analysis in this thesis.

6.1 Empirical Approach

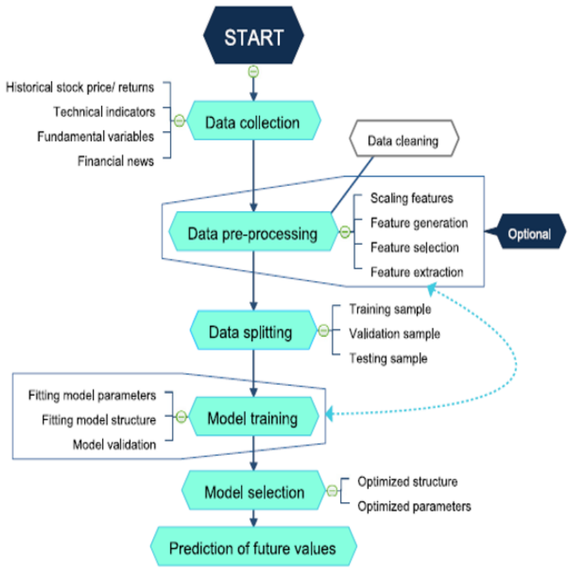
To address the research question in this project, an empirical investigation is conducted through out-of-sample testing. Out-of-sample testing is a crucial component of empirical forecasting that helps determine the accuracy and reliability of a forecasting model. It is essential to evaluate the performance of forecasting models on new data that it has not been trained on, hereby imitate a real-life forecasting situation (Galarnyk, 2022). To develop the ideal model for future predictions, it's necessary to undertake training and testing based on historical periods, before using the models for forecasting in real-time.

In out-of-sample testing, the dataset is divided into two segments, training and testing. The training segment is applied for model training i.e., learning and fitting the parameters of the model architecture. While the testing is applied for evaluating the model's unbiased

[8] performance. The most common train and test split is 80/20, but can vary depending on the research in question. This project will examine these proportions differently as it investigates for different data training lengths against very specific months of forecast, see section 6.4.

Within the training of deep learning models, it is common to use a validation set. Test-set and validation-set definitions are reversed interchangeably in the literature. A validation dataset is a subset of data retained from the models training-set, which is used as an estimate of model performance for model optimization (Brownlee, 2020). This creates opportunities for selection between different model architectures as well as optimizing the models' parameters and other components to improve model performance, before forecasting on the test-set. The general workflow of forecasting stock prices using machine learning is summarized in figure 16.

Figure 16: General workflow of forecasting stock prices using deep learning.



[8]unbiased refers to the objective evaluation of the model's performance, ensuring that the assessment is based on factual data from the testing segment and not influenced by prior knowledge or the training process

The process starts with data collection, in this case stock prices and features. These data have been preprocessed as described in section 5, and further normalized before used for training. Features are selected for the specific stocks using tests, see section 7.3.1. Data has been split into training, validation and test, whereas the length of the training data varies, see section 6.4. The model architecture, and the range hyperparameter are tested have been predefined, see section 6.3. Within the model training, hyperparameter tuning takes place, where many different combinations of hyperparameters are tested. In model selection, the model that demonstrates the best performance by accurately predicting the validation set is chosen as the model for finally predicting the test set.

6.2 Deep Learning Approach

In recent years deep learning has been recognized as a powerful algorithm in handling big data [9]. Data are generated at an accelerated rate, where these sophisticated deep learning algorithms offer a robust tool for processing, analyzing, and deriving insights from raw and large amounts of data that are beyond the capacity of traditional statistical techniques. Originating from the idea of replicating the neurons of the human brain and taking the advantage of the power of today's technology, deep learning has the opportunity to learn complex patterns with high accuracy in a short amount of time. This has become an interesting field of research in many applicational fields, which is why this project aims to make a contribution to this field of science.

Using deep learning for forecasting involves a combination of methodologies, algorithms, and iterative approaches that often resemble a trial-and-error process. This means that designing a deep learning architecture regardless of the deep learning methods[10], is a compre-

[9]Big data refers to extremely large datasets that can be analyzed to reveal patterns, trends, and associations, especially relating to human behavior and interactions. These datasets are so large that they require specialized processing solutions and analytics tools.

[10]Deep learning method here refers to the type of a deep learning layers are used; RNN, CNN, LSTM etc.

hensive application which there are no one clear solution to. It is characterized by repeated, varied attempts, which are continued until success or project abandonment. In all stages of creating a deep learning model, besides choosing the method of deep learning technique and the model architecture itself, there are serval hyperparameters that can be adjusted and model additions to experiment with. Deep learning models can be sensitive to hyperparameters, method and architectural decisions, requiring multiple iterations to find an optimal or satisfactory solution. In the following section, the architecture of the models and the choices of hyperparameters are described.

6.3 Model architecture and hyperparameters

The architecture and choice of hyperparameters of a deep neural network can be considered arbitrary, and it is therefore relevant to specify this design. Estimation based on long-term memory, test for all possible combinations of components and multiple hyperparameter adjustments, all together is a process that requires extensive matrix operations and calculations (Moolayil, 2019). It Is therefore necessary to take precautions for the available computational power and time, therefore limit the number of parameters and establish a specific model architecture to run. The long-short-term-memory models designed in this thesis is therefore based on a balance between performance and the ability to learn quickly. All deep learning models are created in R, using mainly the packages Keras and Tensorflow. Many other packages as seen in appendix D and E, are used for data preparation, data handling, graphs, and performance analytics.

In this subsection, the model's architectures and relevant hyperparameters are presented. All architectural components presented in this subsection are used for both the LSTM and the LSTM-X models. The choice and testing of hyperparameters, choice of optimization algorithm, choice of activation function, train-validation split, number of layers, dropout rate, number of iterations etc. are crucial in deep learning. These significantly impact model performance, including how well the model learns from the data, how quickly it trains, and how well it generalizes to new, unseen data (Torres, 2020).

In this thesis hyperparameter tuning is created through a loop that test out for 72 combinations of hyperparameters, which are done for up to 5 lagged timesteps and 3 training data lengths, for each individual stock. Meaning there are tested for 1,080 LSTM models and 1,080 LSTM-X models per stock. There are architectural parameters that determine the model's complexity and training parameters, which are crucial for the training process. The architectural parameters include the type of layer, number of layers, and number of neurons in each layer, while the training parameters consists of the learning rate, epochs, and batch size (Moolayil, 2019). The hyperparameters chosen in this thesis is based on hyperparameters shown to produce great results in the literature in this field.

6.3.1 Layers and neurons

The type of layers, the combination of layers and the number of layers defines the fundamental structure of the model. Adding more and different type of layers contributes to the complexity of the model which might increase model performance. Although it most likely would also use more computational power and time. Increasing the complexity through the number of layers may decrease the marginal effect, which in literature is argued to be three layers or above. This thesis tests for LSTM layers and the commonly used dense layers, where the number of layers is limited to two LSTM layers and two dense layers. The number of neurons in the model's hidden layers commonly takes on values that is to the power of two. This is because the training algorithm processes data more efficiently in this format (Moolayil, 2019). In this thesis, different combinations of the number of neurons are tested for values in the power of two in the range of [32;256], where the number in subsequent layers is kept constant or reduced.

6.3.2 Learning rate, epochs and batch size

Iterative optimization algorithms in deep learning models frequently focus on adjusting model parameters using gradient descent. Here, "gradient" signifies the slope's inclination rate, and "descent" refers to the act of moving downward, see section 3. Gradient descent is associated with the parameter called learning rate, determining the speed at which the neural network components are trained. Adjusting this rate influences the training pace, affecting how quickly or slowly the model learns (Namini & Namin, 2018). In this process, the model gathers information and learns systematic patterns within the data. The model's weights are consistently updated, enabling the model to adapt information optimally. The learning rate therefore directs the adjustment of the network's weights in relation to the loss gradient. In this thesis we test for the learning rates 0.01 and 0.1.

To ease the model's learning process, the entire dataset may be too large to be fed into a neural network all at once. Therefore, it might be necessary to split the data into smaller batches and train the network over multiple phases. The term batch size denotes the quantity of training data utilized in each segment. Essentially, when a large dataset is broken down into smaller segments, each segment is called a batch. The number of batches equals the number of iterations for a single complete round of training using the whole dataset (Namini & Namin, 2018). In this thesis the batch sizes tested for are 32 and 64, which are widely accepted and reused, often helping to improve the learning process and making the learning curve smoother for the network (Moolayil, 2019).

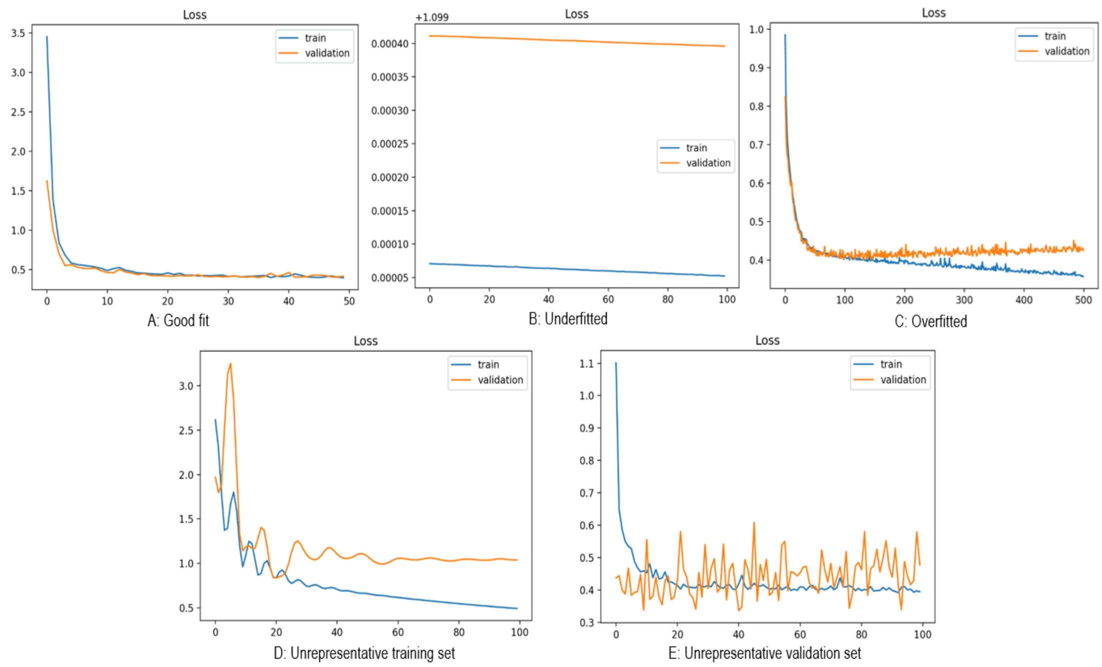
Epochs denotes how many times a dataset is passed through a model for training. Meaning one epoch represents, when the entire dataset is passed forward through the network one single time. In deep learning, since gradient descent is used for optimization, the model benefits from processing the entire dataset multiple times, allowing for weight adjustments to improve accuracy and performance. However, the optimal number of epochs, required for training varies, as different datasets necessitate different approaches to achieve the best results in model training (Namini & Namin, 2018). In this thesis the number of epochs is set to be equal to 100, in order to accommodate balance between performance and the ability to learn quickly, since higher epochs means higher training time. In addition, 'callback_early_stopping' has been used in this thesis for the purpose of stopping the training earlier, if there is no significant improvement of learning has been happened within 5 epochs.

