Norwegian School of Economics

Bergen, Fall 2021

# The Three Musketeers of Portfolio Allocation: Risk, Return, and Machine Learning

*A data-driven approach to portfolio allocation using machine learning and Markowitz in the Norwegian equity market*

**Hans Oscar Morstad**

**Lars Tobias Bock**

**Supervisor: Håkon Otneim**

Master thesis, MSc in Economics and Business Administration

NORWEGIAN SCHOOL OF ECONOMICS

# Abstract

The portfolio selection problem is one of the most discussed topics in financial literature. Harry Markowitz (1952) is recognized as the first to formalize the risk-reward trade-off methodology used in portfolio selection. Through his mean-variance framework, he detailed the importance of diversification and laid the foundation for the modern portfolio theory we know today.

This thesis explores a novel approach to portfolio allocation enabling the mean-variance framework and machine learning. We employ machine learning to predict the quarterly expected return and the associated covariance matrix for stocks trading on Oslo Stock Exchange. To construct the predictions, we deploy the renowned Extreme Gradient Boosting algorithm, also called XGBoost. We investigate the opportunity to use quarterly reports, macroeconomic and economic variables as predictors of quarterly stock returns and covariances. Furthermore, we apply these predictions in the mean-variance framework from Markowitz to construct quarterly portfolios.

The results from the Thesis Model are disappointing. The objective of the quarterly portfolio optimization is to maximize the Sharpe ratio. Unfortunately, the Thesis Model is not able to construct portfolios that reliably aligned with this goal. Nevertheless, the model initially yields an impressive one-year return. However, under new conditions the performance change drastically. The statistical evaluation of the XGBoost prediction models entails that they both deliver highly inaccurate predictions, which propagates further through to the portfolio allocation process. Moreover, there is little evidence that the models can detect any patterns in the data beneficial for portfolio construction. In sum, the model struggles to foresee market developments, which accumulates into a model incapable of consistently performing with satisfying financial results.

# Acknowledgments

This thesis was written during the fall of 2021 as part of the MSc in Economics and Business Administration program at the Norwegian School of Economics (NHH), with majors in Financial Economics and Business Analytics.

The process of writing this thesis has been educational and challenging. However, despite the challenges we have had to overcome, we will reminisce the process as fun and rewarding.

We would like to send our gratitude to our supervisor, Associate Professor Håkon Otneim. During the writing process, his guidance and support have been imperative to this project. Furthermore, we would like to thank Mr. Otneim for enabling us to explore this novel approach to portfolio optimization.

Lastly, we would like to thank each other for an impeccable partnership and a great semester.

# Table of Contents

# List of Figures



# List of Tables

# 1. Introduction

Despite a raging pandemic, financial markets have never been more pertinent. All over the world, stock indexes reach new peaks, cryptocurrencies are prominently advancing, while social media platforms cause margin calls and demolish entire hedge funds. In a Norwegian context, Oslo Stock Exchange is at an all-time high, and private investors have never used the Norwegian equity market more frequently. Per July 2021, around 530 000 Norwegians were investing in the Norwegian stock market, and the current number is expected to be even higher (Aksjenorge, 2021). As such, it is fair to say that stock returns seldom have been more topical than they are today.

Active investors in financial markets usually own several assets. They all desire to combine a portfolio of assets capable of consistently “Beating the Market”. In other words, active portfolio managers seek to compose a portfolio that delivers excess return against a benchmark index with a given amount of risk (Qian, Hua, & Sorensen, 2007). One of the most common frameworks in portfolio selection is the mean-variance model (MV), proposed by Harry Markowitz in 1952. The objective of the MV-model is to either maximize the Sharpe ratio, maximize the return for a preferred level of risk, or minimize the risk for a certain return. This framework aims to select the optimal allocation of weights for assets in a portfolio. The weight allocation is based on the expected returns of the assets and their associated volatility captured in the covariance matrix. Conclusively, estimates of expected returns and the covariance matrix are required to select the optimal portfolio following the MV-model (Markowitz, 1952).

The first part of the Markovitz framework is expected returns. Estimation of expected returns is broadly addressed in literature (Green, Hand, & Zhang, 2013). Commonly, these estimations are predictions of some sort. Still, despite all the research, stock return prediction is not easy. The stock market is, in general, characterized as dynamic, unpredictable, non-stationary, and non-linear (Vijha, Chandolab, Tikkiwalb, & Kumarc, 2020). There are various factors to these dynamics. A non-comprehensive list of influencing factors includes political conditions, global and local economy, company-specific financial reports, macroeconomic factors, and the psychology of investors (Henrique, Amorim, & Kimura, 2019). Consequently, due to these dynamics, stock return prediction is established to be a risky and challenging task. However, the search for the “perfect” prediction model is a constant pursuit among investors and fund managers worldwide.

Although the execution is complex, the concept of stock return prediction is simple. In essence, it conceptualizes the thought of determining whether the value of a stock is about to increase or decrease. Despite the renowned Efficient Market Hypothesis (EMH) Fama (1970), the search for effective financial models has been part of finance for a long time. An uncomprehensive list of approaches deployed to solve this challenge includes the Single Index Model (Sharpe, 1963), Capital Asset Pricing Model (Mossin, 1966), historical average returns (Markowitz, 1952), stock price momentum models (Jegadeesh & Titman, 1993) and fundamental models (Graham, 1949). All have their flaws and weaknesses emphasizing the complexity of stock price prediction.

Nevertheless, creating financial models with predictive abilities is still highly attractive in academia (Henrique, Amorim, & Kimura, 2019). Though, the motivation is obvious. Academically, a model capable of predicting future returns above market indices would provide strong evidence contrary to the Efficient Market Hypothesis, one of the most famous economic theories there is. Financially, such a model would unchain the opportunity to gain significant short-term profits within financial markets before the new information is incorporated into the market.

In broad terms, there are two approaches to stock price prediction: technical analysis and fundamental analysis (Henrique, Amorim, & Kimura, 2019). The foundation of technical analysis is that history tends to repeat itself and that this applies to financial markets and markets patterns (Achelis, 2000). The main principle of technical analysis is to identify and use patterns and indicators from historical prices to predict future prices (Kirkpatrick & Dahlquist, 2016). Furthermore, technical analysis relies on internal market information and assumes all predictive factors of stock price fluctuations to be hidden in the stock price (Chang, Liao, Lin, & Fan, 2011). Hence, technicians argue that stock prices can be predicted using historical patterns and signals. However, assuming the EMH holds on at least a weak form for Norway and Oslo Stock Exchange, any prediction of future returns based on previous returns seems hollow for stocks on Oslo Stock Exchange or any stock exchange for that matter.

The second approach to stock return prediction is fundamental analysis. Introduced by Benjamin Graham in 1949, fundamental analysis is still a widely used method to predict the future value of an asset. In essence, fundamental analysis is a qualitative approach where internal factors such as company-specific financial statements are combined with external, non-company-specific factors like the current market and macroeconomic situation (Graham, 1949). Together, these factors are used to determine the intrinsic value of a firm and identify mispriced securities (Hur, Manoj, & Yohanes, 2006). Thus, fundamentalists estimate the development in stock prices based on financial analyses of companies or industries.

The traditional approach to fundamental analysis is to use publicly available data. Often, this involves the financial statements of companies, which are used to construct financial ratios. These ratios are then used to determine whether a company is undervalued or overvalued based on the historical development and industry peers. Moreover, the ratios can indicate potential growth opportunities and expose the financial health of a company. Nevertheless, there is more to fundamental analysis than just financial statements. For example, economic factors such as interest rates and general macroeconomic factors are important features in fundamental analysis. These additional factors aim to portray the general development in the economy, providing a more nuanced picture of the market, which is important when estimating stock returns.

The second part of Markowitz optimization is volatility, represented by the covariance matrix of the associated assets. Covariance is an imperative concept in finance, especially in portfolio construction. Covariance is used to measure the state of instability between returns. A standard assumption in finance is that the covariances of stock returns are more stable than the returns themselves (Merton, 1980). This implies that the historical covariance could reasonably estimate future covariance. Furthermore, Merton (1980) argues that the impact of the expected returns is more significant than the impact of the covariance estimations. As such, changes in the estimated covariance matrix do not entail a considerable difference in the portfolio composition. In contrast, slight changes in the return estimates can cause significant changes in the portfolio composition (Awoye, 2016). Nevertheless, covariance is still an influential factor in MV portfolio selection.

When predicting the covariance of stock returns, one common approach applies historical covariance as a proxy to estimate future covariance (Markowitz, 1952) (Markowitz, 1999). However, there is clear empirical evidence that the assumption of constant financial covariance is ambitious (Engel & Gizycki, 1999). Although the volatility of financial time series can be clustered and relatively stable in certain periods, extensive research shows that financial covariance varies over time (Engle, Ledoit, & Wolf, 2017). This phenomenon is known as heteroscedasticity. One example is displayed in Figure 1, showing how the monthly returns on Oslo Stock Exchange Benchmark Index (OSEBX) have fluctuated from Q1 2000 to Q2 2021.

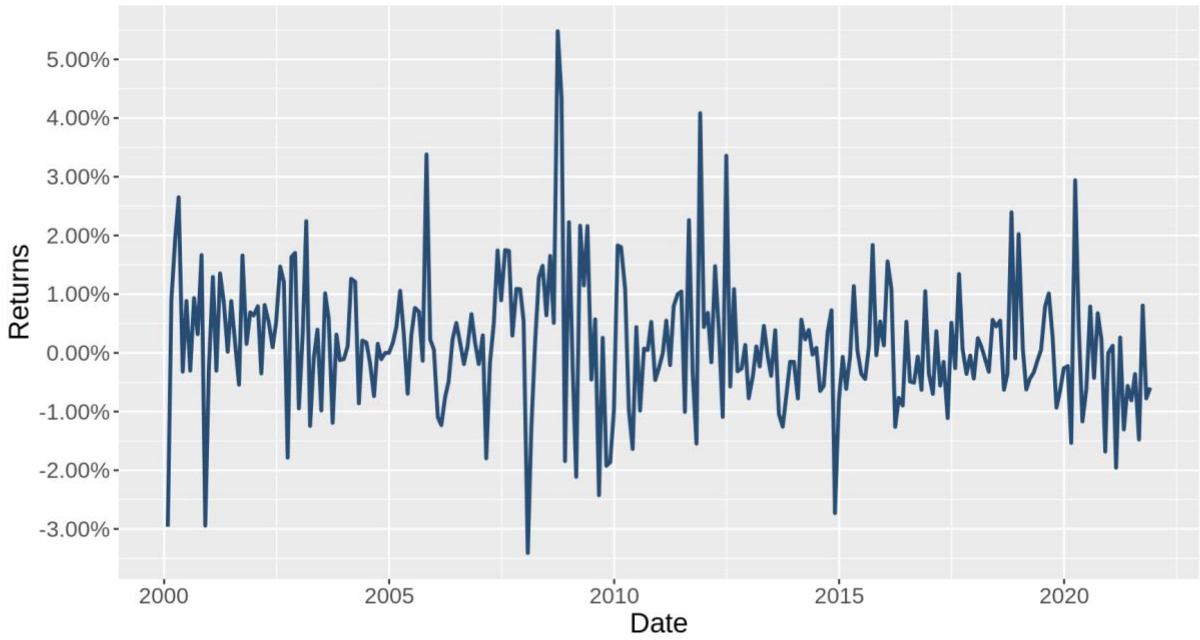


*Figure 1, Monthly returns for OSEBX*

When we study Figure 1 it is evident that the variance is far from constant. The apparent instability of the volatility exemplifies that the historical covariance approach, suggested by Markowitz, is deficient. Moreover, studies have recognized that past financial data influences future data (Engel R. , 1982). In statistical terms, this means that financial data is known to be autoregressive. Fortunately, there are models capable of capturing the mentioned characteristics with great precision. This highlights an interesting paradox. While predictions of first-order moments (stock returns) are especially challenging, predictions of the second-order moments (variance/correlation) are more reliable (Nelson & Foster, 1992). One of the most recognized methods to predict second-order movements is the ARCH/GRACH framework.

Rober F. Engel (1982) introduced the Autoregressive Conditional Heteroscedasticity (ARCH) framework. ARCH models are able to resemble the volatility clustering observed in asset returns and have two important assumptions. First, ARCH assumes that the shock $a_t$ of an asset return is serially uncorrelated with nonconstant variances conditional on the past while having constant unconditional variances. Second, ARCH models assume the dependence of $a_t$ to be described by a quadratic function of lagged values (Tsay, 2005). Specifically, the shock is defined as:

$$\alpha_t = \sigma_t \epsilon_t \tag{1.1}$$

where $\epsilon_t$ is a sequence of random variables with a mean 0 and a variance 1 following the same distribution. Formally, we can write the ARCH($p$) as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \cdots + \alpha_p a_{t-p}^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-1}^2 \tag{1.2}$$

From basic statistics, it is evident that variance must be non-negative. Thus, ARCH requires the coefficients $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i > 0$ to guarantee that the unconditional variance of $a_t$ is finite (Tsay, 2005).

Although the ARCH framework has interesting properties, it is not perfect. One apparent drawback of the ARCH framework is that it requires many parameters to estimate the return volatility of an asset effectively. Tim Bollerslev (1986) introduced the generalized autoregressive conditional heteroskedasticity (GARCH) framework to cope with this. The GARCH($p, q$) is shown in equation 1.3.

$$a_t = \sigma_t \epsilon_t \tag{1.3}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \tag{1.4}$$

where $\beta_j \geq 0$ and $\alpha_1 + \cdots + \alpha_p + \beta_1 + \cdots + \beta_q < 1$. As GARCH is an extension of ARCH, the differences are not huge. In summary, GARCH($p, q$) includes lagged conditional variances, whereas ARCH($p$) only consists of the conditional variance specified as a linear function of past sample variances (Bollerslev, 1986). As such, the GARCH framework corresponds to some adaptive learning mechanisms. Additionally, it enables modeling of conditional change in variance over time and changes in the time-dependent variance (Tsay, 2005).

The introduction of the ARCH and GARCH frameworks revolutionized the estimation of time series volatility in finance and economics. These frameworks can efficiently model volatility of financial assets prices such as bonds, market indices, and stocks. Moreover, they enable forecasts of the entire distribution, not just the mean as in ordinary regression problems (Tsay, 2005). Inspired by these capabilities, other varieties of the GARCH framework have been introduced. A popular approach is to apply multivariate GARCH models. Among these models, the Dynamic Conditional Correlation (DCC) GARCH model is one of the most recognized in the context of modeling financial time series (Engel & Sheppard, 2001) (Fiszeder & Orzeszko, 2021). Nevertheless, although GARCH/ARCH framework provides evidence that we can forecast volatility efficiently, the task is not easy (Chan, Karceski, & Lakonishok, 1999).

Today, different variations of GARCH are widely used in volatility prediction. However, there exist alternative approaches to volatility prediction. Recent development in technology and methodology have brought up novel estimation techniques. These new techniques are applicable for both volatility and return predictions, and subsequently, portfolio optimization. Among these, Machine learning has become one of the most popular. The most apparent explanation is the ability of machine learning algorithms to handle the chaotic and non-linear nature of financial markets. Consequently, these algorithms can identify non-linear patterns and relations that previously were non-disclosable in the financial markets (Fiszeder & Orzeszko, 2021) (Basaka, Kar, Saha, Khaidem, & Sudeepa, 2019). Hence, the application of machine learning in finance has increased drastically in the last couple of years, where the typical approach is to predict either volatility *or* returns.

## 1.1 Scope of the Thesis

This thesis investigates a novel approach to portfolio optimization using machine learning. Instead of estimating either the expected return or the volatility of considered assets, we aim to do both, using two separate models. We want to examine a machine learning approach to quarterly portfolio selection using predictions of quarterly expected returns for the considered stocks and the associated covariance matrices. Moreover, we rely our portfolio allocation on the MV model introduced by Markowitz. In other words, the construction criterium for our portfolio optimization is to maximize the Sharpe ratio. Further, most research on portfolio optimization and stock covariance prediction has focused on markets outside Europe, mainly in the US. Hence, there is little research on the Norwegian stock market. Thus, only stocks traded on Oslo Stock Exchange are considered throughout the thesis.

Both prediction models will enable the concept of fundamental analysis. To predict stock returns, we investigate the use of financial ratios based on company-specific data from quarterly financial reports. In addition, non-company-specific data such as macroeconomic and economic indicators are included as predictors to capture the overall market movements. The same non-company-specific data is also used in a separate model to predict the quarterly covariance between the stocks. Both models will apply the XGBoost algorithm proposed by Chen and Guestrin (2016). XGBoost is, in addition to the mentioned machine learning advantages, renowned for its efficacy, computational speed, model performance, and handling of missing values satisfactory (Nielsen, 2016).

## 1.2 Motivation

The risk aspect is often neglected in the discussion of stock return prediction. Instead, the traditional focus is centered around predicting winners and losers. On the contrary, there is a well-known risk-reward trade-off between the return of an investment and the risk involved. Thus, portfolio selection without accounting for risk is not very accurate. The rational investor always wants to adjust the portfolio to the associated risk of the considered assets. Consequently, the idea of predicting financial returns without accounting for risk appears hollow. One of the best risk-reward trade-off measures in portfolio selection is the Sharpe ratio. Thus, we use the maximation of the Sharpe ratio as the selection criterium for our portfolios.

Second, machine learning and financial time series are known to have matching characteristics. The chaotic and non-linear nature of financial markets aligns well with the capabilities of machine learning to handle and determine non-linear patterns. Moreover, literature shows evidence that machine learning has predictive capabilities for financial time series, both in terms of volatility and returns. However, there is modest research on the topic of machine learning in portfolio optimization. Therefore, an imperative feature to the motivation to this thesis is the unexplored area of portfolio optimization using machine learning predictions for both returns and volatility.

Quarterly portfolio optimization using financial data is partly motivated by an unorthodox valuation method from the Norwegian portfolio manager Thomas Nielsen. He uses a self-developed model that applies fundamental analysis to select the stocks in his portfolio consisting of Nordic companies (Bjergaard, 2020). Moreover, past literature suggests that fundamental analysis, as opposed to technical analysis, is suitable for long-term stock-price movement but not suitable for the short-term stock-price change (Khan, Alin, & Akter, 2011). Also, due to transaction costs and other barriers, most investors have a long-term horizon on their investments. Therefore, we do not consider a short-term approach predicting daily returns.

## 1.3 Document Structure

The continuation of this thesis is divided into six parts. We will continue with some theoretical background on portfolio optimization and the selection approach for contextualization purposes. Then, we present a thorough literature review on the topic of return and covariance prediction using machine learning. Additionally, we assess the linearities in financial markets before discussing the assumptions in the Markowitz framework. Next, we introduce the financial and technical framework of the thesis in the methodology chapter. Following the methodology chapter is an introduction to the data we use and a run-through of the optimization model we apply. Finally, we present and discuss the results from the portfolio optimization before we submit some concluding remarks.

# 2. Introduction to Portfolio Theory

A portfolio might consist of only one type of asset or a combination of different assets like stocks, bonds, real estate, and more. Further, the number of combinations regarding weights of different assets in a portfolio is, in theory, infinite. Even in practice, the number of asset types and weights is close to countless, especially as the number of assets rises and short selling is allowed. As such, portfolio allocation is one of the greatest challenges in finance (Goldfarb & G. Iyengar, 2003).

The challenge of portfolio selection has been carefully discussed in the academic literature. Most discussions rely on the work of Harry Markowitz, known as the father of modern portfolio theory. In his book, *Portfolio Selection,* Markowitz (1952) introduced the modern aspect of risk in portfolio theory. Before the publication, portfolio theory lacked (1) sufficient coverage of the effects of diversification when risks are correlated, (2) the distinguishment between efficient and inefficient portfolios, and (3) analyses of risk-return trade-offs for portfolios (Markowitz, 1999). Markowitz covered these topics and established the modern portfolio theory we familiarize ourselves with today.

Markowitz (1952) argues that any rational investor is risk-averse. This implies that an investor will only take on increased risk if compensated with a higher expected return. Hence, an optimal selection of assets-weights yields the highest feasible expected return for a given level of risk. Such a portfolio is called an efficient portfolio (Markowitz, 1952). Moreover, Markowitz (1952) argues that predictions of security returns follow the same probability postulations as random variables do.

The fact that asset returns follow the same probability postulations as random variables do has two key takeaways. The first is regarding the expected return of the portfolio. The expected return of a portfolio is the weighted average of the expected returns of the individual securities in the portfolio. On this topic, one of the fundamental assumptions of Markowitz is that the returns are multivariate normal, meaning they are symmetric, have short tails, etc. The second takeaway is regarding the variance of the portfolio. The variance of a portfolio is, according to Markowitz, an explicit function of the standard deviation of the individual securities, the covariances between said securities, and their weights in the portfolio. Thus, an investor should obtain a diversified portfolio by avoiding combining highly correlated securities in a portfolio (Markowitz, 1952).

The expected return of a portfolio $P$, consisting of $n$ assets, can be formulated using the definition of the expected value.

$$\widehat{\mu_P} = w_1 \cdot \hat{\mu}_1 + w_2 \cdot \hat{\mu}_2 \ldots w_n \cdot \hat{\mu}_n = \sum_{i=1}^{n} \hat{\mu}_i w_i \tag{2.1}$$

where $w_i$ is the weight for security $i$ in the portfolio, and $\hat{u}_i$ is the expected return. Subsequently, we can write the formula for the portfolio variance as follows

$$\sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov_{ij} \tag{2.2}$$

where $\sigma_i$ is the sample standard deviation of the returns for security $i$ and $\rho_{ij}$ is the correlation of the returns between security $i$ and $j$.

