AALBORG UNIVERSITET

SPECIALEAFHANDLING 2023

SAMFUNDSØKONOMI

Det Samfundsvidenskabelige Fakultet

Forecasting af finansielle aktiver med henblik på at fremme den bæredygtige omstilling

Vejleder: Peter Pommergård Lind

Cecilie Kirk Fabrin, Cecilie Meier Brinkholm & Jakob Slott Kristiansen

Aalborg Universitet - Danmark

01-06-2023

BUSINESS SCHOOL
AALBORG UNIVERSITY

Titelblad

Uddannelse:	Cand.oecon
Uddannelsessted:	Aalborg Universitet
Projekt gruppe:	Gruppe 5
Gruppemedlemmer:	Cecilie Meier Brinkholm, Cecilie Kirk Fabrin & Jakob Slott Kristiansen
Vejleder:	Peter Pommergård Lind
Titel:	Forecasting af finansielle aktiver med henblik på at fremme den bæredygtige omstilling.
Antal anslag:	161.850 tegn
Antal normalsider:	68 sider

Forfattere

Cecilie Brinkholm

Cecilie Meier Brinkholm
Studie nr. 20185407

Cecilie Fabrin

Cecilie Kirk Fabrin
Studie nr. 20185515

Jakob Slott Kristiansen

Jakob Slott Kristiansen
Studie nr. 20185868

Abstract

As environmental, social, and governance, ESG, considerations gain prominence in the investment landscape, accurate forecasting models and investment strategies for ESG assets grow in significance and can incentivize private investors to revise their portfolios.

This thesis presents a comparative analysis of Long Short-Term Memory, LSTM, and Autoregressive Integrated Moving Average, ARIMA, forecasting models for ESG assets compared to the SP 500 index. By leveraging a comprehensive dataset encompassing highly rated ESG assets from the SP 500, the study explores the accuracy of predictions and effectiveness of these two forecasting techniques in the context of portfolio optimization.

The findings highlight the applicability of LSTM and ARIMA models in forecasting ESG assets and assessing their impact on mean-variance portfolio construction compared to the SP 500 and simpler investments strategies. The ARIMA model demonstrate proficiency in capturing short-term fluctuations, in contrast, LSTM models exhibit superior ability in capturing non-linear relationships and long-term dependencies, allowing for two enhanced portfolio optimizations. The study's results provide insights for investment professionals seeking to integrate ESG assets into their portfolio management strategies. By utilizing ARIMA and LSTM models, with further mean-variance portfolio optimization, results in profits exceeding those of the SP 500 benchmark. This knowledge enables and incentivizes investors to constructs portfolios that align with ESG investment objectives while not compromising competitive risk-return profiles.

Indholdsfortegnelse

1 Introduktion **1**

1.1 Problemformulering 4

1.2 Afgrænsning 5

1.3 Specialestruktur 7

2 Litteraturreview **8**

2.1 Bæredygtige investeringer og profitabilitet 8

2.2 Forecasting ved ARIMA og LSTM 9

3 Teoretisk gennemgang af aktiemarkedet **11**

3.1 Den efficiente markedshypotese 11

3.2 Random walk 12

3.3 Udviklingen på aktiemarkedet 13

4 Teoretisk gennemgang af ARIMA og LSTM **14**

4.1 Machine learning og økonometri 14

4.2 Teoretisk gennemgang af ARIMA 16

4.2.1 Autoregressive Integrated Moving Average 16

4.2.2 Box Jenkins Metodologi 19

4.2.3 Box Cox Transformation 21

4.3 Teoretisk gennemgang af LSTM 22

4.3.1 Kunstig intelligens & machine learning 22

4.3.2 Neurale netværk 22

4.3.3 Recurrent neural network 26

4.3.4 Long short term memory 28
4.4 Evalueringsmetoder - Forecast 32
4.4.1 Mean Absolute Error & Root Mean Squared Error 32
4.4.2 Diebold-Mariano test 33
4.4.3 Pesaran-Timmermann test 33

5 Teoretisk gennemgang af Mean-variance **35**
5.1 Evalueringsmetoder - Mean-variance 38

6 Databehandling **39**
6.1 Bæredygtighed og sektorfordeling 39
6.2 Deskriptiv Statistik 44

7 Udvælgelse af ARIMA og LSTM **55**
7.1 Udvælgelse af ARIMA 56
7.2 Udvælgelse af LSTM 59
7.2.1 Neuroner og layers 60
7.2.2 Epochs 62
7.2.3 Læringsrate og regulering 62
7.2.4 Batch Size 64

8 Forecast og evaluering af ARIMA og LSTM **66**

9 Porteføljestrategier **76**
9.1 Antagelser og begrænsninger 76
9.2 Resultater - mean-variance 78

10 Diskussion **83**
10.1 Anvendelse af forecast i mean-variance 83

10.2 Forecasting af bæredygtige aktier 88
10.2.1 LSTM - Netværksarkitektur, reproducerbarhed og fortolkning . . . 89
10.2.2 ARIMA - Udvælgelseskriterier og kompleksitet 91
10.3 Inkorporering af økonomisk teori 92
10.4 Investering med et bæredygtigt formål 93

11 Konklusion **96**

12 Litteraturliste **98**

13 Bilagsoversigt **108**

14 Appendiks A **110**

15 Appendiks B **120**

Figurer

4.1 Deep learning neural network (Kavlakoglu, 2020) 23
4.2 Neuralt netværk - Nodes (Baheti, 2022) 23
4.3 Gradient Descent (Gudimalla, 2021) 26
4.4 Recurrent Neural Network (Olah, 2015) 26
4.5 Cell state (Olah, 2015) 28
4.6 Forget gate (Olah, 2015) 29
4.7 Input gate - Del 1 (Olah, 2015) 29
4.8 Input gate - Del 2 (Olah, 2015) 30
4.9 Output gate (Olah, 2015) 31
5.1 Mean-standard deviation diagram (Hillier m.fl., 2011, kap. 5 s. 136) . . 36
6.1 MSCI ESG Quality Rating (Your Green Wealth, 2023) 40
6.2 Sektorfordeling blandt udvalgte bæredygtige aktier og S&P 500 41
6.3 Udviklingen i aktiekursen for S&P 500 44
6.4 Udviklingen i aktiekursen for udvalgte aktier 45
6.5 Udvikling i aktieafkast for udvalgte aktier 49
6.6 Skewness (AnalystPrep, 2020) 52
6.7 Kurtosis (AnalystPrep, 2019) 52
6.8 Korrelationsmatrice 53
7.1 Illustration af data 55
7.2 Illustration af AIC for Visa Inc. 56
7.3 Ljung Box p-værdier 58
7.4 MSE ved forskellige antal layers og units 61
7.5 Læringsrater (Rakhecha, 2010) 63
7.6 Opbygning af LSTM Visa Inc. 64

8.1 Forecast - ARIMA 66
8.2 Forecast - LSTM 67
8.3 MAPE - ARIMA & LSTM 68
8.4 Diebold-Mariano test 71
8.5 Succesrate - ARIMA & LSTM 72
8.6 Forecastede retninger 73
9.1 Kummulative procentvise afkast 78
10.1 S&P 500, Simpel strategi & Mean-variance - Buy and Hold 93
10.2 Mean-variance, ARIMA & LSTM - Buy and Sell 94

Tabeller

6.1 Specialets bæredygtige aktier 42
6.1 Specialets bæredygtige aktier 43
6.2 Tracking error, information ratio og beta værdier 46
6.2 Tracking error, information ratio og beta værdier 47
6.3 Tracking error, information ratio og beta - ESG 48
6.4 Skewness, Kurtosis og Jarque-Bera test 50
6.4 Skewness, Kurtosis og Jarque-Bera test 51
7.1 ARIMA - Modeludvægelse 57
7.2 LSTM- Modeludvægelse 65
8.1 Forecastfejl - ARIMA og LSTM 69
8.2 Gennemsnitlig MAPE - ESG-rating 70
8.3 Afvigelse i faktisk og forcastede varians 74
8.3 Afvigelse i faktisk og forcastede varians 75
9.1 Porteføljeresultater 79

9.2 Annualiserede porteføljeresultater 81
10.1 Daglig varians & afkast 83
10.1 Daglig varians & afkast 84
A.1 Test for autokorrelation - Ljung Box 110
A.2 Fejlprædikteringer for ARIMA 111
A.3 Fejlprædikteringer for LSTM 112
A.4 Test for bias 113
A.5 Diebold-Mariano test - LSTM & ARIMA 114
A.6 Diebold-Mariano test - Random Walk 115
A.6 Diebold-Mariano test - Random Walk 116
A.7 Pesaran-Timmermann test 117
A.7 Pesaran-Timmermann test 118
A.8 Portefølje gennemsnitlige vægt 119

1 Introduktion

Den bæredygtige omstilling har siden annonceringen af FN's Sustainable Development Goals i 2015 og FN's Parisaftale i 2016, vundet større indpas i den politiske verden (Business for Good, n.d.; KEFM, n.d). Politikere rundt i verden har derfor været ivrige efter at bekendtgøre deres forpligtigelse over for en bæredygtig og klimaneutral økonomi i år 2050, i overensstemmelse med Parisaftalen. Konsulentvirksomheden McKinsey fremsætter, at dette vil kræve en årlig investering på $9.2 trillioner frem til år 2050, fra regeringer, selskaber og individer. Dette anfører vigtigheden i allokeringen af kapital, som implicerer, at investeringsområdet spiller en afgørende rolle i forbindelse med den bæredygtige omstilling. Hertil fremgår omkostninger i forbindelse med den bæredygtige omstilling at påvirke hele verden, men skaber samtidig grundlag for en række vækstmuligheder i forbindelse med bedre udnyttelse af natur-kapital, samt humane og teknologiske ressourcer (McKinsey, n.d). World Economic Forum rapporterer i denne forbindelse at:

"Climate leaders can attract and retain better talent, realize higher growth, save costs, avoid regulatory risk, access cheaper capital and create new sources of value for customers. Done well, this will translate to higher shareholder returns and a sustainable source of competitive advantage (World Economic Forum, 2022)."

Dette fremsætter, at de økonomiske effekter ikke udelukkende kommer til udtryk gennem omkostninger, men ligeledes skaber gode investeringsmuligheder verden over. På trods af det store fokus på bæredygtighed og muligheder herigennem, angiver McKinsey, at det påkrævede investeringsbeløb på $9.2 trillioner, forudsætter en 60% stigning relativt til det nuværende investeringsniveau. Dette kan ikke alene opnås med offentlige midler, men skal samtidig drives af private investorer, hvorfor det fortsat er essentielt at motivere og skabe incitament til bæredygtig investering (Finans Danmark, n.d). Bæredygtige investeringer fremsættes generelt som det etiske valg, men tilsidesættes ofte grundet økonomiske aspek-

ter, da hovedformålet for investorer typisk er at opnå et konkurrencedygtigt afkast. For at opnå den påkrævede investering, er det derfor essentielt, at der ligeledes forekommer et økonomisk incitament, som tager forbehold for afkastmuligheder samt risikoniveauet. En ofte anvendt metode, hvorpå investorer kan opnå indsigt i dette er ved anvendelse af forecasting og porteføljeteori.

Forecasting af finansielle aktiver er en omdiskuteret økonomisk disciplin, som ofte relateres til *den efficiente markedshypotese* fremstillet i 1970 af Eugene Fama (Fama, 1970). Hypotesen angiver, hvorledes de finansielle markeder reflekterer alt tilgængeligt information, hvor ny viden diffunderes og inkorporeres uden forsinkelse.
Den efficiente markedshypotese relateres til *random walk hypotesen* fremstillet i "A Random Walk Down Wall Street" af Burton Malkiel i 1973. Denne angiver, at finansielle aktiver bevæger sig tilfældigt og herved uforudsigeligt (Enders, 2014, kap. 1 s. 4)[1]. Malkiel hævder, at hvis hypotesen ikke kan forkastes, vil en abe med bind for øjnene, som kaster dartpile mod en Wall Street Paper, sammensætte en portefølje, hvis afkast er i overensstemmelse med en portefølje udarbejdet af specialister.

> *A blindfolded monkey throwing darts at a newspaper's financial pages could select a portfolio that would do just as well as one carefully selected by experts (Ferri, 2021).*

Herved fremsætter den efficiente markedshypotese samt random walk hypotesen det irrelevant at forecaste finansielle aktiver. I kontrast hertil er nyere økonomisk forskning præget af en overbevisning om, at det er muligt at forecaste de finansielle markeder. Et merafkast, som førhen blev anset for en anormalitet i markedet, anses nu som delvist forudsigeligt. Anvendelsen af forecast modeller har derfor været bredt anvendt i litteraturen, hvor konkurrencen mellem økonometriske og Machine Learning modeller har været stigende.

[1]Random walk hypotesen er givet ved $y_{t+1} = y_t + \varepsilon_{t+1}$, hvor ε_{t+1} betegnes som white noise, givet ved uhæmmet information på markedet (Enders, 2014, kap. 1 s. 4).

Lineære økonometriske modeller, herunder *ARIMA*, anses som et simpelt og bredt anvendt værktøj til forecasting af finansielle aktiver[2]. Populariteten tager udgangspunkt i modellens implementerbarhed, hvortil eksisterende litteratur finder, at modellen overgår mere komplicerede forecastingsmetoder. Finansielle aktiver karakteriseres dog ofte som ikke-lineære, hvorfor det kan anses som værende fordelagtigt at anvende ikke lineære-modeller[3]. Anvendelsen af Machine Learning modeller, herunder *LSTM*, har i forbindelse med en øget mængde data og teknologiske fremskridt vundet indpas i litteraturen. LSTM anses som abstrakt og er i stand til selv at selektere information, hvilket gør metoden fordelagtig til forecasting af finansielle aktiver (Mallikarjuna og Rao, 2019).

Investeringsbeslutninger påvirkes ligeledes af andre faktorer, eksempelvis risikoniveau og transaktionsomkostninger. Porteføljesammensætning er derfor af væsentlig betydning for investorer, hvis interesse består i at maksimere afkast givet risiko. Harry Markowitzs nobelprisvindende porteføljestrategi, *mean-variance*, tager udgangspunkt i en afvejning mellem afkast og risiko, og er et bredt anvendt investeringsværktøj til optimal porteføljesammensætning. Specialet søger herunder at koble forecasting og mean-variance for bæredygtige aktier. Dette er med henblik på at undersøge, om det forekommer profitabelt for investorer at allokere mere kapital til bæredygtige aktier og hermed skabe incitament til investering i den bæredygtige omstilling.

[2]Linearitet angiver, at tidsserien estimeres som en lineær funktion. For en simpel regression givet ved $y = \beta_0 + \beta_1 x + u$, er den marginale effekt af x på y herved konstant, $\frac{dy}{dx} = \beta_1$ (Wooldridge, 2019, kap 6).

[3]Ikke-lineære modeller relateres til, at den marginale effekt af x på y ikke forekommer konstant over tid. Specialet anvender *LSTM* til ikke-lineær modellering.

1.1 Problemformulering

Ovenstående angiver, at der eksisterer omfangsrigt litteratur inden for forecasting af finansielle aktiver. Herunder fremgår der konsensus om, at der forekommer systematiske mønstre i finansielle aktiver, men ikke, hvorledes ARIMA er LSTM overlegen. Der forekommer hertil kun en begrænset mængde litteratur om anvendelsen af forecast til optimerende porteføljesammensætning, hvorfor specialet ønsker at koble forecast og mean-variance. Da ovenstående fremfører, hvorledes den bæredygtige omstilling stadig kræver en massiv stigning i investeringsniveauet, anses det relevant at tage afsæt i bæredygtige aktier. Specialet undersøger derfor, hvorvidt ARIMA relativt til LSTM modeller forekommer fordelagtige at anvende til forecasting af bæredygtige aktier samt, hvorledes anvendelsen af forecast kan opnå konkurrencedygtige afkast. Dette med henblik på at skabe større incitament til bæredygtig investering, som kan bidrage til at opnå det krævede investeringsniveau på 9.2 trillioner. Det er herved specialets hensigt at undersøge følgende problemstilling:

Formår ARIMA og LSTM modeller at forecaste bæredygtige aktier, og herunder at præstere et overnormalt afkast? Hvorledes kan indsigter herfra anvendes til at skabe incitament til investering med et bæredygtigt formål?