6.3.3 Learning curves

Learning curves shows the learning performance over time in terms of experience. Depending on the specific modifications in code, the learning curves is most commonly represented by a loss function over epochs. The loss function is linked to a minimizing metric, where a value of 0.0 indicates that the training dataset learned perfectly and no mistakes were made. For this thesis the loss function metric is set to MSE for the validation set, to observe how well the trained data predict the validation set. This gives an idea of how well the model is generalizing the model for new unseen data. The epochs represent the experience, meaning the number of times that the whole data has passed through the training, whereas the loss function ideally should decrease and convergence to the lowest possible value. Together the learning curves for training and validation can be used to diagnose the model's performance. Whether it might underfit or overfit, as well as detect if the training or validation dataset might be insufficient. Knowing these, and detecting them early can help the process of model optimization. Underfitting occurs when a model is too simple to learn the underlying structure of the data. The model fails to capture the underlying trends in the data, resulting in poor performance. It is commonly followed by the training loss being high and decreases very slowly over time. In underfitting, the model performs poorly on both the training and validation sets because it cannot capture the complexity of the data.

Overfitting happens when a model learns the training data too well, including its noise and outliers. The model becomes too specialized in the training data, reducing its ability to generalize to new, unseen data. The training loss decreases initially, becoming very low as the model learns the training data completely, whereas the validation loss decreases initially after which it starts to increase (Brownlee, 2019). Figure 17 illustrates examples of cases of good learning curve (A), underfitted (B), overfitted (C), unrepresentative training set (D) and unrepresentative validation set (E). Note that these examples are not exhaustive.

Figure 17: Examples of learning curve diagnostics



To accommodate overfitting this thesis has included dropout layers, which helps with regularization during training by setting a fraction of the input units to zero at each update during training. This helps to prevent any single neuron from becoming overly specialized. However, too excessive use of dropout can result in underfitting. This thesis tests for values 0.1 and 0.2 for the dropout layer.

6.3.4 Optimizer and activation functions

The optimizer adjusts the weights of the network to minimize the loss function. It is within the optimizer the learning rate is specified. In this thesis the Adaptive Moment Estimation called the Adam optimizer is used. It adapts the learning rates of each parameter, combining the benefits of two other popular optimizers, the RMSProp [11] and AdaGrad [12]. It is proven to be very efficient and requires little memory, making it suitable for models with high computational needs (Brownlee, 2021). In addition, activation functions introduce non-linearity to the models, allowing them to learn from the error and make adjustments, which is essential for learning complex patterns. In this thesis the ReLU and Linear activations functions are used, these are explained in appendix A 12.1.1. A summary of all selected hyperparameters is shown in table 7 and an illustration of the model's architectural layers is shown in figure 18. The model in this thesis is designed to predict the stock price of tomorrow given the prices of up to five lagged days. The LSTM-X includes features in the input but only predicts the stock price of tomorrow as output.

[11] From the RMSProp, Adam uses the square of the gradients to normalize the parameter updates, which helps in faster convergence and adjusts the learning rates of each parameter adaptively (Brownlee, 2021)

[12] From the AdaGrad, Adam uses a moving average of past gradients, incorporating the benefits seen in momentum-based methods, providing an additional smoothing effect to navigate the parameter space more efficiently (Brownlee, 2021)

Table 7: Tested hyperparameters

Parameters	Value(s)
Numb. lagged input	[1;5]
Numb. neurons	31, 63, 128, 256
Learning rate	0.01, 0.1
Batch size	32, 64
Epochs	100
Optimizer	Adam
Activation function	ReLU, Linear
Dropout	0.1, 0.2

Figure 18: LSTM and LSTMX model layer architecture

6.4 Varying training data lengths

As an addition, this thesis will review the impact of various training data lengths on the predictive power of the models. Different input lengths allow for a nuanced understanding of a model's adaptability and performance under varied informational scopes. This discussion emerges from analyzing the fluctuations of the stocks, see section 5, questioning how much information is relevant to predict the stock prices of tomorrow. The stock market is under constant development, evolving with new trends, technologies, and global influences that continuously shape market dynamics. This idea not only might improve the robustness of the predictive models, but also disclose potential overfitting or underfitting issues. For comparison reasons, the validation within the training set is remained fixed for the period of 03-22-22 to 08-31-22. The training data lengths considered in the analysis are named short-term, medium-term and long-term, and covers 1 year 8 months, 2 years 8 months and

4 years and 8 months, respectively, starting in the beginning of 2021, 2020 and 2018. For this examination following arguments are set up against each other, defining why each of these might be adding the strongest set of information to the models:

Short-term: Newest data can better capture the current market trends and volatilities. Given that stock markets are exposed to frequent and rapid changes, the latest data often holds more relevance in predicting short-term market movements. This also supports the idea of data degradation. Data ages and its predictive power may diminish, because it no longer reflects the current market conditions and trends.

Long-term: A long dataset captures historical trends and cycles, providing a comprehensive view of the stock's performance over various market conditions. It includes more seasonal cycles, market fluctuations, and reactions to different economic events. Models trained with longer datasets might be more robust, because they have seen a more diverse range of market situations. It helps in generalizing the model better, preventing it from overfitting to short-term market noise and anomalies. Additionally, deep learning models tends to need larger amount of data to capture the complex patterns within.

Medium-term: This length offers a balanced perspective between having a long-enough historical context and maintaining the relevance of recent data, allowing the model to be tuned to both historical trends without moving to far away from current market conditions. The medium-term also includes data from the COVID-19 crisis, which provides valuable insights into how the market behaves during extraordinary global event creating very high volatility, which might be essential for building strong forecasting models.

It is essential to mention that although these are very interesting to test, they might be very stock specific. Different stocks has their own unique degrees of variability and volatility. Certain stocks might have more stable and predictable patterns, requiring less historical data for accurate forecasting. Others, however, might be more volatile, necessitating more extensive historical data for better training and understanding of their underlying trends and patterns. Important historical events, company specific events, sectoral characteristics all include information that might be captured in different data lengths influencing the stock predictability.

Upon closer examination the relationship between the annualized volatility of the test data and the length of training data are presented in table 8. In the event that the volatility of the test data is significantly greater than that of the training data, there is a risk of overfitting. Here, the models will learn the noise in the training data rather than the underlying patterns. Conversely, if the volatility in the test data is significantly less than in the training data, there is a risk of underfitting. If the test data has a greater volatility, there is a probability that the models do not generalize well to these variations. This can lead to poor performance on unseen data in forecasting. Thus, models trained on less variable data may not handle the volatility of the test data effectively. If both training and test data have similar volatility, all other things being equal, the models will have a better foundation to provide more robust and accurate forecasts on unseen data.

Table 8: Annulized variance

	Test Period	Long-term	Medium-term	Short-term
Jyske Bank	0.314	**0.322**	0.370	0.384
Novo Nordisk	0.244	**0.258**	0.277	0.285
Rockwool	0.444	0.397	**0.444**	0.450
Royal Unibrew	0.244	0.299	0.343	**0.296**

From table 8, it would be expected that the long-term annulized volatility for Jyske Bank and Novo Nordisk, medium-term annulized volatility for Rockwool and short-term annulized volatility for Royal Unibrew gets the closest to the annualized volatility of the test period. Thus, understanding the differences in variability is a crucial aspect in evaluating the performance of forecasting models. However, it is important to acknowledge that volatility or annualized volatility are not the only important information that the data lengths include, so the results may vary from these expectations.

6.5 Handling of explanatory variables

In order to answer the problem statement, a main focus is to test for the effect of adding external variables to forecasting models. Forecasting models depend on historical data to predict future developments in the dependent variable. When these models only contain historical data for the dependent variable, they are based on the assumption that future data (data to be predicted) will have a behavior or pattern similar to past data. In most cases, the underlying structure and relationship of the data changes, which can have a significant negative impact on the ability of the forecasting models to predict future fluctuations. If these future fluctuations in the variable occur abruptly and without precedent, which is not uncommon in financial data, then it might not be optimal to have only the historical values of the depended variables in the models.

To meet some of these difficulties, a number of studies have explored the idea of using external factors as features in forecasting models, see section 4. Thus, this project continues to explore the idea of taking into account these external features and apply this idea to the specific area of stock price forecasting. In this way, the purpose of this approach is to investigate whether it is possible to achieve better forecasting predictions in terms of accuracy after incorporating external features.

Once these external variables have been identified it is essential to investigate and test the relation between these and the dependent variables. This makes it possible to examine, which of these variables shows the most significance with a specific stock. For this purpose, Engle-Granger (EG) test is performed, see section 3. This test is widely accepted within the literature (Pesavento, 2004). It is a well-known econometric test used to verify the usefulness of one variable to forecast another, although this test does not provide insight into the true causal relationship between two variables. Thus, a variable is "helpful" for forecasting, when added to the forecasting model, it reduces the forecasting error.

The EG-test is testing the null hypothesis of cointegration. In this thesis it is chosen to only include variables which reject the null hypothesis at the 1% level of significance. This is a relatively more conservative approach, as it reduces the risk of Type 1 errors. This describes the case where there is a false rejection of a true null hypothesis, i.e. an error in seeing an effect that is not actually there. Here, a trade-off exists between Type 1 and Type 2 error, where Type 2 describes an error by not seeing an effect that is actually there. The reasoning behind the choice of the 1% level for is that only the most relevant external variables should arguably be included in the models, and Type 1 errors in this case are given the highest weight. In case there are no external variables at the 1% level, the most significant external variables at the 5% level will be included in the model in order to investigate the effect of external features.

After determining which external variables have a long-run relationship with the price of the individual stocks, using the EG-test, checking for multicollinearity is a prudent next step. If there are many variables significant at the 1% level, it is essential that multicollinearity test has been carried out between all external variables, see section 3. Multicollinearity is not in itself a limitation in forecasting, as deep learning models, like LSTM models, are able to handle some degree of multicollinearity better than more traditional statistical methods. This is partly due to the fact that parameters are fitted using backward propagation, see section 3. However, addressing multicollinearity is still relevant for several reasons when selecting external features to include in deep learning models, as this holds other significant implications for forecasting.

In deep learning models it can be challenging to clarify which feature is genuinely contributing to the model's ability to make predictions. When the features included in the model are highly correlated, the understanding of the importance of these is complicated, and thus the model exhibits a lack of transparency as well as interpretability. Although deep learning models can handle a large number of features, they are still susceptible to overfitting. This can be especially true when there is too much redundant information in the model derived from highly correlated features. In this case, the model may overweight a particular group of features. The reason for this is that a model trained on redundant features is not necessarily able to generalize to new data. To accommodate this, it would be ideal to provide the model with diversified sources of information, i.e. different groups of features, as this would enable the model to learn a more robust representation of the input data. In addition, noise accumulation may occur in the case of highly correlated features. This occurs if one of these correlated features contains noise, where in this case the model may double count this noise. This will naturally have a negative impact on predictions. Hence, reducing feature redundancy would simplify the model and can potentially enable the model to reduce the time it takes the model to train and make predictions. This is an essential aspect, as a delimitation in this project is computational power, see section 1. Thus, a reduction in multicollinearity will ensure that no significant resources are wasted on redundant information. For these reasons, it is considered that it is an advantage for the deep learning forecasting models to take into account and correct for multicollinearity.