The portfolio variance formula implies that returns from different assets covariate. Thus, certain combinations of assets can drag the variance of a portfolio down. On the other hand, some assets covariates, causing the portfolio variance to increase. As such, different combinations of assets provide different portfolio variances. The process of choosing a combination of assets to lower the risk of a portfolio is called *diversification* (Rubinstein, 2002). Diversifying portfolios is the key to obtaining efficient portfolios with an optimal risk-return trade-off (Lohre, Opfer, & Orszag, 2011). Figure 2 exemplifies the value of diversification. The figure shows the expected return and expected volatility (standard deviation) for stocks A, B, and C. Creating portfolios consisting of the three securities makes it possible to generate a variety of expected returns and standard deviations. The portfolio compositions with the highest possible expected return for any given level of risk can be plotted as illustrated with the blue line. This line is called the efficient frontier and represents the optimal combination of assets for any possible level of risk (Markowitz, 1952). As shown in figure 2, a combination of the three stocks can achieve a more optimal risk-return trade-off than any individual stock.

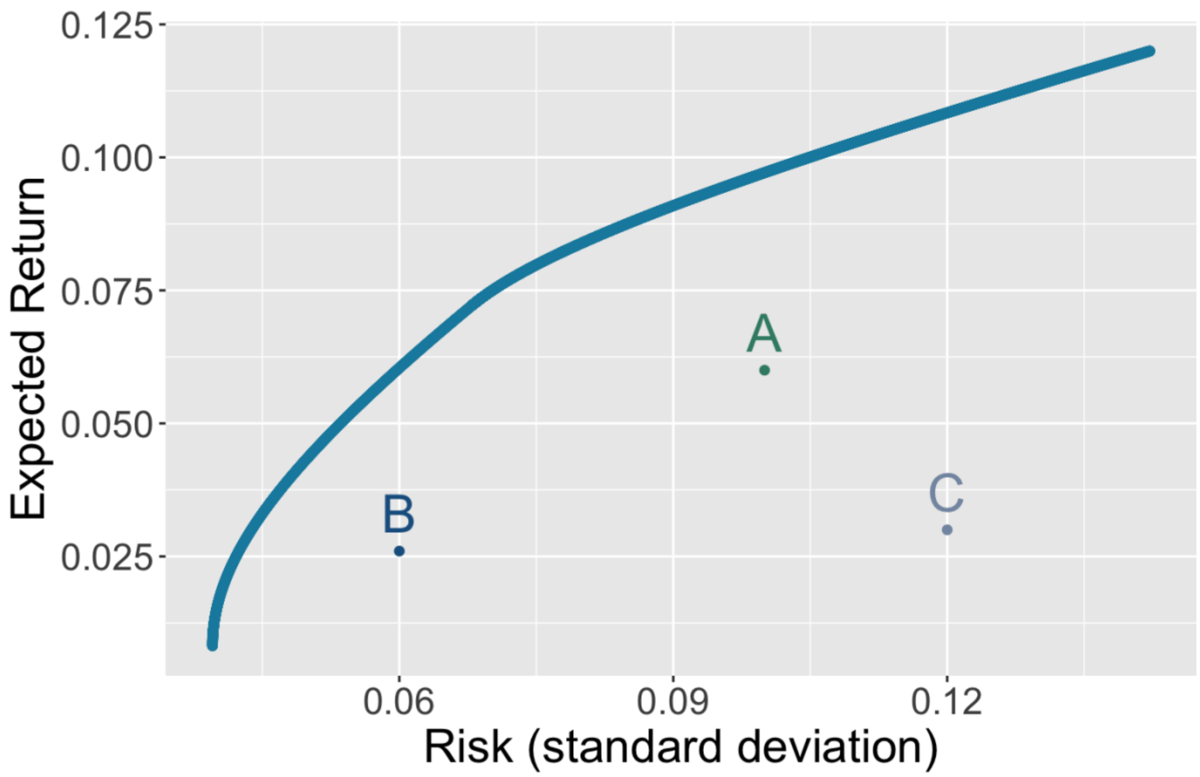


*Figure 2, Efficient Frontier (blue line)*

## 2.1 Sharpe ratio

The Sharpe ratio was introduced in 1966 under the name *reward-to-variability ratio* by William Sharpe. He proposed a ratio to measure the risk-adjusted performance of a portfolio, where risk is represented by the standard deviation of the portfolio (Sharpe W. F., 1966). We can calculate the ratio is calculated the following approach

$$SR = \frac{\mu - r_f}{\sigma} \tag{2.3}$$

where $\mu$ is the return of the portfolio, $r_f$ is the risk-free rate with an equivalent time horizon, and $\sigma$ is the risk of the portfolio denoted as the standard deviation of the portfolio returns, also called volatility (Eling & Schuhmacher, 2007).

Comparing different investment strategies with various risks and returns is a non-trivial task. A strategy might have a higher expected return than other strategies. However, the same strategy might also have a higher expected risk. Hence, joint ratio calculating the risk-adjusted expected return makes comparing different investment strategies easier due to a common ground (Dowd, 2000). This exemplifies how the Sharpe ratio can be used in an investment decision manner, ex-ante. If the Sharpe ratio is used ex-ante, the variables in equation 2.3 are estimated (Sharpe W. F., 1966). Alternativity, the Sharpe ratio can be used as a performance measurement, ex-post. In short, the ex-post Sharpe ratio uses actual returns and volatility to compute the Sharpe ratio of a desired period in the past. In other words, the ex-post Sharpe ratio is backward-looking and is often used for evaluation purposes. Calculating the ex-post Sharpe ratio enables us to compare the performance of different portfolios in terms of the risk-return trade-off (Eling & Schuhmacher, 2007).

The Sharpe ratio varies across the efficient frontier. To locate the maximum Sharpe ratio graphically on the efficient frontier, using the capital allocation line (CAL) is necessary (Bodie, Kane, & Marcus, 2018). When finding the maximum Sharpe ratio, the CAL is created using the risk-free rate as intercept and the following tangency point on the efficient frontier. See figure 3 for illustration, where the black line is the CAL, and the bright blue dot is the maximum Sharpe ratio point. The difference between the expected return on the allocation line and the risk-free rate is called the risk premium and is the premium an investor receives for taking on risk. The tangency point between the efficient frontier and the capital allocation line represents one specific portfolio with specific weights for the associated assets. This portfolio yields the highest Sharpe ratio of all the possible combinations

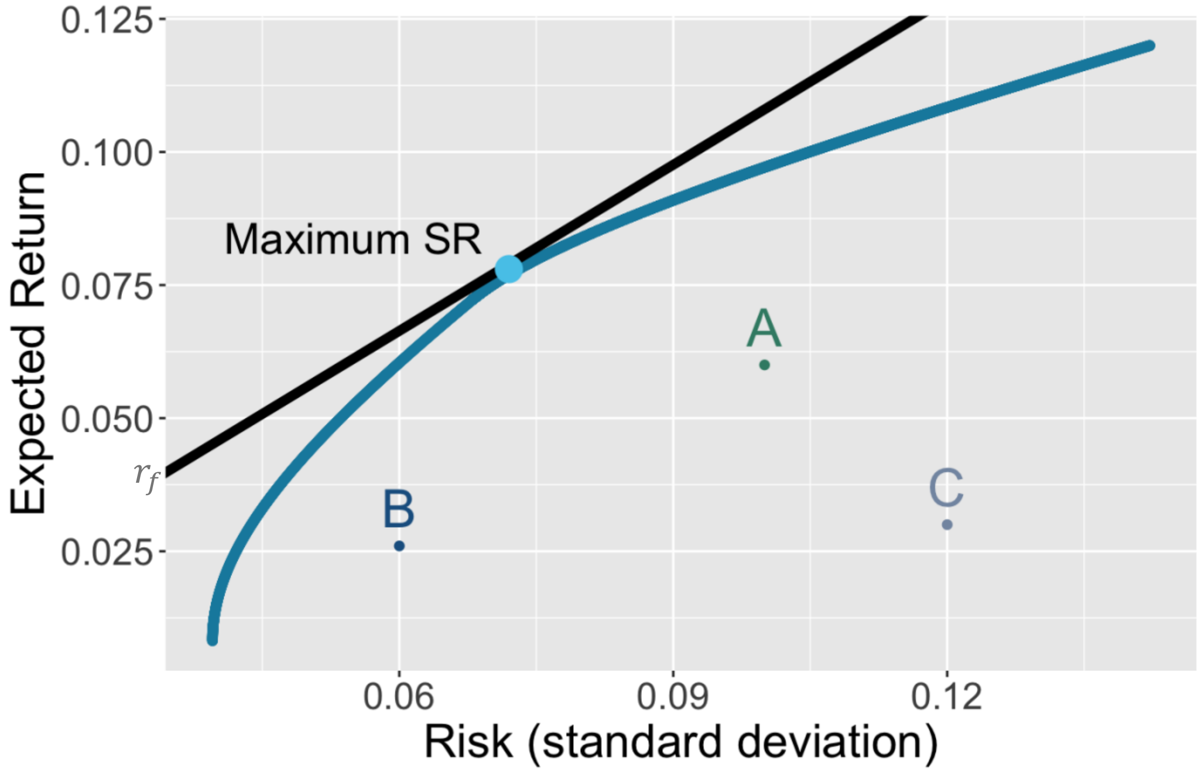


*Figure 3, Efficient frontier (blue line) with CAL (black line) and maximum Sharpe ratio point (bright blue dot)*

## 2.2 Market efficiency

Market efficiency is a vital concept in finance. The term efficiency portrays a capital market in which all relevant information is embedded in the price of a financial asset (Dimson & Mussavian, 1998). A capital market fuels growth and expansion through financing (Coşkun et al., 2017). Furthermore, the primary task of a capital market is to allocate ownership of the capital stock in the economy (Fama, 1970) (Fama, 1965). This task is exercised by providing a market for investors and companies to trade securities affiliated with company ownership. Such assets include, but are not limited to, shares, bonds, real estate, and more. For capital markets to function optimally, these markets need to consist of assets that are priced to reflect all available information at any given time fully. As such, a capital market is efficient if it manages to do so (Fama, 1970).

Eugene Fama proposed in 1970 three levels of market efficiency. *Weak* form, also called random walk theory, portrays a market where all information about previous security prices is impounded in the current price of the security. Hence, excess returns cannot be obtained by analyzing characteristics of previous price development for a security, such as trend, seasonality, and variations. *Semi-strong* form depicts a market where security prices fully and fairly reflect all publicly available information. This includes accounting information, merger and accusation transactions, management structure, etc., in addition to all information mentioned in weak form efficiency. Lastly, *Strong* form of market efficiency states that the price of a security reveals all information affecting said price, public or inside regardless.

In his initial paper on market efficiency, Fama (1970) argues strong evidence for at least a weak form of market efficiency in capital markets. Later research shows that this is partly true, and more so for developed than for emerging equity markets (Chan et al., 1997) (Mobarek & Fiorante, 2014). European developed equity markets such as Norway, Germany, Sweden, Portugal, Ireland, France, United Kingdom, Finland, Spain, and the Netherlands satisfy most of the requirements of a strict random walk regarding daily stock returns. The presence of a strict random walk regarding daily stock returns confirms the presence of weak market efficiency in the mentioned markets, including Norway (Worthington & Higgs, 2004).

# 3. Literature Review

In essence, modern portfolio theory is conceptualized by risk and return. Thus, when we construct portfolios for investment purposes, it is imperative to consider the expected return and the expected risk associated with each portfolio. In the mean-variance framework, the expected return is the weighted aggregation of the expected returns, while the risk is embodied in the covariance between the considered assets. This chapter will present a thorough literature review on stock return prediction using fundamental analysis and covariance prediction in a machine learning context. Furthermore, we will elaborate on some of the fundamental assumptions in the mean-variance model by Markowitz.

## 3.1 Fundamental Analysis and Stock Returns Prediction

The concept of fundamental analysis dates back to Benjamin Graham. In essence, fundamental analysts believe that internal and external factors reflect the stock price of a company (Cavalcante, Brasileiro, Souza, Nobrega, & Oliveira, 2016). Graham argued there were three fundamental measures investors should notice: the size of the firm, the capitalization, and the price-earnings ratio (Graham, 1949). Using the work of Graham as a basis, fundamentalists apply a combination of financial ratios computed from financial statements and stock price, combined with other quantitative and qualitative tools to determine the value of the stock (Lam, 2004).

When evaluating the value of a firm, one of the most widespread methods used is to assess future cash flows. Another popular method is comparing the desired company to other comparable companies. Both these methods rely on the idea that financial data, combined with market data, carry essential information about the value of a firm (Hong & Wu, 2016). Stock returns are thus related to capital investment, earnings yield, growth opportunities, and changes in profitability, as well as changes to the discount rate used for discounting further cash flows (Chen & Zhang, 2007). The importance of financial data on longer-term stock returns has been empirically proven through extensive research (Hong & Wu, 2016).

Ou and Penman (1998) use annual statements from the industrial, utility, and financial stocks trading on NYSE and AMEX from 1965-1972 to estimate yearly predictions for the period 1973-1983. Muhammad (2018) uses on his side 115 non-financial companies trading on the Karachi Stock Exchange to study the relationship between fundamental data and stock returns.

Both studies show that fundamental analysis has predictive power regarding stock returns using historical accounting data from financial statements. Abarbanell and Bushee (1998) use data from NYSE from 1974-1993 and both the relative change and the absolute value of financial ratios to predict returns and select portfolios. The results of this approach were a portfolio that earned an average 12-month cumulative size-adjusted abnormal return of 13.2 percent, indicating that fundamental signals combined with absolute financial ratios from accounting data can be used to predict abnormal returns. Furthermore, Abarbanell and Bushee argue that their findings insinuate that abnormal stock returns can be achieved using fundamental analysis and fundamental signals (Abarbanell & Bushee, 1998)

San and Hancock (2012) study the relationship between accounting data, macroeconomic data, and forecasted returns. Over the period 1990-2000, financial statements and macroeconomic data from 33 countries were analyzed, providing similar results as mentioned above. However, the paper discovers differences in the long- and short-term stock return predictions, highlighting the long-term effect of macroeconomic changes on stock prices. Contributing to this topic, Bertuah & Sakti (2019) argue that using a combination of financial performance and macroeconomic factors influences long-term stock returns. Moreover, studies show that macroeconomic factors correlate with stock returns, with varying impacts (Flannery, Protopapadakis, & Notes, 2002) (Tangjitprom, 2012). In recent decades, macroeconomics and economic factors have been significant variables for mid-to-long-term movements in stock prices. Industrial production, national output, long-term interest rates, exchange rates, and inflation have proven to be important factors and variables (Peiró, 2016) (Mahmood & Dinniah, 2007).

### 3.1.1 Stock Price Prediction Using Machine Learning

Eakins & Stansell (2003) suggest that stock price forecasting using a neural network model and ratios from fundamental analysis yields outperforming returns. Using financial data for all stocks listed on Compustat from 1975 to 1996, filtering out small and highly volatile stocks, their model outperformed the S&P 500 index and Dow Jones Industrials by 5.7% and 5.6%, respectively, over the 20 years. Similarly, Huang, Capretz, & Ho (2019) utilized machine learning and neural networks to predict stock returns. Utilizing quarterly data from Q1 1996 to Q4 2017 for 70 stocks from the S&P 100 index, they construct monthly “Buy” and “Sell” portfolios using financial ratios. The period Q1 1996 to Q2 2013 was used as training data while testing the model on the remaining data. The results show that the model excellently separates winners and losers, thus selecting portfolios outperforming the benchmark index in the test period.

Namdari and Li (2018) compare technical and fundamental analysis in a machine learning context. They explore data consisting of 12 selected financial ratios and stock prices of 578 technology companies on NASDAQ in the period 2012-06 to 2017-02. Two separate models were created to compare the two approaches to predict stock returns. The first model uses a fundamental approach utilizing the 12 financial ratios, while the second model uses technical signs from historical prices for the same companies. The results show that the model based on fundamental analysis outperforms the alternative model (Namdari & Li, 2018).

## 3.2 Covariance Prediction

Forecasting volatility has been interesting for researchers within finance for a long time (Trucíos, Zevallos, Hotta, & Santos, 2019). One of the most popular approaches to covariance prediction is multivariate GARCH models. However, studies show that the results of multivariate GARCH models perform poorly when handling large portfolios (Engle, Ledoit, & Wolf, 2017). Machine learning is an alternative to the common econometric approaches. Research shows that machine learning can outperform econometric models. The most prominent advantages of machine learning are offering a more generalized approach than standard statistical models (Altman, Bzdok, & Krzywinski, 2018) (Makridakis, Spiliotis, & Assimakopoulos, 2018).

Prediction of covariance matrices is still challenging. First, one of the cornerstones of multivariate volatility modeling is that the predicted covariance matrices must be positive definite (Chiriac & Voev, 2011). Second, to limit computational challenges and limit the inflation of the number of estimated parameters, the model dynamics are often limited due to the imposition of parameter restrictions (Fiszeder & Orzeszko, 2021). To cope with these challenges, predicting Cholesky factors decomposed from covariance matrices is renowned for being one of the most recognized solutions (Andersen, Bollerslev, Diebold, & Labys, 2003) (Chiriac & Voev, 2011).

### 3.2.1 Covariance Prediction Using Machine Learning

In the context of portfolio optimization, the literature on covariance prediction using machine learning is limited. Cai, Xianggao, Lai, and Lin (2013) and Bucci (2020) attempt to use a neural network approach to forecast covariance. Both papers utilize Cholesky decomposition of the covariance matrix to predict the covariances and obtain promising results. Based on the results from the two cited papers, Fiszeder, Orzeszko, & Witold (2021) apply Support Vector Machines to predict range-based covariance matrices of returns from the currency pairs EUR/USD, USD/JPY, and GBP/USD in the foreign exchange market. Once again, the Cholesky decomposition was used to ensure the positive definiteness of the predicted covariance matrices. The results show that machine learning can provide more accurate predictions than benchmark models like GARCH-X or DCC GARCH (Dynamic Conditional Correlation) (Fiszeder & Orzeszko, 2021).

## 3.3 Non-linearity of Financial Time-series

The empirical properties of financial time series are extensively discussed in economic literature. An incomplete list involves the autocorrelation of returns, volatility clustering, leverage effects, dependencies between assets, and distribution characteristics, such as fat tails, leptokurtosis, and asymmetry (Fiszeder & Orzeszko, 2021) (Tsay, 2005). Moreover, there is an empirical joint understanding that financial markets can be nonlinear. Examples include energy futures (Mariano, 2007), emerging stock markets (Kian-Ping, Brooks, & Hinich, 2008), currency markets (Sadique, 2011), and equity portfolios (Wey, 2018). Most machine learning algorithms do not presume linearities in the prediction and are thus known to perform well in predicting financial time series.

## 3.4 Assumptions Made by Markowitz

One of the most fundamental assumptions of the mean-variance model by Markowitz (1952) is that security returns follow a joint-Gaussian distribution. Following this assumption, asset returns are assumed to be multivariate normal distributed. Furthermore, the assumption implicates that mean security return and mean associated variance are reliable estimates for the future asset return and asset variance. However, empirical studies have exposed asymmetries in the distribution of financial returns (Sleire, et al., 2021). Asymmetric financial return distributions suffer from characteristics such as positive excess kurtosis and skewness (Blanca, Arnau, López-Montiel, Bono, & Bendayan, 2013). This means that substantial negative returns, which occur in very bearish markets like a recession, are more strongly correlated than returns are in a bullish market. Such a mechanism diminishes the diversification effect when it is needed the most. Therefore, when utilizing historical returns samples to determine expected returns and the associated covariance matrix, the likelihood of estimation error increases (Low, Faff, & Aas, 2016).

Markowitz excludes short selling of shares in his paper on modern portfolio theory (Markowitz, 1952). As such, the weights in the portfolio must be non-negative. Short selling involves selling shares of a stock that are borrowed in anticipation of a decline in the price of said stock. If the price declines, the investor, who is short in said stock, purchases back the borrowed shares at a new lower price and returns the borrowed shares to the lender (Lee & Lee, 2020). Black, Jensen, and Scholes (1972) modify the mean-variance portfolio model from Markowitz, allowing for short-selling of shares and negative weights in the portfolio. Consequently, allowing the weights in the portfolio to be negative enables the portfolio to exploit stocks with both positive and negative returns. Further, a portfolio with the opportunity to short sell shares is also positioned to exploit covariances between the returns on a broader range (Black, Jensen, & Scholes, 1972). The relaxation of the short-selling constraint in the efficient frontier model from Markowitz allows for potentially higher Sharpe ratios.

# 4. Methodology

Before we elaborate on the Thesis Model, we will introduce the most fundamental methodologies in this thesis. One of the most central aspects is machine learning. Consequently, we provide a systematic explanation of the theoretical framework on this topic. In detail, we present a thorough description of tree-based machine learning algorithms, including one of the most common tree-algorithms, XGBoost. Another imperative part of this thesis is portfolio optimization using matrix algebra. Accordingly, we present the details of the portfolio optimization framework and its structural prerequisites in the context of maximizing the Sharpe ratio. Finally, we present the bias-variance tradeoff and times series cross-validation theory.

## 4.1 Machine Learning Algorithms

Machine Learning (ML) methods are algorithms that can learn a specific task without being explicitly programmed. Coming from artificial intelligence, ML systems learn or improve on an automated task through experience (Jordan & Mitchell, 2015). More formally, Tom Michell (1997) provides a concise definition of machine learning algorithms, consisting of the factors experience, task, and performance.

*"A computer program is said to learn from experience $E$ with respect to some class of tasks $T$ and performance measure $P$, if its performance at tasks in $T$, as measured by $P$, improves with experience $E$."*

- *Tom Michell (1997)*

### 4.1.1 The Task – T

Machine learning tasks conceptualize the method of processing the data provided (Goodfellow, Bengio, & Courville, 2016). The adequate method depends on the specific question or problem and the available data in each instance. As such, the task of the machine learning algorithm is *not* the process of learning itself but rather the technique used for learning. There are numerous tasks associated with machine learning models. However, there are two common categories. These are classification and regression. In this thesis, we enable a regression approach to predict stock returns.