Ovenstående problemstilling undersøges indledningsvist ved at estimere og sammenligne en random walk, ARIMA og LSTM modeller til forecasting. Dette er grundet et ønske om at opnå de bedst mulige forecast, da en bedre forecastnøjagtighed, forventes at være en forudsætning for at opnå et overnormalt afkast. For at undersøge, hvorvidt bæredygtige investeringer præsterer en overnormal profit, baseres modellerne på bæredygtige aktier, der sammenlignes med et benchmark givet ved S&P 500. For at specialet kan vurdere om der opnås et overnormalt afkast, inddrages en mean-variance analyse, imens specialets resultater holdes op mod den efficiente markedshypotese, med et formål om at opnå både en praktisk og teoretisk vinkel.

1.2 Afgrænsning

Med et formål om at konkretisere specialet, samt udarbejde en dybdegående og fyldestgørende rapport, anses det vigtigt at undersøge og besvare specialets problemstilling, med afsæt i flere afgrænsende valg. Hertil diskuteres de afgrænsende valg løbende og heraf, hvilken betydning de har for specialets endelige konklusioner.

Det anvendte litteratur omhandler primært den økonomiske styringsramme, hvortil specialet har indsamlet data fra Yahoo Finance, som danner grundlaget for specialets primære genstandsfelt. Indledningsvist tager specialets data afsæt i det amerikanske aktiemarked, hvorved geografiske diversificeringsmuligheder begrænses. Alle specialets aktier er hermed denomineret i samme valuta, hvorfor analysen er fri for varierende valutakurser og konverteringer. Det udvalgte data tager afsæt S&P 500, som er et indeks for de 500 største, målt på markedsværdi, børsnoterede amerikanske selskaber og betragtes i flere kredse som det bedste mål for de Amerikanske finansielle markeder og den generelle økonomiske udvikling (Barnes, 2022). Ydermere inkluderes 20 amerikanske aktier, udvalgt fra iShares ESG Screened SP 500 ETF, som imødekommer en bæredygtig investeringsstrategi. Valget medfører en ekskludering baseret på indeksudbyderens ESG-eksklusionskriterier. Hermed inddrager specialet ikke selskaber, som er involveret i termisk kul, tobak, kontroversielle våben, håndvåben, militærkontrakter og olie. Herudover indeholder specialet ikke enkeltstående aktier med en ESG rating under "A", for at sikre et afkast og investeringsstrategi, som afspejler udviklingen i bæredygtige investeringer. Aktier med ESG ratingen "A"inkluderes hertil med et formål om at opnå en sektor-diversificering, som afspejler S&P 500 for at danne bedre grundlag for en sammenligning af specialets bæredygtige porteføljer og S&P 500. Specialets data er indhentet for perioden primo 2013 til ultimo 2022. Afslutningsvist tager hele specialets datasæt afsæt i lukkekurser, hvilket dog ikke forekommer afgørende for de endelige resultater.

Specialets 20 udvalgte aktier samt S&P 500 angiver den primære litteratur, hvortil den sekundære litteratur har til formål at underbygge argumentationen. Hertil fremgår den anvendte sekundære litteratur, som et samspil mellem eksisterende litteratur og den anvendte teori, herunder ARIMA, LSTM og mean-variance. I forhenværende kurser er de tre teorier introduceret, med udgangspunkt i forskellige fagområder. Dette ligger til grund for interessen i at udarbejde en dybdegående komparativ analyse, som klarlægger praktiske problemstillinger. I forbindelse med forecasting inkluderes udelukkende one-step ahead metoden. Valget herom er grundet, at metoden forekommer mere nøjagtig relativt til multi-step afhead. Dette er med et belæg om, at multi-step ahead konvergerer mod en langsigtet ligevægt og herved opnår en langsigtet trend frem for daglige udsving. Kombinationen mellem den primære og sekundære litteratur er med et formål om at opnå en kvalificeret indsigt i kausaliteten, ved forecasting af aktier og hermed frembringe en overensstemmelse mellem empiri og teori. Indsamlingen af empiri er med et udgangspunkt i at tydeliggøre tendenser og fremstille et retvisende diversificeret billede af det bæredygtige aktiemarked.

Valget om at benytte one-step-ahead har betydning for anvendelsen af mean-variance, da analysen anvender dagligt køb og salg. Valget om at inkludere 20 aktier er for at opnå en diversificeret portefølje. Dette er grundet antagelsen i mean-variance om, at når antallet af aktiver øges i porteføljen, går variansen mod nul. Dette betyder, at den ikke markedsbaserede risiko konvergerer mod nul, som hermed mindskes i en portefølje med flere aktiver. Veldiversificerede porteføljer, indeholder derfor et stort antal værdipapirer, hvilket specialet forsøger at efterleve. Det antages hertil udelukkende profitabelt at investere i tangetporteføljen såfremt, at den forventede sharpe ratio er positiv. Såfremt afkastet i minimum varians porteføljen er mindre end det risikofrie aktiv, kan denne i nogle perioder forekomme negativ. Dette efterkommes i specialet ved at gå kort i tangetporteføljen. Ovenstående afsnit danner grundlag for specialets fremadrettede udformning og konklusioner.

1.3 Specialestruktur

Introduktionen, problemformuleringen samt afgrænsningen danner grundlag for specialets videre udarbejdelse og herved rapportstrukturen. Med henblik på at besvare afhandlingens problemstilling opstilles følgende 11 hovedafsnit:

Kapitel 2 angiver et kort litteraturreview, som klarlægger eksisterende litteratur, med henblik på at kortlægge og anvende tidligere erfaringer til modellering af specialets modeller, samt anvendelse heraf til en mean-variance analyse.

Kapitel 3, 4 & 5 angiver specialets teoretiske fundament, herunder en beskrivelse af aktiemarkedet, ARIMA, LSTM og mean-variance. Specialets udvalgte evalueringsmetoder vil hertil fremgå løbende i afsnittet.

Kapitel 6 beskriver specialets data. Indledningsvist klarlægges metodologien for udvælgelsesprocessen, hvortil der fremgår en præsentation af dataet, som tager afsæt i relevant deskriptiv statistik. Dette er med et formål om at anvende indsiger og forståelse af dataet i forbindelse med specialets analyser.

Kapitel 7, 8 & 9 fremfører udvælgelsen af ARIMA og LSTM modeller, samt specialets udarbejde forecasts. Modellernes relative performance sammenlignes med udgangspunkt i specialets evalueringsmetoder. Endvidere specificeres mean-variance analysen. Herunder klarlægges en række antagelser, som specialets analyse tager afsæt i, hvortil de opnåede resultater beskrives og diskuteres.

Kapitel 10 diskuterer, hvorvidt ARIMA og LSTM modeller formår at forecaste bæredygtige aktier og herunder at præstere et overnormalt afkast. Herunder hvorledes indsigter herfra kan anvendes til at skabe incitament til investering i mere bæredygtige aktier.

Kapitel 11 opsummerer specialets konklusioner udarbejdet i de tidligere afsnit, og besvarer herved afslutningsvist specialets udarbejdede problemstilling.

2 Litteraturreview

Følgende afsnit har til formål at klarlægge samt diskutere eksisterende litteratur. Dette er med henblik på at skabe overblik og forståelse for lignende litteraturs metoder og empiriske resultater samt, hvorledes specialet kan bidrage med ny viden. Indledningsvist fremsættes eksisterende studiers resultater omhandlende relationen mellem bæredygtige investeringer og profitabilitet. Endvidere fremgår et yderligere fokus i forbindelse med ARIMA og LSTM til forecasting, samt anvendelse af forecast i en mean-variance analyse.

2.1 Bæredygtige investeringer og profitabilitet

Bæredygtige investeringer har de senere år været en drivkraft bag ændringer inden for investeringsanalyser, finansiel rådgivning og risikostyring. Dette er grundet den ekstraordinære efterspørgsel efter bæredygtige investeringer, hvor investeringsaktiviteten har været stigende i et bemærkelsesvægtigt tempo og blev mere end fordoblet i 2020 (Broadridge, 2021). Dette har bevirket til omfattende forskning og talrige analyser i forbindelse med forholdet mellem selskabets miljømæssige og økonomiske præstation. Herunder finder en række studier, at forholdet er positivt og, at bæredygtige investeringer er traditionelle investeringer overlegne (Hart og Ahuja, 1996; Dowell m.fl., 2000; King m.fl., 2002). I relation hertil fremsætter Milonas, Rompotis & Moutzouris i en komparativ analyse af bæredygtige og ikke bæredygtige investeringer, at mange bæredygtige investeringer performer bedst ved en aktiv strategi, da langsigtede forecasts ofte vil performe dårligt på bæredygtige investeringer (Milonas m.fl., 2022). I kontrast hertil er der ligeledes en række studier, som finder et ubetydeligt eller negativt forhold (Filbeck og Gorman, 2004; Telle, 2006; Ziegler og Nogareda, 2009). Herved fremgår der i nyere studier en tendens til, at der er et positivt forhold mellem bæredygtige investeringer og profitabilitet, hvor det ikke endegyldigt kan vurderes, at der forekommer konsensus på området.

2.2 Forecasting ved ARIMA og LSTM

Inden for økonomiske kredse er der de seneste år fremkommet en konsensus om, at der eksisterer systematiske mønstre i aktiekurser, som bør være mulige at forecaste. Der forekommer dog stadig diskussioner og manglende konsensus i forbindelse med, hvilke modeller, som opnår den bedste forecastperformance (Mallikarjuna og Rao, 2019).

Balsara, Chen og Zhang sammenligner herunder ARIMA relativ til en random walk, med henblik på at teste, hvorvidt ARIMA modeller er i stand til at forecaste finansielle aktiver. Studiet tager afsæt i kinesiske aktiekurser, hvor resultaterne angiver, at ARIMA opnår et mere nøjagtigt forecast. Det fremgår hertil, at de udarbejdede forecasts er fordelagtige med henblik på at generere et positivt afkast (Balsara m.fl., 2007). Endvidere forecaster Adebiyi, Adewumi og Ayo børsnoterede aktier på NYSE og NSE ved anvendelse af ARIMA modeller. Studiet finder, at modellerne forekommer fordelagtige ved kortsigtede forecast, hvortil ARIMA kan anses som en konkurrent til andre forecastmetoder (Adebiyi m.fl., 2014).

I forbindelse med machine learning modeller, sammenligner Lee og Yoo tre forskellige recurrent neurale netværk til forecasting af aktieafkast, hvor resultaterne angiver, at LSTM opnår den bedste performance (Lee og Yoo, 2018). Namini og Namin sammenligner herunder ARIMA og LSTM til forecasting af finansielt samt generelt økonomisk data. Undersøgelsen tager afsæt i seks aktiver med månedligt data fra 1985 til 2018. Studiet finder, at sofistikerede deep-learning modeller generelt opnår den bedste forecastperformance, hvor LSTM, med udgangspunkt i RMSE, var ARIMA overlegen med 84-87% (Siami-Namini og Namin, 2018). I forlængelse heraf finder Tavakoli, at LSTM i gennemsnit er 88% bedre end ARIMA. Her benyttes det samme data, hvor det undersøges ud fra daglige, ugentlige og månedlige observationer. De opnår overordnet samme konklusioner, hvor frekvensen af data ikke anses at have betydning (Siami-Namini m.fl., 2019). Der fremgår herved en

konsensus i eksisterende litteratur i forbindelse med, at LSTM modeller performer bedre relativt til ARIMA (Yamak m.fl., 2019; Ma, 2020; Xiao m.fl., 2022).
Shumway finder hertil, at begge modeller har styrker og svagheder. Herunder er LSTM mere effektiv til komplekse forecasts, der kræver forståelse af langsigtede tendenser og mønstre, mens ARIMA er mere egnet til kortsigtede forecasts af stationære tidsserier (Shumway og Stoffer, 2017). Valget af metode afhænger derfor af det specifikke problem, anvendte data og påkrævede nøjagtighedsniveau (Siami-Namini og Namin, 2018).
Det fremgår hertil i Campbell og Thompson, at forecasts på trods af en lav forklaringsgrad, stadig kan forekomme økonomisk fordelagtig for mean-variance baserede investeringsbeslutninger (Campbell og Thompson, 2008). Dette er særligt gældende i forbindelse med investeringsstrategier, som anvender løbende køb og salg, da forecasting ofte vil være i stand til bedre at fange udsvingene relativt til en gennemsnitlig historiske tendens.

Ovenstående gennemgang af eksisterende litteratur klarlægger generelle tendenser i empiriske studiers resultater. Der er herunder en tendens til, at LSTM er ARIMA overlegen, hvor det dog ikke kan vurderes, hvorledes dette er gældende for bæredygtige aktier. Dette forekommer som en konsekvens af, at studier omhandlende forecasting af bæredygtige investeringer, er af meget begrænset omfang. Endvidere er forecastingsprocessen ofte det primære fokus i litteraturen, hvortil anvendeligheden af de udarbejde forecast til porteføljesammensætning sjældent analyseres, på trods af implikationer om, at dette kan forekomme profitabelt. Specialet har herved til formål at bidrage til eksisterende litteratur i forbindelse med at undersøge, hvorvidt LSTM er ARIMA overlegen til forecasting af bæredygtige aktier. Hertil har specialet et yderligere perspektiv, hvor forecasting implementeres i en mean-variance analyse. Dette er med henblik på at undersøge, hvorvidt anvendelse af forecasts kan medvirke til at generere et bæredygtigt merafkast relativt til markedet.

3 Teoretisk gennemgang af aktiemarkedet

Nedenstående afsnit indeholder en gennemgang af aktiemarkedet som helhed, hvortil underliggende dynamikker klarlægges. Dette leder over i en forklaring af random walk, den efficiente markedshypotese og deres implikationer.

Aktiemarkedet kendetegnes ved et reguleret miljø, hvor markedsdeltagere kan handle aktier og andre godkendte værdipapirer, hvilket sikrer prisgennemsigtighed, likviditet og fair handel. Dette er eftersom alle markedsdeltagere bidrager til en fair og gennemsigtig prisfastsættelse, hvor prisen på enhver aktie bestemmes kollektivt af alle købere og sælgere (Y.-L. Chen m.fl., 2022). Aktiemarkedet er gennem tiden forsøgt at blive forklaret gennem en række fremsatte teorier, herunder den efficiente markedshypotese og random walk.

3.1 Den efficiente markedshypotese

Den efficiente markedshypotese beskriver, hvordan de finansielle markeder reflekterer alt tilgængeligt information, hvorved ny viden diffunderes og inkorporeres i kursen uden forsinkelse. Dette forhindrer investorer i at opnå en fordel relativt til markedet, hvorved det udelukkende forekommer muligt at opnå systematiske højere afkast gennem mere risikable investeringer. Dette kan udtrykkes ved:

$$E(\tilde{p}_{j,t+1}|\phi_t) = [1 + E(\tilde{r}_{j,t+1}|\phi_t)]p_{jt}$$

Herved udtrykkes den forventede pris af et aktiv j i $t + 1$, givet det tilgængelige informationssæt $E(\tilde{p}_{j,t+1}|\phi_t)$, som prisen, p_{jt} og det forventede afkast givet det tilgængelige informationssæt $E(\tilde{r}_{j,t+1}|\phi_t)$. Indfries dette har investorer ikke med udgangspunkt i informationen ϕ, mulighed for at opnå systematiske overnormale afkast, hvorfor eventuelle spekulationer, kan angives som 'fair game'. Differensen mellem afkastet og det forventede afkast er hermed nul, hvilket fremfører et efficient marked, hvor mængden af tilgænge-

lig information illustrerer formen for efficiens (Fama, 1970). Der eksisterer tre former af efficiens. Svag efficiens angiver, at al tidligere information inkorporeres i prisen, hvorfor fundamentale analyser kan medvirke til, at der i perioder produceres afkast over markedsgennemsnittet. Semi-stærk efficiens implicerer hertil, at ingen analyser vil skabe mulighed for overnormal profit, da al ny viden diffunderes i kurserne øjeblikkelig. Hvor svag og semi-stærk efficiens tager afsæt i offentlig information, angiver stærk efficiens, at al information, både offentlig og privat, inkorporeres i prisen. (Thune, 2021). Specialet tager herunder afsæt i semi-stærk efficiens i henhold til at undersøge den efficiente markedshypotese.