By definition, external features are being added to potentially improve the models. Therefore, variables with a correlation coefficient above 0.8 will not be included in the same model in order to minimize multicollinearity, leading to potentially more accurate and robust forecasts, see section 3. Both EG-test and multicollinearity test give an initial indication of which features may be important for forecasting, although the final test of the effect of these variables is done empirically.

7 Model selection and analysis

7.1 Selection and forecasts of ARIMA models

In this section, a selection of the best ARIMA model for each of the four stocks will be made. First, each dependent variable's augmented Dickey-Fuller p-value is investigated to see if they are stationary. It appears from the ADF test that the four stocks are not stationary, which is particularly relevant to the ARIMA models, see section 3. Therefore, these can be differentiated, after which the ADF test shows that these are all integrated of the first order, and thus stationary after differentiation. The selection criterion for best ARIMA model is based on AIC, which implies that the model explains the largest part of the variation in the time series, by using the fewest possible number of parameters, see section 3. As the ARIMA model for each of the four stocks only serves as a baseline model in this thesis, the models are selected using the longest possible data length. This is done in order to make the baseline model as standardized as possible. The training period for the models will thus be from 01-02-2018 to 08-31-2022. In the selection of the best ARIMA model, all ARIMA combinations up to and including 5 previous values are tested for both AR-lags and MA-lags. Here, the 5 lags represent the closing prices for one trading week back in time. Park & Ryu (2021) finds that their Bi-LSTM models with shorter time lags (i.e., five or ten days) predicted volatility more accurately, thus one week (i.e., five trading days) should be a sufficient number of lagged parameters.

It is essential that the time series of the stock's shows non-stationarity, after testing for unit-root. To accommodate this, the closing prices are differentiated so that they follow an ARIMA(p,1,q) process that is stationary in first-difference. The ARIMA model, which exhibits the lowest AIC value for each of the four stocks, is then implemented to forecast the test period, see table 9.

Table 9: Selected ARIMA models

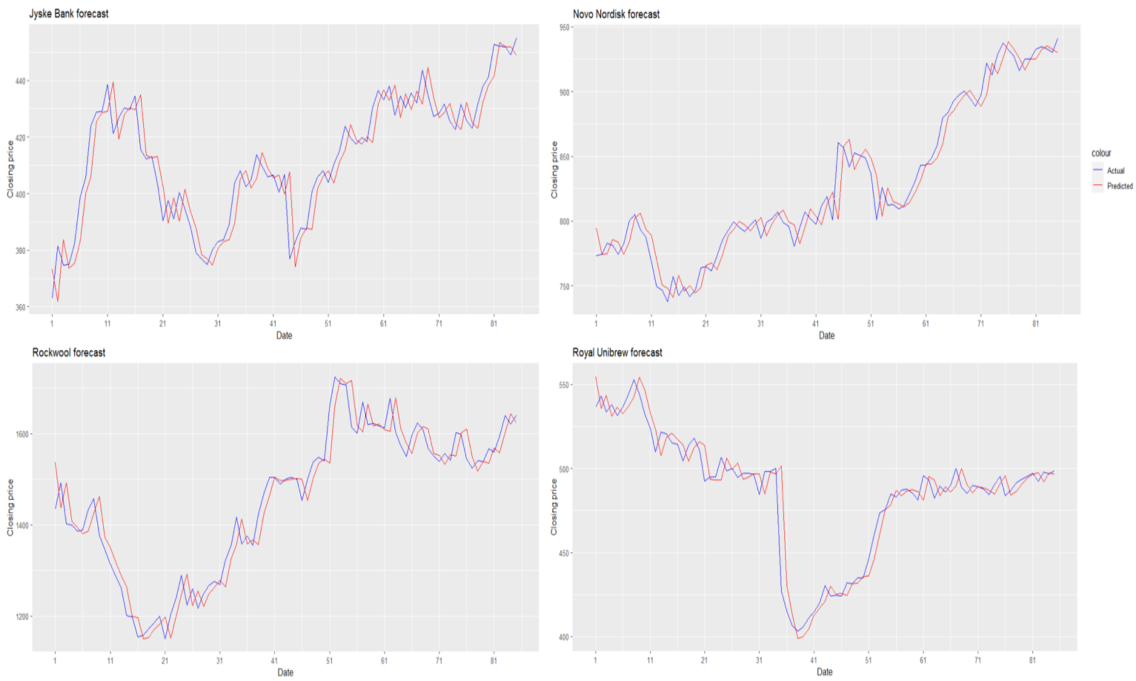
	ARIMA model	AIC Value
Jyske Bank	(1,1,1)	8762.07
Novo Nordisk	(3,1,3) with drift	-14629.64
Rockwool	(5,1,4) with drift	5259.18
Royal Unibrew	(5,1,4)	-18915.07

For each of these selected ARIMA models, a number of diagnostics are also performed to ensure the usefulness of these for forecasting. The ARCH test is applied as test of no autocorrelation, see appendix C. The ARCH-test, shows that the residuals are heteroscedastic, thus the squared residuals are autocorrelated. This means that the variance of the residuals are autocorrelated through time. It does not affect the consistency of the estimators, however like any form of heteroscedasticity leads to inefficient standard errors. This might impact the fitting of the forecasting models, as non-constant variance can impair the learning process of some algorithms. This is consistent with the skewness values indicating that the distribution for each of the four stocks was not symmetrical around the mean, see section 5.

ARIMA models assume normal-distributed residuals. The application of both Shapiro and Jarque-Bera test shows that models do not meet the assumption of normal-distributed residuals. This is to be expected, as significant outliers occur in the time series. These outliers can be removed by correcting for structural breaks, however, these may add relevant information to the forecast. Removing these values, would limit the models possibility to train on very volatile periods and therefore limits its ability to forecast during periods of significant macroeconomic fluctuations. Therefore, it is not considered optimal in this thesis to remove these extreme values from the time series, as these potentially impairs the models' ability (especially LSTM) to give realistic and reliable forecast results. Since the ARIMA models only function as baseline models, the performance of the later LSTM models is weighted higher, which is why outliers are not removed in either ARIMA or LSTM, as these must be comparable with exactly the same data.

To accommodate the non-normally distributed residuals and the inclusion of outliers, the Box-Cox transform is instead used for ARIMA. Here, non-normally distributed data is transformed into relatively more normal-distributed data, see appendix C. A graphic illustration of the forecast for the four ARIMA models is shown in figure 19. This illustrates how the forecasted values move compared to the actual values.

Figure 19: ARIMA forecasts



Here, a significant observation is that forecast values capture the actual fluctuations in stock prices with delay. During periods of relatively large fluctuations in actual stock prices, the models are less able to forecast the fluctuations sufficiently. Here, a relatively larger gap occurs between the lines for actual data and forecast values, respectively. This is also the reason why this project uses LSTM models as an alternative, where these are expected to be better suited to handle periods of large fluctuations.

7.2 Selection and forecasts of LSTM models

Section 6 has clarified the framework and the considerations of the following model selection and analysis. The selection of LSTM models is focused on determining, which lengths of training data results in the best performance for the four specified stocks, and discussing the rationale and implications of these outcomes. Subsequently, the best performing LSTM models and their forecasts for each stock will be presented. This is followed by a comprehensive analysis focusing on the models' ability to learn and generalize to unseen data, as well as discussing their potential effectiveness in predicting future stock movements.

In model selection, using the MSE of the validation set is a widely known practice. MSE quantifies the average squared differences between predicted and the actual values within the validation set, which offers estimate of the model's predictive accuracy before using it to forecast. A lower MSE value indicates a better performance. This metric is used in determining the best model between different number of lagged values, since all models have been tested for up to 5 lags. The MSE of validation values of the best models for each stock and training data length are presented in table 10.

Table 10: MSE of Validation Values

Stock	Jyske Bank	Novo Nordisk	Rockwool	Royal Unibrew
Short-term	**1.912**	**1.572**	**0.863**	0.956
Medium-term	2.758	2.174	1.035	**0.909**
Long-term	4.190	3.363	1.426	1.340

For comparison purposes, a consistent validation set was maintained across different models. From the results in table 10, it seems that the best performing data length based on the validation set is the short-term for all stocks except for Royal Unibrew, which is the medium-term. These results appeal to the argument of the newest data being the most relevant, described in section 6. The newest data allows models to be more aware of current realities, thus develop forecasts that are more reflective of imminent market patterns. The MSE values of validation of Royal Unibrew for the short-term and medium-term are not that far from each other. This could indicate minor significant information within the timeframe, captured by the medium-term, have been relevant for the Royal Unibrew stock. This information is unknown, and could be significant events, market patterns or just the model being too sensitive to recent fluctuations, while maintaining more relevant historical context. It is additionally observed from table 10 that the MSE of validation increases with the length of training data with the exception of Royal Unibrew. A reason for the increase in MSE of validation, due to increase in training data lengths, could be the addition of more data, potentially including outdated or irrelevant information. Such data introduces confusion, disrupting the model's learning process and possibly leading the model towards overfitting.

While the validation set has provided insights into the model's performance, for the evaluation and comparison of the different data lengths, it is crucial to also assess the model's efficacy directly on the test set. Evaluating the model on the test set reflects how the model will perform on unseen, real-world data, providing a more accurate representation of its generalization capability. This step is essential to determine the practical applicability and robustness of the model. For this purpose, this thesis uses RMSE as a metric. Table 11 presents the models' performance on forecasting the test set, where the lowest RMSE values represents the models with the better performance.

Table 11: RMSE Values of Test Period

Stock	Jyske Bank	Novo Nordisk	Rockwool	Royal Unibrew
Short-term	8.335	12.803	41.723	10.925

Table 11: RMSE Values of Test Period

Stock	Jyske Bank	Novo Nordisk	Rockwool	Royal Unibrew
Medium-term	**8.170**	12.704	**41.688**	**10.856**
Long-term	8.176	**12.652**	41.797	10.871

When selecting a model between the different training data lengths, the results form table 11 differs from the results from table 10. In table 11 the lowest RMSE values shows for the medium-term instead of the short-term for all stocks expect Novo Nordisk that has the best performance with the long-term length. A reason for the variation in results is due to how different lengths of training data resonates differently with the validation and test periods. The inconsistencies or noise within the validation and test datasets leads to varied outcomes. From table 11, it is important to notice that there is no grater difference in the results between the different data lengths. The RMSE values for the data lengths for each stock are remarkably close to each other. This indicates that training with different data lengths did not have an impact on the predictive accuracy for these models. This seems unexpected and to understand this there is a need for a more in-depth analysis of the models and their training.