### 4.1.2 The Performance Measure – P

There are several measures of performance to assess the capabilities of a machine learning algorithm (Goodfellow, Bengio, & Courville, 2016). These measures are used to evaluate the results from the machine learning algorithm. The measure of use varies with the task of learning. Regardless of the performance measure, it is essential to study how the machine learning algorithm performs on unseen data when determining its performance. Otherwise, the results will come out as biased. Hence, when evaluating the performance of the algorithm, it is imperative to use a test set of the data separate from the data used to train the algorithm (Goodfellow, Bengio, & Courville, 2016). We elaborate on the performance measures we use in this thesis in section 4.3.

### 4.1.3 The Experience - E

Put simply, experience involves how the learning process unfolds and how the algorithm experiences the data applied during the learning process. Typically, a machine-learning algorithm is classified according to three broad learning approaches: unsupervised learning, reinforced learning, and supervised learning (Goodfellow, Bengio, & Courville, 2016). See figure 4 for illustration. In this thesis, we apply supervised learning. Still, to complete the methodical framework, we present all three approaches.

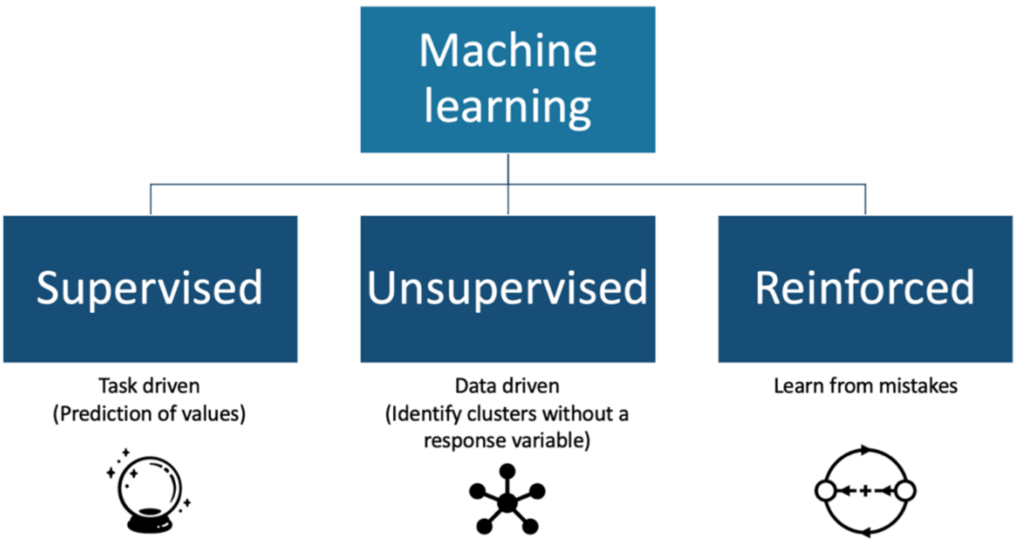


*Figure 4, Machine learning experiences*

### Supervised Learning

The most common machine learning algorithms enable supervised learning methods (Jordan & Mitchell, 2015). Supervised learning algorithms exploit a traditional data structure with a series of inputs or explanatory variables, $x_1, x_2, x_3 \ldots x_n$ and a sequence of corresponding outputs, or response variables, $y_1, y_2, y_3 \ldots y_n$. The algorithm is trained on a set of predictors $(x)$ with an associated response variable $(y)$. The concept of supervised learning is that the algorithm shall learn to extrapolate the response to data to be able to produce the correct output, given a new input (Ghahramani, 2004). The output is either a class label, classification, or a numeric regression (Goodfellow, Bengio, & Courville, 2016).

### Unsupervised Learning

In the occurrence of unlabeled data where a response variable is missing, supervised learning algorithms are not applicable (Jordan & Mitchell, 2015). In the presence of such data, there are methods to analyze the structural property or clustering hidden in the collection of unlabeled data points (Sutton & Barto, 1998). This is described as unsupervised learning (Ghahramani, 2004) (Goodfellow, Bengio, & Courville, 2016). In broad terms, unsupervised learning discovers patterns in data points with no pre-existing labels. (Hinton & Sejnowski, 1999). As such, unsupervised learning enables classification without a clear idea of the basis of the classification. Instead, the data points are organized into groups that are not previously defined, with the intention that the unsupervised learning algorithms discover these groups (Angarita-Zapata, Alonso-Vicario, Masegosa, & Legarda, 2021).

### Reinforced Learning

The third approach to the machine-learning paradigm is reinforcement learning (Jordan & Mitchell, 2015). In contrast to the former approaches, unsupervised learning algorithms do not experience a fixed dataset. These algorithms operate in an interactive environment where a feedback loop enables them to learn from their previous actions by trial and error through a reward/punishment structure. Instead of training the model to indicate the correct output for a given input, reinforced learning algorithms are trained to discover which actions return the highest reward. In essence, the goal of reinforced learning is not to determine hidden patterns in the data structure but rather to maximize the reward signal from its predictions (Sutton & Barto, 1998).

### 4.1.4 Decision Trees

Some of the most prominent machine learning algorithms apply tree-based methods (Rokach & Maimon, 2014). A basic approach to tree models involves partitioning the predictor space into several simple regions. The segmentation is based on hierarchal splitting rules specified to each model. Accordingly, it is common to summarize the splitting rules in a tree structure. Hence, tree models are described as decision tree methods (James, Witten, Hastie, & Tibshirani, 2013). Below, we present a general approach to the tree algorithm.

**Algorithm (1.1):** *Decision trees*

1 Consider the entire predictor space

2 Choose the predictor $X_j$ and the cutpoint $s$ that splits $X$ into two pieces

$$R_1(j,s) = \{X \mid X_j < s\} \text{ and } R_2(j,s) = \{X \mid X_j \geq s\}$$

by minimizing

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

3 Repeat for each new region until the convergence criteriums are met

Small decision trees are known to be easy to interpret. However, for large datasets, they do not provide effective prediction schemes (Yom-Tov, 2004). As the size of the datasets increase, decision trees become harder to interpret. Furthermore, another drawback of decision trees is their relative sensitivity to noise, especially if the training data size is small. Subsequently, decision trees often suffer from high variance, meaning that minor changes in the training data cause drastically different results. Consequently, decision trees tend to perform poorly on out-of-sample predictions (Yom-Tov, 2004).

### 4.1.5 Ensemble Methods

There are ways to alienate the drawbacks of decision trees. By combining several decision trees, using ensemble methods, the variance and predictive performance of trees can substantially improve (James, Witten, Hastie, & Tibshirani, 2013). Ensemble methods combine predictions from different models to produce more reliable estimates. Many ensemble methods exist. However, there are mainly two methods recognized as the standard (Zhang & Ma, 2012).

One way to reduce the potential high variance problem of decision trees is *bootstrapped aggregation*, or *bagging* (James, Witten, Hastie, & Tibshirani, 2013). A general statistic approach assumes $n$ independent observations $x_1 \dots x_n$ to have variance $\sigma^2$. As such, the variance of the mean becomes $\frac{\sigma^2}{n}$, meaning that averaging observations reduces the variance. The same concept can be applied to statistical learning methods such as decision trees. However, instead of creating several training sets from the total population, the training set is bootstrapped into $B$ different training sets. Then, the decision tree is applied to each training set to obtain the predictions $f^{*b}\,(x)$. Lastly, after training the model on all $B$ training sets, the mean of the predictions is computed. We present a generalized objective function for bagging in equation 4.1.

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{\,b}(x) \tag{4.1}$$

Another ensemble method is boosting. Tree boosting is a popular and highly effective ensemble method (Chen & Guestrin, 2016). Similar to bagging, boosting is a general approach that can be applied to most statistical learning methods. However, whereas bagging relies on bootstrapped training data samples, boosting applies the concept of sequentially growing. This means that each tree is grown on an adapted training data set, modified based on the ability of previous trees to predict the different outcomes. Thus, when the model returns incorrect predictions, the misclassified samples are assigned larger weights for the next tree. Meanwhile, the samples that are correctly classified are assigned lower weights. Below we present a generalized boosting algorithm formulated by James, Witten, Hastie, & Tibshirani (2013).

Algorithm 1.2: Decision tree boosting

1. Set $\hat{f}(x) = 0$ and $r_i = y_i$, for $\forall i \in 1 \dots N$
2. For $\forall b \in 1 \dots B$
   a. Fit a tree $\hat{f}^b$ with $d$ splits ($d + 1$ terminal nodes) to the training data $(X, r)$
   b. Update $\hat{f}$ by adding a shrunken version of the new tree:
   $$\hat{f}(\mathrm{x}) \leftarrow \hat{f}(x) + \lambda \hat{f}(x)$$
   c. Update the residuals
   $$r_i \leftarrow r_i + \lambda \hat{f}^b(x_i)$$
3. Ultimately, the output of the boosted model is,
$$\hat{f}(\mathrm{x}) = \sum_{b=1}^{B} \lambda \hat{f}^b(x)$$

$B$ is the number of trees, $\lambda$ is the shrinkage parameter controlling the learning rate of the algorithm, whereas $d$ is the number of splits.

One extension of boosting is called gradient boosting. Gradient boosting follows the same approach of sequentially adding trees based on previous performance. However, instead of assigning weights based on the previous classification, gradient boosting adjusts the weights using gradients (Friedman J. , 2001). Hence, gradient boosting aims to minimize the loss function in the model by consecutively adding trees applying a Gradient Descent algorithm when adding the new trees (Friedman, Hastie, & Tibshirani, 2000). As such, the new trees are constructed to be maximally correlated with the negative gradient of the loss function, creating a more generalized approach, expanding the opportunity set of boosting.

### 4.1.6 XGBoost

One instance of gradient boosting is extreme gradient boosting, or XGBoost. The method was first introduced by Chen and Guestrin (2016) and has become one of the most popular approaches to machine learning. XGBoost is renowned for efficacy, computational speed, and model performance and is one of the best performing algorithms in supervised learning (Chen & Guestrin, 2016) (Osman, Ahmed, Chow, Huang, & El-Shafie, 2012). Moreover, XGBoost is a highly adaptive method that carefully accounts for the bias-variance tradeoff in almost every aspect of the learning process (Nielsen, 2016)

The main idea of XGBoost is to continuously add weak learners, in the form of new trees, with different weights to a set of regression trees. Each additional tree attempts to resemble the previous predictions residuals (Zhang, Bian, Qu, Tuo, & Wang, 2021). For the data set provided, with $n$ observations and $m$ features, XGBoost aims to predict the output:

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \; f_k \in F \tag{4.2}$$

Here, $\hat{y}_i$ express the predicted value. $F$ represents the space of regression trees in the model, while $K$ is the number of regression trees. Hence, $f_k$ corresponds to a tree regression with an independent tree structure and leaf weights. In the process of learning, XGBoost aspires to minimize the following regularized objective function:

$$(X)\ Obj(\phi) = \sum_{i=1}^{n} l(\hat{y}_i^t, y_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{4.3}$$

$$\text{where } \Omega(f_k) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{4.4}$$

In the objective function, $l(y_i, \hat{y}_i^t)$ represents the loss function measuring the difference between the prediction and the observed value $y_i$. This can be any loss function as long as it is second-order derivable (Chen & Guestrin, 2016). The second term in the objective function **Ω**, represents the regularization part of the algorithm and penalizes complexity in the model. As such, a lower value of $\Omega(f_k)$, indicates a generalization ability.

Although XGBoost resembles a standard gradient boosting algorithm, there are differences, some more prominent than others. One crucial distinction is the use, and computation, of second-order gradients. As mentioned in section 4.1.5, gradient boosting applies a Gradient Descent to minimize the loss function of the mode. In contrast, XGBoost uses the second-order derivative as an approximation. This provides more precise information about the direction of the gradients, and thus better prerequisites to obtain the minimum of the loss function (Chen & Guestrin, 2016). Another superiority of XGBoost is the advanced regularization captured by $\Omega(f_k)$. XGBoost has a built-in L1 (Lasso Regression) and L2 (Ridge Regression) regularization, which improves model generalization while smoothing the final learned weights to avoid over-fitting and to handle the bias-variance trade-off. We describe the trade-off in section 4.3.1. Lastly, XGBoost is sparsity aware, meaning it handles sparsities, or missing data, by only considering non-missing observations when making a prediction. Furthermore, the algorithm maintains a tree structure model. Thus, normalized data is not a requirement (Chen & Guestrin, 2016).

## 4.2 Portfolio Optimization

In this thesis, we construct portfolios by maximizing the expected Sharpe ratio. To do this, we apply a matrix multiplication method presented by Eric Zivot (2013). This method requires short selling shares to be allowed. As such, we allow short selling of shares in the portfolio optimization process. Below, we explain the mathematical formulations associated with the matrix multiplication method.

From section 2, we know that the expected return of a portfolio $P$ with $n$ assets is

$$\widehat{\mu_P} = w_1 \cdot \widehat{\mu_1} + w_2 \cdot \widehat{\mu_2} \ldots w_n \cdot \widehat{\mu_n} = \sum_{i=1}^{n} u_i w_i \tag{4.5}$$

Moreover over, the variance of the same portfolio is

$$\sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij} \tag{4.6}$$

When handling portfolios consisting of many assets, the algebra of embodying portfolio variances and expected returns becomes heavy. However, many of the computations can be significantly simplified by using matrix algebra (Zivot, 2013). The expected return of $n$ assets in a portfolio and the related weights can be rewritten as:

$$\mu_n = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}, \qquad w_n = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \tag{4.7}$$

Furthermore, the covariance and variance of the $n$ assets in the portfolio can be illustrated by the $n * n$ matrix $\sum$

$$\sum = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix} \tag{4.8}$$

With these simplifications in mind, the expression of expected portfolio return and covariance can be reformulated to matrix notation as we illustrate below:

$$\widehat{\mu_P} = \sum_{i=1}^{n} u_i w_i = (\mu_1 \quad \cdots \quad \mu_n) \times \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = u^T w \tag{4.9}$$

$$\sigma_P^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j Cov_{ij} \; = \; (w_1 \quad \cdots \quad w_n) \times \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{bmatrix} \times \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = w^T \textstyle\sum w \tag{4.10}$$

Subsequently, the maximation of the Sharpe ratio can be reformulated as a maximation problem where the ratio is maximized by determining the optimal composition of weights.

$$\underset{w}{Max} \; \frac{w'\mu - r_f}{(w'\sum w)^{\frac{1}{2}}} = \frac{\mu_P - r_f}{\sigma_P}, \qquad s.t\; w'1 = 1 \tag{4.11}$$

Further, solving the Sharpe ratio maximation problem has the following Lagrangian:

$$L(\mathrm{t}, \lambda) = \left(\mathrm{w}'\mu - r_f\right)(\mathrm{w}'\textstyle\sum \mathrm{w})^{-\frac{1}{2}} + \lambda(\mathrm{w}'1 - 1) \tag{4.12}$$

As such, after applying the chain rule, the first-order conditions are

$$\frac{\partial L(\mathrm{w}, \lambda)}{\partial \mathrm{w}} = \mu(\mathrm{w}'\textstyle\sum \mathrm{t})^{-\frac{1}{2}} - \left(\mathrm{w}'\mu - r_f\right)(\mathrm{w}'\textstyle\sum \mathrm{t})^{-\frac{3}{2}} \cdot \textstyle\sum \mathrm{w} + \lambda 1 = 0 \tag{4.13}$$

$$\frac{\partial L(w, \lambda)}{\partial \lambda} = \mathrm{w}'1 - 1 = 0 \tag{4.14}$$

Conclusively, after some tiresome algebra, we can express the weights $w$ of the tangency portfolio as follows in equation 4.12 below.

$$w_{Max\ Sharpe} = \frac{\sum^{-1}\left(\mu - r_f \cdot 1\right)}{1^{\mathrm{T}} \sum^{-1}\left(\mu - r_f \cdot 1\right)} \tag{4.15}$$

Note that if the maximation problem would include no short selling of shares, the optimization problem could be written as equation 4.11 plus an inequality constraint $w_i \geq 0, i{=}1, \dots, N$. However, the inequality constraint prevents the use of Lagrange multipliers to obtain an optimal solution, and therefore, matrix algebra (Zivot, 2013).

### 4.2.1 Positive Definiteness in Covariance

One necessity for portfolio optimization using matrix multiplication is positive definite covariance matrices. This is an apparent challenge in multivariate volatility modeling of covariance matrices (Chiriac & Voev, 2011). A matrix is positive definite if the eigenvalues in the matrix are strictly positive. In algebraic terms, a $n * n$ matrix $V$, is positive definite if $x'Vx$ is strictly positive for any non-zero vector $x$. Positive definiteness permits economy and numerical stability in the solution of linear systems (Higham, 1988). Subsequently, one of the prerequisites of Sharpe ratio optimization using matrix algebra, is thus, that the covariance matrix is positive definite (Kwan & Clarence, 2010).

A positive definite matrix is invertible. Correspondingly, the inverted matrix, $V^{-1}$, is also positive definite. The implications of a non-positive definite covariance matrix for the optimization problem in equation 4.11 are essential. Without a positive definite covariance matrix, equation 4.12 will not provide a minimum of the Lagrangian. As such, the weights from the optimization problem will provide a portfolio different from the portfolio that maximizes the Sharpe ratio. Thus, we must ensure positive definite matrixes when performing portfolio optimization.

Covariance matrix predictions from a forecasting model are not guaranteed to be positive definite (Fiszeder & Orzeszko, 2021). One way to ensure positive definiteness is to use Cholesky decomposition and predict Cholesky factors instead of predicting each element in the covariance matrix. This method was suggested by Andersen, Bollerslev, Diebold, and Labys (2003) and later implemented in machine learning by Fiszeder and Orzeszko (2021). Cholesky decomposition, or Cholesky factorization, maps a $n * n$ positive definite matrix $A$ to the product of $A_t = K_t K_t'$ $(t = 1,2, \ldots T)$ . Here, $K_t$ is the lower triangular matrix and $K_t'$ is the conjugate transpose. As such, the Cholesky factors $K_t$ can be interpreted as the square root of $A$, emphasizing the requirement of non-negative eigenvalues to enable Cholesky decomposition. We illustrate Cholesky decomposition in equation 4.16.

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} = \begin{pmatrix} K_{11} & 0 & 0 \\ K_{21} & K_{22} & 0 \\ K_{31} & K_{32} & K_{33} \end{pmatrix} * \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ 0 & K_{22} & K_{23} \\ 0 & 0 & K_{33} \end{pmatrix} \tag{4.16}$$

In a covariance matrix context, this means the following

$$\begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2 & \sigma_1\sigma_3 \\ \sigma_1\sigma_2 & \sigma_2^2 & \sigma_2\sigma_3 \\ \sigma_1\sigma_3 & \sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix} = \begin{pmatrix} K_{11} & 0 & 0 \\ K_{21} & K_{22} & 0 \\ K_{31} & K_{32} & K_{33} \end{pmatrix} * \begin{pmatrix} K_{11} & K_{12} & K_{13} \\ 0 & K_{22} & K_{23} \\ 0 & 0 & K_{33} \end{pmatrix} \tag{4.17}$$

The prediction of Cholesky factors ensures the final covariance matrix to be positive definite without imposing any restrictions on the predicted values. Consequently, this thesis enables Cholesky decomposition proposed by Andersen, Bollerslev, Diebold, and Labys (2003) when predicting covariance matrices.

## 4.3 Model Performance

Different statistical learning models perform differently on different data sets. Uncovering the best model for a given data set is essential for accurate predictions. Hence, it is crucial to assess the model performance of a statistical learning model in any prediction attempt. However, selecting the best model is one of the utmost challenging parts of performing statistical learning in practice (James, Witten, Hastie, & Tibshirani, 2013).

A common way of assessing model performance is by measuring the accuracy of predictions. When applying a statistical model to a previously unseen test set, the measured accuracy is referred to as *testing error*. We can distinguish between the past testing error and the future testing error. It is not interesting to focus on how a statistical model would have performed in the past, but rather how well the model will perform in the future (James, Witten, Hastie, & Tibshirani, 2013).

### 4.3.1 Bias-variance Trade-off

*Mean squared error* (MSE) is a frequently used method for quantifying the extent to which the predicted value is close to the true value (James, Witten, Hastie, & Tibshirani, 2013). MSE is given by:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2 \tag{4.18}$$

where $y$ is the true value, and $\hat{f}$ is an estimated function for the predicted value $x$. The expected test MSE for any given value $x_0$ is composed as the sum of three elemental magnitudes: the variance of $\hat{f}(x_0)$, the squared bias of $\hat{f}(x_0)$ and the variance of the error term $\varepsilon$ (James, Witten, Hastie, & Tibshirani, 2013). Therefore, we can write the expected test MSE as:

$$E(MSE) = E\left(y_0 - \hat{f}(x_0)\right)^2 = Var\left(\hat{f}(x_0)\right) + \left[Bias\left(\hat{f}(x_0)\right)\right]^2 + Var(\epsilon) \tag{4.19}$$

$E(MSE)$ is the average test MSE, obtained if $f$ was repetitively estimated using many training sets, before testing each $f$ at $x_0$. To minimize the expected test error, we must choose a statistical learning model that simultaneously accomplishes low bias and low variance. This is often referred to as finding the optimal bias-variance trade-off. Note that the variance of the error term $\epsilon$ is irreducible and is thus unalterable and represents the minimum expected MSE (James, Witten, Hastie, & Tibshirani, 2013).