3.2 Random walk

Den efficiente markedshypotese er forbundet med ideen om en random walk. Logikken bag random walk er, at kursen ikke følger nogen mærkbar tendens, hvorfor ændringerne i kurserne forekommer tilfældige og hermed uforudsigelige. Tidligere information, har herved ikke indflydelse på fremtidige ændringer og kan derfor ikke anvendes til at forudsige fremtidige kurser. En random walk kan udtrykkes ved:

$$y_t = y_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim IID(0, \sigma^2)$$

y_t afhænger herved perfekt af sin forsinkede værdi y_{t-1}, samt en white noise proces ε_t. Dette betyder, at ændringerne i en random walk kun er afhængig af white noise processen, som angiver en tilfældig ændring (Enders, 2014).

En white noise proces angives herunder som en samling af ukorrelerede variable ε_t, med en middelværdi på nul og en konstant varians givet ved (Pfaff, 2008 kap. 1. s. 6):

$$E[\varepsilon_t] = 0 \qquad E[\varepsilon_t^2] = \sigma^2$$

$$E(\varepsilon_t, \varepsilon_{t-k}) = cov(\varepsilon_t, \varepsilon_{t-k}) = 0$$

Her angiver $k \geq 0$ laggede værdier af fejlleddet. En white noise proces er således givet som uafhængig og identisk fordelt, defineret ved følgende:

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

Random walk-teorien blev som tidligere nævnt fremført i, *A Random Walk Down Wall Street* af Malkiel i 1973, som argumenterer for, at det ikke forekommer muligt at slå markedet. Frem for at anvende algoritmer og teknisk analyse med et formål om at forudsige aktiekurser, hævder Malkiel herved, at investorer er bedre stillet i besiddelse af en diversificeret portefølje (Smith, 2023).

3.3 Udviklingen på aktiemarkedet

Med tiden har investorer på aktiemarkedet oplevet, at mængden af information har været stigende, hvilket skaber mulighed for andre investeringsstrategier. I forbindelse med denne udvikling er spørgsmålet om, hvorvidt der forekommer gennemskuelige mønstre i aktiekurserne kun blevet større, hvortil tiltroen til den efficiente markedshypotese og random walk hypotesen er reduceret. Dette bidrager til, at flere markedsdeltagere øger kompleksiteten af deres strategier, hvortil de søger at drage fordele af den øgede mængde tilgængelig data. Kompleksiteten øges ikke udelukkende for en investor i forhold til viden og erfaring, men også for de eksisterende algoritmer. Disse anvendes til at guide og informere med et formål om at udnytte mønstrene på markedet og hermed skabe en overnormal profit.

4 Teoretisk gennemgang af ARIMA og LSTM

Følgende afsnit angiver en teoretisk gennemgang af ARIMA modeller, herunder økonometriske tests og deres fremgangsmåde. Endvidere angives LSTM modeller samt en beskrivelse af machine learning, deep learning og måden, hvorpå LSTM opsættes og optimeres. Afslutningsvist klarlægges relevante evalueringsmetoder, som anvendes med henblik på evaluering af specialets opnåede forecasts.

4.1 Machine learning og økonometri

Machine learning og økonometri angiver to centrale områder inden for økonomiske kredse og kan defineres ved:

"Econometrics is the application of statistical methods to economic data in order to give empirical content to economic relationships (Wrg, 2019)."

"Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead (Wrg, 2019)."

Begge modeller anvendes herved med henblik på at beskrive, analysere og forklare data med afsæt i statistiske modeller og relevante variable. Hertil beskæftiger økonometri sig udelukkende med økonomiske aspekter, mens machine learning anvendes til et bredere sæt af problemstillinger. I specialet anvendes modellerne til forecasting, hvor de to modeller har et fælles mål om at opnå den bedste forecast model.

En essentiel forskel i machine learning og økonometri er i forbindelse med det teoretiske fundament. Økonometri har et solidt fundament i matematik, statistik og sandsynlighedsteori. Økonometriske modeller har herved påviselige og attraktive egenskaber, og vurderes

primært ud fra deres robusthed. Hertil anvendes AIC eller andre evalueringskriterier til at evaluere hver model relativt til andre modeller med afsæt i træningsdata.
Machine learning inddrager ligeledes matematiske egenskaber, men primært til at forklare modellens adfærd, frem for at klarlægge pålidelighed og egenskaber. Modellen vurderes primært ud fra dens empiriske effektivitet, hvilket tager afsæt i en antagelse om, at hvis modellen er i stand til at opnå et godt forecast, vil den succesfuldt have lært de skjulte mønstre i dataet. For at vurdere modellerne og undgå overfitting, opdeles data i træningsdata og testdata. Modellens forecastperformance vurderes heraf med afsæt i MAE eller andre mål for forskelle i faktiske og forecastede værdier (Wrg, 2019).

Økonometriske modeller har endvidere kun én løsning. Den bedste lineære unbiased estimater, BLUE, for koefficienterne findes ved brug af den ordinære mindste kvadrerede, OLS, estimater. Den bedste model er herved givet ved modellen med den laveste varians af estimatet. Machine learning modeller bestemmes herimod ved en iterativt metode under træningen. Herunder inddrages en vis mængde tilfældighed, som medfører, at to identisk opbyggede modeller, kan opnå forskellige løsninger for den bedste model. Herved er løsningen for en machine learning model et estimat, der kun sandsynligvis er optimalt.
Ydermere er økonometriske modeller ofte parametriske og herved underlagt en række antagelser. Machine learning modeller er i kontrast hertil ikke-parametriske, og baseres i stedet på data uden underliggende antagelser om fordelingen (Wrg, 2019).

Ovenstående klarlægger en række ligheder og forskelle mellem modeller inden for machine learning og økonometri. Specialet ønsker herunder at undersøge, hvorvidt det er muligt at optimere profit ved forecasting og porteføljesammensætning, hvorfor det anses relevant at undersøge de to modeltyper relativt til hinanden. Nedenstående afsnit vil herved uddybe specialets økonometriske model, ARIMA, og machine learning model, LSTM.

4.2 Teoretisk gennemgang af ARIMA

Forecasting har til formål at prædiktere fremtidige værdier, ved at afdække systematiske variationsmønstre i data over tid. I specialet anvendes tidsserie regression, med henblik på at specificere og beskrive statistiske sammenhænge i data. En tidsserie kan udtrykkes ved:

$$y_t = seasonal_t + trend_t + cyclical_t + \varepsilon_t$$

Tidsseriedata har herved en intern struktur, bestående af trend, en cyklisk komponent, sæsonvariation og en uregelmæssigt komponent. Trendkomponenten beskriver dataets langsigtede forhold, hvortil den cykliske komponent angiver regelmæssige periodiske bevægelser. Ydermere angiver sæson komponenten hyppigere og kortvarige regelmæssige bevægelser, imens den uregelmæssige stokastiske komponent afslutningsvist antages at være en white noise proces ε_t. Med henblik på at estimere tidsseriedata, anvender specialet lineære autoregressive integrated moving average modeller, der beskrives i nedenstående afsnit.

4.2.1 Autoregressive Integrated Moving Average

Autoregressive Moving Average, ARMA, anvendes til at analysere og forecaste data. Tidsserier kan forecastes med afsæt i egne tidligere værdier, givet ved autoregressive processer, AR, samt fejlledets tidligere værdier, givet ved moving average processer, MA.

Autoregressive processer, AR(p), angiver den afhængige variabel, som en lineær funktion af sine egne forsinkede værdier og en white noise proces. En AR proces, af ordenen p, kan udtrykkes ved:

$$y_t = \mu + \theta_1 y_{t-1} + \theta_2 y_{t-2} +, .., + \theta_p y_{t-p} + \varepsilon_t$$

$$\varepsilon_t \sim IID(0, \sigma^2) \qquad \theta_p \neq 0; \sigma_\varepsilon^2 \geq 0$$

θ_k angiver korrelationskoefficienten mellem den afhængige variabel y_t og AR processer-

ne y_{t-p}, hvorfor parameteren anses afgørende for, hvorvidt AR processen er stationær. En grundforudsætning for benyttelsesværdige estimater er stationært data, da anvendelse af ikke-stationært data medfører spuriøse resultater og sammenhænge mellem variable (Enders, 2014 Kap. 4 s. 184-185). Stationaritet uddybes derfor yderligere senere i dette afsnit. I tilfælde af, at $|\theta| = 1$, indeholder tidsserien unit root og forekommer herved ikke stationær. Hertil indikerer $|\theta| < 1$ en stationær tidsserie, som konvergerer mod den langsigtede middelværdi $\mu/(1-\theta)$ når $n \to \infty$ (Pfaff, 2008). *Unit Root* betegner hertil en stokastisk tendens af en AR proces, som kan indgå i en ARMA model. Forekommer der unit root, vil tidsserien være en ikke-stationær proces, hvorfor test for unit root er et essentielt værktøj.

Moving Average processer, MA(q), angiver den afhængige variabel, som en lineær funktion af fejlledets forsinkede værdier og en white noise proces. En MA proces, af ordenen q, kan udtrykkes ved:

$$y_t = \mu + \varepsilon_t + \alpha_1\varepsilon_{t-1} + \alpha_2\varepsilon_{t-2} +, .., +\alpha_q\varepsilon_{t-q}$$

$$\varepsilon_t \sim IID(0, \sigma^2) \qquad \alpha_q \neq 0; \sigma_\varepsilon^2 \geq 0$$

α_q betegner korrelationskoefficienten mellem y_t og MA processerne, ε_{t-q}. ε_{t-q} antages at være identisk og uafhængigt fordelt, hvorfor MA processen altid er stationær (Pfaff, 2008).

Autoregressive Moving Average modeller, ARMA(p, q), angiver en fusion af AR(p) og MA(q) processer, og kan udtrykkes ved:

$$y_t = \mu + \theta_1 y_{t-1} +, .., +\theta_p y_{t-p} + \varepsilon_t + \alpha_1\varepsilon_{t-1} +, .., +\alpha_q\varepsilon_{t-q}$$

$$\varepsilon_t \sim IID(0, \sigma^2) \qquad \alpha_q \neq 0; \theta_p \neq 0; \sigma_\varepsilon^2 \geq 0$$

Ovenstående defineres som en ARMA model, når tidsserien er stationær i I(0). I angiver

ordenen af integration, som i dette tilfælde er nul, hvilket betyder, at tidsserien er en stationær proces, uden brug af differentiering.

Stationaritet defineres som en fælles sandsynlighedsfordeling for en samling af stokastiske variable, der ikke ændres over tid. Givet, at en tidsserie er en stationær proces, forekommer det muligt at estimere middelværdien, variansen og autokorrelationen med afsæt i et langsigtet gennemsnit[4]. En proces betegnes svagt stationær, såfremt nedenstående antagelser opfyldes (Enders, 2014 kap. 2 s. 51-53):

$$E(y_t) = E(y_{t-1}) = \mu$$

$$E[(y_t - \mu)^2] = E[(y_{t-k} - \mu)^2] = \sigma_y^2$$

$$E[(y_t - \mu)(y_{t-k} - \mu)] = cov(y_{t-j} - y_{t-j-k}) = \gamma_k$$

Ovenstående angiver, at gennemsnittet og variansen er konstant over tid. Kovariansen mellem hvilke som helst to tidspunkter t og t-k afhænger hertil kun af k, for alle $k > 1$. Ikke stationære tidsserier indeholder ingen konsekvente mønstre over tid, hvorved det ikke er muligt at generalisere eller forecaste. Stationaritet anses derfor i specialet som altafgørende for de endelige resultater (Newbold, 1975). Hovedparten af tidsserier er ikke-stationære processer, som ofte kan gøres stationære ved differentiering. Dette resulterer i en integreret ARMA model, angivet som Autoregressive Integrated Moving Average, ARIMA(p, d, q), hvor d angiver graden af integrering (Prabhakaran, 2021).

[4]Autokorrelation angiver enhver korrelation mellem ε_t og ε_{t-k} for alle $k \geq 0$, og er herved en regelmæssig bevægelse, der ikke tages forbehold for i koefficienternes værdier (Pfaff, 2008, kap. 3 s. 61).

4.2.2 Box Jenkins Metodologi

Box Jenkins metodologien er en metode, som anvendes til at udvælge en passende model til at estimere samt forecaste en univariate tidsserie[5]. Metoden indeholder tre faser, givet ved identifikation, estimering og diagnostik.

Identifkationsfasen har til formål at opnå en stationær tidsserie og ligeledes identificere parametre i ARIMA modellen. Herunder fremføres et plot af det valgte data samt autokorrelationsfunktionen, ACF, og den partielle autokorrelation, PACF. Plottet af tidsserien klarlægger brugbare oplysninger, om mulige afvigelser, manglende værdier samt, hvorvidt dataet fremstår stationært. ACF og PACF benyttes hertil med henblik på at estimere antallet af parametre p og q til forecasting (Newbold, 1975). Denne fase er hermed elementær for at opnå brugbare estimater. Dette er gældende da opnåelsen af stationært data er en nødvendighed for, at data kan anvendes til generalisering og forecasting. I tilfælde af ikke-stationært data vil modellen ikke kunne afspejle de underliggende mønstre i dataet, som resulterer i fejlestimater. En ikke-stationær tidsserie kan eksempelvis indeholde mønstre, der i virkeligheden er korrelationer grundet trend eller en sæsonbestemt komponent i dataet.

Estimeringsfasen har til formål at udvælge den stationære og parsimoinious model, der bedst beskriver det udvalgte data[6]. Specialet optimerer modellernes parametre ved brug af maximum likelihood estimering. Der estimeres herunder en række alternative modeller, som sammenlignes baseret på deres sandsynlighedsværdier. Specialet tager udgangspunkt i Akaike information kriterie, AIC, der fremgår som en af de mest gennemgående udvælgelseskriterier i litteraturen. AIC kan udtrykkes ved (Difference Between, n.d):

[5]Univariate angiver, at den forklarende variabel udelukkende er en funktion af sine egne tidligere værdier, samt fejlled (Enders, 2014 kap. 2 s. 51-53).

[6]En parsimonious model er givet ved den model, som forklarer mest muligt ved mindst mulig input.

$$AIC = 2K - 2ln(L)$$

K og L angiver antallet af uafhængige variable og et log-likelihood estimat for den specifikke model. Inddragelsen af flere paramentre, K, medvirker generelt til, at forklaringsgraden øges og, at AIC mindskes. Andet led angiver en strafparameter, som sikrer at inkluderingen af parametre, der ikke reelt øger modellens forklaringsgrad, vil medføre en øget AIC. Modellen med den laveste AIC angiver herunder den parsimonious model.