Since this thesis examines different training data lengths it is relevant to discuss whether MSE might be the best choice for evaluating the model performance. MSE is highly sensitive to outliers, and therefore may increase due to the added volatility and outliers with larger data lengths. Since the goal aligns more with evaluating the model's real-world applicability and performance, using RMSE on the test set is more aligned with this objective. Therefore, several considerations have led to the decision of using the results of the test set instead of the validation set as a more suitable technique for model selection. Using the test set helps ensure that the model evaluation is unbiased and not overfitted to the validation set, especially when there is a noticeable increase in MSE of validation with longer training data. Therefore, the selected models and forecasts for each stock based on the results in table 11 are presented in figures 20 and 21.

Figure 20: LSTM models architecture.

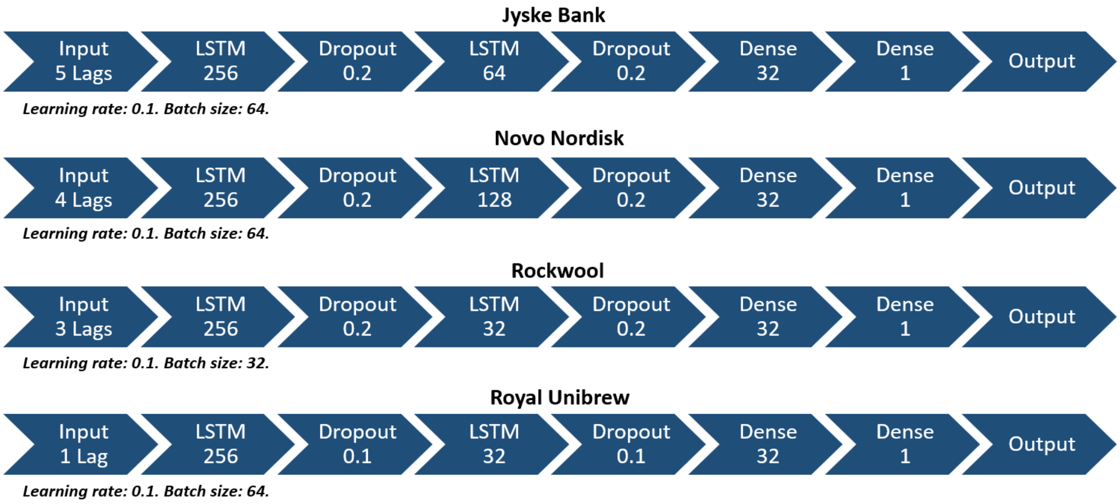
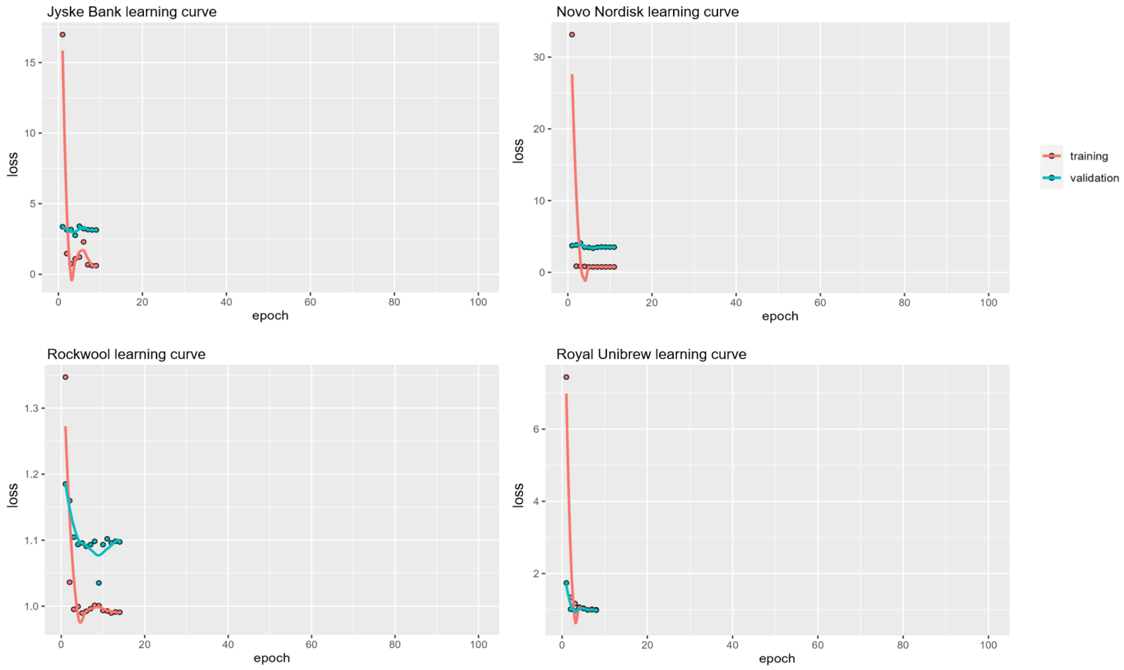


Figure 21: LSTM forecasts.

From figure 21, it seems the fluctuations of all four stocks has a delay and are very similar to the actual price the day before. It seems that the LSTM models poses as a naive model, setting the guess for tomorrows price equal to todays price. The forecasts of all four stocks seems to show almost exact same values from day before with minor deviations. This could indicate that the models have not been complex enough to learn from the data and capture the systematic patterns. Instead, it might have found that maintaining the same price is the safest prediction, possibly due to a lack of detecting patterns or due to noise in the data. Understanding this, it is essential to look deeper into the training of the models, presenting the learning curves of the four selected LSTM models in figure 22.

Figure 22: XX LSTM Learning Curves



From figure 22, it is observed that the training curves drops very quickly and then increases a bit again. For Novo Nordisk the training seems to be stabilized after the increase, whereas for Rockwool and Jykse Bank the training seems to fluctuate in loss, which indicates difficulties in learning and stabilizing. As described in section 6, these learning curves indicates clear underfitting. A underfitted model can be detected when validation does not decrease in loss and the distance between training and validation is large, which is the cases for Jyske Bank, Novo Nordisk and Rockwool. On the contrary Royal Unibrew is difficult to diagnose. Ignoring the peak in the decrease and then increase of training, this could almost indicate a good fit, since both training and validation are decreasing and end up at the same rate. Common for all of them is that training ends under less than 20 epochs, and none of the values decrease enough. This clearly indicates that none of the models have trained optimally, which has led to more naive and simple models, affecting the model's performance. Meaning that the models were not complex enough to capture the systematic patterns within the data. The occurrence of underfitting might explain the naive forecasts and the relatively close RMSE values between all data lengths. Concluding that no matter the data lengths, these models where too simple to produce significant forecasts.

Underfitting can occur due to several reasons. For instance, if the data used is noisy or inconsistent, it can make it challenging for the model to identify and learn underlying patterns. The architecture of the model itself is important. A model that is too simple, with only a few layers or neurons, might not be complex enough to grasp the complexities of the data. Within the model architecture, hyperparameters are another crucial aspect. Choices such as high regularization, like dropout, can overly simplify the model. From figure 20, it can be observed that common across all four stocks the highest dropout of 0.2 has been selected, which appears to support this idea. The models selected 0.2 over 0.1 which could indicate that higher dropout was needed to cope with very noisy data. Additionally, how the model is trained is vital. If the model is not trained for a sufficient number of epochs, it may not get enough iterations to learn effectively. Tools like early stopping can be used to monitor this, and in setting it to equal to 5 in this thesis might have been to restrictive. Similarly, choosing an unsuitable batch size can also hinder the model's learning process, preventing it from adapting to various data patterns properly. Adjustments in these areas are essential to improve the model's performance and allow it to learn the data's underlying patterns more effectively.

It is clear from the analysis above and understanding of model components, that a more sophisticated architecture or test of greater amount of hyperparameters is required to accommodate this forecasting objective. In conclusion, these models and their results are not significant enough to answer, which data length are the optimal for forecasting of these stocks. Their general forecasting performance can further be determined when compared to a baseline model in section 8. The exact reason for these results is unclear, however, if the models are too simple and unable to catch the complex patterns in the data, it might be beneficial to add more features to aid in improving their learning capacity.

7.3 Feature selection

Engle-Granger tests are performed for each of the three data lengths. The external variables included are all described in section 5. The variables included in the tests thus represent stock indices, commodity prices, exchange rates and bond rates. In Table xx in Appendix F, test results of EG-test can be observed. A value less than -2.58 indicates which of the external variables that has a long-run (LR) relation with the respective stocks. In general, it can be observed from these results that the stock indices are frequently occurring, whereas bond rates are only represented for Jyske Bank and Rockwool, in the case of the short data length. For Jyske Bank, the external variables that have a LR-relation, when using the short data length, are FTSE100, G10Y, Gas, Oil, Silver, Gold and all three exchange rates. Contrary, for Rockwool, none of the variables are significant, when using the long data length. To include external variables into this model, it is chosen to incorporate variables at the 5% level of significance, see section 6. Thus, the bond rates for 2-year government bonds (G2Y), commodity prices for silver and the exchange rates for USD and EUR are significant for this model.

In order to make the final selection of external variables, tests have been carried out for multicollinearity between all the external variables. The reasons for this are set out in section 6 and relate in particular to the models' ability to produce potentially more accurate and robust forecasts. When two or more of the external variables are multicollinear, only the variable with the highest LR-coefficient is included in the later LSTM-X model. From appendix F, it can also be observed which of the external variables that granger-causes the respective stocks after taking into account multicollinearity. In conclusion, the final selected external variables for the stocks and different data training lengths are summarized in table 12[13].

Table 12: Selected features

	Jyske Bank	Novo Nordisk	Rockwool	Royal Unibrew
Long	Gas	Gas	G2Y (5%)	DAX
	Oil		R.USD (5%)	Silver
			R.EUR (5%)	Gold
				R.EUR
Medium	FTSE100	FTSE100	Silver	DAX
	Gas	Gas		Silver
	Oil	Oil		
	R.GBP			
Short	FTSE100	FTSE100	DAX	DAX
	Gas	Gas	G10Y	
	Oil	R.EUR	R.USD	
	Silver			
	Gold			
	R.USD			
	R.GBP			
	R.EUR			

For Royal Unibrew the results shows that the LSTM-X models would include DAX, Silver, Gold and R.EUR, when using the long data training period. It is DAX and Silver, when using the medium data training period. It is only DAX, when using the short data period. Given the sector-specific considerations in section 2, it was described that households reduce their consumption ratio when they are pessimistic about the development of economic

[13]Table abbreviations explanation. G2Y: 2 Year Government Bond Rates, G10Y: 10 Year Government Bond Rates, R.USD: Exchange rate between DKK and USD, R.EUR: Exhange rate between DKK and EURO, R.GBP: Exchange rate between DKK and GBP.

activity. In this context, it makes sense that DAX is correlated, as this variable describes the economic activity in Germany, which is Denmark's largest trading partner. At the same time, the price of gold and silver are also indications of economic activity, as these represent a relatively safer investment compared to the stock market. Even though the company primarily sells stable consumer goods, and the development in the stock price should thus be less affected by cyclical fluctuations, it can be deduced from the test that this may still have an impact. This also supports the observations from section 5, which examined how stock prices react to fluctuations in GDP. Here it is observed that Royal Unibrew, which belongs to a defensive sector (see section 2), is actually relatively sensitive to economic fluctuations. It is argued by Kumbure et al. (2022), see section 4, that exchange rates are among the most commonly used types of variables for forecasting stock prices in the literature. Contrary, this is a surprising result, as it is argued in section 5 that the exchange rate for the euro should have limited effect. Reasoning behind this might be that the DAX is shown significant, which is the primary euro market. However, in some cases empirical results are not always in line with initial theoretical expectations.