The variance of a statistical learning model refers to the amount $\hat{f}$ would change if it was estimated using different training sets. As $\hat{f}$ is fitted on the training set, separate training sets will result in separate $\hat{f}$. Preferably the estimated function does not vary excessively between training sets. Nevertheless, if a statistical learning model has high variance, $\hat{f}$ is sensitive to small fluctuations in the training set. A high variance may result from the statistical learning method modeling the random noise in the training data. This is referred to as overfitting (James, Witten, Hastie, & Tibshirani, 2013). Furthermore, James, Witten, Hastie, and Tibshirani (2013) define bias as the error caused by approximating a complicated problem with a simpler statistical model. High bias can cause a statistical model to overlook relevant relations between predictors and predictions. Consequently, the model cannot capture the variability in the training data. This phenomenon is referred to as underfitting (Aalst, et al., 2008).

### 4.3.2 Thesis Performance Measures

To evaluate the performance of the prediction models in this thesis, we deploy two performance measures: RMSE and MAPE. To provide more intuitive power, MSE can be rooted. By doing so, the error rate, called *root-mean-squared-error* (RMSE), is in the same unit as the true values the model is trying to estimate. The formula for RMSE is:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}(x_i)\right)^2} \tag{4.20}$$

However, MSE, and thus RMSE, are scale-dependent error rates and can therefore not be used for comparisons across different statistical models (Hyndman & Athanasopoulos, 2018). In contrast, *mean-absolute-percentage-error* (MAPE) is not scale dependent and can therefore be used to compare statistical performance across different statistical models. In equation 4.21, we show the MAPE formula.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{f}(x_i)}{y_i}\right| \tag{4.21}$$

### 4.3.3 Time Series Cross-validation

For most real-life prediction problems, the true $f$ is not observable. This opposes an issue when computing the future test error, which, as mentioned earlier, is the desirable test error. Thus, an alternative approach must be chosen to unfailingly estimate the test error (James, Witten, Hastie, & Tibshirani, 2013).

Different techniques exist to overcome the obstacle mentioned above. Due to its simplicity and universality, a commonly used method is cross-validation (Arlot & Celisse, 2010). Cross-validation strives to reliably estimate the test error by designating a subset of the training observations to be held out from the fitting process, called the training process. The fitted (trained) model is then applied to the designated observations held out of the original training process. Such approaches try to simulate the situation where a statistical model is used on previously unseen future observations to measure the accuracy of the model. Hence, cross-validation is used to estimate the future test error (James, Witten, Hastie, & Tibshirani, 2013). However, although cross-validation strives to estimate the future test error, this does not automatically result in a reliable estimate. We can illustrate this with a simple example.

Consider the following function for the true values for observations $x_1, x_2 \dots x_n$:

$$Y_i = f(x_i) = x_i^2 \tag{4.22}$$

If a statistical method estimates $\hat{f}(x_i) = x_i$, the test error for a test set consisting of cross-validation using only one observation $x_1 = 1$ will be $E(MSE) = 0$. Because $x_i^2 \neq x_i$ for all values $x > 1$, the test error estimate of $E(MSE) = 0$ is not a reliable estimate. This is a consequence of a very small test set. As such, cross-validation does not automatically result in a reliable estimate.

$k$-fold cross-validation is a version of cross-validation involving splitting the set of observations into $k$ folds or groups (James, Witten, Hastie, & Tibshirani, 2013). Time series $k$-fold cross-validation splits a data set into a training set and a test set, with a time used as a divider (Hyndman & Athanasopoulos, 2018). We illustrate a time series 5-fold cross-validation in figure 5.

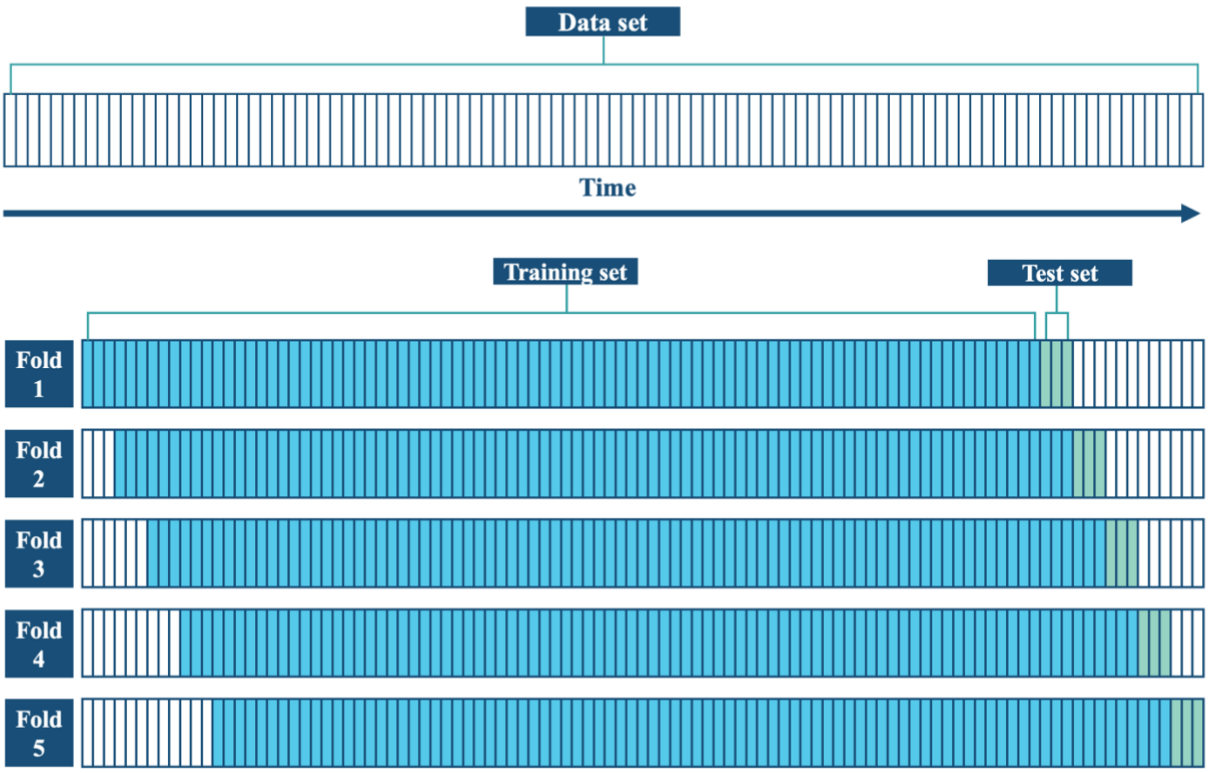


*Figure 5, Time-Series 5-fold Cross-Validation*

From figure 5, we can see that the data is split into a training and a test set in each fold. None of the test sets overlap and are located after the training sets timewise. Next, the standard cross-validation method is then applied to each fold. The training sets are used to fit, or train, the statistical model before the fitted model, is applied to the associated test set. Model performance is then assessed by measuring accuracy on each fold and averaging the test error (James, Witten, Hastie, & Tibshirani, 2013).

By utilizing time-series $k$-fold cross-validation with $5 < k < 10$ we can achieve the simulated effect of previously unseen test data as with other cross-validation methods. At the same time, a more optimal bias-variance trade-off is obtained. Using *k*-fold cross-validation reduces the estimated test error rate variation compared to the standard cross-validation. With *k*-fold cross-validation, the estimated test error is less dependent on which observations are contained in the training set and which observations are included in the test set. In other words, the k-fold approach shrinks the bias of the estimate. Further, k-fold cross-validation reduces variance as noise in the data set has a higher chance of being categorized as noise and left out of the statistical prediction model. This is because the approach applies multiple training and test sets (James, Witten, Hastie, & Tibshirani, 2013).

# 5. Data and Modeling

We classify the data we use into two groups: (1) non-company-specific data, and (2) company-specific data. The company-specific is divided into two sub-categories: *Stock prices* and *Accounting data*. Likewise, we divide non-company-specific data into *Macroeconomic* data and *Economic variables*. The structure of the data set is summarized in figure 6.

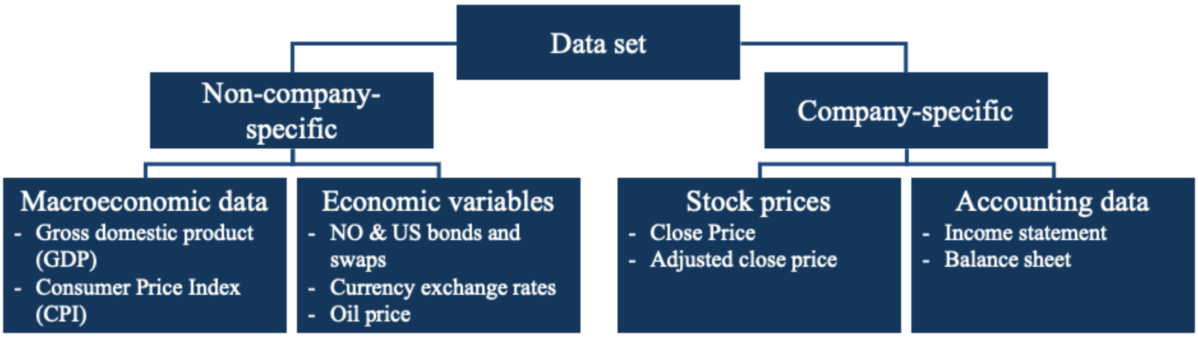


*Figure 6, Data set overview*

The foundation of the data selection has been laid out in the literature review. We recall that fundamental data is assumed to impact stock prices, and therefore stock returns. Moreover, we argue that the non-company-specific data not only impact stock returns but can also predict stock return covariances directly. As such, we believe that the four components shown in figure 6 could provide reliable predictive power in a portfolio selection context.

We collect the data from numerous sources. The data is either observed quarterly or adjusted and converted into quarterly data. Furthermore, we create two separate prediction models, one for covariance prediction and one for prediction of expected return. For both models, the response variables are observed from Q2 2000 to Q3 2021. In contrast, the explanatory variables are observed from Q1 2000 to Q2 2021. That way, the explanatory variables lag one quarter, and we train the models using the response variables at time $t$ and the corresponding explanatory variables at time $t - 1$.

### Company-Specific Data

This thesis focuses on stocks listed on Oslo Stock Exchange (OSE). Consequently, we collect the company-specific data through Refinitive Eikon. The data consists of quarterly financial statements from 180 companies listed on OSE. Only a few companies in the data have been listed since 2000. Thus, the number of quarterly observations increases as the data set approaches 2021. This is evident from figure 7. Furthermore, the data set is limited to observations from 2000 onwards due to a rapid decline in data quality. Also, one constraint regarding the use of accounting data is that most listed companies publish reports every quarter. As such, a maximum of four data observations per year is available for each company. Note that the company-specific data is panel data, meaning it varies between observations in the same period and across different periods.

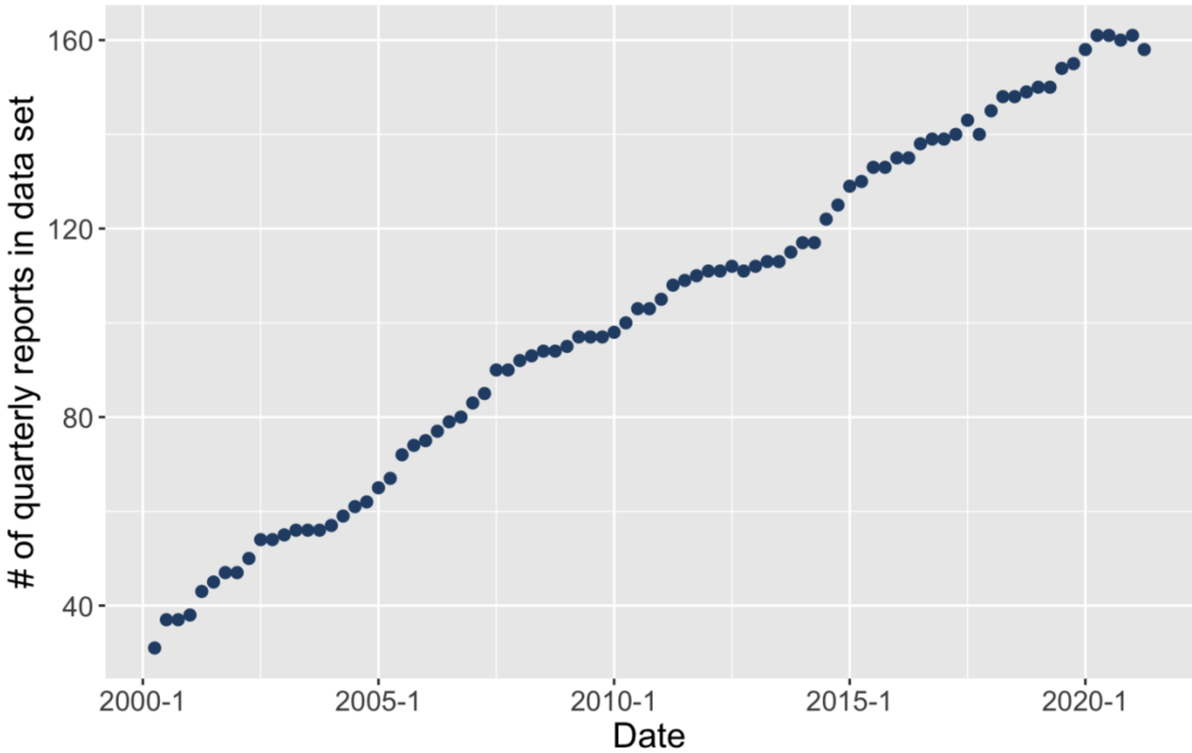


*Figure 7, Number of quarterly reports per quarter in the data set*

One of the key assumptions in this thesis is that accounting data is available as soon as the reporting quarter is ended. For instance, we assume accounting data for the second quarter in 2021 to be available 01-07-2021. This simplifies the reality, as the quarterly reports are usually released later in the upcoming quarter. Ideally, to make the model more realistic, we should incorporate the exact publication date of each report into the data. However, Refinitive Eikon does not provide publication dates for quarterly reports for Norwegian companies as part of their service. Moreover, the publication dates for quarterly reports published before 2010 are hard to obtain, even with a manual approach. Hence, to maintain consistency throughout the dataset, we assume that all accounting data is available by the end of each associated quarter.

To collect stock price-related information, we choose to use Yahoo Finance. We found Yahoo to be the most reliable data source. Most importantly, it was the only source with reliable data on adjusted stock prices. Adjusted stock prices are adjusted for splits, splices, dividends, and other adjustments. When calculating quarterly returns, it is imperative to use adjusted prices to project the actual return from the stocks. To avoid any confusion, let it be clear that the model presented in this thesis does not use any momentum in historical stock prices to perform any prediction. According to the literature review, such an approach would have modest predictive power. The stock prices are included in the data set to calculate returns (adjusted), current market capitalization (close price), and covariance matrices.

#### Non-Company-Specific Data

We collect the non-company-specific data using a Bloomberg Terminal. Bloomberg LP is considered one of the leading financial data providers globally (Scott, 2010). We include macroeconomic factors and economic variables in the model to capture the shape of the economy and the changes in these states. Non-company-specific data does not vary across companies, only across periods. The timeline for the non-company-specific data reflects the company-specific data and spans from Q1 2000 to Q2 2021. Encouraged by the literature review, we assume non-company-specific data to be of importance when predicting stock returns and covariances.

## 5.1 Model Structure

This thesis uses machine learning for portfolio optimization by utilizing three different components. As presented earlier, the approach used in this thesis for portfolio optimization is to maximize the expected Sharpe ratio of the constructed portfolio. We formulate the expected Sharpe ratio as follows.

$$\widehat{SR} = \frac{\widehat{\mu_P} - r_f}{\widehat{\sigma_P}} \tag{5.1}$$

where $\widehat{\mu_P}$ is the expected portfolio return, $\widehat{\sigma_P}$ is the expected portfolio volatility, and $r_f$ is the risk-free rate. Note that we do not need to estimate the risk-free rate because we can observe it in the market. We utilize the 3-month NIBOR as the risk-free quarterly rate. Further, from the methodology chapter, we know that the portfolio return is given by multiplying the return of every stock with each associated weight in the portfolio. Similarly, the portfolio volatility is found by multiplying the said weights with the associated covariance matrix. By combining this knowledge with equation 5.1, it is evident that to calculate the expected Sharpe ratio, we must estimate three quantities. These quantities are (1) the weight for each stock in the portfolio, (2) the return for each stock, and (3) the associated covariance matrix.

The estimation of the three quantities can be thought of as three different sub-models, or components, in the final model. We provide a graphical illustration of the model structure in figure 8. From here on this model goes by the name *Thesis Model*. The *Returns Model* estimates the expected returns, while the *Covariance Model* estimates the expected covariance matrix. The *Weights Allocation Model* uses the estimates from the two other component models and calculates the optimal weights maximizing the expected Sharpe ratio. In the upcoming sections, we provide a thorough discussion of each sub-model.

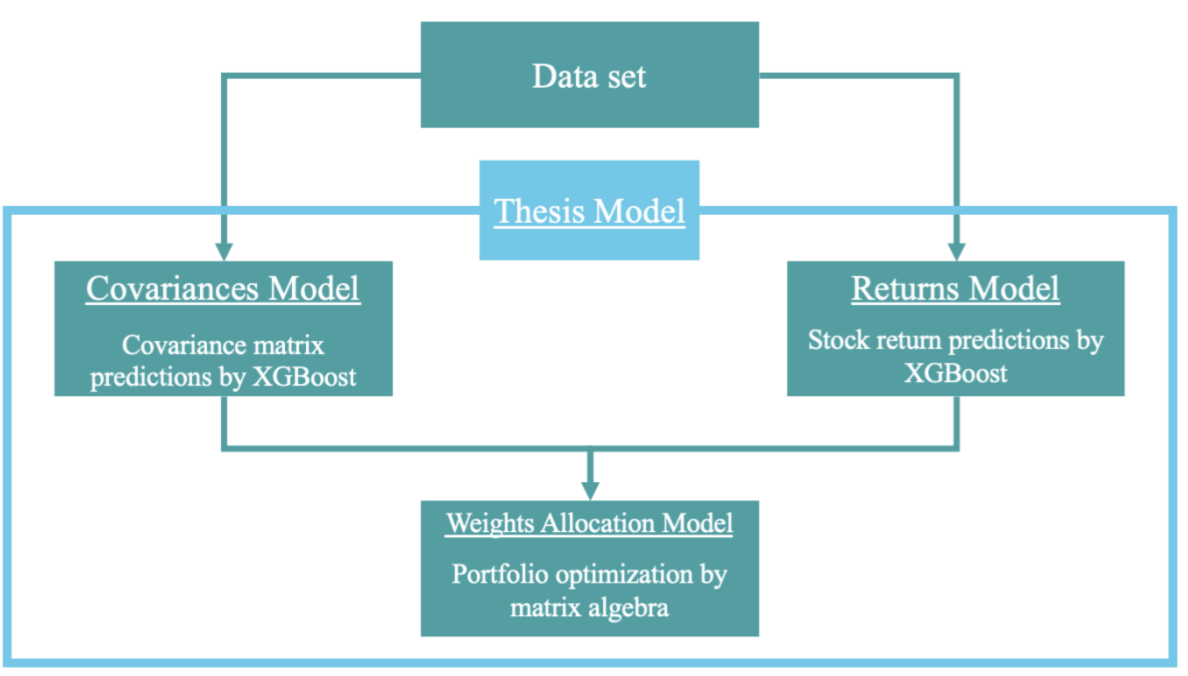


*Figure 8, Graphic illustration of the model structure*

### 5.1.1 Returns Model

The Returns Model is an XGBoost regression model to predict the expected return for each stock in the subsequent quarter. However, a renowned difficulty in any portfolio optimization problem is to predict the expected return (French, Schwert, & Stambaugh, 1987). Inaccurate predictions of the expected returns can make the portfolio behave differently than foreseen, and small changes in the predictions can cause a significant impact on the final portfolio (Merton, 1980). We suggest using quarterly financial data and macroeconomic indicators to predict expected returns using machine learning and XGBoost.

The response variable in the Returns Model, is quarterly logarithmic stock returns. Logarithmic returns, often called log-returns, are additive across time. This implies that the annual return of a stock or a portfolio can be calculated by summarizing the last four quarterly log-returns. We calculate log-returns of financial securities using the following formula:

$$r_{log} = \log \frac{price_t}{price_{t-1}} = \log price_t - \log price_{t-1} \tag{5.2}$$

However, using log-returns have some drawbacks as well. Log-returns are not additive across assets, meaning they cannot be summarized to obtain the combined log-return for the portfolio as a whole. On the other hand, discrete returns, often called simple returns, are additive across assets. Discrete returns are the most popular returns used in finance and represent the relative change in the price of a security. We calculate discrete returns using the following formula:

$$r_{discrete} = \frac{price_t}{price_{t-1}} - 1 \tag{5.3}$$

Consequently, we must convert the predicted log-returns to discrete returns to calculate the return of a portfolio. We convert log-returns to discrete returns using the following formula:

$$r_{discrete} = \frac{price_t}{price_{t-1}} - 1 = e^{r_{log}} - 1 \tag{5.4}$$

We present the predictors for the Returns Model in section 5.3.

### Data Preprocessing of Returns

The stock prices from Yahoo Finance are provided on a daily basis, based on official trading days for each stock. As such, the raw stock data does not contain observations from weekends or other days when the Oslo Stock Exchange was closed. Furthermore, if a stock is left untraded for an entire day, Yahoo Finance treats this as a missing value. Thus, when we merged the stock information with the financial data, missing values occurred if the last day of the month was a nonofficial trading day or if the stock was not traded that day. To account for this, we assume that if the price of a stock is missing on any date, the price is equal to the last available price.