Diagnosticeringsfasen er det sidste trin i Box Jenkins metodologien, og omfatter en vurdering af, om den estimerede ARIMA model beskriver dataet. Dette trin har, ved test for autokorrelation, normalitet og unit root, til formål at sikre, at de residuale i modellen optræder som en white noise proces. Indeholder en model autokorrelation vil OLS have en tendens til at underestimere standardafvigelsen af parametrene og herved øges risikoen for, at insignifikante parametre vurderes signifikante (Pfaff, 2008). Dette angiver, at Box Jenkins er en iterativ proces, da nye informationer opnået ved diagnostik, kan indebære en tilbagevenden til identifikationsfasen, hvor informationen herfra kan anvendes til at estimere en mere passende model (Newbold, 1975). Diagnostikfasen omfatter følgende trin:

Argumented Dicky-Fuller, ADF, er en bredt anvendt test for unit root. Udover at teste for stationaritet tager ADF højde for autokorrelation, ved at inkludere differentierede autoregressive processer. ADF er givet ved:

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \sum_{i=1}^{k} \delta_i \Delta y_{t-i} + \varepsilon_t$$

Δy_t beskriver ændringen i y_t, $\Delta y_t = y_t - y_{t-1}$, hvortil α og β_t angiver en konstant og en trend komponent. Endvidere viser δ i hvor høj grad y_t afhænger af ændringerne i sine tidligere perioder. γ er givet ved $\theta - 1$ og er parameteren af primær interesse. Nullhypotesen opstilles som $\gamma = 0$, hvilket betyder, at en eller flere af tidsseriens rødder tager den

absolutte værdi én og herved, at modellen indeholder unit root (Box m.fl., 2015).
Ljung Box anvendes efterfølgende som en test af goodness of fit. Dette har til formål at vurdere, hvor godt den estimerede model beskriver dataet. Ljung Box tester for autokorrelation i de residuale, hvor nullhypotesen angiver, at der ikke forekommer autokorrelation. Testen kan herved identificere eventuelle mangler i modellen, som er nødvendige at tage højde for i forbindelse med brugbare resultater og forecasting (Pfaff, 2008, kap. 3 s. 61).

I forbindelse med de to trin, kan det forekomme nødvendigt at optimere den estimerede model. Dette er grundet, at det ved estimering af ARIMA modeller er essentielt, at de residuale følger en white noise process. I tilfælde af, at antagelsen brydes, kan det vurderes fordelagtigt med design af en ny model ved transformation af data (Sakia, 1992). Diagnosticeringsfasen er afgørende for at sikre, at modellen beskriver dataet korrekt.

4.2.3 Box Cox Transformation

Box Cox transformationen er en metode, som anvendes til at omdanne data til at følge en normalfordeling. Metode anvendes ofte til at forbedre en model ved at sikre, homoskedasticitet, additivitet og normalitet. Box Cox transformationen tager afsæt i parameteren λ, der bestemmer transformationsgraden (Galili, 2016), som for en tidsserie, y_t, udtrykkes ved (Otexts, n.d):

$$y_t = \begin{cases} log(y_t) & \text{hvis } \lambda = 0; \\ (y_t^\lambda - 1)/\lambda & \text{hvis ikke} \end{cases}$$

λ kan tage enhver positiv eller negativ værdi. Det er gældende, at $\lambda = 0$ tilsvarer en naturlig log transformation, hvortil $\lambda = 1$ ikke implicerer nogen væsentlig transformation af dataet. Afslutningsvist tilsvarer $\lambda = -1$ en invers transformation. Til estimering af λ anvender specialet Guerrero metoden, som minimerer koefficienterne af variansen for underserier af den aktuelle tidsserie (Sakia, 1992)[7].

[7]Der eksisterer to metoder til at udvælge λ, navngivet "Guerrero" og "Loglik". En undersøgelse fra In-

4.3 Teoretisk gennemgang af LSTM

I forbindelse med forecasting, tager specialet yderligere afsæt i long short term memory, LSTM, der er en underkategori af kunstig intelligens, AI. Følgende afsnit klarlægger derfor indledningsvist machine learning og kunstig intelligens. Dette leder over i en gennemgang af neurale netværk, aktiveringsfunktioner og læringsalgoritmer. Endvidere klarlægges recurrent neural network, hvortil LSTM afslutningsvist uddybes.

4.3.1 Kunstig intelligens & machine learning

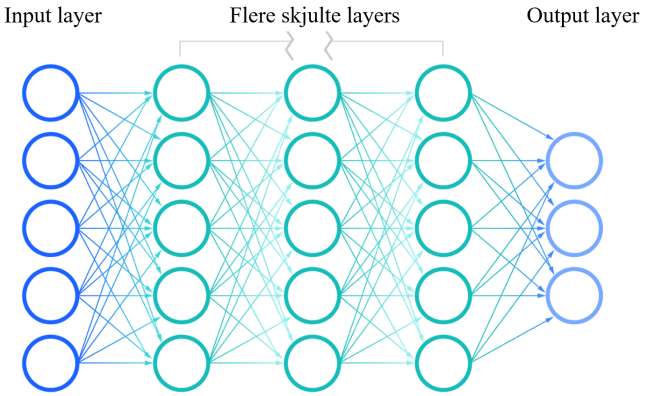
Kunstig intelligens, AI, sigter mod en simulering af menneskelig intelligens i maskiner. AI anvendes med et formål om at gengive komplekse opgaver, der associeres med det menneskelige sind, såsom læring og rationalisering (Su og Yang, 2022). En vigtig drivkraft i udviklingen af AI er machine learning, ML. ML anvender statistiske teknikker til at identificere mønstre i data, hvilket forbedres gennem erfaring (Ayodele, 2010). Specialet tager afsæt i supervised machine learning, SML, hvor algoritmen udarbejder en funktion, der kortlægger input til det ønskede output. Denne metode er hyppigst anvendt til træning af neurale netværk, med en forudsætning om at klassificere eller udarbejde regressionsmodeller til forecasting af numeriske værdier (Delua, 2021).

4.3.2 Neurale netværk

Specialet tager afsæt i Deep Learning. DL er en statistisk teknik, som klassificerer mønstre i data ved anvendelse af neurale netværk med flere skjulte layers. Input, output og skjulte layers i DL består af neuroner, der visualiseres som cirkler i figur 4.1.

ternational Institute of Forecasters fremfører, at Guerrero relativt til Loglik, opnår en værdi af λ, som skaber grundlag for et bedre forecast (Thombre, 2018).

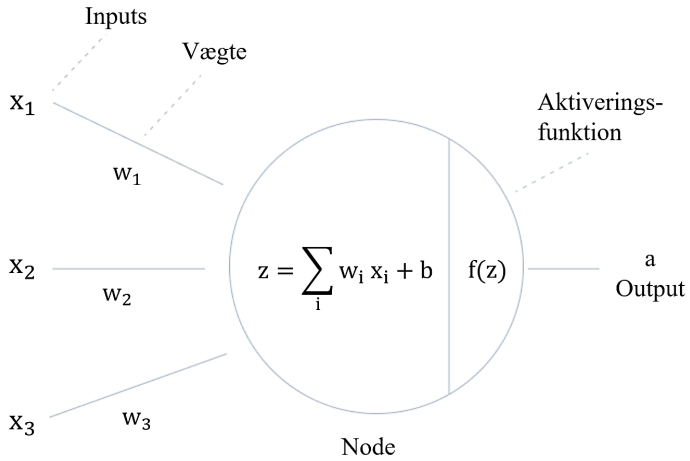
Figur: 4.1 Deep learning neural network (Kavlakoglu, 2020)



Figuren illustrerer et eksempel på et deep neural network med tre skjulte layers og fem neuroner i hvert layer.

Hver neuron har tilhørende vægte og input imens transformeringerne foregår i modellens skjulte layers. Det er modellens skjulte layers, der defineres som en black box, da det ikke er muligt at kortlægge processen og dynamikkerne fra input til output (IBM, 2023a)[8]. De skjulte elementer er en væsentlig kontrast til den økonometriske ARIMA model, hvor det forekommer muligt at afdække og tolke på alle parametrene (Molnar, 2023). NN genererer mange forskellige vægte og kombinationer mellem layers, som illustreret i figur 4.2.

Figur: 4.2 Neuralt netværk - Nodes (Baheti, 2022)



Figuren illustrerer et eksempel på funktionen i en neuron. Neuronen modtager input fra tidligere layers, som summeres og tillægges et biasudtryk. Dette føres gennem en aktiveringsfunktionen, som resulterer i et output.

[8]Dette er i forbindelse med, at det er umuligt at tolke på egenskaberne af modellen og de sammenhænge, som den måtte finde.

Ligningen i figur 4.2 angiver processen, hvorved ethvert input bearbejdes i modellen. Efter et input layer er etableret, tildeles vægte, som angiver vigtigheden heraf. Hvert input multipliceres med deres respektive vægte og summeres, før der tilføjes et bias udtryk. En højre vægt er ensbetydende med et større bidrag til outputtet. NN tilpasses og forbedres herved alt andet lige ved justering af modellens vægtede input (IBM, 2023a).

Aktiveringsfunktioner

Et essentielt komponent i NN er aktiveringsfunktioner, som spiller en integreret rolle, ved at indføre ikke-linearitet til modellen. Dette gør det muligt for NN at udvikle komplekse repræsentationer og funktioner baseret på input, som ikke ville være muligt med en lineær regressionsmodel (Baheti, 2022). Uden en aktiveringsfunktion ville data kun passere gennem netværkets nodes og layers gennem lineære funktioner $a \cdot x + b$. Sammensætningen af disse lineære funktioner er igen en lineær funktion, så uanset hvor mange layers dataet passerer, vil output altid være et resultatet af en lineær funktion. Derfor er aktiveringsfunktioner af afgørende betydning til at inddrage ikke-linearitet. Hertil indeholder hver neuron i NN en aktiveringsfunktion, hvis rolle er at transformere det summerede vægtede input til en outputværdi. Specialet benytter den ikke-lineære aktiveringsfunktion Rectified linear unit, ReLU, som er den mest anvendte til NN og udtrykkes ved (Brodtman, 2021):

$$g(z) = max(0, z) \qquad g(z) = \begin{cases} 0 & z < 0 \\ z & z \geq 0 \end{cases}$$

ReLU ligger i intervallet $[0, \infty]$, hvilket implicerer, at neuronerne kun aktiveres såfremt outputtet er større end 0. Hovedpointen er, at ReLU kun aktiverer et mindre antal neuroner, og er derfor mere beregningsmæssigt effektiv relativt til andre aktiveringsfunktioner [9]. ReLU er hertil en afledt funktion, hvilket muliggør backpropagation.

[9]ReLU har dog også en ulempe, såfremt gradientværdien er nul. Vægtene og bias for nogle neuroner vil her ikke opdateres under backpropagation, hvilket reducerer modellens evne til at tilpasse eller træne fra data (Baheti, 2022).

Backpropagation & Stokastisk gradient descent

Backpropagation er en metode til at beregne gradienten af tabsfunktionen, der efterfølgende anvendes i stokastisk gradient descent, som angiver læringsalgoritmen for neurale netværk. Formålet er at opnå indsigt i, hvordan ændringer af vægte og bias påvirker netværkets overordnede adfærd. Specialet anvender neurale netværk med henblik på at opnå en model, der er i stand til at approksimere testdata med afsæt i det valgte træningsdata. Træningsalgoritmen har til formål at minimere tabsfunktionen $C(w, b)$, givet ved en funktion af vægte, w, samt bias, b. Hertil anvendes læringsalgoritmen for DL, stochastic gradient descent, som estimerer gradienten ∇C, ved at beregne ∇C_X for en række prøver af stokastiske træningsinput, hvorefter gennemsnittet estimeres. Dette resulterer i et estimat tilnærmelsesvist lig den sande gradient, hvor gradient descent hastigheden øges og læringen af modellen accelereres (Nielsen, 2015).

Backpropagation benyttes for hvert træningssæt, til at estimere vektoren af tab i output layer, hvorefter vektoren af tab i hvert layer beregnes. Med udgangspunkt heri beregnes gradientværdierne, som sammensættes til gradienten af tabsfunktion. Denne angiver vektoren af de partielle afledte i forhold til enhver vægt samt bias i netværket og er givet ved:

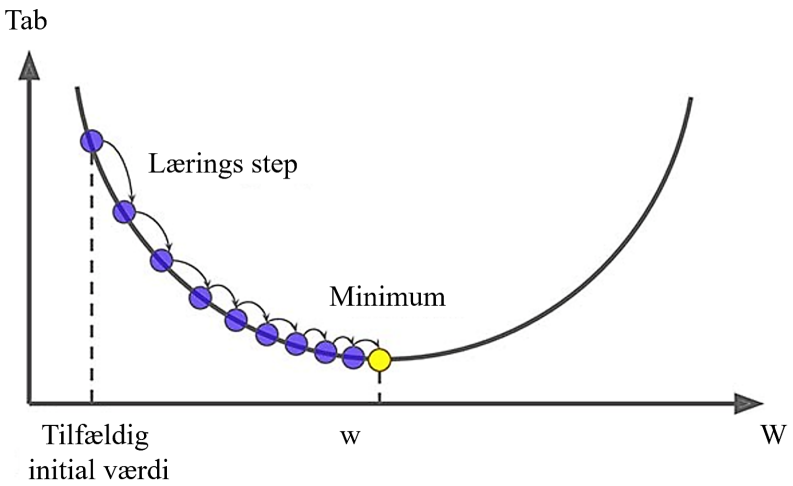
$$\nabla C = (\frac{\partial C}{\partial w}, \frac{\partial C}{\partial b})^T$$

Gradient descent anvendes med henblik på at minimere tabsfunktionen. Algoritmen opdaterer for hvert layer vægte og bias i henhold til følgende:

$$\Delta C \approx \nabla C \cdot (\Delta w, \Delta b)^T$$

Processen illustreres i figur 4.3, hvor tabsfunktionens initiale værdi, fremstår visuelt som det øverste punkt på kurven. Over flere iterationer tilnærmes det lokale minimum.

Figur: 4.3 Gradient Descent (Gudimalla, 2021)



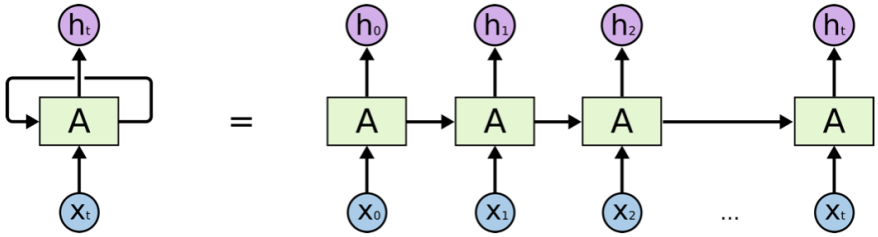
Figuren illustrerer læringsprocessen i et neuralt netværk. Det øverste punkt på kurven angiver initialværdien af tabsfunktionen, som over flere iterationer tilnærmes et minimum givet ved de nedadgående pile og lilla. Det lokale minimum er markeret med gul og angiver, at algoritmen har fundet de optimerede vægte og bias.

Processen afmærkes med nedadgående pile og lilla, hvor det lokale minimum markeres med gul. Optimeringsalgoritmen har således fundet vægte og bias, som medfører minimumsværdien for tabsfunktionen (Gudimalla, 2021).

4.3.3 Recurrent neural network

Recurrent neural network, RNN, anvendes på sekventiel data og kendetegnes ved at benytte information fra tidligere sekvenser for at påvirke det aktuelle input samt output. Opbygningen bag RNN illustreres i figur 4.4.

Figur: 4.4 Recurrent Neural Network (Olah, 2015)



Figuren illustrerer et udrullet recurrent neural network. Hvert input x behandles i et skjult layer A, som hernæst resulterer i et output h. Informationen loopes herunder i hvert layer gennem hele sekvensen.

Som illustreret ovenfor behandles hvert input x, i et skjulte layer A, som hernæst resulterer i et output h. Evnen til at inkorporere tidligere information, fremføres i loopet i modellens skjulte layer, hvori konstruktionen og aktiveringsfunktionen selekterer information, som bidrager til modellens forecasts. Loopet illustreres til højre i figur 4.4, som fremhæver, at informationen opdateres mellem skjulte layers gennem hele sekvensen. Hermed er tidligere sekvenser af relevant betydning for den fremtidige udvikling (Olah, 2015). En simpel RNN som illustreret i ovenstående figur, kan angives matematisk ved:

$$h_{t+1} = f(w_x x_t + w_h h_t + b_h)$$

Hvor output y på tidspunktet t beregnes ved:

$$y_t = f(w_y h_t + b_y)$$

$\cdot$ er et dot produkt imens x_t og y_t angiver input og output for netværket til tiden t. Endvidere er h_t en vektor, som gemmer værdierne af units i de skjulte layers til tiden t. Herudover er w_x og w_h vægte forbundet med input og skjulte units, hvor w_y er vægte forbundet med units i de skjulte layers til output. Afslutningsvist er b_h og b_y bias forbundet med hidden og output layers (Saeed, 2022).