For Novo Nordisk, it is only Gas, when using the long data training period. It is FTSE100, Gas and Oil, when using the medium data training period. It is FTSE100, Gas and R.EUR, when using the short data period. Given the sector-specific considerations in section 2, it was described that Danish exports in 2022 benefited from the recovery in export markets. The increase in exports of goods was supported in particular by exports of pharmaceuticals. Therefore, the reason why the FTSE100 index is significant may be that both Danish exports and FTSE100, representing economic activity in the UK, are affected by the same global economic conditions. During periods of global economic growth, the demand for Danish exports, such as pharmaceuticals, may therefore increase. Since economic activity in the UK is important for Danish exports, this may therefore have an impact on the stock price of Novo Nordisk, which is argued to be a major contributor to the development in exports of goods, see section 2. Additionally, it can be argued that the price of medicine is generally very elastic, as demand will remain somewhat unchanged, the prices of gas and oil may have an impact on the company's production costs, see section 5, which may be reflected in the stock prices. Rising oil and gas prices may also contribute to inflationary pressures in the economy, which means that the Danish central bank can respond by adjusting monetary policy, such as raising interest rates. These changes in interest rates can influence investment decisions and thus the stock price of Novo Nordisk. In this context, market psychology plays a role in relation to risk aversion, see section 2. Thus, it is natural that geopolitical tensions or supply disruptions, such as in natural gas, can also have an impact on stock prices through oil and gas prices. The influence of the exchange rate of the euro is expected to be limited, however, currency movements can reflect economic stability and investor confidence in a country. Here, a stable or strengthening EUR against the DKK can be seen as a sign of economic stability in the Eurozone. This may also have a positive impact on investor sentiment towards Danish companies, which will naturally also affect Novo Nordisk and thus the stock price.

For Jyske Bank, it is Gas and Oil, when using the long data training period. It is FTSE100, Gas, Oil and R.GBP, when using the medium data training period. It is FTSE100, Gas, Oil, Silver, Gold, R.USD, R.GBP and R.EUR, when using the short data period. Thus, Jyske Bank appears to have been affected by the largest number of external variables. However, each of the significant variables has a unique influence, as these are tests for multicollinearity. It can be observed that all three exchange rates tested are significant, which adds in line with the results and arguments given by Nti et al. (2019). In their research, they discovered that the importance of different macroeconomic factors varied across sectors, indicating that the impact of each macroeconomic variable is different and varies between different sectors. For the banking sector, a clear relationship with exchange rates was indicated. However, it is crucial to recognize that these results are specific to Ghanaian stocks and sectors and do not necessarily reflect the dynamics observed in Danish stocks and sectors. Jyske Bank may have investments or assets denominated in other currencies, which is why changes in exchange rates may affect the value of these holdings. This naturally affects the bank's balance sheet, which can thus be reflected in the stock price. It is also expected that the FTSE100 index is significant, as indices give a good indication of the general development of stock markets and thus economic activity, see section 5. It is somewhat surprising that this is more significant than the OMXC25 index. This is an example of when empirical results does not always aligns with theoretical reasoning, as it should be expected that the economic activity in Denmark would impact the banking sector in Denmark more relative to economic activity in the UK. At the same time, the price of gold and silver are also indications of economic activity, as these represent a relatively safer investment compared to the stock market, see section 5. As economies grow, there is generally greater demand for oil and gas, which is used in various industries. It is therefore natural that economic growth can therefore also have a positive impact on the banking sector. Here, increased economic activity will often lead to more borrowing, lending and investment, which helps to increase banks' earnings. In addition, rising oil and gas prices may contribute to higher inflation, which may prompt central banks to raise interest rates. These changes in interest rates may affect the banking sector and affect loan demand and may thus indirectly affect Jyske Bank's stock price through monetary policy decisions. Conversely, increased oil and gas prices can create increased operating costs, which affects the creditworthiness of some customers, which has an impact on loan quality.

For Rockwool, none of the variables were significant at the 1% level, when using the long data training period. Thus, the significant variables at the 5% level are included instead. After testing for multicollinearity, only G2Y as well as the exchange rates for USD and EUR are significant for this model. When using the medium data training period, Silver is the only significant variable. Furthermore, DAX, G10Y and R.USD are significant, when using the short data period. In section 2 sector-specific considerations states that the development of industrial production is closely linked to exports. It is therefore to be expected that the exchange rates for EUR as well as USD are significant, as Rockwool generates a significant part of its turnover from exports. In addition, Rockwool will be able to report higher revenues, after converting from the other currencies, if EUR or USD strengthens against DKK. This can increase investor confidence and thus positively affect the stock price. Changes in exchange rates can also have an impact on production costs, as the company may source raw materials, equipment or other resources from the United States or Germany, which may affect the stock price. These exchange rate fluctuations can further affect competitiveness, where weaker DKK can increase competitiveness in international markets, which can also positively affect the stock price. In addition, it makes sense that DAX is significant, as this variable describes the economic activity in Germany, which is Denmark's largest trading partner, which relates to exports. Furthermore, section 2 described that interest rate increases have an impact on investments in purchases and loans for construction and renovation of real estate. Therefore, it is also expected that the rates on 2-year and 10-year government bonds are significant. The reason for this is that these Danish government bonds can be used as a proxy for money interest and should therefore mimic market expectations, see section 2. In addition, government bonds can be categorized as an alternative investment to the stock market. Thus, an increase in the rates on any of these government bonds will be an indicator of negative expectations for stock developments.

This shows which variables are included in the models and on the basis of these results it is expected that the addition of these macroeconomic features can to some extent optimize the stock price forecast of the LSTM-X models compared to the LSTM models.

7.4 Selection and forecast of LSTM-X

The selection of LSTM-X models will be similar to LSTM. Since the goal of thesis aligns more with evaluating the model's real-world applicability and performance, the Root Mean Square Error (RMSE) of the test set serves as our primary criterion for model selection. Furthermore, our selection process for the LSTM-X models also includes varying data lengths, since the models including features might also potentially benefit from the different amount of information. This creates equal terms for both the LSTM and LSTM-X. The performance of the LSTM-X models is presented in table 13.

Table 13: RMSE Values of Test Period

Stock	Jyske Bank	Novo Nordisk	Rockwool	Royal Unibrew
Short-term	8.369	14.318	44.506	13.691
Medium-term	10.365	28.810	**43.708**	**10.889**
Long-term	**8.120**	**12.734**	46.2137	11.242

It is clear from table 13, that the features have had an impact on the forecasting, resulting in more varied outcomes than for LSTM. Keeping in mind the results obtained from LSTM, the incorporation of additional features can have opposite effects on the LSTM-X models. From one perspective, these features support the models capture trends and patterns with additional information, improving the predictability. Conversely, these features may contribute to a greater increase in data complexity and noise, making it more challenging for the simple models to navigate. This will result in worsen predictive power, and more noisy forecasts.

It is noticeable, that Jyske Bank long-term has shown to produce great results. From section 7.3, it is known that this model has the most features (eight) included compared the other models with a maximum of four features. This suggests that a greater number of features may contribute to improved training of LSTMX models. In addition, Novo Nordisk medium-term appear to stand out, deviating significantly from the outcomes observed in other data lengths. Comparing the included features on the data lengths, R.GBP is the only variable in medium-term that is not included in the other data lengths of Novo Nordisk. This could indicate that R.GBP could potentially have been influencing the forecasts adversely, causing them to be less accurate. These analytics, while insightful, cannot be considered entirely reliable, since the data lengths may also have had an impact. However, considering that in LSTM the data lengths had no notable impact, it seems plausible to argue that the features mainly contribute to the observed varied results. For a more accurate comparison of the features' impact, the analysis should have been structured differently, by holding data lengths constant while testing various feature frequencies and combinations. This thesis relies heavily on the guidance of the EG-test, based on the arguments presented in section 6. The model architectures and forecasts of the best performing LSTMX models are shown in figure 23 and figure 24.

Figure 23: XX LSTMX models architecture.

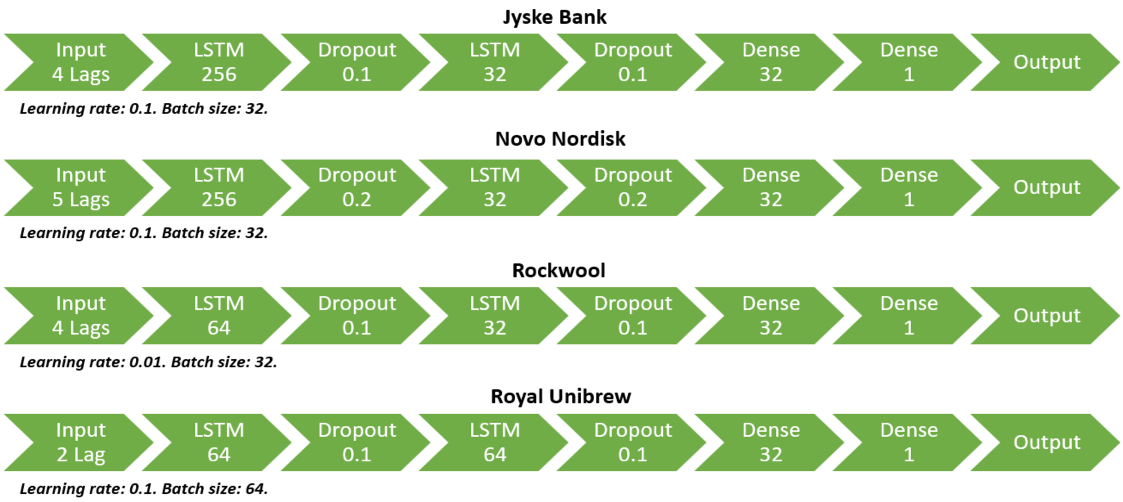


Figure 24: XX LSTMX forecasts.

From figure 24, it appears that all forecasts, with the exception of Rockwool, operate similarly to the naive models, setting today's price equal to tomorrows. This indicates that no additional information from the features was used in forecasting these stocks. Rockwool, on the other hand, demonstrates minor deviations from the naive model, but they are not necessarily good. The spikes appear more volatile, and they do not accurately capture the actual market fluctuations. Contrary to expectations, the dropout rate has decreased for Jyske Bank, Rockwool, and Royal Unibrew, indicating that the models utilize much more information compared to LSTM. These results indicates that the models have not been complex enough to learn from the data and capture the systematic patterns. Instead, they might have deduced that maintaining the same price is the safest prediction.