This method is called Last Observation Carried Forward (LOCF). We include a practical example of the LOCF process in figure 9.

| Company | Date | Price |
|---|---|---|
| ABG | 01.04.2013 | 8.3 |
| ABG | 02.04.2013 | NA |
| ⋮ | ⋮ | ⋮ |
| ABG | 27.04.2013 | 8.5 |
| ABG | 28.04.2013 | NA |
| ABG | 01.05.2013 | 8.4 |

| Company | Date | Price |
|---|---|---|
| ABG | 01.04.2013 | 8.3 |
| ABG | 02.04.2013 | 8.3 |
| ⋮ | ⋮ | ⋮ |
| ABG | 27.04.2013 | 8.5 |
| ABG | 28.04.2013 | 8.5 |
| ABG | 29.04.2013 | 8.5 |
| ABG | 30.04.2013 | 8.5 |
| ABG | 01.05.2013 | 8.4 |

*Figure 9, LOCF process*

An essential clarification is that we use forward-looking quarterly returns to express the log-return for the next quarter. This is necessary as the model aims to predict quarterly returns. See figure 10 for clarification. Accordingly, the data we use to train the model include forward-looking returns. Meanwhile, the predictors are not forward-looking but express conditions at the end of the current quarter.

| Company | Date | Price |
|---|---|---|
| ABG | 31.03.2013 | 8.3 |
| ABG | 01.04.2013 | 8.3 |
| ⋮ | ⋮ | ⋮ |
| ABG | 27.06.2013 | 8.5 |
| ABG | 28.06.2013 | 8.5 |
| ABG | 29.06.2013 | 8.5 |
| ABG | 30.06.2013 | 8.5 |

$$\text{Return}_{2013\ Q2\ ABG} = \log(Price_{new}) - \log(Price_{old})$$

| Company | Date | $Log\ Return_{t+1}$ |
|---|---|---|
| ABG | 2013 Q2 | 2.38 % |

*Figure 10, Log-return calculation*

### 5.1.2 Covariance Model

The Covariance Model is an XGBoost regression model to predict stock return covariance matrices. While expected return is a central part of the Markowitz portfolio optimization problem, the expected portfolio variance is an equally significant share of the optimization problem. We argue that portfolio variance estimation has often been neglected in the portfolio optimization discussion. Portfolio risk and return are equally interesting aspects of portfolio optimization. As we accent in the introduction, the prediction of covariance matrices using machine learning on the Norwegian stock market has never been conducted on such a large scale.

The name of the response variable for the Covariance Model is *Variance/Covariance*. The associated predictors for the Covariance Model are presented in section 5.3. We calculate the quarterly stock return variance for each stock by using daily returns to obtain the daily variance of each stock, following equation 5.5:

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \tag{5.5}$$

where $x_i$ is the return on day $i$, $\overline{x}$ is the mean of the returns and $n$ is the number of observations. The variance is then scaled up to represent quarterly variance by multiplying the daily variance with the number of trading days in the quarter. We approximate the average number of trading days to be 60.

The approach for calculating stock returns covariances is similar to the one for calculating the variance of stock returns. We calculate the daily covariance of a pair of stock returns using equation 5.6.

$$Cov(x,y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) \tag{5.6}$$

where $x_i$ is the return on day $i$ for stock $X$, $\overline{x}$ is the mean of the returns for stock $X$, $y_i$ is the return on day $i$ for stock $Y$, $\overline{y}$ is the mean of the returns for stock $Y$, and $n$ is the number of observations. The covariances of the returns are scaled up from a daily frequency to a quarterly frequency using 60 as the number of trading days.

### Data Preprocessing

Because we wish to predict quarterly covariances, we group the daily adjusted close prices into quarterly time series. Then, we calculate the quarterly covariance matrix for each quarter. However, the issue of missing values is also apparent when calculating covariances. To calculate the covariance, the number of observations, $n$, must be equal for all pairs. This is evident from equation 5.6. Hence, covariance calculation is impossible if $n$ is different for $x$ and $y$. In the calculation of the quarterly covariance matrices, we use the official trading days. Accordingly, missing values only occurred when a stock did not trade on an official trading day. Once again, we utilize the LOCF analogy. Consequently, if a missing value occurred due to a stock not being traded that day, the missing value was replaced with the last valid price from the previous valid trading day. Figure 11 exemplifies the process.

| Company | Date | Price |
|---|---|---|
| ABG | 01.04.2013 | 8.3 |
| ABG | 02.04.2013 | NA |
| YAR | 01.04.2013 | 200.2 |
| YAR | 02.04.2013 | 201.4 |
| ⋮ | ⋮ | ⋮ |
| ABG | 27.06.2013 | 8.5 |
| ABG | 28.06.2013 | NA |
| YAR | 27.06.2013 | 195.1 |
| YAR | 28.06.2013 | 195.0 |

| Company | Date | Price |
|---|---|---|
| ABG | 01.04.2013 | 8.3 |
| ABG | 02.04.2013 | 8.3 |
| YAR | 01.04.2013 | 200.2 |
| YAR | 02.04.2013 | 201.4 |
| ⋮ | ⋮ | ⋮ |
| ABG | 27.04.2013 | 8.5 |
| ABG | 28.04.2013 | 8.5 |
| YAR | 27.04.2013 | 195.1 |
| YAR | 28.04.2013 | 195.0 |

*Figure 11, Arbitrary representation of LOCF for ABG and Yara in April 2013*

#### Definiteness Problems when Predicting Covariance

As mentioned in section 4.2.2, a positive definite covariance matrix is imperative to maximize the Sharpe Ratio using matrix multiplication. However, a calculated covariance matrix from stocks is not always positive definite by nature. Often, this is due to high dimensionality and linear dependencies between the stocks, causing the matrix to be multicollinear. In matrix terms, this is reflected by one or more eigenvalues of 0. According to Andersen, Bollerslev, Diebold, and Labys (2003), linear dependencies can occur even in low-dimensional cases with only three or four assets, with an increasing probability as the number of assets increases. As such, we enable Cholesky decomposition to ensure that the predicted covariance matrices are positive definite. We separate the prediction process into four steps.

1. Calculate the quarterly $n * n$ covariance matrices $A_t$, $t = 1,2 \dots T$, where $T$ is the number of quarters available.
2. Decompose the covariance matrices $A_t$ to Cholesky factors using Cholesky decomposition. The decomposed matrices have the form $\mathrm{A_t} = K_t K_t'$.
3. Predict the each Cholesky factor and construct the Cholesky decomposition matrix $K_{t+1}$
4. Construct the predicted covariance matrix by reversing the Cholesky decomposition, $\mathrm{A}_{t+1} = K_{t+1} K_{t+1}'$

Note that the Covariance Model does not forecast the entire Cholesky decomposition matrix at once but rather builds the matrix cell by cell. The model iterates through the matrix and predicts the factors accordingly. Finally, the entire covariance matrix is built by reversed Cholesky decomposition.

### 5.1.3 Weights Allocation Model

The third model in this thesis, the Weights Allocation Model, aims to find the optimal portfolio weights. The optimal weights are characterized as the weights that maximize the expected Sharpe ratio of the portfolio. We obtain the optimal weights utilizing the predictions generated by the XGBoost models presented above, combined with the matrix algebra described in the methodology chapter. We recall that the formula for finding the optimal weights is:

$$w_{Max\ Sharpe} = \frac{\widehat{\sum}^{-1}\left(\hat{\mu} - r_f \cdot 1\right)}{1^{\mathrm{T}}\widehat{\sum}^{-1}\left(\hat{\mu} - r_f \cdot 1\right)} \tag{5.7}$$

where $\hat{\mu}$ is a vector with the expected quarterly returns, and $\widehat{\sum}$ is the expected associated covariance matrix. As such, the portfolio rebalancing is done quarterly.

## 5.2 Descriptive Statistics

Table 2 shows the descriptive statistics for the two response variables *Log.Return* and *Variance/Covariance*. As described previously, we calculate *Log.Return* once each quarter for each company. Meanwhile, we calculate the *Variance/Covariance* once each quarter for every stock and every pair of stocks.

| $\boldsymbol{Measure}$ | $\boldsymbol{Log.Return}$ | $\boldsymbol{Variance/Covariance}$ |
|---|---|---|
| $Observations$ | 8648 | 560321 |
| $Sampling\ frequency$ | $Quarterly$ | $Quarterly$ |
| $Start\ period$ | $Q2$ 2000 | $Q2$ 2000 |
| $End\ period$ | $Q3$ 2021 | $Q3$ 2021 |
| $Mean$ | $-0.0029$ | 0.04 |
| $Std.Dev$ | 0.3018 | 0.3184 |
| $Variance$ | 0.0911 | 0.1014 |
| $Skewness$ | $-1.1818$ | 5.436492 |
| $Kurtosis$ | 16.2677 | 397.0881 |
| $Jarque-Bera\ (p\ value)$ | $<\ 2.2e-16$ | $<\ 2.2e-16$ |
| $Min$ | $-3.9548$ | $-19.7549$ |
| 1 $Quartile$ | $-0.1174$ | $-0.0372$ |
| $Median$ | 0.0092 | $4.4674e-07$ |
| 3 $Quartile$ | 0.1350 | 0.1021 |
| $Max$ | 2.5402 | 27.0729 |

*Table 1, Descriptive statistics*

The table above clearly shows that the returns and covariances are skewed with relatively high kurtosis. Further, we can reject normality with a Jarque–Bera test on a 1% significance level for both response variables. The normality rejection for the covariance variable implies that multivariate normal distribution does not sufficiently describe the dependence between stock returns, particularly in the tails of the distribution (Sleire, et al., 2021). This observation supports the arguments made in this thesis about how a standard Markowitz approach does not sufficiently cover the estimation of covariance between stock returns. However, as we can see from figure 12, the distribution of the *Log.Return* variable resemble a normally distributed shape to some extent.

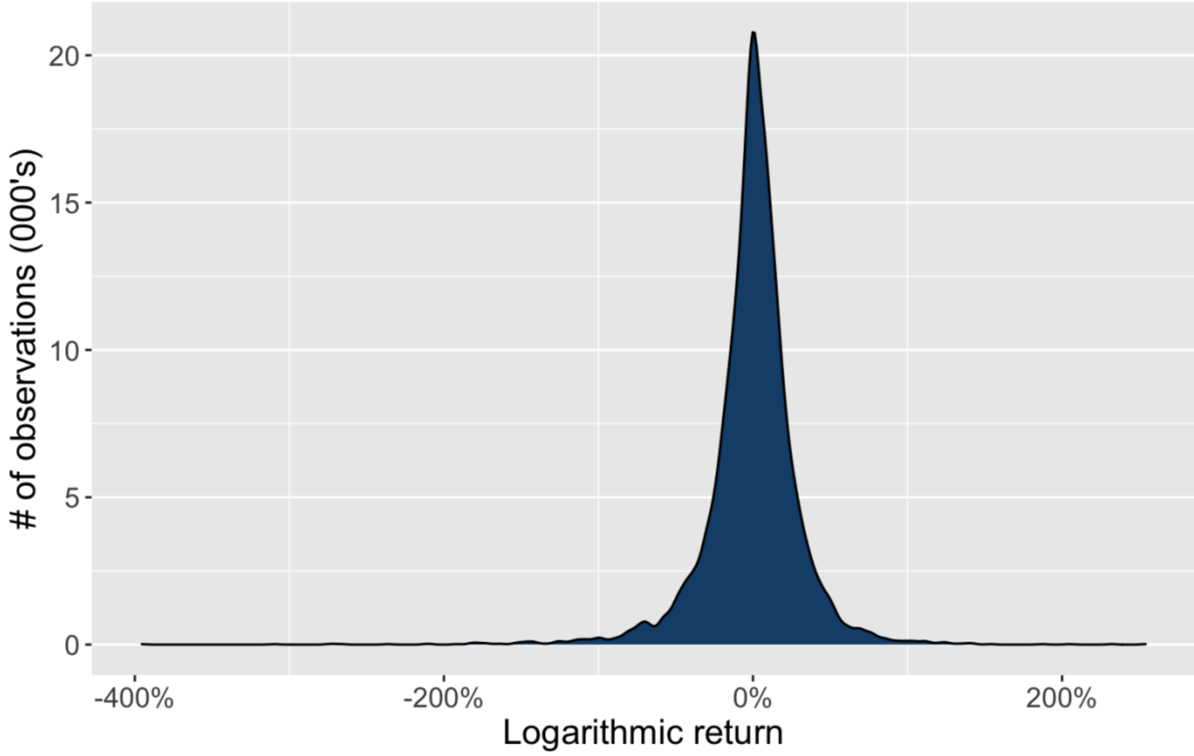


*Figure 12, Distribution of logarithmic quarterly returns in the data set*

## 5.3 Predictors

From the literature review, it is evident that financial ratios can provide some predictive insight. However, to capture the inner dynamical changes of capital structures, we must study the relative company-specific change from quarter to quarter. Still, the absolute level of each ratio is interesting as it describes the financial status of the company. Further, both relative change and absolute values are included for non-company-specific values. The arguments are comparable to the arguments used for the financial ratios. Fluctuations in the non-company-specific indicators can disclose directional changes in the economy. On the other hand, the absolute values provide an overview of the economy in general. Consequently, quarterly change and absolute values are included to capture both dynamics. We compute the quarterly change for all variables with the formula in equation 5.8.

$$Relative\ change = \frac{Measure_t - Measure_{t+1}}{Measure_{t+1}} \tag{5.8}$$

### 5.3.1 Returns Model Predictors

The selection of the explanatory variables included in the stock return model is based on the literature presented in chapter 3. Each predictive is constructed and included in the model under the assumption that they capture the effects laid out in the literature review. The predictors are divided into non-company-specific and company-specific variables.

#### Non-Company-Specific Predictors

The non-company-specific are summarized in table 2. These variables are a combination of macroeconomic and economic variables thought to contextualize the current economic situation. Note that these variables only vary between periods and not between companies within the same period. Further, we include both the absolute value and quarterly change in each quarter.

| **Economic variables** | **Currency rates** |
|---|---|
| NIBOR 3-month | EUR/NOK |
| Gros National Product (GDP) | USD/NOK |
| Consumer Price Index (CPI) | |
| North-Sea Oil Price | |
| **Stock indexes** | **Bonds & swaps** |
| OSEBX | Norwegian 10-year nominal swap rates |
| S&P 500 | Norwegian 5-year nominal swap rates |
| Dow Jones | Norwegian 10-year bond yield |
| Nasdaq | US 30-year bond yield |
| Nikkei 225 | US 10-year bond yield |
| DAX | US 5-year bond yield |

*Table 2, Non-Company-Specific predictors*

*Norwegian GDP*

As mentioned in the literature review, macroeconomic factors impact stock returns, especially industrial production indicators. Based on this, we include the general level of the Norwegian Gross Domestic Product (GDP). This is thought to capture parts of said impact through the relationship between GDP and stock returns (Chaudhuri & Smiles, 2004).

*CPI*

The Norwegian consumer price index (CPI) reflects the current inflation. Inflation is important to investors because investors aim to increase their purchasing power. However, inflation works in the opposite direction of this goal, creating lower purchasing power. Therefore, we include CPI as a predictor to reflect the current inflation and general purchasing power.

*NIBOR*

We include the 3-month Norwegian interbank offered rate (NIBOR) to mirror the general interest rate level in Norway. Further, we include NIBOR to reflect a risk-free investment with a quarterly time horizon.

*North Sea Oil Price*

Approximately 35% of the variance in the Norwegian GDP is linked to the Norwegian petroleum industry (Bjørnland & Thorsrud, 2013). We assume that the oil price carries information about the world economy, therefore contributing with predictive power.

*Currency exchange rates*

We feature exchange rates in the model to capture market predictions about the Norwegian economy. The considered exchange rates are between Norwegian kroner (NOK), US dollars (USD), and Euros (EUR). We assume the uncertainty and risk, such as future interest rate predictions, associated with the Norwegian economy to be captured in the exchange rates.

*Global stock market indexes*

Global stock market indexes capture the state of the world financial markets. Further, it is fair to assume that the global financial markets influence companies on Oslo Stock Exchange and that changes in these markets give a pointer for how companies on OSE will behave forwards. With this in mind, we include stock market indexes from Norway, Germany, Japan, and the US as predictors.

*US & NO bonds and swaps*

The US economy influences the Norwegian economy significantly (Qvigstad, 2011). Therefore, we incorporate the US Treasury bonds and swaps in the model. This is to capture the current and future interest rates in the US and the inflation expectations. The same applies to why we include bonds and swaps issued by Norges Bank.

### Company-Specific Predictors

The next set of predictors in the Returns Model is company-specific. These variables are created using quarterly published financial and the closing price for each stock. The majority of the variables are ratios. The chosen ratios, inspired by the literature review, are frequently used by investors and financial analysts. When using ratios as predictors, the variables are scaled, creating a more common foundation. Furthermore, companies can more manageably be compared based on ratios. We present an overview of the company-specific ratios in table 3.

| **Pricing ratios** | **Leverage & liquidity ratios** |
|---|---|
| Price-to-book value (P/B)<br>Price-to-earnings (P/E)<br>EV-to-EBIT (EV/EBIT)<br>EV-to-EBITDA (EV/EBITDA) | Debt-to-Equity (D/E)<br>Debt-to-Cash (D/C)<br>Debt-to-Sales (D/S)<br>Liabilities-to-Assets |
| **Profitability ratios** | **Size measurements** |
| Return on equity (ROE)<br>EBIT-to-Assets | Revenue<br>Enterprise Value |

*Table 3, Categorization of company-specific predictors*

*Price-to-book value (P/B)*

Price-to-book is a ratio given by the following formula:

$$P/B = \frac{Price\ per\ share}{Equity\ book\ value\ per\ share}$$

Price-to-book is used to uncover assets in a firm that are on such a high level of intangibly that they are hard to value, and therefore not present on the balance sheet. Such assets could be human capital, trademark value, and customer relationships. In theory, a company with a low, but positive, price-to-book ratio shall be more favorable than a company with a high price-to-book. A low price-to-book ratio insinuates that a more significant part of the investment in the company can be recovered by selling some of the assets in the company. However, this presumes that the company assets are priced correctly. As such, a low price-to-book value might suggest that the company assets are overvalued.

*Price-to-earnings (P/E)*

Price-to-earnings (P/E) is used to analyze at what rate a company is earning back the equity invested by the shareowners. The inverse of the price-to-earnings ratio is the equity discount rate. The formula for the P/E is given by:

$$P/E = \frac{Price\ per\ share}{Earnings\ per\ share}$$

A low P/E implies a rapid payback time at the current earnings level. As such, a high P/E implies the opposite.

*Return on equity (ROE)*

ROE measures the return on the shareholder equity. Shareholder equity is the paid-up equity injected into the company when the company is established or later through share issues and capital increases. Return on equity is calculated with the following formula

$$ROE = \frac{Earnings}{Shareholders\ equity}$$

ROE measures how well a company deploys shareholder equity and is linked to what return an investor can expect on an investment.

*Revenue*

Despite that, almost all the other company-specific predictors are ratios, *revenue* is an absolute value. The current revenue level is used to classify the size of a company based on sales. According to the Small-form Effect, smaller companies tend to achieve higher long-term excess returns than larger firms (NBIM, 2012). The absolute value of revenue is included in the model to catch some of this effect and other dynamics.

*Sector*

Companies in different sectors are valuated differently. An example of such differences is the high pricing of tech companies in the later years. The high pricing is accepted because most of the future cash flows from such tech companies lay far into the future (Kim, Pukthuanthong-Le, & Walker, 2008). Furthermore, we assume that macroeconomic changes and shocks affect sectors differently. Therefore, we include a sector categorization in the model.

### Variables Exclusively for Non-Financial Companies

Commercial banks and insurance companies, categorized as financial companies, have a somewhat different capital structure than other companies. For instance, the balance sheets of commercial banks do not include loans in the traditional manner like other companies. Instead, commercial banks borrow money from the central banks and through customer deposits. Moreover, the debt of insurance companies also has some unique characteristics. A large portion is associated with insurance liabilities to customers and therefore is not linked to the financing of the company. This considered, we do not include the following predictors for the financial companies in the model.

*Enterprise value (EV)*

We include the enterprise value (EV) of a company as an absolute value in the model. The reasoning for this is similar to argumentation for including revenue as an absolute value. In short, we use EV to incorporate the size of a company. The following formula calculates EV:

$$EV = Market\ capitalization\ (MC) + Debt - Cash,$$
$$Where\ MC = Number\ of\ shares\ \times share\ price$$

*EV-to-EBIT (EV/EBIT)*

EBIT, or *Earnings Before Interest and Taxes,* represents the earnings a company can produce before any stakeholders get their share. EV/EBIT is regarded as a broader version of the P/E ratio. Instead of focusing solely on the equity holders, EV/EBIT focuses on a wider group of stakeholders. This can provide an idea of the value of the company, regardless of the capital structure. A company will likely change its capital structure multiple times throughout its lifetime. Therefore, using a valuation ratio that focuses on the full earnings potential through EBIT might give an investor a more holistic overview of a company. EV/EBIT is given by:

$$EV/EBIT = \frac{EV}{EBIT}$$

*EV-to-EBITDA (EV/EBITDA)*

EBITDA, *Earnings Before Interest, Taxes, Depreciation, and Amortization*, represent the result a company can generate solely on its operating activities. This can be thought of as how effectively a company generates value from its input factors. The reasons we include the

EV/EBITDA in the model are similar to those mentioned for EV/EBIT as both represent earning potential. The formula for the EV-to-EBITDA ratio is given by

$$EV/EBITDA = \frac{EV}{EBITDA}$$

*Debt-to-Equity (D/E)*

There is a linear relationship between the portion of debt to equity and the perceived risk for the equity in a firm. Traditionally, this is quantified as the discount rate. Furthermore, a higher portion of debt raises the risk for equity holders (Modigliani & Miller, 1958). The nature of debt claims causes this. If a company goes bankrupt, debt holders have priority for their claim above equity holders. Therefore, a firm taking on higher leverage is equivalent to increasing the risk for equity holders. This effect is captured by adding D/E to the model. The D/E ratio is given by:

$$Debt/Equity = \frac{Capitalized\ debt}{Market\ capitalization}$$

*Debt-to-cash (D/C)*

As mentioned above, debt imposes a risk. However, high amounts of cash might counteract parts of this risk associated with debt. A firm with a low debt-to-cash ratio will have better conditions to handle potential market declines. Therefore, the risk imposed by debt is tied to the debt-to-equity ratio and the debt-to-cash ratio. We formulate the D/C ratio as:

$$Debt/Cash = \frac{Debt}{Cash}$$

*Debt-to-Sales (D/S)*

We include D/S in the model to indicate how much dept a firm can handle, independent of its effectiveness of recourses. The ratio is given by:

$$Debt/Sales\ = \frac{Debt}{Sales}$$

*Liabilities-to-Assets*

The liabilities of a firm might consist of various accounting items, such as debt, in the form of bank loans or bond issues, accounts payable, or pension obligations. Common for them all is that they impose some sort of claim on the assets of the company, with priority above the invested capital from shareholders. Liabilities-to-Assets measures the size of these claims and is given by:

$$Liabilities/Assets = \frac{Liabilities}{Assets}$$

*EBIT-to-Assets*

Investors use EBIT-to-Assets to determine how effective a company is at generating profit on its assets. EBIT-to-Assets has some of the capabilities as EV/EBIT. However, by including cash, a more comprehensive picture of the effectiveness of the company is captured. The ratio is given by the formula:

$$EBIT/Assets = \frac{EBIT}{Assets}$$

### 5.3.2 Covariance Model Predictors

As for the Returns Model, the Covariance Model also includes macroeconomic and economic variables. The variables incorporated in the Covariance Model are: *Norwegian GDP* and *CPI, 3-months NIBOR, North Sea Oil price, Currencies, Global stock indexes and US & NO bonds and swaps*. For a thorough explanation of these predictors, please referrer to the previous section and table 2. Furthermore, to separate the observations from each other, we include company variables indicating which covariance/variance the model is ought to predict.