Da RNN klassificerer sekventiel data benyttes en udvidelse af back propagation, defineret back propagation through time, BPTT, til estimering af gradienterne. Principperne bag BPTT er identiske med traditionel back propagation, men adskiller sig ved summering af fejl for hvert t. Tiden udtrykkes, som en veldefineret ordnet række af beregninger, som forbinder det ene tidstrin til det næste (IBM, 2023b). Matematisk beskrives BPTT.

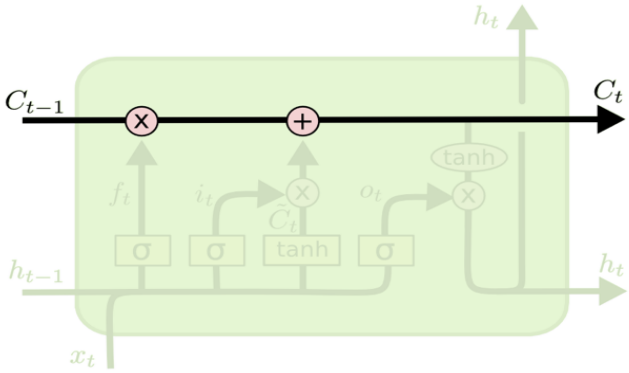
$$h_t = W_{hx} x_t + W_{hh} h_{t-1}$$

Et appellerende aspekt ved RNN er, at modellen formår at inkorporere information over tid. Hvis tidsintervallet bliver for stort, vil modellen dog ikke være i stand til at lære sammenhænge i dataet. Dette betegnes som langsigtede afhængighedsproblemer (Olah, 2015).

4.3.4 Long short term memory

Specialet anvender en underkategori af RNN, long short term memory, LSTM. Grundessensen ved LSTM er givet ved cell state, der medfører, at modellen kan lære af information på kort og lang sigt. Hermed er modellen designet til at undgå langsigtede afhængighedsproblemer. Cell state, C_t, som illustreres i figur 4.5, kan forklares som et transportbånd, der løber gennem hele kæden med nogle mindre lineære interaktioner. Dette er årsagen til, at information kan passere gennem C_t uden større ændringer (Olah, 2015).

Figur 4.5 Cell state (Olah, 2015)

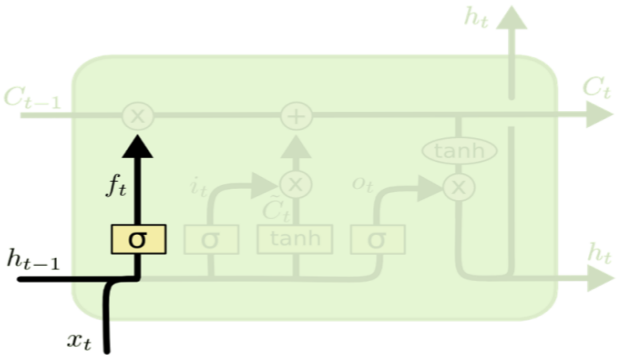


Figuren illustrerer cell state i LSTM, der kan forklares som et transportbånd, som løber gennem hele kæden med nogle mindre lineære interaktioner.

Figur 4.5 illustrerer, at LSTM kan fjerne eller tilføje information til C_t, der er reguleret af strukturer, defineret gates. Størstedelen af gates, sammensættes af et sigmoid neuralt net layer, σ, samt en punktvis multiplikationsoperation, angivet $\times$. Et sigmoid layer tager en værdi i intervallet [0;1] og fremfører, hvor meget information, der passerer gaten. Værdien nul implicerer, at intet information passerer gaten og vice versa. LSTM indeholder tre gates, der selekterer informationen, som videreføres til C_t.

Første step i modellen er selekteringspocessen i C_t. Denne proces sker i et sigmond layer, givet ved forget gate layer, som illustreret i figur 4.6 (Olah, 2015).

Figur: 4.6 Forget gate (Olah, 2015)



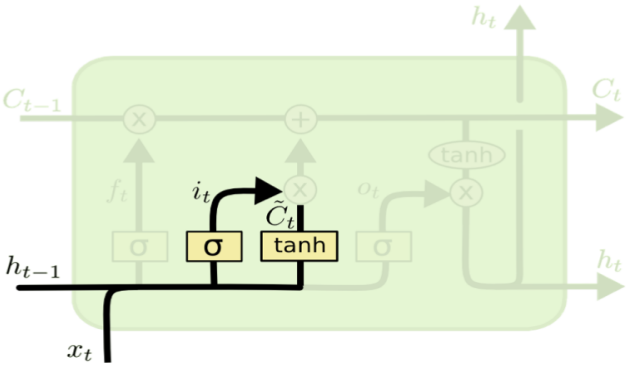
Figuren illustrerer forget gate i LSTM, som medvirker til at selektere informationen i C_t .

Modellens skjulte layers modtager information fra det tidligere layer, C_{t-1}, tidligere output, h_{t-1} og nye input, x_t. Informationen passerer modellens gates, som bearbejder informationen forskelligt. Informationen, som tilføjes C_t fra forget gaten, betegnes f_t og er givet ved:

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

W_f er en vægtmatrice, imens b_f er en bias vektor[10]. Begge estimater fremstår som et yderligere sæt af vægte (DeepAI, n.d). Hernæst multipliceres det bearbejdede information med σ, som fremfører andelen af information som tilføjes C_t. Næste step er todelt og afgørende for, hvilken ny information, der gemmes i C_t. Første step illustreres i figur 4.7.

Figur: 4.7 Input gate - Del 1 (Olah, 2015)



Figuren illustrerer første del af input gate i LSTM, som er afgørende for, hvilken ny information, der gemmes i C_t. Første del er givet ved input gate layer, i_t, der fastlægger, hvilke værdier, som opdateres.

[10]Vægtmatricen og bias vektoren tager afsæt i tilfældige værdier og opdateres løbende med ny information.

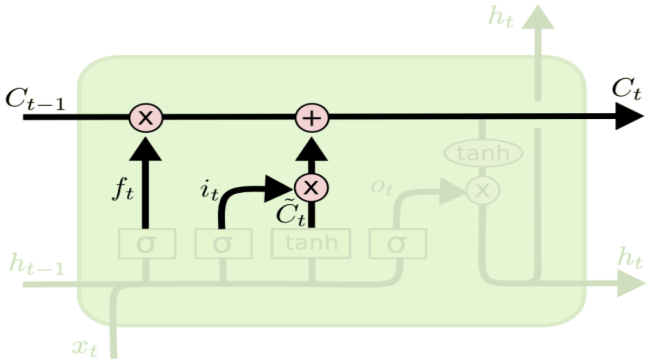
Første del er givet ved input gate layer, i_t, der fastlægger, hvilke værdier, som opdateres. Anden del er et tahn layer, som estimerer en vektor bestående af nye kandidatværdier, $\tilde{C}_t$. Modellens input gate layer samt tahn layer er givet ved:

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

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

Konstruktionen af formlerne fremgår identisk med forget gate, hvor de respektive vægtmatricer, W_i og W_c multipliceres. Hernæst adderes de respektive bias vektorer b_i og b_c for hver af disse processer. Graden af information som videreføres, estimeres ved multiplicering med σ for input gate layer og tahn funktionen i tahn layer. Afslutningsvist integreres input gate layer og tahn layer, som illustreret i figur 4.8, hvorved C_t opdateres.

Figur: 4.8 Input gate - Del 2 (Olah, 2015)



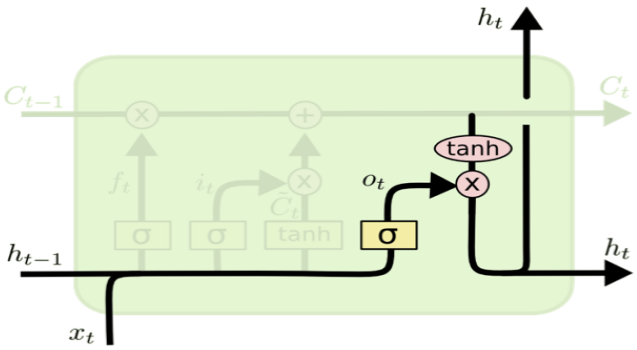
Figuren illustrerer anden del af input gate. Input gate- og tahn layer integreres, hvorved C_t opdateres.

Ovenstående illustrerer, hvordan LSTM opdaterer C_{t-1} til C_t. Dette fremføres med afsæt i nedenstående ligning, hvor C_{t-1} multipliceres med f_t og adderes med $i_t\tilde{C}_t$. Dette angiver modellens nye kandidatværdier skaleret efter, hvor meget LSTM opdaterer hver enkelt værdi (Olah, 2015):

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

Den sidste gate har til formål at klargøre, hvad modellens output skal være. Outputtet baseres på modellens cell state, men er en filtreret version, hvorved der udelukkende skabes output baseret på de udvalgte informationer. Denne proces illustreres i figur 4.9.

Figur: 4.9 Output gate (Olah, 2015)



Figuren illustrerer output gate i LSTM. Denne har til formål at klarlægge, hvad modellens output skal være, med afsæt i den udvalgte informationen. Dette tager afsæt i et sigmoid layer σ og en tanh funktion, som multipliceres med output fra sigmoid gaten.

Herunder benyttes et sigmoid layer, σ, som angiver, hvor meget af den nuværende cell state, som anvendes til output. Første step er givet ved:

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

-10 Hertil føres den nuværende cell state gennem tanh funktionen for at opnå værdier i intervallet [-1;1]. Herefter multipliceres det med output fra sigmoid gaten, o_i, givet ved:

$$h_t = o_t tanh(C_t)$$

Det endelige output for den enkelte sekvens er givet ved h_t imens C_t angiver informationen, som overføres til skjulte layers i næste sekvens, hvorefter hele processen gentages.

4.4 Evalueringsmetoder - Forecast

Følgende afsnit beskriver specialets evalueringsmetoder, som benyttes til at sammenligne, evaluere og diskutere de udvalgte ARIMA og LSTM modeller. Litteraturen anvender en bred vifte af statistiske metoder, hvor specialet har udvalgt de mest anvendte og relevante.

4.4.1 Mean Absolute Error & Root Mean Squared Error

Tabsfunktioner angiver metoder, som estimerer differencen mellem forecastede og faktiske værdier. De angiver herved forecastnøjagtigheden, hvor en lavere værdi implicerer et mere nøjagtigt forecast (Kotze, n.d.). For at sikre, at positive og negative forecastfejl ikke udlignes, benyttes kun tabsfunktioner, som anvender absolutte eller kvadrerede fejl.

Mean Absolute Error, MAE, angiver gennemsnittet af de absolutte forskelle mellem faktiske og forecastede værdier. MAE skalerer alle forecastfejl lineært, hvorfor fejlene vægtes ligeligt i gennemsnittet. Målet kan udtrykkes ved (Kampakis, 2020):

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

Mean Aboslute Percentage Error, MAPE, er hertil bredt anvendt grundet metodens intuitive resultat. Denne beregnes tilsvarende MAE, men i % af de faktiske værdier:

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

Root Mean Squared Error, RMSE, er givet ved kvadratroden af de gennemsnitlige kvadrerede forskelle mellem faktiske og forecastede værdier:

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

Kvadreringen af fejlene medvirker til, at forecastfejlene pålægges en straf, som stiger når forecastfejlene øges (Kampakis, 2020). Dette angiver, at RMSE skalerer større fejl eksponentielt og derfor altid tager en værdi, som er lig eller større end MAE. MAE og RMSE kan herved kombineres med henblik på at forklare variationen i forecastfejlene. En større forskel mellem de to værdier indikerer en større spredning i fejlene og vice versa.

4.4.2 Diebold-Mariano test

Diebold-Mariano testen anvendes til at teste, hvorvidt et givet forecast er signifikant mere nøjagtig relativt til et andet (Zaiontz, 2021). Testen anvender følgende tabsfunktion:

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

e_{1i} og e_{2i} angiver forecastfejlene for model 1 og 2, $g(e_{1i})$ og $g(e_{2i})$ angiver tabsfunktionerne, hvortil $\bar{d}$ angiver middelværdien af forskellen mellem tabsfunktionerne. Nullhypotesen er givet ved $\bar{d} = 0$, hvilket betyder, at der ikke er signifikant forskel mellem modellerne. Der angives flere alternative hypoteser; den første angiver, at de to forecast er signifikant forskellige, den anden, at forecast to er signifikant mere nøjagtig og den sidste angiver, at forecast to er signifikant mindre nøjagtig (McLeman-Hasselgaard og Nielsen, 2017).

4.4.3 Pesaran-Timmermann test

Pesaran-Timmermann testen anvendes til at vurdere, hvorvidt en forecast model er i stand til korrekt at forecaste retningsskift i en given tidsserie. Testen er givet ved:

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

S_n antages at være normalfordelt og de enkelte led er givet ved:

$$\hat{p} = n^{-}1\sum_{t=1}^{n} I(y_t, x_t)$$

$$\hat{p}_* = \hat{p}_y\hat{p}_x + (1 - \hat{p}_y)(1 - \hat{p}_x)$$

$$\hat{var}(\hat{p}) = n^{-1}\hat{p}_*(1 - \hat{p}_*)$$

$$(\hat{p}_y) = n^{-1}\sum_{t=1}^{n} I(y_t)$$

$$(\hat{p}_x) = n^{-1}\sum_{t=1}^{n} I(x_t)$$

$$I(.) = \begin{cases} 1 =, & \text{if } (.) \geq 1 \\ 0 =, & \text{otherwise} \end{cases}$$

$\hat{p}_*$ angiver sandsynligheden for, at tidserien, y_t, og forecastet, x_t, begge er større end nul summeret med sandsynligheden for, at de ikke er. $\hat{p}$ angiver hertil andelen af gange den faktiske og forecastede værdi har samme retning. $\hat{p}$ ligger i intervallet [0;1], hvor $\hat{p} = 1$ implicerer, at alle forecasts er af korrekt retning. Nullhypotesen angiver, at y_t og x_t er uafhængigt fordelt og herved, at x_t ikke er i stand til at forecaste retningsskift i y_t (Pesaran og Timmermann, 1992).

5 Teoretisk gennemgang af Mean-variance

Mean-variance er et populært finansielt værktøj, som specialet anvender med et formål om at udarbejde en optimeret portefølje af grønne aktier. Mean-variance udarbejder porteføljer baseret på det procentvise forventede afkast, variansen samt en kovarians-matrice, beregnet på baggrund af historisk data. Teorien tager herunder afsæt i diversifikationsbegrebet, hvor det antages muligt at diversificere risikoen i en portefølje, såfremt aktiverne ikke korrelere perfekt. Dette betyder, at de enkelte aktiver, er af central betydning for en porteføljes samlede risiko og afkast, hvilket afspejles i kovariansen mellem aktivernes forventede afkast (Hillier m.fl., 2011, kap. 4 s. 88-109).

Mean-variance antager, at investeringsbeslutninger alene tager afsæt i det forventede afkast og variansen af en samlet portefølje. Det antages hertil, at investorer har præferencer om at opnå porteføljer, der genererer højest muligt afkast givet risiko. Herudover antages data at følge en normalfordeling, imens markedet betragtes som friktionsløst (Hillier m.fl., 2011, kap. 5 s. 124-125)[11]. Teorien beregner den optimale portefølje med afsæt i tre informationssæt, givet ved det forventede afkast, variansen og kovariansen mellem *N* aktiver. Det forventede afkast af en portefølje kan udtrykkes ved:

$$\mu_P = E(\tilde{R}_P) = w^\top \mu$$

μ og w er vektorer for aktivernes forventede afkast og de enkelte vægte i den samlede portefølje, som er givet ved:

$$\mu = \begin{pmatrix} E[\tilde{R}_1] \\ E[\tilde{R}_2] \\ \vdots \\ E[\tilde{R}_N] \end{pmatrix} \qquad w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix}$$

[11] Friktionsløst marked angiver, at der handles til enhver kurs, uden hensyntagen til transaktionsomkostninger, regler eller skattemæssige konsekvenser (Hillier m.fl., 2011, kap. 5 s. 124-125).