While these models exhibit the lowest RMSE values, examining the other LSTMX models provides additional insights into the inclusion of features in the LSTM models. From appendix A 12.1, it is clear that the additional features make the predictions more volatile, particularly noticeable in the larger spikes observed, as for Jyske Bank medium-term, Novo Nordisk medium-term, Royal Unibrew short-term, and Rockwool long-term, which contribute to the increase in RMSE. Looking closely, the variations added by the features have actually made the models predict almost accurately in a few instances, marking an improvement over the naive model. This is the case for the Novo Nordisk short-term, Jyske Bank medium-term, Royal Unibrew short-term, Royal Unibrew long-term and Rockwool long-term. This could arguably be caused by random behavior, and it is therefore still important to consider the strong indications that the models are too simple and require optimization. However, some feature management or regularization on the features might accommodate this problem slightly more by reducing the very aggressive spikes.

Based on above considerations and analysis, it is not possible to evaluate the performance between ARIMA, LSTM and LSTM-X. To do this, certain metrics and statistical tests are required, which will be utilized in the next section.

8 Comparison of models

In this evaluation, a comparison of ARIMA, LSTM and LSTM-X models' performance are carried out. Table 14, presents the estimation errors with the use of RMSE, MAE and MAPE, starting with assessing forecast accuracy.

Table 14: Evaluation metrics for all stocks

		Jyske Bank	Novo Nordisk	Rockwool	Royal Unibrew
RMSE					
	ARIMA	8.3298	12.8016	42.5044	11.1453
	LSTM	8.1698	**12.6517**	41.6879	10.8563
	LSTM-X	**8.1200**	12.7344	43.7085	10.8895
	NAIVE	8.1576	12.6794	**40.6407**	**10.7003**
MAE					
	ARIMA	6.4830	9.6027	33.1847	6.9279
	LSTM	6.3632	9.5338	34.7543	6.8492
	LSTM-X	6.4029	9.4371	32.8755	6.7262
MAPE					
	ARIMA	1.5838	1.1686	2.3063	1.4440
	LSTM	1.5547	1.16127	2.417751	1.430027
	LSTM-X	1.5633	1.14875	2.283529	1.40374

Table 14 reveals that no model appears to be substantially more accurate than the others. In previous analyses in section 7, it was observed that forecasts where similar to the naive model. Hence, the metrics for the naive model was included to determine whether the other models exhibited a more improved performance. The results indicate that the accuracy of the models is not far from the naive model. For Rockwool and Royal Unibrew, the naive model surprisingly outperformed the ARIMA, LSTM, and LSTM-X models, which is quite disheartening. This aligns with the arguments presented in section 7, that models may have been too simple to capture the underlying patterns and therefore adopt the approach of a naive model with a few modifications that seems more away from the true values. However, it's worth noting that the best-performing model for Novo Nordisk is the LSTM, while for Jyske Bank, it is the LSTM-X model. Despite this, the accuracy of these models does not deviate significantly from that of the naive model, indicating a need for improvement in our predictive models. It's noteworthy that the LSTM and LSTM-X models outperforms the naive model in the case of Jyske Bank and Novo Nordisk stocks and not for Rockwool and Royal Unibrew. In section 5, it is observed that Jyske Bank and Novo Nordisk both exhibit a positive trend during the forecast periods. Conversely, Rockwool and Royal Unibrew, where the naive model demonstrated superior performance, display negative trends in their forecast periods. This observation could imply that the LSTM and LSTM-X models tend to lean towards a slightly more optimistic or positive prediction. Such a tendency seems beneficial, when forecasting stocks that exhibit an upward trend, as it aligns more with their actual performance.

This thesis includes the Diebold-Mariano test to evaluate the significance of the difference in accuracy between forecasting models. The test offers insights into whether the forecasting errors of one model are statistically distinct from another. A significant difference identified by the DM test implies that one model has superior forecasting accuracy relative to the other. Essentially, the DM test aids in determining whether variations in the performance of two predictive models are attributed to random chance or if they are statistically significant. The results of the Diebold-Mariano test are presented in table XX, where the first column outlines the alternative hypotheses. This presentation facilitates a clearer understanding of whether the observed differences in model accuracy's are of statistical significance.

Table 15: Diebold Mariano test results

	Jyske Bank	Novo Nordisk	Rockwool	Royal Unibrew
LSTM $\neq$ ARIMA	**0.054**	0.413	0.128	0.0827
LSTM < ARIMA	0.973	0.793	0.936	0.958
LSTM > ARIMA	**0.027**	0.206	0.0641	**0.041**
LSTM $\neq$ ARIMA	**0.062**	0.731	0.410	0.166
LSTM < ARIMA	0.969	0.634	0.205	0.916
LSTM > ARIMA	**0.031**	0.365	0.794	0.083
LSTM $\neq$ ARIMA	0.688	0.222	0.151	0.475
LSTM < ARIMA	0.344	0.888	0.924	0.762
LSTM > ARIMA	0.655	0.111	0.075	0.237

This test uses a null hypothesis asserting that the models have the same predictive accuracy. This is tested against three alternative hypotheses: the forecasts of the first model differ in accuracy compared to the second model; the first model is less accurate than the second model; or the first model is more accurate than the second model. The test is conducted with a 5% significance level.

Overall, the results indicates that no one model outperforms the others in terms of accuracy. However, an exception is observed in the case of Jyske Bank. Here, when comparing LSTM and ARIMA, a p-value of 0.05 for LSTM $\neq$ ARIMA and 0.027 for LSTM > ARIMA suggests that, at a 5% significance level, the LSTM may be slightly more accurate than ARIMA. This is supported by a high p-value of 0.973 for the alternative hypothesis LSTM < ARIMA, dismissing the alternative hypothesis that LSTM is less accurate than ARIMA in forecasting Jyske Bank's stocks. Similar results are also observed when comparing LSTM-X and ARIMA for Jyske Bank, with LSTM-X performing better. However, it is essential to highlight that there is not a significant difference in the predictive accuracy of LSTM and LSTM-X concerning Jyske Bank's forecasts.

The distance between the forecast and the actual values indicates the errors of the forecast. These errors do not take into account whether the model forecasts a too high increase relative to an actual increase in the price, or an increase where in the actual data there is a decrease in the price. Therefore, it is interesting to use the Pesaran-Timmermann test to test whether the different forecasting models tend to predict the actual direction of the stock price. The reason why this may be relevant is that financial gains are dependent on the sign of the forecast the following day, as the purpose of stock trading is to make a profit. Here, correct forecasting of the sign will, all things being equal, be crucial for profitable stock trading. The Pesaran-Timmermann test tests the null hypothesis, which states an independent distribution of signs between forecast and actual data. It tests for 85 forecasted signs for each of the four stocks, indicating their test period. Table 16 shows the results of this test.

Table 16: Persian Timmermand test results

		True of 85	Rate of succes	P-value
Jyske Bank				
	ARIMA	36	42.35%	0.95
	LSTM	36	42.35%	0.94
	LSTM-X	**37**	43.53%	0.92
Novo Nordisk				
	ARIMA	43	50.59%	0.50
	LSTM	**45**	**52.94%**	0.31
	LSTM-X	43	50.59%	0.48
Rockwool				
	ARIMA	39	45.88%	0.78
	LSTM	40	47.06%	0.71
	LSTM-X	**42**	49.41%	0.54
Royal Unibrew				
	ARIMA	38	44.71	0.86
	LSTM	**41**	48.24	0.65
	LSTM-X	39	45.88	0.80

From this, both the number of correctly predicted signs for each of the four stocks and the associated success rates are observed. In general, there does not seem to be a significant difference between the three types of models, as the difference in the number of correctly predicted signs is a maximum of three between the best and worst performing model. Since the model with the highest success rate is only 52.94 percent, observed for the LSTM model for Novo Nordisk, it can generally be concluded that none of the models are able to adequately predict the signs correctly. Instead, this indicates a more random prediction of signs, which is consistent with the associated p-values. Thus, the null hypothesis that forecast and actual data are independently distributed cannot be rejected for any of the models. Based on this, it is estimated that the LSTM and LSTM-X models are only slightly better than the ARIMA models in all cases. Here it can be observed that LSTM is best for Royal Unibrew and Novo Nordisk, whereas LSTM-X is better for Jyske Bank and Rockwool.

Overall, the results confirm that neither the LSTM nor the LSTM-X models fully realized their potential in this thesis. While existing literature, as presented in section 4, demonstrates various successes with both LSTM and especially LSTM-X models, this thesis presents a somewhat contradicting outcome. As described in section 7, there is a need for further optimization in architecture, hyperparameters and additionally feature management to accommodate an analysis of this nature.

9 Discussion

The overall underperformance of the LSTM and LSTMX-X models could suggest that utilizing these models to predict the Danish stock market in this thesis has not yielded significant insights. However, this is not entirely true. This section aims to discuss the implications and nuances of the thesis results, offering a more comprehensive understanding of their impact and relevance.

9.1 Navigating theoretical expectations and empirical realities

There is not always alignment between theory and empirical findings when it comes to determining the variables that impact outcomes. Theoretically proposed variables do not always correspond with the variables that exhibit a practical impact in real-life scenarios. Furthermore, different empirical models may yield varying results, leading to inconsistencies in determining which variables are genuinely influential. This variation highlights the complexity of accurately identifying influential variables and the potential discrepancy between theoretical expectations and empirical realities.

In this thesis, the results from the selection of the external variables showed that almost all of the initially selected variables were relevant in one or more of these forecasting models. It was not expected that all 15 external variables would prove to be significant. However, among these variables, only the development in copper prices did not show any significant impact or correlation. This is a somewhat surprising result, as copper prices are generally considered by many financial investors to be a leading indicator of the state of the global economy (Horowitz, 2022). This is because copper prices tend to be cyclical, following economic activity. The reason for this is that copper is, by nature, a metal that is widely used in various industries and sectors, such as electronics and construction. Copper is also a highly non-substitutable metal, hence changes in supply and demand within the global economy frequently have a direct impact on copper prices.

Therefore, it was naturally expected that copper prices in this thesis could be used as a determining feature for the development in stock prices. It is essential to be aware that even if a variable is not found significant by econometric tests, this variable may still have a relevance and therefore have a real effect that is just not captured in these tests. In this thesis, it was decided to follow the instructions from the econometric tests, even though copper prices could potentially be used in an alternative model scenario. In this regard, not all combinations of external variables are possible to test in the scope of this thesis and could therefore be a relevant scenario to test in a further study.

A decision was made to include only the most relevant external variables in the models, rather than incorporating all potential external variables. An alternative approach would be to include as many external variables as possible and then let the model itself select between this information. In this thesis, however, it was decided to prioritize issues such as handling of redundant information and transparency. For further study it is questionable whether 'all at once' or a selection of external variables through tests would be the most optimal approach.

An essential aspect to discuss in relation to the empirical study in this thesis is uniform vs unique model architecture. The same model architecture is used for both the LSTM and LSTM-X models. Thus, there is a consistent general structure in the architecture for a model with only historical data and a model that also includes external variables. In this context, it can be argued that the way the two types of models process the data information should ideally be different, and thus have different architectures. It would be expected that LSTM-X would require more complex architectures. Models that include external variables must be specified to account for these additional dimensions in the data to prevent overfitting. To optimize the learning in this model with external features, it will therefore potentially be necessary to add additional input layers and make other adjustments to the model architecture.