In addition to the macroeconomic and economic predictors, we incorporate a variable called *Variance (0/1)* and in the Covariance Model. *Variance (0/1)* incapsulates whether the predicted value represents a variance or a covariance in the covariance matrix. The variance of stock returns tends to be higher than the associated covariance with other returns. Moreover, the variance of stock returns is always positive due to the squared feature in the variance formula. On the other hand, covariance can be both positive and negative due to the covariance coefficient in the covariance formula. In figure 13, we illustrate the two clarifications.

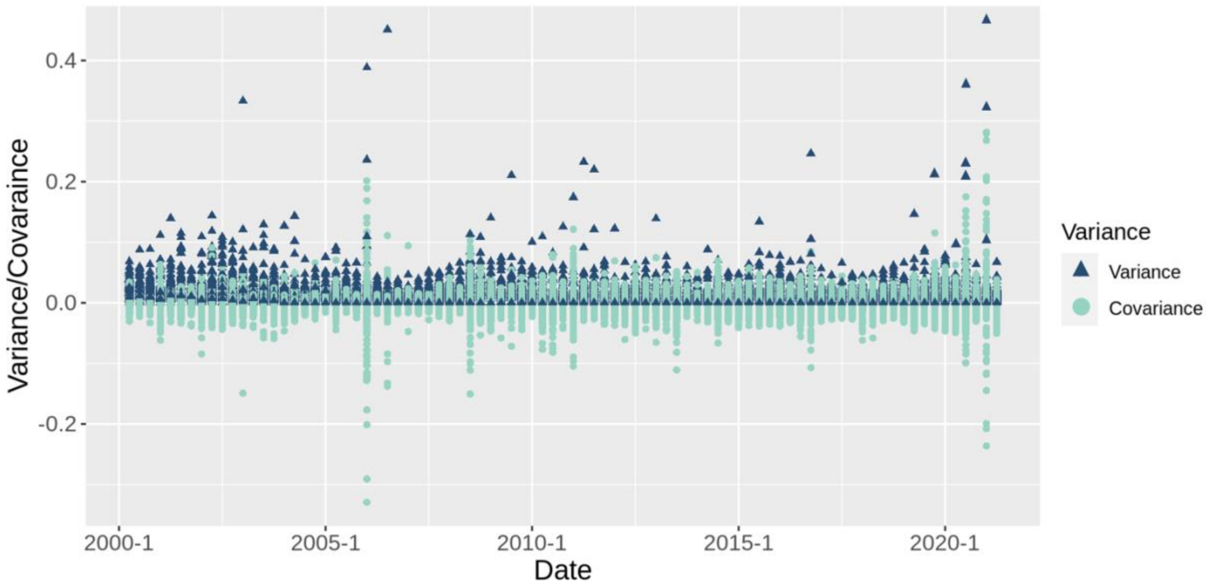


*Figure 13, Distribution of the response variable variance/covariance in the data set*

Figure 13 shows that the variances of stock returns generally have a larger magnitude than the covariance of stock returns. Further, the plot shows how the variances only have positive values while the covariances fluctuate between positive and negative values. Additionally, the figure shows periodic prominent increases in magnitude for the covariances. These increases clearly show that stock returns covariance more during certain periods. It is fair to say that this is especially true for bearish markets because we can observe the covariances increasing during both the financial crises in 2008 and the corona pandemic in 2020.

## 5.4 XGBoost Thesis Specifications

The XGBoost algorithm relies on hyperparameters. In short, hyperparameters are predetermined values or weights used in the learning process of an algorithm. These parameters must be tuned to make the algorithm learn as adequately as possible and find the optimal bias-variance trade-off. We only use the training sample data to tune the models in this thesis. Furthermore, to avoid overfitting, we utilize time-series cross-validation. For both models, we tune a total of six parameters. The parameters are described in table 1.

| Parameter | Description |
| --- | --- |
| Number of trees | The number of trees grown in the model |
| Minimum child weight | The minimum number of data points in a node required for the node to be split further |
| Tree dept | The maximum number of splits for each tree |
| Learn rate (shrinkage) | The rate at which the model adapts from iteration-to-iteration |
| Loss reduction | The required reduction in the loss function to split further |
| Sample size | The proportion of data exposed to the fitting routine |

*Table 4, Parameters tuned in for the XGBoost*

Tuning an XGBoost model can be a tedious process. The process involves assessing the performance of different combinations of parameter values to see which combination performs best based on a selected performance measure. We apply RMSE as a performance measure, as this is considered best practice. To obtain an unbiased selection of the parameters, we apply only the training data in the parameter selection. Furthermore, to measure the performance of the different combinations of parameters, we apply a 10-fold time-series cross-validation in the selection process.

Moreover, we create a grid of parameters for tuning, using a grid search to cover the whole parameter spectrum. The grid is created using a maximum entropy distribution approach to obtain an unbiased grid search for the parameters. Due to high dimensional data and extensive computational power requirements, we use Google Cloud Platform and a virtual computer to tune the models. Further, XGBoost requires all input to be of class numeric. As such, characters and factors are not applicable. Consequently, we must remove or replace all non-numeric variables with binary variables, as suggested by Chen, He, Benesty, and Tang (2021). In practice this means we must replace columns containing company names and sector names with binary columns.

# 6. Results and Discussion

To assess the Thesis Model, we use the last four quarters in the data set as a test period. As such, we study the quarterly results from Q4 2020 to Q3 2021. We desire to predict for one whole year to get a more comprehensive understanding of the predictability. Note that we train the model on all available data from the past. Hence, the prediction process can be considered a four-fold time-series cross-validation, with a cumulative growing training set as time progresses.

First, we find the expected optimal weights in the portfolios using the methodology described in section 5.1. In short, we obtain these weights based on the criteria to maximize the expected Sharpe ratio of the portfolios. Further, we use the actual returns from the stocks in the portfolio, and the associated actual covariance matrix, to calculate the actual Sharpe ratio for the portfolio for each quarter in the test period. We summarize the process in figure 14.

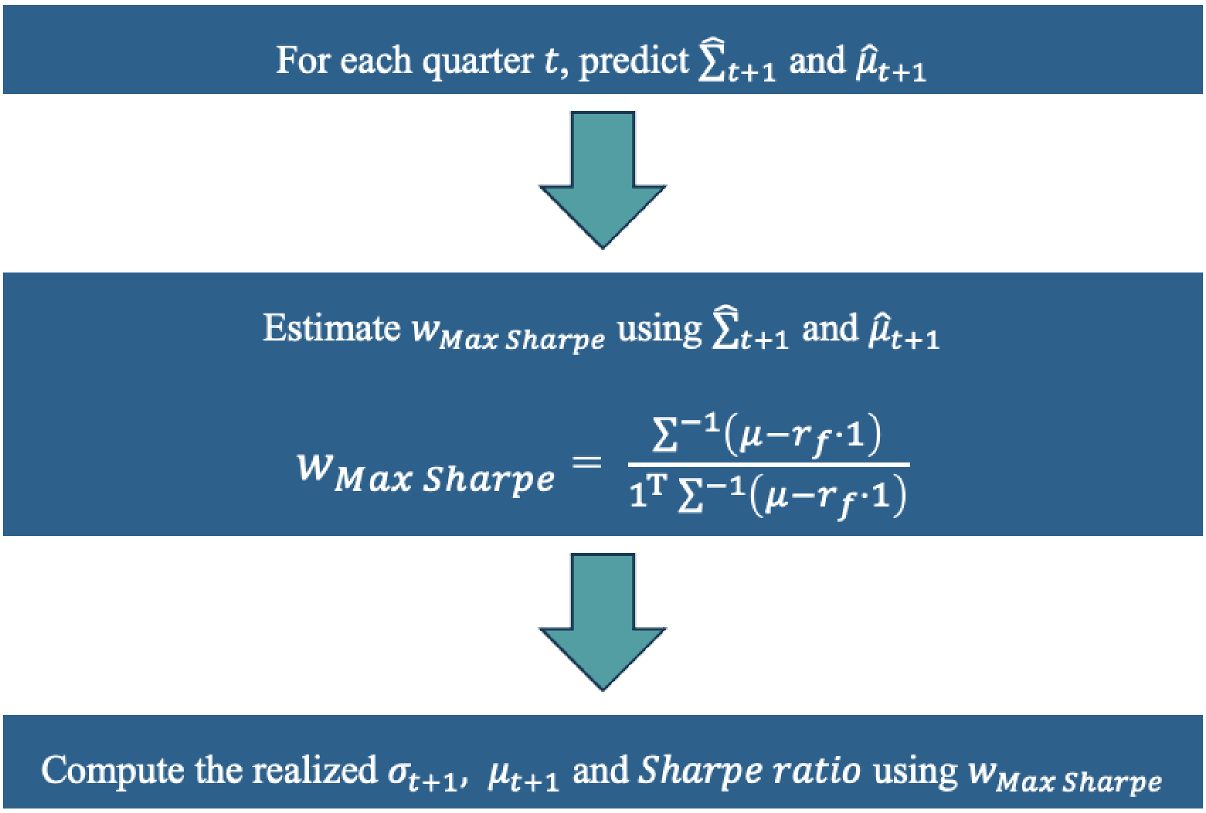


*Figure 14, Process of finding the optimal weights in the Thesis Model*

Following the recipe from figure 14, we obtain the results shown in table 5. In the table, $\mu$ is the realized quarterly discrete return, $\sigma$ is the realized quarterly volatility, and $Sharpe$ is the realized quarterly Sharpe ratio.

| $Quarter$ | $r_f$ | $\mu$ | $\sigma$ | $Sharpe$ |
|---|---|---|---|---|
| $Q4$ 2020 | 0.0028 | 2.03 | 0.67 | 3.05 |
| $Q1$ 2021 | 0.0039 | 0.13 | 0.42 | 0.30 |
| $Q2$ 2021 | 0.0038 | 0.00 | 0.18 | (0.04) |
| $Q3$ 2021 | 0.0020 | 0.06 | 0.18 | 0.35 |
| $Mean$ | 0.0031 | 0.56 | 0.36 | 0.92 |

*Table 5, Quarterly realized financial performance Thesis Model*

From table 5, we observe that the Thesis Model performs with varying financial results between the quarters. Both the realized return and the associated risk vary significantly, causing heavy fluctuations in the Sharpe ratio. For instance, in Q4 2020, the model achieves an extraordinary return of $203\%$. However, the volatility associated with this return is $67\%$, higher than in any other quarter. Still, due to the tremendous return, the Sharpe ratio is divine. On the contrary, in Q2 2021, the model obtains a return of $0\%$ for the whole quarter with an associated volatility of $18\%$. Due to the impact of the risk-free rate, the model yields a negative Sharpe in the third test quarter. Hence, investing in the risk-free option would have been preferable. In total, throughout the test period, the model achieves an average Sharpe ratio of $0.92$. However, the high Sharpe ratio in Q4 2020 undoubtedly increases the average. When excluding Q4 2020, the Thesis Model achieves an average Sharpe ratio of only $0.20$, demonstrating the skewed contribution to the average Sharpe between the test quarters.

To better understand the performance of the Thesis Model, we can compare the model to other portfolio selection methods. Table 6 below compares the Thesis Model financial performance to the classical Markowitz portfolio and the equally weighted $(1/N)$ portfolio. The classical Markowitz model we apply utilizes historical returns to calculate both expected returns and the covariance matrix.

| | *Thesis model* | | | *Markowitz* | | | *1/N* | | |
|---|---|---|---|---|---|---|---|---|---|
| $Quarter$ | $\mu$ | $\sigma$ | $SR$ | $\mu$ | $\sigma$ | $SR$ | $\mu$ | $\sigma$ | $SR$ |
| $Q4$ 2020 | 2.03 | 0.67 | 3.05 | 0.33 | 0.13 | 2.55 | 0.37 | 0.11 | 3.36 |
| $Q1$ 2021 | 0.13 | 0.42 | 0.30 | 0.13 | 0.13 | 1.03 | 0.14 | 0.09 | 1.46 |
| $Q2$ 2021 | −(0.00) | 0.18 | (0.04) | 0.11 | 0.12 | 0.83 | 0.03 | 0.10 | 0.23 |
| $Q3$ 2021 | 0.06 | 0.18 | 0.35 | 0.06 | 0.05 | 1.09 | 0.03 | 0.07 | 0.34 |
| $Mean$ | 0.56 | 0.36 | 0.92 | 0.16 | 0.11 | 1.37 | 0.14 | 0.09 | 1.35 |

*Table 6, Quarterly financial performance of Thesis Model compared to other portfolio selection methods*

Table 6 shows that the Thesis Model outperforms both the classical Markowitz method and the $(1/N)$ portfolio in terms of average portfolio return. Once again, this is due to the deviant performance in Q4 2020. Nevertheless, despite an impressive Sharpe ratio of 3.05 in the first quarter, both the classical Markowitz and the $(1/N)$ portfolio outperforms the Thesis Model when it comes to the average Sharpe ratio. However, the explanation is apparent. The volatility of both comparison methods is much smaller and more stable. This indicates that the risk of the portfolios from these models is much lower. Hence, even though the mean return of the Thesis model is ample compared to the other models, the mean Sharpe ratio is lower due to the high risk in the Thesis model.

To further contextualize the performance, we compare the Thesis Model to a stock market index. Commonly, portfolios are measured using an index as a benchmark. Thus, we find it intuitive to follow the same approach. We compare the model to the Oslo Stock Exchange Benchmark Index (OSEBX) as this thesis only considers Norwegian stocks. The comparison is shown in table 7.

| | *Thesis Model* | | | *OSEBX* | | |
|---|---|---|---|---|---|---|
| $Quarter$ | $\mu$ | $\sigma$ | $Sharpe$ | $\mu$ | $\sigma$ | $Sharpe$ |
| $Q4$ 2020 | 2.03 | 0.67 | 3.05 | 0.11 | 0.10 | 1.17 |
| $Q1$ 2021 | 0.13 | 0.42 | 0.30 | 0.09 | 0.07 | 1.22 |
| $Q2$ 2021 | 0.00 | 0.18 | (0.04) | 0.06 | 0.07 | 0.83 |
| $Q3$ 2021 | 0.06 | 0.18 | 0.35 | 0.04 | 0.07 | 0.51 |
| $Mean$ | 0.56 | 0.36 | 0.92 | 0.07 | 0.07 | 0.93 |

*Table 7, Quarterly realized financial performance of thesis model and OSEBX*

The interpretation of the results in table 7 is comparable to the interpretation of table 6. The Thesis Model is superior to the index when considering average portfolio return, thanks to the high portfolio return in Q4 2020. However, OSEBX is performing considerably better in terms of the amount of risk associated with the realized return. This results in OSEBX exceeding the model on average Sharpe ratio. Interestingly, the index has such a high diversification that the average volatility is lower than the $(1/N)$ portfolio.

Portfolios from the Thesis Model are constructed with the objective to maximize the Sharpe Ratio. However, from the discussion above, it is evident that the model struggles to be consistent on this topic. Still, before we condemn the model, we assess the accumulated wealth. This method evaluates a portfolio based on the total return of the period, translated into monetary wealth. As such, we compute the accumulated wealth for the Thesis Model, the classical Markowitz, the $(1/N)$, and OSEBX. In Figure 15, we display the wealth plot, showing the development in accumulated wealth for the test period.


*Figure 15, Generated wealth*

The wealth plot shows a phenomenal wealth generated by the Thesis Model, especially in the first test quarter. Over the four-quarter test period, the Thesis Model has a one-year return of 264%. In comparison, the classical Markowitz, the $(1/N)$, and OSEBX yield 77%, 65%, and 36%, respectively. From a portfolio return point of view, this is an extraordinary result. Nevertheless, as we have discussed, the abnormal return comes at the price of high portfolio volatility. The considerable differences in volatility between the Thesis model and the other methods and OSEBX are even more apparent when we study figure 15. Especially, the portfolio volatility for the Thesis Model is extreme in Q2 2020. From January 18th 2021 to March 15th 2021 the accumulated wealth decreases by 46%. In contrast, the other, less risky alternatives do not suffer from similar fluctuations.

We find the results above to be rather interesting. The Thesis Model has a higher average quarterly return than any other method by far. Nonetheless, this high average return comes with the cost of a high amount of risk, which the model is punished for when calculating the Sharpe ratio. One of the most surprising dynamics is the fluctuations within each quarter. Such fluctuations do not characterize a well-diversified portfolio. One cause of these dynamics could be that the model relies heavily on a few shares with abnormal returns. To study these dynamics, we first investigate the distribution of stock returns for the test period in figure 16.

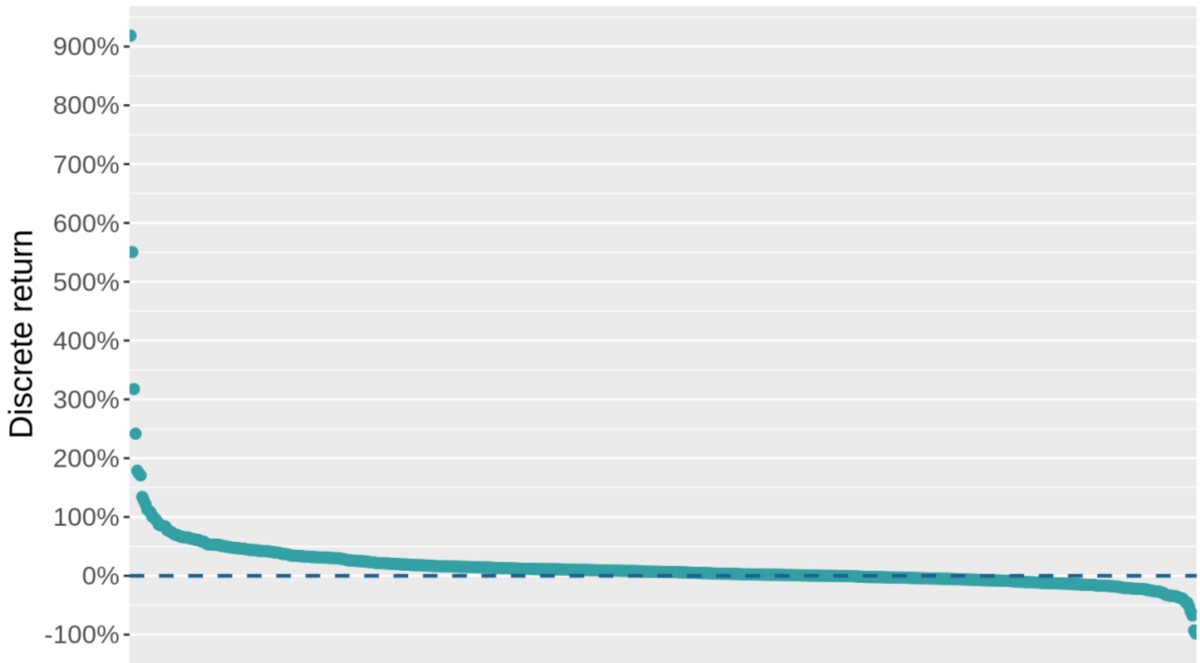


*Figure 16, Distribution of discrete quarterly stock returns*

By studying figure 16, we observe some outliers regarding quarterly returns, both positive and negative. The realized portfolio return presented earlier in this section indicates that the Thesis Model might invest heavily in some of these outlier returns, considering the enormous result of 264% one-year return. Another indication of this phenomenon is the high average volatility for the Thesis Model shown in table 6, which is far higher than the volatility for the $(1/N)$ portfolio and OSEBX. This indicates that the Thesis Model has a far lower degree of diversification than the $(1/N)$ portfolio and OSEBX. To investigate if the Thesis Model stakes heavily on a small fraction of the shares, we can study the allocated portfolio weights in the first test quarter, the quarter with the highest return for the Thesis Model. We present the weights in figure 17.