Endvidere kan variansen af en sammensat porteføljе udtrykkes ved:

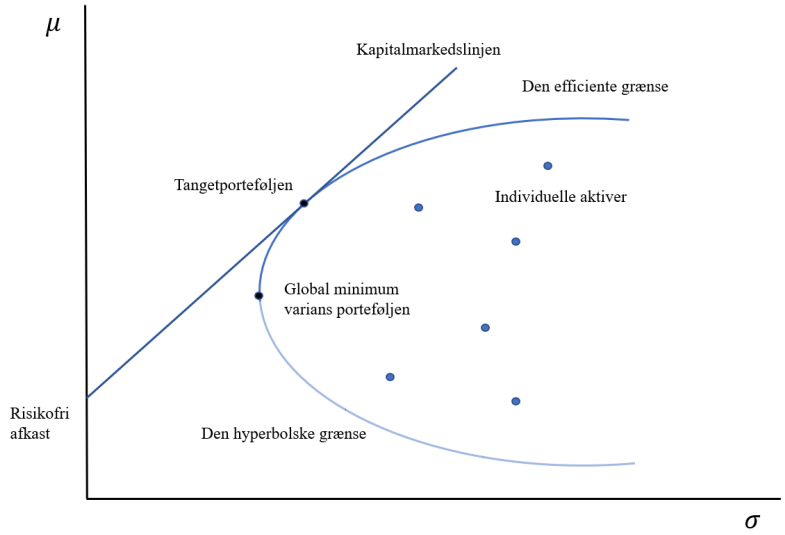
$$\sigma_P^2 = w^\top \sum w$$

$\sum$ angiver en $N \times N$ varians-kovarians matrice mellem afkastene af porteføljens N aktiver:

$$\sum = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1N} & \sigma_{2N} & \cdots & \sigma_N^2 \end{bmatrix}$$

Med udgangspunkt i teoriens antagelser, samt forskellige niveauer af afkast og risiko, illustreres et mean-standard deviation diagram i figur 5.1.

Figur: 5.1 Mean-standard deviation diagram (Hillier m.fl., 2011, kap. 5 s. 136)



Figuren illustrerer et mean standard deviation diagram, hvor x-aksen angiver standardafvigelsen af afkastet og y-aksen det forventede afkast. De enkelte markører angiver individuelle aktiver, som sammenfattes til den hyperbolske grænse. Den øvre del angiver den efficiente grænse, med alle efficiente porteføljer. Markedsligevægten er givet ved tangetporteføljen, som angiver porteføljen, med det højeste afkast givet risiko.

Figuren illustrerer sættet af mulige porteføljesammensætninger, angivet ved den hyperbolske grænse. Den øvre del betegnes som den efficiente grænse, der markerer de efficiente porteføljer. Den efficiente grænse starter fra den globale minimum varians portefølje, som angiver porteføljekombinationen, der tilsvarer den mindste forventede risiko. Tangentpor-

teføljen angiver hertil det punkt, hvor kapitalmarkedslinjen tangerer med den efficiente grænse. Den globale minimum variansportefølje og tangentporteføljen, beregner tilsammen den efficiente grænse. Vægtningen for de to porteføljer kan udtrykkes ved:

$$w_{min} = \left[\mathbf{1}^\top \sum^{-1} \mathbf{1}\right]^{-1} \sum^{-1} \mathbf{1} \qquad w_{tan} = \frac{\sum^{-1}(\mu - R^f \mathbf{1})}{\mathbf{1}^\top \sum^{-1}(\mu - R^f \mathbf{1})}$$

1 angiver en N × 1 vektor af 1-taller, imens R_f angiver det risikofrie afkast.

Kapitalmarkedslinjen angiver sættet af effektive porteføljer, som kombinerer risikoprægede aktiver med en risikofri investering og er givet ved (Hillier m.fl., 2011, 2011, kap. 5 s. 129):

$$\bar{R}_P = R_f + \frac{\bar{R}_T - R_f}{\sigma_T}\sigma_P$$

$\bar{R}_T$ og σ_T angiver middelværdien og standardafvigelsen for tangentporteføljens afkast.

Givet, at teoriens antagelser indfries og, at der foreligger et risikofrit aktiv, ønsker investorer en portefølje placeret på kapitalmarkedslinjen. Hældningen på kapitalmarkedslinjen, betegnes som værende prisen for risiko og måler graden af efficiens i en portefølje. Hældningen defineres Sharpe ratio og er givet ved (Hillier m.fl., 2011, kap. 5 s. 128):

$$Sharpe\ Ratio = \frac{\bar{R}_p - R_f}{\sigma_p}$$

R_p angiver afkastet for porteføljen, hvor σ_p angiver standardafvigelsen for porteføljens merafkast i relation til det risikofrie afkast.

Tangetporteføljes hældning viser den maksimale Sharpe ratio, hvor investorer opnår det maksimale afkast, givet risiko[12]. Den enkelte investors placering relativ til tangetporteføljen afhænger af investorens risikovillighed. Risikoaverse investorer ønsker en portefølje placeret tæt på det risikofrie afkast, hvorimod risikovillige investorer ønsker en portefølje placeret til højre for tangetporteføljen (Hillier m.fl., 2011, kap. 5 s. 128).

[12]Givet, at afkastet af minimum varians porteføljen overgår det risikofrie afkast, vil tangentporteføljen placeres på den øvre del af den hyperbolske grænse.

5.1 Evalueringsmetoder - Mean-variance

Med henblik på at sammenligne og vurdere de udarbejdede porteføljers profitabilitet og risikoniveau, anvendes porteføljernes merafkast, gennemsnitlige afkast, standardafvigelse, Sharpe ratio, information ratio, totale afkast og maximum drawdown.

Information ratio, relateres til Sharpe ratio, da begge angiver et merafkast givet risiko. Information ratio reflekterer afkastet relativt til et givet benchmark og kan udtrykkes ved:

$$Information\ Ratio = \frac{R_p - R_b}{\sigma_p}$$

R_p angiver afkastet for porteføljen, R_b er afkastet for et givet benchmark og σ_p er standardafvigelsen for porteføljens merafkast (Han m.fl., 2021).

Maximum drawdown angiver det maksimale fald i en porteføljes værdi fra et top- til lavpunkt, før et nyt toppunkt opnås. Dette kan udtrykkes ved:

$$Maximum\ drawdown = \frac{net\ v_l - net\ v_p}{net\ v_p}$$

$net\ v_l$ og $net\ v_p$ angiver porteføljens netværdi ved det givne lav- og toppunkt. Maximum drawdown angiver herved porteføljens største tab for perioden, hvorfor denne anvendes som et mål for den nedadgående risiko. Der tages herunder ikke forbehold for frekvensen af tab, samt hvor hurtigt eller, hvorvidt porteføljen genvinder den tabte værdi (Hayes, 2022).

6 Databehandling

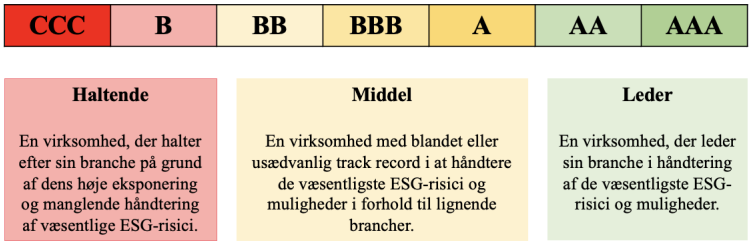
Følgende afsnit konkretiserer metodologien for udvælgelse af specialets aktier, som tager udgangspunkt i bæredygtighed, herunder ESG kriteriet, samt sektorfordelingen. Afslutningsvist angives en beskrivelse af specialets bæredygtige aktier, med udgangspunkt i relevant deskriptiv statistik, som har til formål at konkretisere dataets karakteristika.
Med henblik på at vurdere, hvorvidt forecasting og porteføljesammensætning kan være med til at skabe incitament til mere bæredygtig investering anvendes 20 bæredygtige aktier fra iShares ESG Screened S&P 500 ETF. Hertil inddrages S&P 500 som et benchmark for den generelle udvikling på aktiemarkedet (Blackrock, 2023).

6.1 Bæredygtighed og sektorfordeling

Specialet anvender environment, social and governance kriteriet, ESG, til at identificere selskaber, der er engageret i bæredygtig praksis. De tre faktorer angiver i hvor høj grad selskaber imødegår vejen til en fremtid, som opfylder nuværende generationers behov, uden at bringe fremtidige generationers behov i fare (FN‘s Verdensmål, 2016). Den miljømæssige faktor har til formål at vurdere et selskabs indvirkning på miljøet, herunder dens CO2 aftryk, energieffektivitet og brug af naturressourcer. Den sociale faktor vurderer et selskabs indvirkning på mennesker, herunder selskabets arbejdspraksis samt politikker for mangfoldighed. Afslutningsvist undersøger ledelsesfaktoren et selskabs ledelsesstruktur, interne kontroller, aktionærrettigheder og etiske standarder (Broadridge, 2021).
Faktorerne anvendes herved som en målestok for, hvor ansvarligt et selskab er på de tre områder, og anses som et værktøj til at kommunikere et selskabs bæredygtighedsstrategi (Erhvervstyrrelsen, 2021). Selskaberne vurderes af eksterne selskaber, hvor aktierne anvendt i specialet alle har en rating udarbejdet af MSCI ESG Quality Rating. Denne måler et selskabs evne til at håndtere vigtige mellem- til langsigtede risici og muligheder, der op-

står fra de tre ESG-faktorer. Denne rating dannes for individuelle selskaber indenfor sammenlignlige brancher, hvorfor ratingen ikke kan sammenlignes på tværs. På baggrund heraf repræsenterer MSCI ESG quality Rating en AAA-CCC-skala, hvor "AAA"og "CCC"er de respektive højeste og laveste ESG-ratings, som illustreret i figur 6.1.

Figur. 6.1 MSCI ESG Quality Rating (Your Green Wealth, 2023)

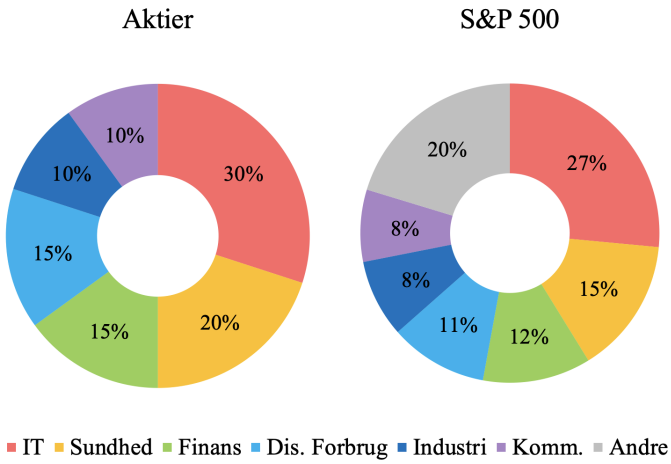


Figuren illustrerer mulige ESG-ratings for selskaber. Den dårligste rating er givet ved "CCC"og den bedste "AAA". Ethvert selskab vurderes herunder relativt til andre selskaber i samme branche.

Med udgangspunkt i denne vurdering, inkluderer specialet udelukkende selskaber med en MSCI ESG Quality Rating på "A"eller derover, samt ekskluderer sektorerne fossile brændstoffer, tobak og militær. Dette betyder, at når specialet refererer til bæredygtig investering er dette implicit med selskaber, der har en ESG rating på "A"eller derover (BlackRock, 2023). Specialets valg er grundet et mål om at opstille en portefølje, der afspejler S&P 500´s vægtning af sektorer. Dette forekommer uopnåeligt ved udelukkelsen af selskaber med ratingen "A", da antallet af selskaber med ratingen "AA "eller "AAA"er for begrænset til at opnå en pålidelig afspejling af diversificeringen og udviklingen i S&P 500.

Specialets aktier er endvidere udvalgt med henblik på at opnå en diversitet i porteføljen, så denne bliver mere attraktiv for en investor. Aktierne i diverse sektorer er herunder udvalgt ud fra på selskabets størrelse, baseret på markedsværdi. De udvalgte sektorer er givet ved IT, sundhed, finans, diskretionære forbrugsgoder, industri og kommunikation. Sektorfordelingen for specialets aktier og S&P 500 illustreres i figur 6.2.

Figur. 6.2 Sektorfordeling blandt udvalgte bæredygtige aktier og S&P 500



Figuren illustrerer den procentvise sektorfordeling for specialets bæredygtige aktier og S&P 500 afrundet til hele værdier. IT og sundhed vægtes højest, hvortil S&P 500 har 20% aktier i sektorer som specialet ikke inddrager.

Ovenstående angiver, at fordelingen af aktier tilnærmelsesvis afspejler fordelingen i S&P 500. Specialet har herunder taget udgangspunkt i de seks sektorer, som har den største procentvise vægtning. Det er hertil essentielt for en investor at diversificere en portefølje relativt til sektorer, da disse reagerer forskelligt på økonomiske udsving. Dette kan komme til udtryk ved aktier inden for cykliske og defensive sektorer. Cykliske aktier følger ofte udviklingen i økonomien, hvorved aktiekursen i høj grad påvirkes af udsving i konjunkturen. Defensive aktier er i kontrast hertil ofte relativt ufølsomme over for konjunkturudsving. I perioder med vækst er cykliske aktier derfor fordelagtige, mens defensive aktier foretrækkes i perioder med lavkonjunktur. Specialets udvalgte sektorer kan uddybes ved følgende:

- **IT sektoren** omfatter selskaber, der arbejder med udvikling af IT løsninger. Sektoren kendetegnes som cyklisk, med høje vækstrater, store volatile udsving og høj risiko.
- **Sundhedssektoren** omfatter selskaber, der arbejder med udvikling samt markedsføring af sundhedsrelaterede produkter. I forbindelse med, at dette forekommer som en væsentlig og resistent del af samfundet, defineres denne som en defensiv sektor.
- **Finanssektoren** tager afsæt i finansielle institutioner, herunder banker, forsikrings-

selskaber samt kapitalforvaltere og betegnes som en cyklisk sektor.

- **Diskretionære forbrugsgoder sektoren** er præget af selskaber, som producerer goder og tjenester, der er målrettet skønsmæssigt forbrug. Denne sektor er herved ligeledes kendetegnet som cyklisk.

- **Industrisektoren** består af selskaber, som er involveret i produktion af goder til private og offentlige sektorer. Aktier i denne sektor påvirkes herved af den økonomiske situation og betegnes derfor som cykliske.

- **Kommunikationssektoren** angiver selskaber, som arbejder med kommunikationstjenester, herunder eksempelvis mobil, bredbånd eller satellit-tjenester. Aktier i denne sektor kan være følsomme over for regulering og teknologisk udvikling, hvor sektoren som helhed stadig kategoriseres som defensiv (Vorgaard, 2020).

Med udgangspunkt i ESG-kriteriet samt sektorfordelingen, tager specialet afsæt i nedenstående aktier illustreret i tabel 6.1.