9.2 Transparency in deep learning and the advantages for investors

The transparency of deep learning models is something that researchers are working to achieve, as these models can in many ways be considered a 'black box'. More transparency could provide greater insight into the conditions and features that drive the models' predictions. In this context, it would be possible to identify which of the macroeconomic variables included in the model have the greatest influence on stock price predictions. By identifying the most relevant variables, it will also be possible to optimize the accuracy of the model and increase the understanding of these effects. Understanding their impact will enable investors and researchers to better understand how the model adapts to external changes. More transparency will therefore make it possible to identify forecast errors more effectively, which can be used to improve and fine-tune model forecasts.

This would potentially have changed the analysis in this thesis if it was possible to isolate the effect of each external variable in a manageable way. By observing the development of the stock price forecast in the illustration in appendix A, it would potentially be optimizing if it was possible to isolate the information from the external variables that create large deviations in the forecast. In this case, these models would potentially have performed better, as it can be observed that they capture the fluctuations in actual stock prices relatively better than the naive-like predictions in figure xx (analysis). Thus, making such adjustments to the features in the network could be relevant for future forecast models that include information from external variables. In general, enhancing transparency in deep learning models could improve the accuracy of the forecast and the architectural integrity. A more transparent model allows for a clearer understanding of how input variables influence predictions, facilitating more insightful interpretations. Additionally it increases the reliance and confidence in the model's forecasts, as users have a better understanding of the model's decision-making process. Most importantly, increased transparency could aid in identifying potential biases or errors in the model, contributing to the optimization of its predictive performance and architectural robustness.

The LSTM forecasting models proposed in this thesis are designed with the intention to serve as a tool that assists financial investors in making improved decisions in stock trading. In practice, for a given type of investor, it would be ideal if it were possible to forecast developments in the stock market with high accuracy. The reason for this is that more accurate predictions will allow for continuous maximization of returns cet. par. More accurate forecasts also allow for a better response to market risks and can thus increase confidence in the financial markets. In relation to forecasting with external variables, more accuracy can provide more insight into what conditions and factors actually affect stock prices. This can give investors, as well as researchers, a better understanding of the dynamics that drive stock prices. Researchers use these forecasts to form impressions about general market conditions, which in turn form the basis for expectations about the future. Therefore, a natural motivation is that deep learning models, e.g. LSTM architectures, can be constructed to achieve this higher degree of accuracy.

9.3 Capturing volatility and further model development

The general purpose of the LSTM models in this thesis is that they should be able to capture patterns in training data to predict the macroeconomic fluctuations observed in the autumn of 2022. It is relevant to emphasize that it is not the macroeconomic shock itself, such as in September 2022, the models are designed to predict. Rather it is the fluctuations in the stock prices, following the shock, which the models are intended to capture. Here, training data naturally does not contain information that can be used to predict geopolitical tensions or sudden supply and demand shocks. However, it is possible to optimize a model so that it can handle expected high volatility in the test data, i.e. the period after a shock to the macroeconomy.

It is also relevant to discuss how LSTM models handle volatility. The LSTM models in this thesis cannot directly capture and handle volatility in the test data. Alternatively, it can be discussed how integrating a generalized autoregressive conditional heteroskedasticity (GARCH) model could optimize the LSTM model's ability to handle volatility. GARCH models are specifically designed to handle and predict volatility. By including such a model in the context of LSTM, it would be possible to construct and train a GARCH model separately to predict volatility and use this measure of volatility as an additional feature in the LSTM model. It could also be possible to construct a combined hybrid model, with the strengths of both models, that can create a trade-off between accurately predicting volatility and the accuracy of the actual forecast (Kim & Won, 2018). Integrating GARCH models could thus potentially generate relatively more robust and accurate forecasts when dealing with large fluctuations in the test data. Such integration would be a possible choice for further development of the models in this thesis.

In continuation of the above, it can be discussed whether the LSTM models, with external variables included, could potentially be optimized by constructing so-called hybrid models, as described by (Niu et al., 2020). These models have an advantage in being able to utilize the properties of different deep learning models as well as traditional statistical models. Here, statistical models can act as regularization in the architecture of a hybrid model. By combining the properties of these models, a hybrid model will be able to generate more robust and accurate forecasts. Hybrid models have the ability to process information such that only the most relevant and influential features are taken into account. Thus, it will be possible to reduce noise from redundant information, which is especially essential when data appears more volatile.

10 Conclusion

The results, analyzes and discussions in this thesis have been prepared with the aim of answering the following problem statement:

Does inclusion of macroeconomic features increase the performance of LSTM models in forecasting Danish stock prices? How is this applied to a period of macroeconomic distress during 2022?

In this thesis, a variety of stock price forecasting models for four specific stocks in the Danish OMXC25 index was developed. The thesis set out to explore whether LSTM models and LSTM-X models could be fine-tuned to predict volatility more effectively, especially in periods marked by economic distress. The initial focus was on determining the ideal lengths of training data for LSTM models and comprehending the implications of these choices. Through the analysis of model components, it became evident that a more sophisticated architecture and an exploration of a broader set of hyperparameters were necessary to meet the ambitious forecasting goals. While the models did not provide conclusive insights into the optimal data length for forecasting these stocks or whether they contained specific information for predicting periods of high stock price fluctuations, the findings highlighted promising paths for further exploration and advancements in the field of stock price forecasting through the utilization of LSTM models.

Engle-Granger tests were conducted for different data lengths, revealing the external variables to be included in the LSTM-X models. It was anticipated that incorporating these macroeconomic features could partially optimize the stock price forecast of the LSTM-X models compared to the LSTM models. A comparison of ARIMA, LSTM, and LSTM-X models was performed, but no model demonstrated a substantial advantage over the others. Surprisingly, for Rockwool and Royal Unibrew, the naive model outperformed the more complex ARIMA, LSTM, and LSTM-X models, indicating potential shortcomings in the complexity of the models to capture underlying patterns. Notably, the LSTM model performed best for Novo Nordisk, while the LSTM-X model performed better for Jyske Bank. However, these models did not significantly outperform the naive model, highlighting the need for improvement in predictive accuracy. Overall, the results indicated that neither the LSTM nor the LSTM-X models fully realized their potential in this study. Despite the successes reported in existing literature, this thesis presented somewhat contradicting outcomes. As outlined earlier in this thesis, further optimization in architecture, hyperparameters, and feature management is necessary to conduct a comprehensive analysis in this context.

In the thesis, it was argued that more transparency could facilitate the identification of forecast errors, enabling more effective improvements and fine-tuning of model predictions. Such improvements might have the potential to alter the analysis of handling volatility. The analysis did not directly establish whether LSTM models significantly contributed to predicting volatility resulting from macroeconomic distress. However, it was discussed that isolating the effect of each external variable in a manageable manner might address this issue. After observing the stock price forecast's development it appeared that isolating information from external variables, causing significant deviations in the forecast, could potentially optimize the models. In this scenario, these models could have performed better, as they demonstrated a relatively better ability to capture the fluctuations in actual stock prices compared to naive-like predictions. Hence, making adjustments to the network's features, enabling the isolation of relevant information from external variables, could be optimizing for future forecast models.

11 References

Ayodele, T. (2010). Types of Machine Learning Algorithms. DOI:10.5772/9385.

Baheti, P. (2022). Activation Functions in Neural Networks [12 Types & Use Cases]. Hentet fra V7Labs: https://www.v7labs.com/blog/neural-networks-activation-functions

BankInvest. (2022). Morgenkommentaren: Aktiemarkedet har kurs mod comeback i 2023. Hentet fra BankInvest:https://bankinvest.dk/nyheder/nyhedsartikler/2022/12/morgen kommentaren-aktiemarkedet-har-kurs-mod-comeback-i-2023/

Brownlee, J. (2019). How to use Learning Curves to Diagnose Machine Learning Model Performance. Hentet fra Machine Learning Mastery: https://machinelearningmaste ry.com/learning-curves-for-diagnosing-machine-learning-model-performance/

Brownlee, J. (2020). What is the Difference Between Test and Validation Datasets? Hentet fra Machine Learning Mastery: https://machinelearningmastery.com/difference-test-validation-datasets/

Brownlee, J. (2021). Gentle Introduction to the Adam Optimization Algorithm for Deep Learning. Hentet fra Machine Learning Mastery: https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/

Callesen, P., Thomsen, C. K., & Krogstrup, S. (2023). Årsrapport 2022. Hentet fra Danmarks Nationalbank: https://www.nationalbanken.dk/media/pn5bl1qs/2023-03-15-aarsrapport-2022-da.pdf

Chen, J. (2021). Neutral Network. Hentet fra Investopedia: https://www.investopedia.co m/terms/n/neuralnetwork.asp

DeepAI. (n.d.). Bias Vector. Hentet fra DeepAI: https://deepai.org/machine-learning-glossary-and-terms/bias-vector? fbclid=IwAR3KRwBkbM11cewMUIrxYdBG1GSVrv NOp-PrROCxoNXWOOFqu PKbF8atN9wA

Delua, J. (2021). Supervised vs. Unsupervised Learning: What's the Difference? Hentet fra IBM: https://www.ibm.com/blog/supervised-vs-unsupervised-learning/

DORS. (2022). Kapitel II Konjuktur OG Offentlige Finanser. Hentet fra https://dors.dk/fil es/media/rapporter/2022/e22/endelig_rapport/e22_kap._ii.pdf

DST. (2023). Nøgletal for nationalregnskabet (BNP). Hentet fra Danmarks Statistik: https://www.dst.dk/da/Statistik/emner/oekonomi/nationalregnskab/noegletal-for-nationalregn skabet-bnp

Du, J., Liu, Q., Chen, K., & Wang, J. (2019). Forecasting stock prices in two ways based on LSTM neural network. DOI:10.1109/ITNEC.2019.8729026.

ECB. (2023). ECB Staff Macroeconomic Projections for the Euro Area. Hentet fra ECB: https://www.ecb.europa.eu/pub/projections/html/ecb.projections202303_ecbstaff 77c 0227058.da.html

Enders, W. (2014). Applied Econometric Time Series. Wiley.

Euroinvestor. (2023). Jyske Bank. Hentet fra Euroinvestor: https://www.euroinvestor.dk/markeder/aktier/jyske-bank/75221

Euroinvestor. (2023). Novo Nordisk B. Hentet fra Euroinvestor: https://www.euroinvestor.dk/markeder/aktier/novo-nordisk-b/67613

Euroinvestor. (2023). Rockwool B. Hentet fra Euroinvestor: https://www.euroinvestor.dk/markeder/aktier/rockwool-b/291234

Euroinvestor. (2023). Royal UNIBREW. Hentet fra Euroinvestor: https://www.euroinvestor.dk/markeder/aktier/royal-unibrew/70095

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. DOI:10.1016/j.ejor.2017.11.054.

Galarnyk, M. (2022). Understanding Train Test Split. Hentet fra builtin: https://builtin.com/data-science/train-test-split

Gers, F., Schmidhuber, J., & Cummins, F. (2000). Learning to Forget: Continual Prediction with LSTM. DOI:10.1162/089976600300015015.

Gudimalla, V. (2021). Concept of Gradient Descent in Machine Learning. Hentet fra Medium: https://varshithagudimalla.medium.com/concept-of-gradient-descent-algorithm-in-machine-learning-44f587ac16ac

Hami, A., & Pougnet, P. E. (2015). Embedded Mechatronic Systems. Elsevier Ltd.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. DOI:10.1162/neco.1997.9.8.1735.