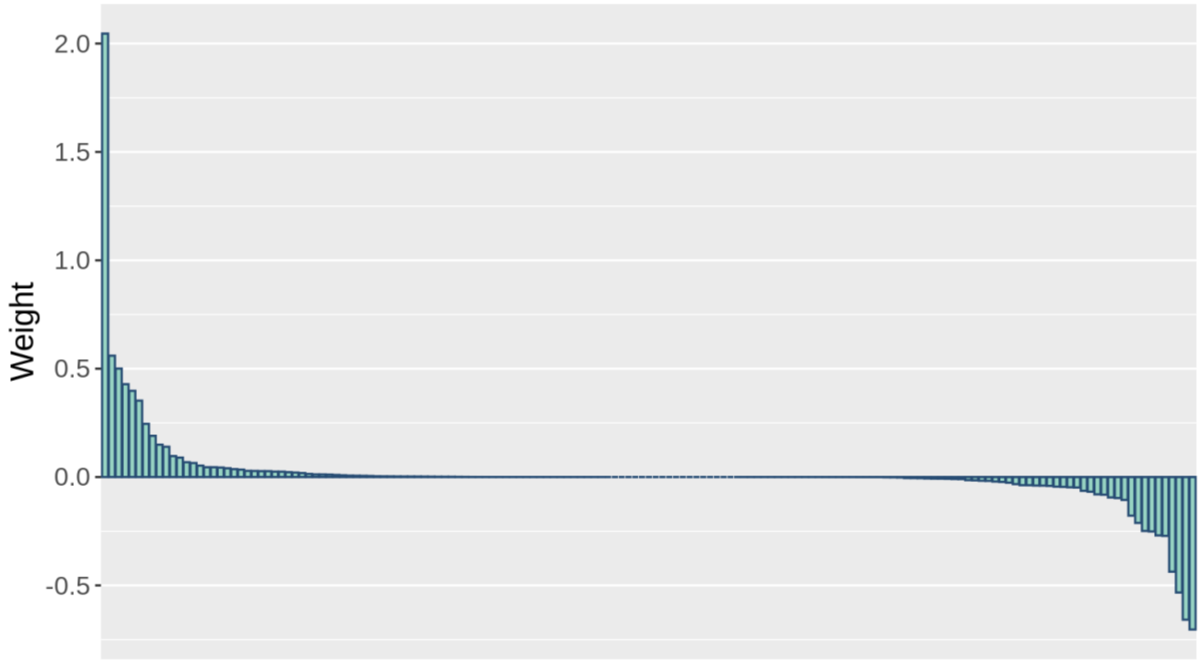


*Figure 17, Allocated weights by the Thesis Model in Q4 2020*

The figure above shows that the Thesis Model strongly emphasizes a low number of shares in Q4 2020. The highest invested position is weighted 340 times higher than the average weight. Such allocation shows a low degree of diversification. This could indicate that the model bets on a few shares and that these bets have a highly positive outcome for the Thesis Model. If so, this is a worrying observation, implying that the results presented earlier might not represent the true model performance. Thus, we find it necessary to study this matter further to validate the robustness of the Thesis Model.

To investigate if a few shares cause the Thesis Model performance presented above, we can study the contribution of each share in each quarter. We calculate the portfolio contribution by multiplying the weight of a share with the return for the same share. Figure 18 below shows the distribution of the portfolio contribution of the stocks in each quarter for the Thesis Model. We use box plots where observations inside the box are within the range of the 95% and 5% percentile.

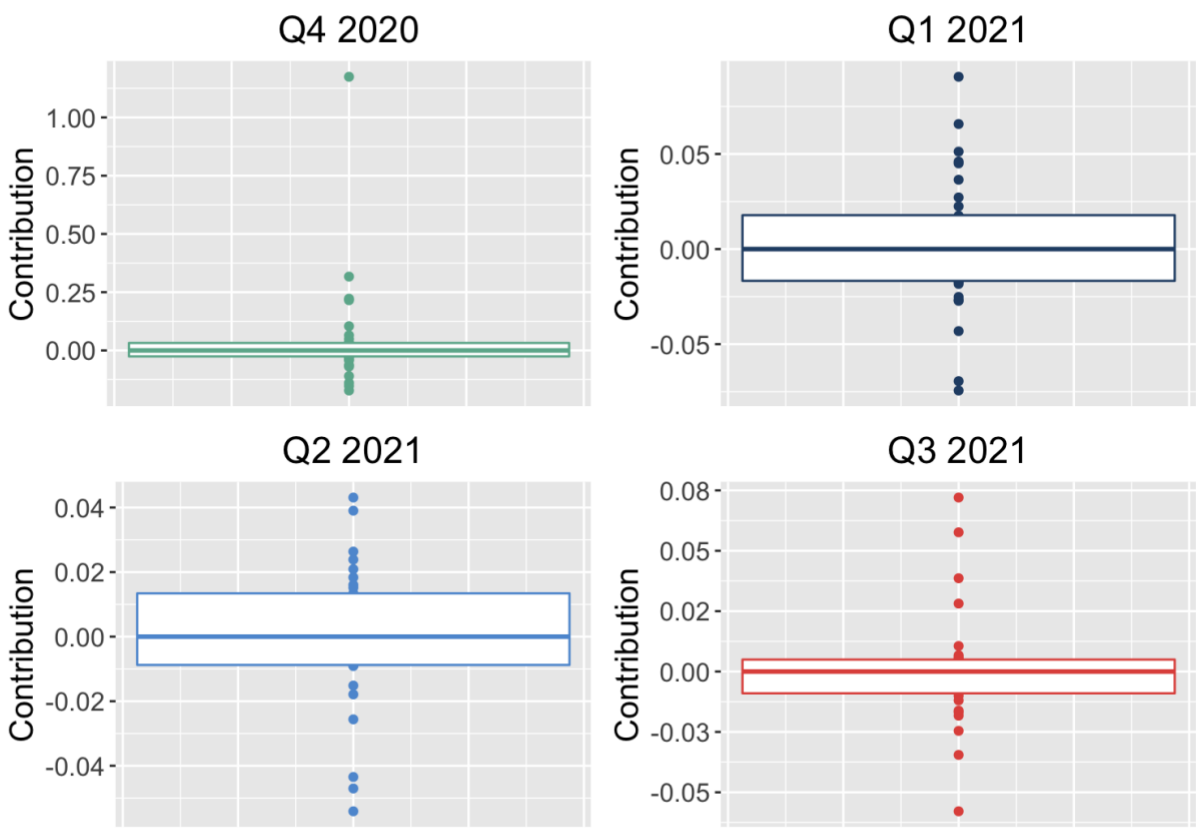


*Figure 18, Distribution of portfolio contribution each quarter*

Figure 18 shows some distinct outliers. Q4 2020 is especially striking, where one investment position constitutes almost 60% of the total portfolio return for the said quarter. This remark, combined with the previous discussion, confirms the suspicion that the Thesis Model relies heavily on a small number of shares, which is the opposite of diversification. As such, we raise the question of whether the Thesis Model would show staggering results if we were to run the model with different conditions.

We use the initial results as a baseline to test the Thesis Model under different conditions. Then, we remove the stocks with portfolio contributions outside the lower 5 % percentile and upper 95 % percentile for each quarter. As such, we obtain a new data set we can use to test the robustness of the model and the associated results. Thus, we rerun the Weight Allocation Model for the Thesis Model with the new conditions. We also rerun the classical Markowitz method and the (1/N) portfolio. The results of the reruns are shown below in table 8.

| | ***Thesis model*** | | | ***Markowitz*** | | | ***1/N*** | | | ***OSEBX*** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $Quarter$ | $\mu$ | $\sigma$ | $SR$ | $\mu$ | $\sigma$ | $SR$ | $\mu$ | $\sigma$ | $SR$ | $\mu$ | $\sigma$ | $SR$ |
| $Q4$ 2020 | 0.42 | 0.68 | 0.61 | 0.23 | 0.16 | 1.40 | 0.29 | 0.10 | 2.82 | 0.11 | 0.10 | 1.17 |
| $Q1$ 2021 | 0.43 | 0.26 | 1.63 | 0.15 | 0.12 | 1.23 | 0.14 | 0.10 | 1.40 | 0.09 | 0.07 | 1.22 |
| $Q2$ 2021 | (0.49) | 0.55 | (0.91) | 0.08 | 0.08 | 0.91 | 0.02 | 0.11 | 0.19 | 0.06 | 0.07 | 0.83 |
| $Q3$ 2021 | (0.09) | 0.25 | (0.38) | 0.03 | 0.05 | 0.64 | 0.02 | 0.07 | 0.22 | 0.04 | 0.07 | 0.51 |
| $Mean$ | 0.07 | 0.43 | 0.24 | 0.12 | 0.10 | 1.04 | 0.12 | 0.09 | 1.16 | 0.07 | 0.07 | 0.93 |

*Table 8, Financial performance without using stock of top and bottom 5% contributors for the Thesis Model*

By changing the initial conditions, the characteristics of the portfolios constructed by the Thesis Model change considerably. There is a noteworthy decrease in average quarterly return, dropping from 56% to 7% with the new conditions, a decrease of almost 90%. However, the average volatility for the Thesis Model does not decrease. Instead, the average volatility increases from 36% initially to 43% with the new conditions. Consequently, the vast reduction in average return for the Thesis Model combined with the slight increase in average volatility demolishes the Sharpe ratio of the selected portfolios.

Furthermore, the Thesis Model now performs worse than any other presented portfolio selection methods or index, regardless of measurement. This diverges greatly from the initial results presented earlier in the thesis. Likewise, we recall that the Thesis Model initially annihilates any competition when comparing the different portfolio selection methods and OSEBX in terms of generated wealth. The same comparison can be made with the new conditions. This is shown in figure 19.


*Figure 19, Generated wealth with new conditions*

The graphics from figure 19 supports the findings we have previously addressed. The plot clearly illustrates a new hierarchy in terms of portfolio return between the different portfolio selection methods and OSEBX. The Thesis Model is no longer the undisputed winner regarding portfolio return. Instead, with the new conditions, it ends up losing wealth at the end of the test period. Note that the accumulated wealth also decreases for both the classical Markowitz and the $(1/N)$ portfolios. However, the decrease is negligible compared to the decrease for the Thesis Model. Furthermore, the Thesis Model is still the most volatile of the four time-series, with vast fluctuations throughout the test period supporting the high average volatility presented in table 8.

The results from running the Thesis Model on new conditions are alarming. One aspect is that the Thesis Model, under new conditions, performs very poorly. However, the most important takeaway is the inconsistencies in terms of results. Initially, we addressed the fact that the Thesis Model struggles to construct portfolios with Sharpe ratios that outperform the market. Nevertheless, due to the huge returns in the first test quarter, the model outperformed the other models and OSEBX in terms of generated wealth. However, the performance argument expressed in accumulated wealth is no longer applicable when changing the conditions. Instead, the development in generated wealth is the complete opposite. Moreover, the realized ex-ante Sharpe ratio for the Thesis Model is much lower. This supports the suspicion that the model is consistently failing to select portfolios with a market-beating Sharpe ratio and that the initial results we obtain do not represent the true performance of the model. Nonetheless, to explore the cause of the mentioned inconsistencies, we are intrigued to examine the statistical performance of prediction models.

## 6.1 Statistical Performance

To assess the two statistical prediction models, the Returns Model and the Covariance Model, we use RMSE and MAPE as performance measurements. As the models are run once for each of the four test quarters, the performance measures presented below in table 9 represent the averages for the test period. This resembles a four-fold time-series cross-validation.

| $\boldsymbol{Measure}$ | $\boldsymbol{Returns\ Model}$ | $\boldsymbol{Covariance\ Model}$ |
|---|---|---|
| $Mean$ | $-0.0029$ | $0.0401$ |
| $RMSE$ | $0.4573$ | $0.3281$ |
| $MAPE$ | $584\%$ | $151800\%$ |

*Table 9, Statistical performance of the prediction models*

The table above shows that statistically, the Returns Model and Covariance Model perform poorly. Especially the Covariance Model has a high MAPE, suggesting that the Covariance Model is not capable of finding a reliable approximation of the estimation function $\hat{f}$. The Returns Model also incurs a high MAPE and an average quarterly log-return misprediction of 46%. Again, this implies that the Returns Model struggles to find a reliable approximation of the estimation function. It appears that neither of the prediction models is able to identify beneficial patterns or relationships between the predictors and the response variables.

The poor statistical performance of the two prediction models explains the inconsistent financial performance of the Thesis Model presented in the previous section. Unreliable predictions of stock returns and covariance matrixes make it nearly impossible for the Weight Allocation Model to consistently find weights that yield good financial performance. The inaccuracy of the Returns Model and the Covariance Model indicates that the portfolio allocation process in the presented Thesis Model is close to random. With predictions not remotely resembling reality, the model is not likely to perform portfolio allocation with good financial results.

The lack of prediction accuracy clarifies why the Thesis Model struggles to perform consistently satisfying results. The initial financial result is, without doubt, impressive, and the model generates a one-year return most investors only can dream about. However, this result is undoubtedly nothing else than a good portion of luck. The inconsistency and inaccuracy of the Thesis Model make the model highly unreliable. It is fair to say that using the Thesis Model in its current form is close to doing portfolio allocation randomly. This is unmistakably illustrated with the exceptional poor statistical performance of the two prediction models.

## 6.2 Variable Importance

The main focus of this thesis is prediction performance. As described in the previous section, the statistical models perform with close to zero accuracy. However, discussing the variable importance in the two prediction models might provide useful insight. Below we present the variable importance for the Returns Model and the Covariance Model. We apply Shapley values to quantify the importance of the different variables in the model. The idea behind Shapley values is to give each variable credit based on its marginal contribution. Shapley value is the average of all the marginal contributions from a variable to all possible combinations of variables (Shapley, 1953).

### 6.2.1 Returns Model

Figure 20 illustrates the average variable importance for the Returns Model. The figure shows the top ten most important variables in terms of average Shapley values for the test period. The numerical average of the Shapley values can be interpreted as the average contribution to the predicted log-return for the different variables.

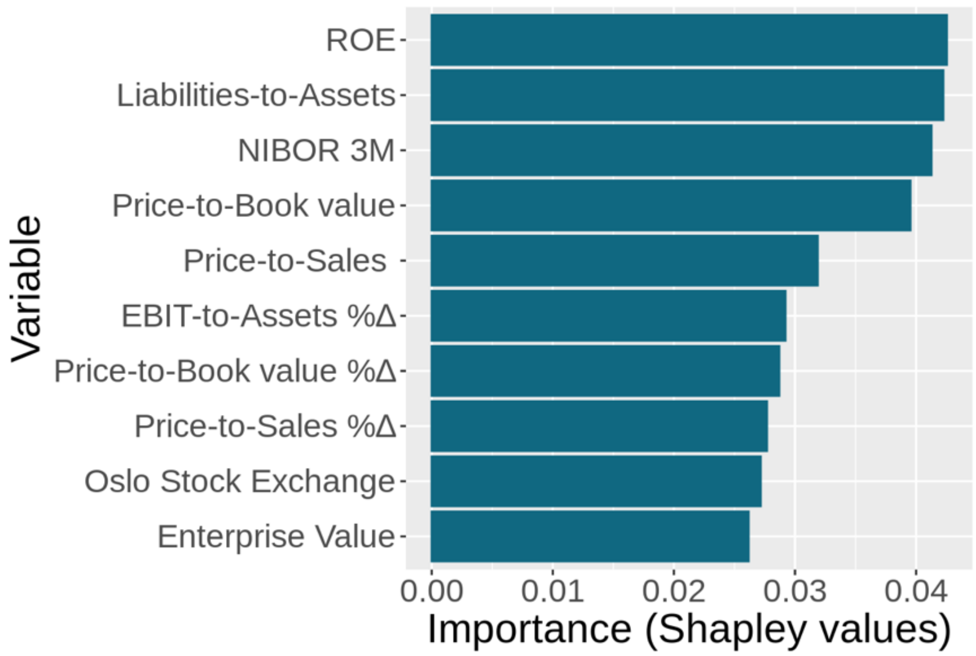


*Figure 20, Variable importance Returns Model*

Figure 20 shows that eight out of the ten most influential variables are company-specific. This is consistent with the literature review, claiming that company-specific characteristics influence long-term stock returns. Furthermore, we are not surprised that the 3-month NIBOR, as an absolute value, contributes relatively much. A low interest level entails a lower discount factor of future cash flows. This is particularly beneficial for companies like tech firms, where a substantial part of the cash flow has a long-term horizon.

The variable importance varies naturally between the top ten variables, ranging from 0.026 for *Enterprise Value* to 0.043 for *ROE*. Nevertheless, the Shapley values are not very large and vary to some extent between quarters. An interesting observation from figure 20 is that the majority of the ten most important variables are absolute values and not relative change. This is interesting as the response variable *Log.Return* is on a relative change format. We would expect to observe more variables to be on the relative change format.

### 6.2.2 Covariance Model

We exhibit the variable importance for the Covariance Model in figure 21. The interpretation of the Shapley values is the same as for the Returns Model. The only difference is that the Shapley values now express the average contribution to the covariance predictions for the different variables.

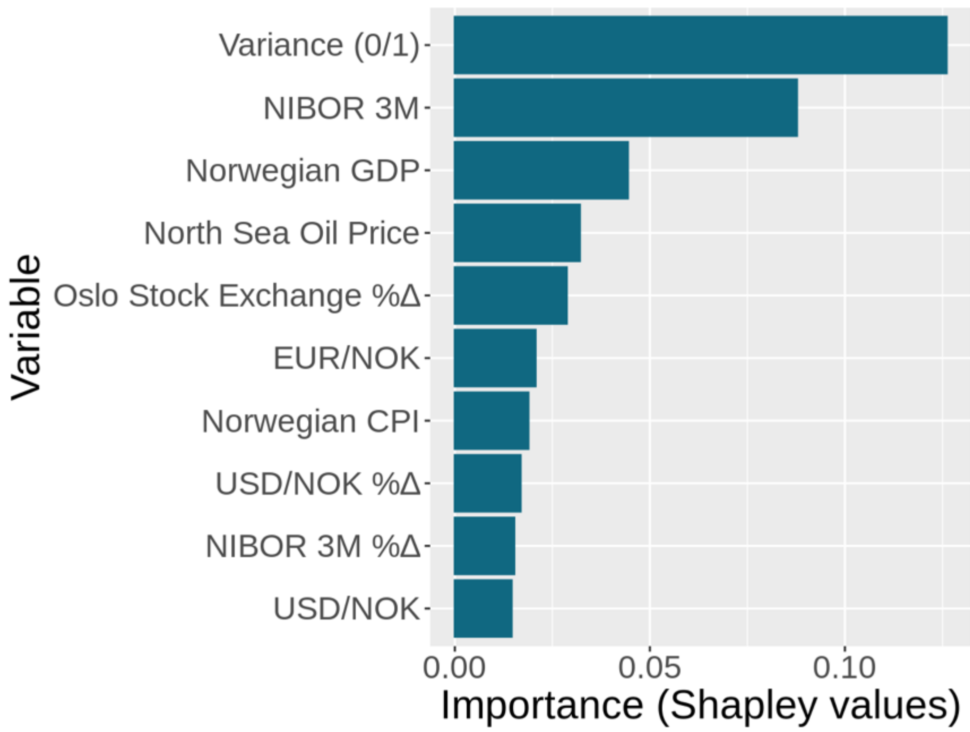


*Figure 21, Variable importance Covariance Model*

Figure 20 shows more variation in the average variable importance for the Covariance Model than the Returns Model. The *Variance (0/1)* variable is almost nine times as influential towards the predictions compared to *USD/NOK*. However, the strong influence from *Variance (1/0)* is not surprising considering the discussion in section 5.3.2. The second most influential variable is *NIBOR 3M*. Central banks commonly use interest rate reduction during a recession and other situations when the market is bearish. The opposite is common in bullish markets. Thus, we argue that *NIBOR 3M* functions as a thermometer for the Norwegian economy. Moreover, from the literature review, we know that returns are more correlated in very bearish markets. Hence, *NIBOR 3M* might indicate when returns are strongly correlated.

## 6.3 Limitations

There are numerous limitations to this thesis. First, the nature of the data limits the Thesis Model. The time horizon of the quarterly data ranges from Q1 2000 to Q2 2021. This, combined with the fact that there is a limited number of stocks traded on OSE, strangles the number of observations available for both prediction models. Also, as shown in Figure 7, Number of quarterly reports per quarter in the data set the availability of observations is skewed. As we approach the present, the number of quarterly observations increases drastically. Part of the explanation is poor data quality from the early years. Further, from 2000 to 2021, the number of companies listed on Oslo Stock Exchange has multiplied, causing a natural increase in data availability. Regardless, the lopsidedness in the data distribution could be considered a weakness.

Second, financial reports are not available immediately after the quarter-end. One of the key assumptions in this thesis is that accounting data is available as soon as the reporting quarter is ended. On one side, we could argue that the financial information, in theory, is available when a quarter ends. However, in practice, this argument is hollow. In reality, we know that the publication of quarterly reports is spread throughout the next quarter. Subsequently, the macroeconomic and economic situation at the beginning of the quarter does not reflect the current economic state when the quarterly report is released. Conclusively, the model is not ideal for operationalization purposes.

Alternatively, we could lag the model one quarter. This would make the model more realistic as the information from the reports would be public by the end of the next quarter. Nevertheless, the issue of inaccurate publication dates is still apparent. Furthermore, the likelihood that the market has already exploited and adjusted the stock price for the information in the quarterly report is high. That said, as machine learning algorithms potentially can discover hidden patterns, there is possible that this approach could perform better than the existing model. Nevertheless, to make the model more realistic, we should have used the exact release date when collecting non-company-specific information and stock prices. Unfortunately, the availability of these dates is highly limited, mainly for older observations. Consequently, reports where the release date is unavailable would have to be discarded, and the number of observations would have been even more limited.