Tabel: 6.1 Specialets bæredygtige aktier

Aktie	Sektor	ESG-Rating	Kredit-Rating
Apple Inc.	IT	A	AAA
Broadcom Inc.	IT	AA	BBB-
Mastercard Inc.	IT	AA	AA-
Microsoft Corp	IT	AAA	AAA
Nvidia Corp	IT	AAA	A-
Visa Inc.	IT	A	AA-
Abbvie Inc	Sundhed	A	BBB+
Eli Lilly	Sundhed	AA	A

Tabel: 6.1 Specialets bæredygtige aktier

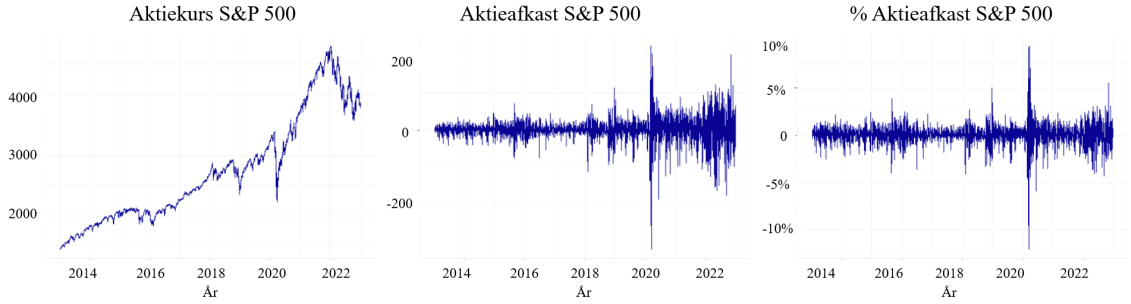
Aktie	Sektor	ESG-Rating	Kredit-Rating
Merck & Co.	Sundhed	AAA	A+
U. Health Group	Sundhed	AA	A-
Bank of America	Finans	A	A
Berkshire Hathaway	Finans	A	AA-
JP Morgan	Finans	A	A+
Amazon.com	Dis. Forbrug	A	AA
Home Depot Inc.	Dis. Forbrug	AA	A
Tesla Inc.	Dis. Forbrug	A	BBB
Caterpillar Inc.	Industri	A	A
U. Parcel Service	Industri	A	A
Alphabet Inc. C	Komm.	A	AA
Alphabet Inc. A	Komm.	A	AA+
Kilde: (MSCI, 2023)			

Tabellen angiver specialets 20 aktier, samt deres tilhørende sektorer. Hertil inddrages aktiernes ESG- og kredit-rating. Alle aktierne har en ESG-rating på "A"eller højere.

6.2 Deskriptiv Statistik

Specialet anvender 20 bæredygtige aktier til forecasting og porteføljesammensætning, hvor det anses væsentligt at konkretisere dataets karakteristika med udgangspunkt i relevant deskriptiv statistik. Specialets bæredygtige aktier analyseres med udgangspunkt i daglige lukkekurser for perioden primo 2013 til ultimo 2022. Udviklingen i S&P 500 illustreres indledningsvist i figur 6.3, da S&P 500 anvendes som specialets benchmark, hvorved resultater og udvalgte aktier sammenlignes løbende med denne.

Figur. 6.3 Udviklingen i aktiekursen for S&P 500

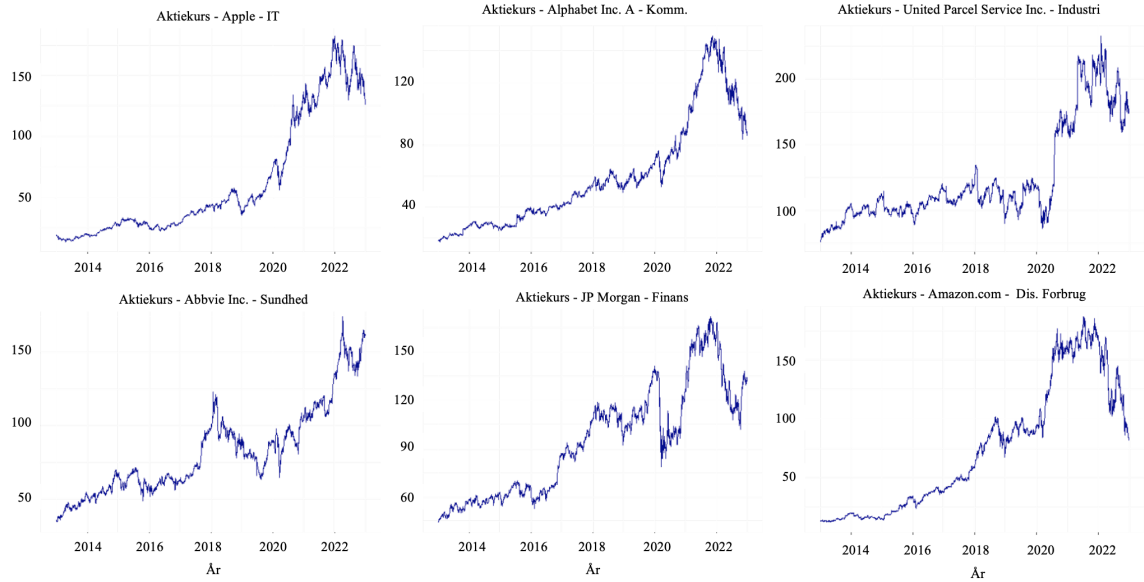


Aktiekursen og det nominelle samt procentvise afkast for S&P 500 illustreres for perioden primo 2013 til ultimo 2022. Her angiver x-aksen tiden og y-aksen kursen.

Det fremgår, at S&P 500 har oplevet en general og stabil vækst, med en række små til moderate fald i kursen. Herunder var der i 2020 både et stort fald i kursen på omtrent 20% og et aktierebound. Udviklingen i 2020 var præget af en række faktorer, herunder den økonomiske nedgang grundet COVID-19, den stærke støtte fra den amerikanske centralbank og den massive støttepakke til den amerikanske økonomi (Kvilhaug, 2023). Endvidere var 2022 præget af et generelt fald i kursen, hvortil de procentvise afkast var mere volatile end tidligere. Udviklingen i S&P 500 kan herved anses relativ stabil gennem hele perioden, hvor der dog forekommer større udsving i specialets valgte testperiode, givet ved primo 2020 til ultimo 2022 (Maverick, 2023). Dette kan være af relevant betydning i forbindelse med senere konklusioner, hvorfor der vendes tilbage til dette senere i specialet.

For at undersøge, hvorledes specialets bæredygtige aktier bevæger sig og relateres til udviklingen i det generelle marked, illustrerer figur 6.4 udviklingen i én udvalgt aktie for hver af de inkluderede sektorer. Illustrationen inkluderes med henblik på at opnå indsigter i specialets data, som senere kan være af betydning i forbindelse med opnåede resultater og udarbejdelse af specialets forecast modeller. Det antages herunder, at hver aktie er repræsentativ for den generelle udvikling i den tilhørende sektor[13].

Figur: 6.4 Udviklingen i aktiekursen for udvalgte aktier



Aktiekurserne illustreres for perioden primo 2013 til ultimo 2022. Her angiver x-aksen tiden og y-aksen kursen. Det skal her bemærkes, at y-aksen varierer for de individuelle aktier.

Det illustreres i figur 6.4, at kurserne udvikler sig forskelligt, hvilket er i overensstemmelse med, at sektorerne er præget af forskellige strukturer. Alle aktierne er hertil præget af perioder, der kan betegnes som henholdsvis bear- og bull-markeder[14]. Det amerikanske aktiemarked er i perioden 2013 til 2019 kendetegnet ved et bull-marked præget af høj vækst, imens 2022 betegnes som et bear-marked med store fald i kurserne (Maverick, 2023). Dette er ligeledes en tendens i specialets bæredygtige aktier. Sundhedssektoren forekommer

[13]Specialet udvælger seks aktier med henblik på at kunne fremsætte overskuelige illustrationer. Det vil derfor fremadrettet være disse seks aktier, som anvendes i specialets figurer.

[14]Bear- og bull-markeder er kendetegnet ved henholdsvis et marked i nedgang og et marked i vækst fra seneste top til bund på 20% (Madsen, 2022).

hertil relativt uafhængig af den generelle udvikling på aktiemarkedet, hvilket er i overensstemmelse med, at denne er en defensiv sektor.

Med henblik på at analysere og sammenligne de enkelte aktier relativt til S&P 500, anvender specialet tracking error, information ratio og beta koefficienter, som angivet i tabel 6.2. Disse er vigtige og bredt anvendte værktøjer til investeringsanalyser (Gupta m.fl., 1999).

Tabel: 6.2 Tracking error, information ratio og beta værdier

Aktie	Sektor	Tracking error	Information ratio	Beta værdier
Apple Inc.	IT	1,31%	5,57%	1,2
Broadcom Inc.	IT	1,66%	6,52%	1,3
Mastercard Inc.	IT	1,10%	3,63%	1,2
Microsoft Corp	IT	1,09%	4,20%	1,2
Nvidia Corp	IT	2,24%	5,11%	1,6
Visa Inc.	IT	1,02%	3,40%	1,1
Abbvie Inc.	Sundhed	1,51%	0,45%	0,8
Eli Lilly	Sundhed	1,46%	4,77%	0,7
Merck & Co.	Sundhed	1,22%	4,76%	0,6
U. Health Group	Sundhed	1,23%	2,05%	0,9
Bank of America	Finans	1,37%	1,19%	1,3
Berkshire Hathaway	Finans	0,71%	1,05%	0,9
JP Morgan	Finans	1,13%	1,47%	1,1
Amazon.com	Dis. Forbrug	1,64%	3,12%	1,1
Home Depot Inc.	Dis. Forbrug	1,03%	5,41%	1,0
Tesla Inc.	Dis. Forbrug	3,28%	2,97%	1,4
Caterpillar Inc.	Industri	1,37%	-0,04%	1,0
U. Parcel Service	Industri	1,19%	0,65%	0,8
Alphabet Inc. A	Komm.	1,17%	2,84%	1,1
Udarbejdet i RStudio				

Tabel: 6.2 Tracking error, information ratio og beta værdier

Aktie	Sektor	Tracking error	Information ratio	Beta værdier
Alphabet Inc. C	Komm.	1,18%	2,82%	1,1
Udarbejdet i RStudio				

Tabellen angiver tracking error, information ratio og beta værdier for alle specialets aktier.

Tracking error angiver i hvor høj grad aktien følger udsvingene i benchmarken, i henhold til standardafvigelsen af forskellen i afkastene. Tracking error kan udtrykkes ved:

$$\text{Tracking error} = \sqrt{Var(R_s - R_b)}$$

R_s og R_b angiver det procentvise afkast for den givne aktie og benchmarken. Tracking error beregnes for daglige afkast og er for alle aktierne mellem 0,71% og 3,28%, som angivet i tabel 6.2. Dette implicerer, at de procentvise daglige afkast forventes at afvige med plus/minus den givne værdi. Værdierne er højest i sektorerne diskretionært forbrug og IT, hvilket er i overensstemmelse med, at sektorerne kendetegnes som meget volatile. Tracking error tager ikke forbehold for, hvorvidt aktien konsekvent har over- eller underperformet relativt til S&P 500. Det forekommer derfor relevant at inkludere information ratio.

Information ratio reflekterer, som angivet i *5.1 Evalueringsmetoder - Mean-variance*, et merafkast relativt til S&P 500. En højere information ratio, indikerer et højere merafkast og herved en højere performance relativt til benchmarken. Denne fremgår negativ for Caterpillar Inc., hvilket implicerer et mindre afkast givet risiko. De resterende aktier opnår en positiv information ratio og formår herved at præstere et merafkast relativt til S&P 500.
For at undersøge en aktie i relation til det generelle marked, er det ligeledes relevant at undersøge, hvorvidt aktien korrelerer positivt eller negativt med markedet. Herunder anvendes

beta koefficienter, som angiver en akties systematiske risiko i forhold til markedsrisikoen. Beta kan udtrykkes ved:

$$\beta = \frac{Cov(R_e, R_m)}{Var(R_m)}$$

$Cov(R_e, R_m)$ angiver kovariansen mellem det procentvise afkast for en aktie og markedet, hvortil $Var(R_m)$ angiver variansen af det procentvise markedsafkast. $|\beta| = 1$ implicerer, at aktiens afkast korrelerer perfekt med markedet, mens $|\beta| > 1$ og $|\beta| < 1$ indikerer, at aktien er henholdsvis mere og mindre volatil. Herunder angiver $\beta > 0$, at aktien følger markedsudviklingen, hvortil $\beta < 0$ angiver, at aktien bevæger sig modsat markedet. Beta kan herved anvendes til vurdering af en akties risiko, da den implicerer, hvorvidt en aktie vil falde eller stige, når det generelle marked falder eller stiger. Alle aktierne har positive beta værdier, hvilket implicerer, at de har en tendens til at følge markedet. Aktier i sundhedssektoren har endvidere de laveste beta værdier, imens IT har de højeste. Dette er i overensstemmelse med sektorernes generelle kendetegn (Kenton, 2023).
Med henblik på at relatere ovenstående til bæredygtighedselementet, angives den gennemsnitlige tracking error, information ratio og beta værdi for, hver ESG-rating i tabel 6.3.

Tabel: 6.3 Tracking error, information ratio og beta - ESG

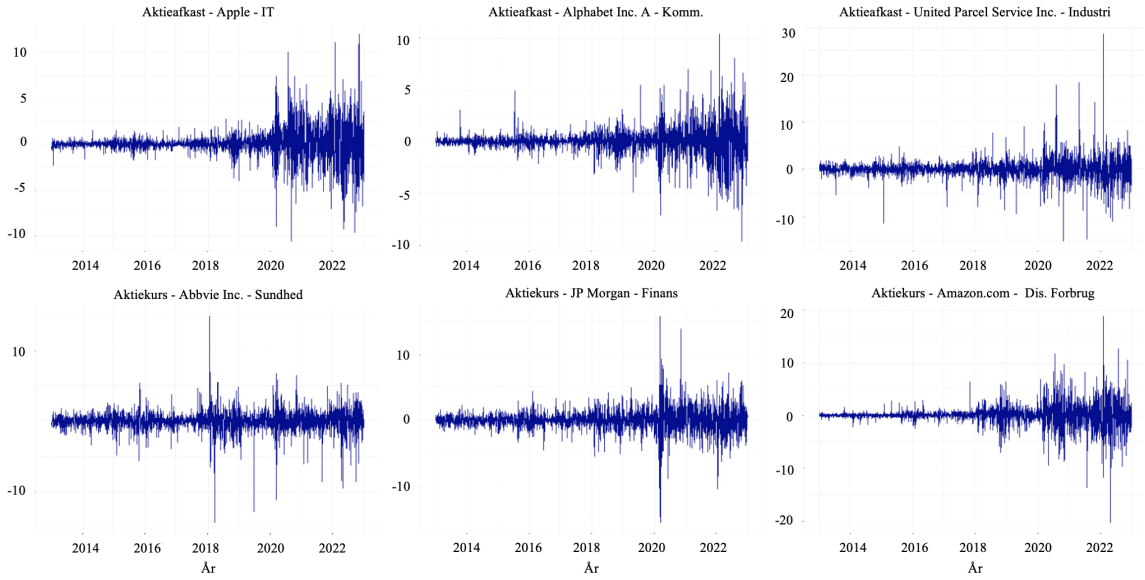
ESG-rating	Tracking error	Information ratio	Beta værdier
Middelværdi ESG "A"	1,30%	2,30%	1,08
Middelværdi ESG "AA"	1,30%	4,45%	1,03
Middelværdi ESG "AAA"	1,52%	4,03%	1,15
Udarbejdet i RStudio			

Tabellen angiver de gennemsnitlige tracking errors, information ratios og beta værdier, hvor alle værdier er afrundet til to decimaler. De mest bæredygtige aktier med ESG-ratingen "AA"og "AAA"opnår den højeste information ratio og herved det højeste merafkast relativt til S&P 500.

Det fremgår af ovenstående, at specialets aktier med ESG-ratingen "AA"opnår den højeste gennemsnitlige information ratio og laveste tracking error samt beta værdi. Dette implicerer, at aktierne herunder opnår det højeste afkast givet risiko relativt til S&P 500, hvorved det forventes, at disse aktier vil vægtes højt i en mean-variance optimeret portefølje. Hertil opnår "AAA"aktierne en højere information ratio end "A"ratede aktier, hvor de dog ligeledes forekommer mere volatile givet ved den højere tracking error og beta værdi.