Horowitz, J. (2022). Another recession warning: Falling copper prices. Hentet fra CNN Business: https://edition.cnn.com/2022/06/24/investing/premarket-stocks-trading/index.html

IBM. (14. September 2020). What are recurrent neural networks? Hentet fra IBM: https://www.ibm.com/topics/recurrent-neural-networks

IBM. (15. July 2020). What is machine learning? Hentet fra IBM: https://www.ibm.com/topics/machine-learning

IBM. (12. March 2021). Supervised vs Unsupervised Learning: What's the Difference? Hentet fra IBM: https://www.ibm. com/cloud/blog/supervised-vs-unsupervised-learning

Jyske Bank. (2022). Et usædvanligt aktieår går på hæld. Hentet fra Jyske Bank: https://www.jyskebank.dk/nyt/privatoekonomi/et-usaedvanligt-aktieaar-gaar-paa-haeld

Kampakis, S. (2020). Performance Measures: RMSE And MAE. Hentet fra The Data Scientist: Performance Measures: RMSE And MAE

Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. DOI:10.1016/j.eswa.2022.116659.

Kumbure, M. M., Lorhmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. DOI:10.1016/j.eswa.2022.116659.

Lu, W., Li, J., Li, Y., & Sun, A. (2020). A CNN-LSTM-based model to forecast stock prices. DOI:10.1155/2020/6622927.

Markets Insider. (2023). Markets Insider. Hentet fra Markets Insider: https://markets.businessinsider.com

MarketWatch. (2023). MarketWatch. Hentet fra MarketWatch: https://www.marketwatch.com

Moolayil, J. (2019). Learn Keras for Deep Neural Networks. Apress: DOI:10.1007/978-1-4842-4240-7.

Namini, S. S., & Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM.

Nasdaq. (2023). Historiske priser - aktier. Hentet fra Nasdaq: https://www.nasdaqomxnordic.com/aktier/historiskepriser?languageId=5

Nationalbanken. (2023). Valutakurser. Hentet fra Nationalbanken: https://www.nationalbanken.dk/da/vores-arbejde/stabile-priser-pengepolitik-og-dansk-oekonomi/valutakurser

Newbold, P. (1975). The Principles of the Box-Jenkins Approach. DOI:10.2307/3007750.

Nielsen, M. A. (2015). Neural Networks and Deep Learning. Hentet fra http://neuralnetworksanddeeplearning.com/ chap2.html

Niu, H., Xu, K., & Wang, W. (2020). A hybrid stock price index forecasting model based on variational mode decomposition and LSTM network. DOI:10.1007/s10489-020-01814-0.

Nti, K. O., Adekoya, A., & Weyori, B. (2019). Random Forest Based Feature Selection of Macroeconomic Variables for Stock Market Prediction. DOI:10.3844/ajassp.2019.200.212.

Olah, C. (2015). Understanding LSTM Networks. Hentet fra Colah's blog: http://colah.github.io/posts/2015-08-Understanding-LSTMs/?fbclid=IwAR176XT5 ZJZzdBrubmE_cGTTFkbBr8S5t6ETZWBkAhI8xdBzFMMnLUEmDfo

Park, D., & Ryu, D. (2021). Forecasting Stock Market Dynamics Using Bidirectional Long-Short Term Memory. Romanian Journal of Economic Forecasting.

Persaran, M. H., & Timmermann, A. (1992). A Simple Nonparametric Test of Predictive Performance. DOI:10.2307/1391822.

Pesavento, E. (2004). Analytical evaluation of the power of tests for the absence of cointegration. DOI:10.1016/j.jeconom.2003.10.025.

Prabhakaran, S. (2019). ARIMA Model - Complete Guide to Time Series Forecasting in Python. Hentet fra Machine Learning Plus: https://www.investopedia.com/terms/n/neural network.asp

S&P Global. (2023). S&P500. Hentet fra S&P Global: https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. DOI:10.1109/ICMLA.2018.00227.

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The Performance of LSTM and BiLSTM in Forecasting Time Series. DOI:10.1109/BigData47090.2019.9005997.

Su, J., & Yang, W. (2022). Artificial Intelligence in Early Childhood Education: A Scoping Review. DOI:10.1016/j.caeai. 2022.100049.

Thune, K. (2021). Efficient Markets Hypothesis (EMH). Hentet fra The Balance: https://www.thebalancemoney.com/efficient-markets-hypothesis-emh-2466619 Voergaard, P. (2020). Hvad er sektorer? Hentet fra Euroinvestor: https://www.euroinvestor.dk/nyheder/hvad-er-sektorer

Xiao, R., Feng, Y., Yan, L., & Ma, Y. (2022). Predict stock prices with ARIMA and LSTM. DOI:10.48550/arXiv.2209.02407.

Zaiontz, C. (2021). Diebold-Mariano Test. Hentet fra Real Statistics Resources: https://www.real-statistics.com/time-series-analysis/forecasting-accuracy/diebold -marianotest/

12 Appendix A

12.1 LSTMX forecasts

Figure 25: XX LSTMX forecasts for Jyske Bank.

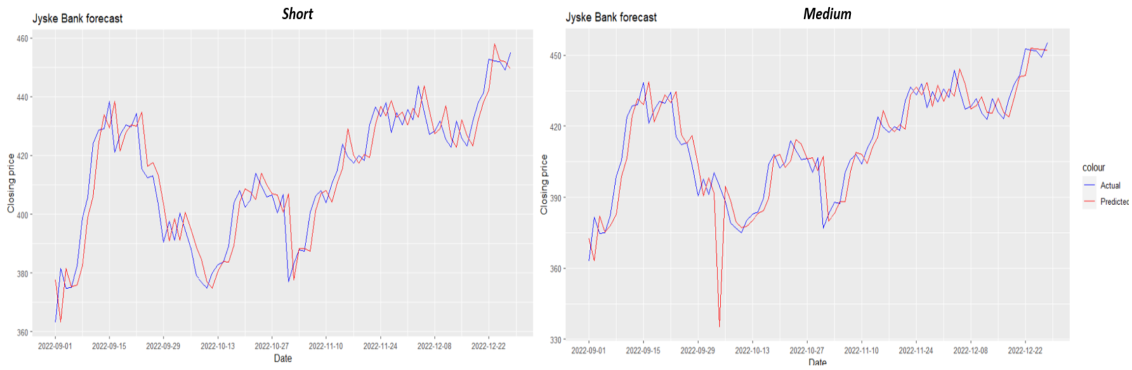


Figure 26: XX LSTMX forecasts for Novo Nordisk

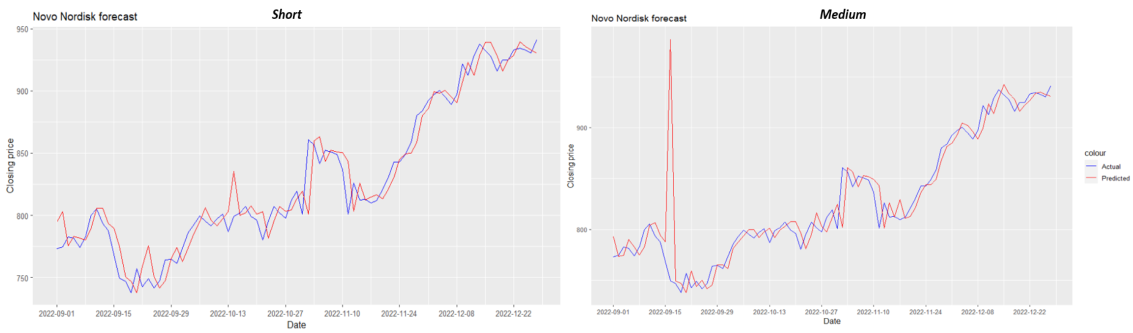
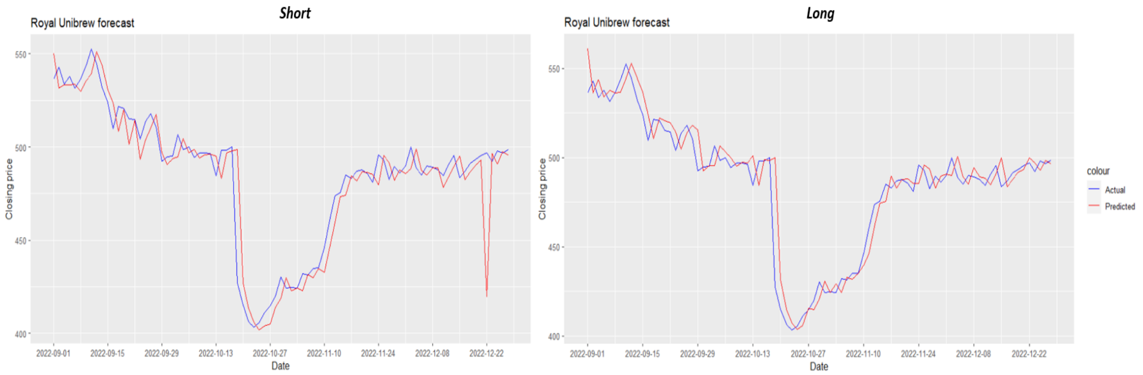


Figure 27: XX LSTMX forecasts for Rockwool.

Figure 28: XX LSTMX forecasts for Royal Unibrew.



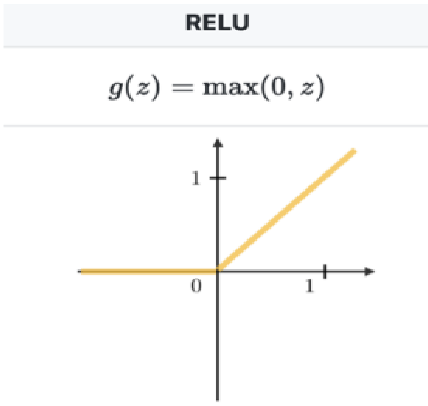
12.1.1 Activations functions

Activation functions introduce non-linearities between layers of the neural network, allowing the model to learn complex patterns. Thus, each neuron contains an activation function, which aims to transform the summed weighted input into an output value. Common activation functions include ReLU (Rectified Linear Unit), Sigmoid and Tanh. This project uses ReLU, which is one of the most widely used in deep learning due to its ability to effectively reduce the vanishing gradient problem. In addition, it is essential that the desired output must return a value between $[0, \infty]$, which is why ReLU is considered to be the most optimal. This is also the most frequently used for this purpose (Baheti, 2022), and is given by the following:

$$g(z) = max(0, z)$$

This describes that when the input z is positive, ReLU returns z, and if the input is negative, it returns 0, as illustrated in figure 28.

Figure 29: XX Rectified Linear Unit funktionen (Amidi & Amidi, 2019).



Thus, ReLU does not activate all neurons at the same time, only if the output of the linear transformation is greater than 0, making the function computationally efficient. The reason for this is that many neurons will remain inactive, i.e. return zero, resulting in efficient use of network capacity. These non-linear transformations are crucial, as networks consisting only of linear transformations between layers would collapse into a single linear transformation. This would greatly reduce the models' ability to learn and represent relationships in the data the network is trained on.