Third, we have made some assumptions regarding the financial trading aspect. In the model, we assume no transaction costs. This assumption is unrealistic. All share brokers require a commission when trading stocks through their platforms. In particular, short sales of stocks are costly. This is due to high initiation costs and interest costs on borrowed shares. Secondly, we have assumed no margin calls. A margin call occurs when a portfolio equity value falls below a certain threshold. This threshold is established by the individuals lending out the stocks used for a short sale. In the occurrence of a margin call, these lenders will demand the portfolio to be liquefied (all assets sold) to limit their losses. Thirdly, we have assumed no taxes on capital gains. While investment firms are exempted from taxes on capital gains in Norway, private investors are not. These assumptions are likely to overestimate the Sharpe ratios we have achieved with the presented model.

The fourth limitation involves the selection of variables. Both prediction models rely on assumptions about which predictors affect stock fluctuations. These variables were selected using previous literature and general conceptions in finance. Although there is evidence from other markets, there is limited evidence that these variables sufficiently reflect the changes in the Norwegian stock market. Therefore, other variables might better predict the expected return, and the covariance between the stock returns analyzed in the thesis. Further, we do not deploy cash flow statements when creating predictors. This is due to low data quality on cash flow statements from Eikon Refinitive and other sources for the considered stocks. This prevents us from using cash flow statements to create predictors.

Lastly, there are limitations related to predictions of financial markets in general. As stated in the literature review, predicting financial fluctuations is difficult due to various reasons. Further, the task becomes even more troublesome as the time horizon increases. In practice, many factors influence financial markets, not only the financial position of companies and macroeconomic development. Some of the most prominent are political conditions, the bounded rationality of humans, and the psychological aspect of financial markets (Henrique, Amorim, & Kimura, 2019). As the time horizon extends, these factors will cause unexpected changes in the market our model struggles to account for.

## 6.4 Further Research

We have applied the machine learning technique XGBoost to predict both stock returns and covariance between said returns. However, using other machine learning methods could have generated more accurate predictions. Thus, it might be interesting to compare the model performance obtained in this thesis with other methods. Examples could be other decision tree methods like random forest or Adaboost, support vector machines, and neural networks. Another approach would be to compare the model performance to other statistical methods such as Generalized Linear Models (GML), Nonlinear Regression Models, or a multivariate GARCH such as a GARCH-X or a DCC GRACH model.

Furthermore, the methodology presented in this thesis cloud be applied to a broader range of stocks and markets. The model might increase its predictive power by extending the data set with data from more stock markets. As mentioned in section 6.3, a more extensive data set might provide a more precise estimate of the true future test error for the prediction models. Additionally, a more extensive data set might offer more variety, enabling the model to discover patterns more clearly and better exclude noise in the data.

The third suggestion for further research relates to the limitation associated with the variable selection. More extensive research and experimentation regarding the impact of variables on Norwegian stock returns could improve the performance of the Thesis Model. As such, the prediction models could benefit from a broader and deeper variable selection process. For example, company-specific variables might provide supplementary predictive power to the Covariance Model. Likewise, if the cash flow statements from listed companies were of good data quality, variables building on these statements could be included in the prediction models.

# 7. Concluding Remarks

In this thesis, we have deployed the portfolio selection theory from Markowitz on the Norwegian equity market using predictions created by the XGBoost algorithm. The portfolio allocation process in the Thesis Model consisted of two steps. First, we predicted expected quarterly returns and coherent covariance matrices for stocks traded on Oslo Stock Exchange. Second, these predictions were then utilized to optimize the weights in the portfolios using matrix algebra. Such an approach has never previously been explored, to such an extent, in the Norwegian equity market.

We tested the Thesis Model through four quarters from Q4 2020 to Q3 2021. Initially, the Thesis Model delivers promising results with an impressive 264% one-year return. Compared to a classical Markowitz method, a $(1/N)$ portfolio, and OSEBX, the Thesis Model showed superior performance in terms of return. However, the Thesis Model simultaneously obtained higher average portfolio volatility than any other selection method. The high portfolio volatility causes the Sharpe ratio of the Thesis Model to become lower than any other presented portfolio allocation method or index.

Moreover, further investigation revealed that the Thesis Model relied heavily on a few stocks, which acquired abnormal returns. This result strongly contrasts with our initial desire to detect a well-diversified portfolio. Furthermore, the results changed drastically when we changed the conditions for the Thesis Model by removing the stocks with the highest portfolio contribution. After rebalancing the model, the one-year return dropped to -7%, far worse than any presented portfolio allocation method or index.

The results from the Thesis Model are disappointing. The objective of the quarterly portfolio optimization is to maximize the Sharpe ratio. Unfortunately, the Thesis Model is not able to construct portfolios that reliably aligned with this goal. Nevertheless, the model initially yields an impressive one-year return. However, under new conditions the performance change drastically. The statistical evaluation of the XGBoost prediction models entails that they both deliver highly inaccurate predictions, which propagates further through to the portfolio allocation process. Moreover, there is little evidence that the models can detect any patterns in the data beneficial for portfolio construction. In sum, the model struggles to foresee market developments, which accumulates into a model incapable of consistently performing with satisfying financial results.

To sum up, this thesis shows that predicting stock returns and covariances is a difficult task. From a quarterly perspective, there are many factors influencing the developments in the stock market. The combination of predictors applied in this thesis appears unsuitable to detect these developments. As such, for further research, we suggest investigating a more thorough variable selection with different variables. Furthermore, although XGBoost is one of the most renowned machine learning algorithms, there is no guarantee that the algorithm is optimal in this prediction context. As such, another suggestion for further research is to examine different machine learning methods to a broader range of stock markets and investigate whether this can provide better predictions. These suggestions might accumulate in a model proficient at predicting both risk and return of stocks sufficiently.

# Bibliography

Abarbanell, S. J., & Bushee, J. B. (1998). Abnormal Returns to a Fundamental Analysis Strategy. *The Accounting Review*, 19-45.

Achelis, S. (2000). *Technical Analysis from A to Z.* McGraw-Hill Education: New York.

Aksjenorge. (2021, July 21). *Nordmenn og Akjser*. Retrieved November 2021 from Aksjenorge: https://aksjenorge.no/wp-content/uploads/2021/07/Statistikk-Andre-Kvartal-2021_Final.pdf

Aalst, W. M., Rubin, V., Verbeek, H. M., Dongen, B. F., Kindler, E., & Günther, C. W. (2008). Process mining: a two-step approach to balance between underfitting and overfitting. *Software & Systems Modeling*, 87-111.

Altman, N., Bzdok, D., & Krzywinski, M. (2018). Statistics versus machine learning. *Nat Methods*, 233–234.

Andersen, T. G., Bollerslev, T., Diebold, F., & Labys, P. (2003). Modeling and Forecasting realized volatility . *Econometrica*, 579–625.

Angarita-Zapata, J. S., Alonso-Vicario, A., Masegosa, A. D., & Legarda, J. (2021). A Taxonomy of Food Supply Chain Problems from a Computational Intelligence Perspective. *Sensors*, 21(20):6910, 21(20), 6910. MDPI AG. Retrieved from http://dx.doi.org/10.3390/s21206910

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4*, 40-79.

Awoye, O. (2016). *Markowitz Minimum Variance Portfolio Optimization using New Machine Learning Methods.* London: University College London.

Basaka, S., Kar, S., Saha, S., Khaidem, L., & Sudeepa, R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 552-567.

Bertuah, E., & Sakti, I. (2019). The financial performance and macroeconomic factors in forming stock return. *Jurnal Riset Manajemen dan Bisnis (JRMB) Fakultas Ekonomi UNIAT, 4*(1), 511 - 522.

Bjørnland, H. C., & Thorsrud, L. A. (2013). Ringvirkninger, Norsk økonomi og olje. *Centre for Monetary Economics BI Norwegian Business School, Working Paper Series, 13*(7).

Bjergaard, A. P. (2020, 04 13). *Finansavisen.* Retrieved from https://finansavisen.no/nyheter/finans/2020/04/13/7515952/first-forvalter-thomas-nielsen-knuser-borsen-med-uortodoks-investeringsstrategi

Black, F., Jensen, M. C., & Scholes, M. S. (1972). The Capital Asset Pricing Model: Some Empirical Tests. *Studies in the Theory of Capital Markets*.

Blanca, M. J., Arnau, J., López-Montiel, D., Bono, R., & Bendayan, R. (2013). Skewness and Kurtosis in Real Data Samples. *Methodology, 9*(2), 78-84.

Bodie, Z., Kane, A., & Marcus, A. (2018). *Investments.* McGraw Hill.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 307-327.

Bucci, A. (2020). Cholesky–ANN models for predicting multivariate realized volatility. *Journal of Forecasting.*, 865-876.

Cai, Xianggao, Lai, G., & Lin, X. (2013). Forecasting large scale conditional volatility and covariance using neural network on GPU. *The Journal of Supercomputing* , 490–507 .

Cavalcante, R. C., Brasileiro, R., Souza, V., Nobrega, J., & Oliveira, A. (2016). Computational Intelligence and Financial Markets: A Survey and Future Directions,. *Expert Systems with Applications*, 194-211.

Chan, K. C., Gup, B. E., & Pan, M.-S. (1997). International Stock Market Efficiency and Integration: A Study of Eighteen Nations. *Journal of Business Finance and Accounting, 24*(6), 803-813.

Chan, K. L., Karceski, J., & Lakonishok, J. (1999). On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model. *The Review of Financial Studies*, 937-974.

Chang, P.-C., Liao, T. W., Lin, J.-J., & Fan, C.-Y. (2011). A dynamic threshold decision system for stock trading signal detection,. *Applied Soft Computing,* , 3998-4010.

Chaudhuri, K., & Smiles, S. (2004, Jan). Stock market and aggregate economic activity: evidence from Australia. *Applied Financial Economics*, *14*(2), 121-129.

Chen, P., & Zhang, G. (2007). How do accounting variables explain stock pricemovements? Theory and evidence. *Journal of Accounting and Economics, 43*, 219–244.

Chen, T., & Guestrin, C. (2016). XGBoost A Scalable Tree Boosting System. *ACM*, 785–794.

Chen, T., He, T., Benesty, M., & Tang, Y. (2021, 11 20). *Understand your dataset with XGBoost.* Retrieved from CRAN: https://cran.r-project.org/web/packages/xgboost/vignettes/discoverYourData.html

Chiriac, R., & Voev, V. (2011). Modeling and forecasting multivariate realized volatility . *Journal of applied econometrics* , 922-947.

Coşkun, Y., Seven, Ü., Ertuğrul, H. M., & Ulussever, T. (2017). Capital market and economic growth nexus: Evidence from Turkey. *Central Bank Review, 17*(1), 19-29.

Dimson, E., & Mussavian, M. (1998). A brief history of market efficiency. *European Financial Management, 4*(1), 91-103.

Dowd, K. (2000). Adjusting for risk:: An improved Sharpe ratio. *International Review of Economics & Finance, 9*(3), 209-222.

Eakins, S., & Stansell, S. R. (2003). Can value-based stock selection criteria yield superior risk-adjusted returns: an application of neural networks. *International review of financial analysis* , 83-97.

Eling, M., & Schuhmacher, F. (2007). Does the choice of performance  the evaluation of hedge funds? *Journal of Banking & Finance 3, 31*(9), 2632-2647.

Engel, F. R., & Sheppard, K. (2001). Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH.

Engel, J., & Gizycki, M. (1999). Forecast Performance| RDP 1999-04: Value at Risk: On the Stability and Forecasting of the Variance-Covariance Matrix. *Reserve Bank of Australia Research Discussion Papers*.

Engel, R. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 987-1007.

Engle, R., Ledoit, O., & Wolf, M. (2017). Large dynamic covariance matrices. *Jounal of Business and Economic Statistics*.

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business, 38*(1), 34-105.

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

Fiszeder, P., & Orzeszko, W. (2021). Covariance matrix forecasting using support vector regression. *Applied intelligence*, 7029–7042.

Flannery, M. J., Protopapadakis, A. A., & Notes, A. (2002). Macroeconomic Factors Do Influence Aggregate Stock Returns. *The Review of Financial Studies, 15*(3), 751–782.

Florou, C., & Chalevas, C. (2010). Key accounting value drivers that affect stock returns: evidence from Greece. *Managerial Finance, 36*(11).

Francq, C., & Zakoian, J.-M. (2019). *GARCH models: structure, statistical inference and financial applications.* John Wiley & Sons,.

French, K., Schwert, W., & Stambaugh, R. (1987). Expected stock returns and volatility. *Journal of Financial Economics, 19*(1), 3-29.

Friedman, J. (2001). Greedy boosting approximation: a gradient boosting machine. *The Annals of Statistics*, 1189-1232.

Friedman, J. H., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *Ann. Stat*, 337-407.

Ghahramani, Z. (2004). An Introduction to Pattern Classification. I O. Bousquet, U. v. Luxburg, & G. Rätsch, *Advanced Lectures on Machine Learning* (ss. 72-113). Canberra: Springer-Verlag Berlin Heidelberg.

Goldfarb, D., & G. Iyengar. (2003). Robust Portfolio Selection Problems. *Mathematics of Operations Research, 28*(1), 1-38.

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

Graham, B. (1949). *The Intelligent Investor.* New York : Harper & Brothers.

Green, J., Hand, J. R., & Zhang, F. (2013). The Supraview of Return Predictive Signals. *Review of Accounting Studies, Forthcoming*, 798-812.

Gurav, U., & Sidnal, N. (2018). Predict Stock Market Behavior: Role of Machine Learning Algorithms. *Springer*, 383-394.

Hanock, J. R., & Seng, D. (2012). Fundamental Analysis and the Prediction of Earnings . *International Journal of Business and Management*.

Henrique, B., Amorim, V. S., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 226-251.

Higham, N. (1988). Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, 103-118.

Hinton, G., & Sejnowski, T. (1999). *Unsupervised Learning: Foundations of Neural Computation.* Cambridge, MA, USA: The MIT Press.

Hong, K., & Wu, E. (2016). The roles of past returns and firm fundamentals in driving US stock price movements. *International Review of Financial Analysis*, *43*, 62-75.

Huang, Y., Capretz, L., & Ho, D. (2019). Neural Network Models for Stock Selection Based on Fundamental Analysis. *2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE)*, 1-4.

Hur, J., Manoj, R., & Yohanes, E. (2006). "Finance and trade: A cross-country empirical analysis on the impact of financial development and asset tangibility on international trade." . *World Development*, 1728-1741.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice (*. OTexts.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning.* New York: Springer.

Jegadeesh, N., & Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 65-91.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 255-260.

Khan, Z. H., Alin, T., & Akter, H. (2011). Price prediction of share market using artificial neural network. *International Journal of Computer Applications*, 42-47.

Kian-Ping, L., Brooks, R. D., & Hinich, M. J. (2008). Nonlinear serial dependence and the weak-form efficiency of Asian emerging stock markets. *ournal of International Financial Markets, Institutions and Money*, 527-544.

Kim, J., Pukthuanthong-Le, K., & Walker, T. (2008). Leverage and IPO under-pricing: high-tech versus low-tech IPOs. *Management Decision, 46*(1), 106-130.

Kirkpatrick, C., & Dahlquist, J. R. (2016). *Technical Analysis: The Complete Resource for Financial Market Technicians, 3rd Edition.* London: Pearson.

Kwan, & Clarence, C. (2010). The Requirement of a Positive Definite Covariance Matrix of Security Returns for Mean-Variance Portfolio Analysis: A Pedagogic Illustration. *Spreadsheets in Education (eJSiE)*.

Lam, M. (2004). Neural network techniques for financial performance prediction:integrating fundamental and technical analysis. *Decision Support Systems*, 567-581.

Lee, C. F., & Lee, J. C. (2020). *Handbook Of Financial Econometrics, Mathematics, Statistics, And Machine Learning.* World Scientific.

Lohre, H., Opfer, H., & Orszag, G. (2011). Diversifying Risk Parity. *Journal of Risk, 16*(5), 53-79.

Low, R. K., Faff, R., & Aas, K. (2016). Enhancing mean–variance portfolio selection by modeling distributional asymmetries. *Journal of Economics and Business, 85*, 49-72.

Mahmood, W. M., & Dinniah, N. M. (2007). Stock Returns and Macroeconomic Influences: Evidence from the Six Asian-Pacific Countries. *Financial Economics, and Futures Market Research Group*.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS One*.

Mariano, M.-G. (2007). Nonlinear Dynamics in Energy Futures. *The Energy Journal*, 7–29.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7*(1), 77-91.

Markowitz, H. (1999). The Early History of Portfolio Theory: 1600-1960. *Financial Analysts Journal, 55*(4), 5-16.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics, 8*(4), 323-361.

Mitchell, M. T. (1997). *Machine learning.* McGraw hill.

Mobarek, A., & Fiorante, A. (2014). The prospects of BRIC countries: Testingweak-form market efficiency. *Research in International Businessand Finance, 30*(C), 217-232.

Modigliani, F., & Miller, M. H. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *The American Economic Review, 48*(3), 261-297.

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica,* 768–783.

Muhammad, S. (2018). The Relationship Between Fundamental Analysis and Stock Returns Based on the Panel Data Analysis; Evidence from Karachi Stock exchange (KSE). *Research Journal of Finance and Accounting*.

Namdari, A., & Li, Z. (2018). Integrating Fundamental and Technical Analysis of Stock Market through Multi-layer Perceptron. *2018 IEEE Technology and Engineering Management Conference (TEMSCON)*.

NBIM. (2012). *A survey of the small-firm Effect.* Retrieved from Norges Bank Investment Management (NBIM): https://www.nbim.no/globalassets/documents/dicussion-paper/2012/discussionnote_12-12.pdf

Nelson, B. D., & Foster, P. D. (1992). Filtering and Forecasting With Misspecifified ARCH Models II: Making the Right Forecast With the Wrong Model. *Journal of Econometrics,* 61-90.

Nielsen, D. (2016). *Tree Boosting With XGBoost - Why Does XGBoost Win "Every" Machine Learning Competition?* Trondheim: NTNU.

Osman, A., Ahmed, A. N., Chow, M. F., Huang, F. Y., & El-Shafie, A. (2012). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, 1545-1556.

Ou, J. A., & Penman, S. (1989). Financial statement analysis and the prediction of stock returns. *Journal of Accounting and Economics*, 295-329.

Peiró, A. (2016). Stock prices and macroeconomic factors: Some European evidence. *International Review of Economics & Finance, 41*, 287-294.

Qian, E. E., Hua, H. R., & Sorensen, H. E. (2007). *Quantitative Equity Portfolio Management.* CRC Press.

Qvigstad, J. F. (2011). *Uroen i verdensøkonomien påvirker også Norge.* Retrieved from Norges Bank: https://www.norges-bank.no/aktuelt/nyheter-og-hendelser/Artikler-og-kronikker/kronikk-2011-08-15-qvigstad/

Rokach, L., & Maimon, O. (2014). *Data Mining with Decision Trees.* Singapore: World Scientific Publishing Co., Inc.

Rubinstein, M. (2002). Markowitz's "Portfolio Selection": A Fifty-Year Retrospective. *The Journal of Finance, 57*(3), 1041-1045.

Sadique, M. S. (2011). Testing for neglected nonlinearity in weekly foreign exchange rates. *Review of Economics & Finance* , 77-88.

Samuelson, P. A. (2015). Proof that Properly Anticipated Prices Fluctuate Randomly. I *The World Scientific Handbook of Futures Markets* (ss. 25-38). WSPC.

Scott, R. H. (2010). Bloomberg 101. *Journal of Financial Education, 36*(1).

Shapley, L. S. (1953). A Value for n-Person Games. *Contributions to the Theory of Games (AM-28), Volume II, 2*, 307-318.

Sharpe, W. (1963). A Simplified Model for Portfolio Analysis. *Management Science*, 277–293.

Sharpe, W. F. (1966). Mutual Fund Performance. *The Journal of Business, 39*(1), 119-138.

Sleire, A., Støve, B., Otneim, H., Berentsen, G. D., Tjøstheim, D., & Haugen, S. H. (2021). Portfolio allocation under asymmetric dependence in asset returns using local Gaussian correlations. *Finance Research Letters*.

Sutton, R., & Barto, A. (1998). *Reinforcement Learning: An Introduction.* The MIT Press.

Tangjitprom, N. (2012). The Review of Macroeconomic Factors and Stock Returns. *International Business Research, 5*(8), 107-115.

Trucíos, C., Zevallos, M., Hotta, L. K., & Santos, A. A. (2019). Covariance Prediction in Large Portfolio Allocation. *Econometrics*.

Tsay, R. (2005). *Analysis of Financial Time Series.* John wiley & sons.

Vijha, M., Chandolab, D., Tikkiwalb, V., & Kumarc, A. (2020). Stock Closing Price Prediction using Machine Learning Techniques. *Procedia Computer Science*, 599-606.

Wey, M. (2018). Nonlinear dynamics of US equity factor portfolios. *Chaos: An Interdisciplinary Journal of Nonlinea*.

Worthington, A., & Higgs, H. (2004). Random walks and market efficiency in European equity markets. *Global Journal of Finance and Economics, 1*(1), 59-78.

Yom-Tov, E. (2004). An Introduction to Pattern Classification. I O. Bousquet, U. v. Luxburg, & G. Rätsch, *Advanced Lectures on Machine Learning* (ss. 1-21). Canberra: Springer.

Zhang, C., & Ma, Y. (2012). *Ensemble Machine Learning: Methods and Applications.* New York: Springer .

Zhang, L., Bian, W., Qu, W., Tuo, L., & Wang, Y. (2021). Time series forecast of sales volume based on XGBoost . *Journal of Physics: Conference Series* .

Zivot, E. (2013). *Portfolio Theory with Matrix Algebra.* Retrieved from Amath 462/Econ 424, University of Washingtion: https://faculty.washington.edu/ezivot/econ424/portfolioTheoryMatrix.pdf