Da specialet ønsker at anvende de bæredygtighed aktier til forecasting er det essentielt, at tidsserierne forekommer stationære. Dette er tidligere beskrevet i afsnittet *4.2 Teoretisk gennemgang af ARIMA*. Hertil fremgår det i figur 6.4, at de udvalgte aktier er præget af trend og drift, hvorved tidsserierne ikke forekommer stationære. Specialet søger at opnå stationaritet ved differentiering, hvortil figur 6.5. illustrerer de differentierede aktiekurser.

Figur: 6.5 Udvikling i aktieafkast for udvalgte aktier



Aktieafkastene illustreres for perioden primo 2013 til ultimo 2022. Her angiver x-aksen tiden, hvor y-aksen angiver aktieafkast. Det skal her bemærkes, at værdierne på y-aksen varierer for de individuelle aktier.

Afkastet for alle aktierne bevæger sig omkring en middelværdi på nul, hvortil variansen er varierende over tid. Det illustreres, at sundhedssektoren har en relativ konstant vari-

ans, hvortil variansen for de resterende aktier er stigende i perioden. Der tages forbehold for den tidsvarierende varians ved anvendelse af Box Cox transformation, som beskrevet i *4.3 Teoretisk gennemgang af ARIMA*. Udviklingen i variansen er forskellig for de enkelte aktier, hvorfor Box Cox transformationen bestemmes individuelt og tilpasses den enkelte tidsserie. Aktiekurserne, integreret af første orden, er herved stationære og kan anvendes til forecasting med ARIMA. Dette understøttes i en ADF test angivet i bilag B3.
I forbindelse med mean-variance, er det vigtig, at de procentvise afkast forekommer tilnærmelsesvist normalfordelt. Med henblik på at undersøge dette, beregnes skewness og kurtosis, for aktiernes procentvise afkast. De to mål sammenfattes efterfølgende i Jarque-Bera testen for normalfordeling, hvor resultaterne illustreres i tabel 6.4.

Tabel: 6.4 Skewness, Kurtosis og Jarque-Bera test

Aktie	Sektor	Skewness	Kurtosis	JB: p-værdi
Apple Inc.	IT	-0,13	8,72	0,00
Broadcom Inc.	IT	-0,14	11,25	0,00
Mastercard Inc.	IT	0,32	12,34	0,00
Microsoft Corp	IT	0,01	11,69	0,00
Nvidia Corp	IT	0,50	12,78	0,00
Visa Inc.	IT	0,25	12,40	0,00
Abbvie Inc.	Sundhed	-0,56	12,82	0,00
Eli Lilly	Sundhed	0,91	15,36	0,00
Merck & Co.	Sundhed	0,16	10,19	0,00
U. Health Group	Sundhed	-0,02	15,66	0,00
Bank of America	Finans	0,27	13,27	0,00
B. Hathaway	Finans	0,63	14,04	0,00
JP Morgan	Finans	0,34	17,06	0,00
Amazon.com	Dis. Forbrug	0,27	9,85	0,00
Udarbejdet i RStudio				

Tabel: 6.4 Skewness, Kurtosis og Jarque-Bera test

Aktie	Sektor	Skewness	Kurtosis	JB: p-værdi
Home Depot Inc.	Dis. Forbrug	-0,90	22,76	0,00
Tesla Inc.	Dis. Forbrug	0,38	8,14	0,00
Caterpillar Inc.	Industri	-0,24	7,68	0,00
U. Parcel Service	Industri	0,68	16,19	0,00
Alphabet Inc. C	Komm.	0,53	11,77	0,00
Alphabet Inc. A	Komm.	0,50	11,71	0,00
S&P 500	-	-0,55	18,47	0,00
Udarbejdet i RStudio				

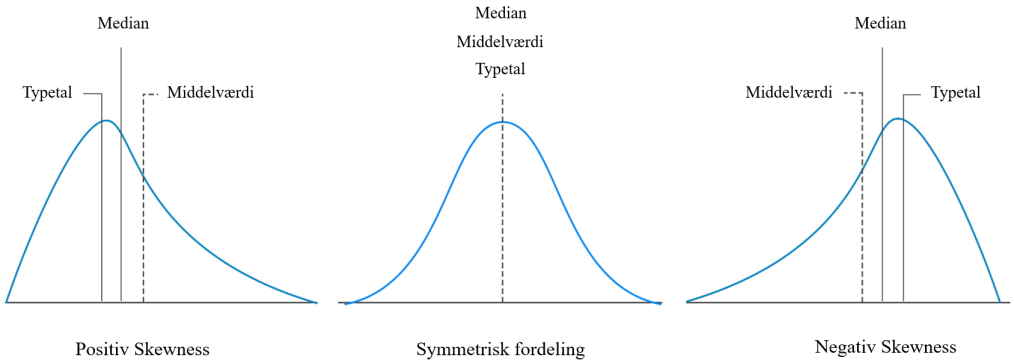
Tabellen angiver skewness, kurtosis og p-værdierne for Jarque-Bera testen beregnet i henhold til de procentvise afkast for alle specialets aktier.

Skewness beskriver fordelingens symmetri omkring middelværdien og kan beregnes ved:

$$S = \frac{E[(y-\mu)^3]}{E[(y-\mu)^2]^{\frac{3}{2}}}$$

y angiver den faktiske værdi og μ angiver middelværdien. Skewness illustreres hertil i figur 6.6. $S = 0$ indikerer, at fejlledende er perfekt symmetrisk distribueret omkring middelværdien, som angivet i midten af figur 6.5. $S > 0$ indikerer en højreskæv fordeling, som kendetegnes ved flest positive afkast, angivet til venstre. Modsat implicerer $S < 0$ en venstreskæv fordeling, og herved en overvægt af negative afkast, som angivet til højre. Det fremgår af tabel 6.4, at størstedelen af de procentvise afkast er præget af skæve fordelinger, hvor der forekommer en forskel i retningen af skævheden. Da investorer generelt foretrækker positive afkast, har de ofte en præference for en højreskæv fordeling (J. Chen, 2023).

Figur: 6.6 Skewness (AnalystPrep, 2020)



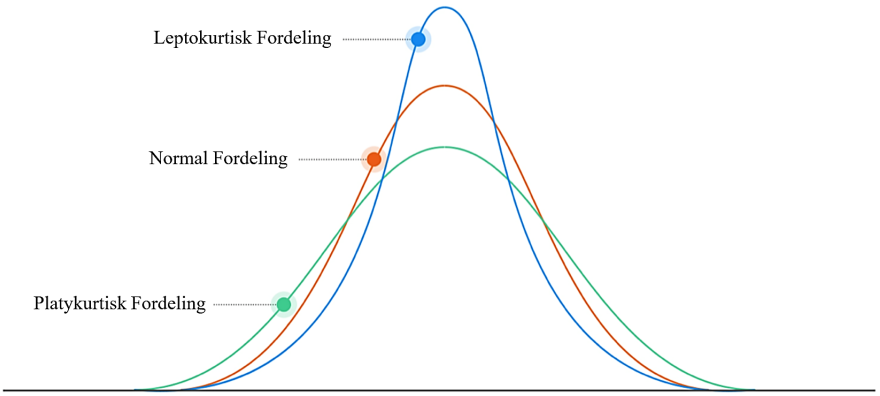
Figuren illustrerer skewness og herved symmetrien omkring middelværdien. Den højre figur angiver en positiv skewness, midten en symmetrisk fordeling og venstre en negativ skewness. Middelværdien angiver gennemsnittet, medianen værdien af den midterste observation og typetal den værdi, som oftest forekommer.

Kurtosis beskriver hertil, hvorledes halerne i en given fordeling adskiller sig fra en standard normalfordeling og er givet ved:

$$K = \frac{E[(y-\mu)^4]}{E[(y-\mu)^2]^2}$$

En standard normalfordeling har en $K = 3$, som benævnes en Mesokurtisk fordeling. $K > 3$ indikerer tunge haler, og dermed store outliers, benævnt en leptokurtisk fordeling, hvortil $K < 3$ indikerer flade haler, og hermed små afvigelser, benævnt en platykurtisk fordeling. De tre fordelinger visualiseres i figur 6.7.

Figur: 6.7 Kurtosis (AnalystPrep, 2019)



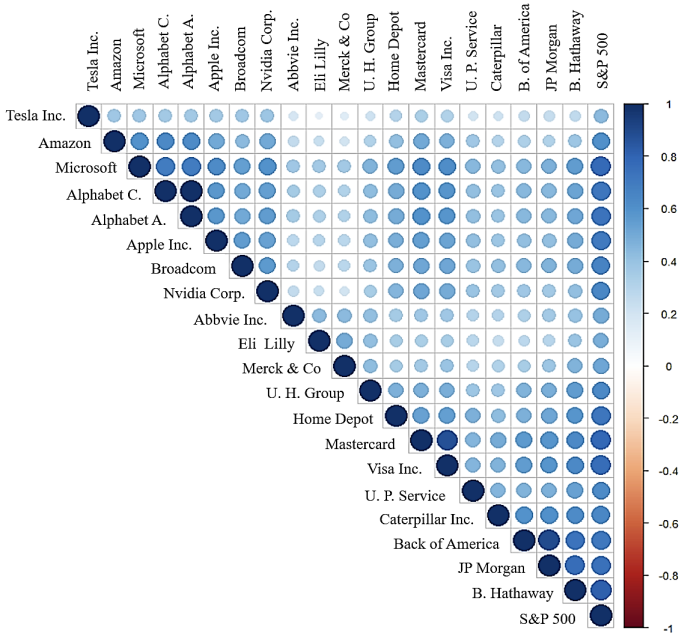
Figuren illustrerer kurtosis. En kurtosis lig tre implicerer en standard normalfordeling, en værdi under tre implicerer en leptokurtisk fordeling, hvortil en værdi over tre angiver en platykurtisk fordeling.

Relateres dette til aktieafkast, vil en leptokurtisk fordeling forbindes med høj risiko, mens en platykurtisk fordeling forbindes med et mindre risikoniveau (AnalystPrep, 2019). Det fremgår af tabel 6.3, at det procentvise afkast for specialets aktier alle har en $K > 3$.
Skewness og kurtosis kombineres afslutningsvist i en Jarque-Bera test, hvor nullhypotesen angiver, at data forekommer normalfordelt. Testen kan udtrykkes ved:

$$JB = \frac{n}{6}(S^2 + \frac{(K-3)^2}{4})$$

Med en p-værdi, som tilnærmes nul for alle aktier, forkastes nullhypotesen om normalfordelt data (Dwight, 2019). Dette kan medvirke til problematikker i forbindelse med usikre resultater og estimater i mean-variance analysen, hvilket uddybes senere i specialet.
I forbindelse med anvendelse af aktier og forecastede værdier i mean-variance, er korrelationen mellem de procentvise afkast af relevant betydning. Dette er grundet, at det er essentielt, at en portefølje er diversificeret, hvilket antages muligt, såfremt de bæredygtige aktier ikke korrelerer perfekt. Korrelationen illustreres i figur 6.8.

Figur: 6.8 Korrelationsmatrice



Figuren er udarbejdet i RStudio og er beregnet for perioden primo 2013 til ultimo 2022. Korrelationen varierer herunder fra 0,12 til 0,99 mellem de enkelte aktier og S&P 500.

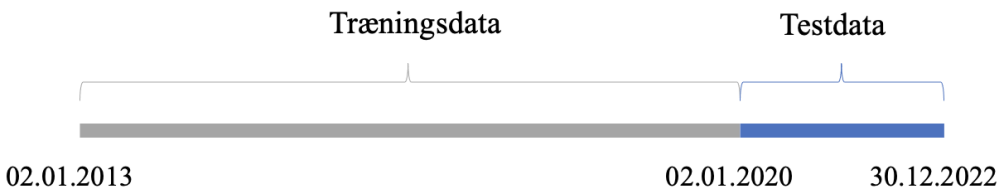
Ovenstående illustrerer, at korrelationen mellem aktierne og med S&P 500 varierer mellem 0,12 og 0,99, hvortil alle aktier korrelerer positiv. Dette implicerer, at udviklingen for hver given aktie følger udviklingen på det generelle marked og i specialets andre aktier. Alle aktierne har en høj korrelation med S&P 500, hvilket ikke er overraskende, da alle specialets aktier vægtes højt i S&P 500 og herved har stor indflydelse på udviklingen i indekset (Hillier m.fl., 2011, kap. 4 s. 88).

Specialets deskriptive statistik har klarlagt dataets karakteristika. Det kan ikke med udgangspunkt heri endeligt vurderes, hvorledes dataet bedst muligt kan forecastes. Følgende afsnit har hermed til formål at præcisere dette ved en gennemgang af udvælgelsen af specialets forecast modeller.

7 Udvælgelse af ARIMA og LSTM

Følgende afsnit har til formål at klargøre udvælgelsesprocessen af ARIMA og LSTM modeller. Specialet udvælger hertil én passende ARIMA og LSTM model for hver af de 20 aktier, hvortil de specifikke metoder og valgte udvælgelseskriterier fremføres nedenfor. I forbindelse med udvælgelsen af de specifikke modeller opdeles tidsserierne i henholdsvis trænings- og test data. Testdata dækker perioden 2020 til ultimo 2022 og er givet ved 753 observationer. Figur 7.1 illustrerer opdelingen af specialets data, i forbindelse med forecasting og porteføljesammensætning.

Figur. 7.1 Illustration af data



Figuren angiver specialets trænings- og testdata. Træningsdata omfatter perioden primo 2013 til ultimo 2019 og er angivet med grå i figuren. Testdata er angivet med blå og omfatter perioden primo 2020 til ultimo 2022.

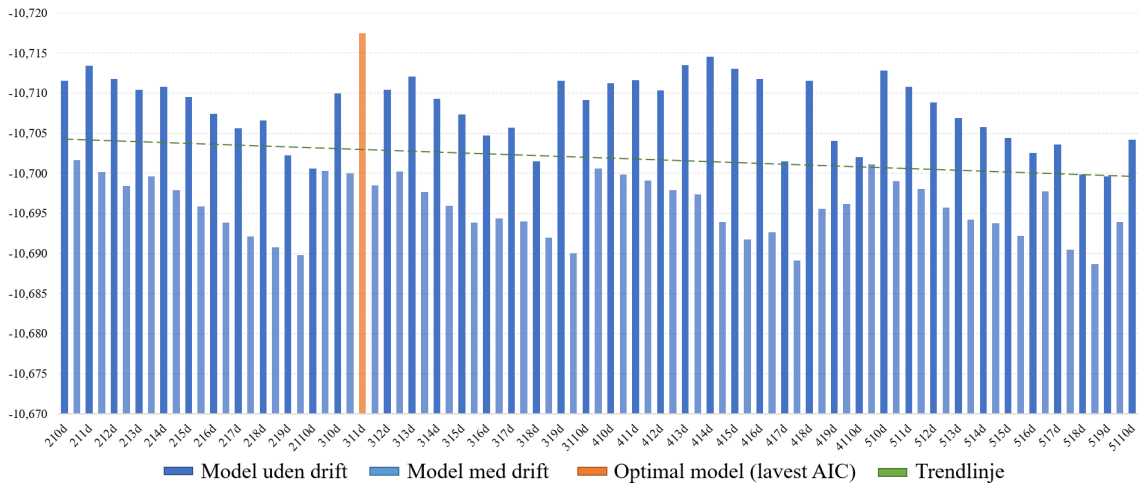
Ovenstående illustrerer tidsperioden for specialets data, hvor træningsdata er givet ved den grå linje, imens den blå linje angiver testdata. Endvidere anvendes 20 % af træningsdata som valideringsdata til estimering af specialets LSTM modeller. Perioden er valgt med henblik på at opnå det højest mulige antal observationer i forbindelse med tilgængeligt data, hvorfor der inkluderes 2518 observationer for hver aktie. Valget om at opnå det højst mulige antal observationer, er grundet en konsensus i litteraturen om, at et højt antal observationer er en vigtig faktor for succes af machine learing modeller (Maheswari, 2019).

7.1 Udvælgelse af ARIMA

Specialets lineære ARIMA modeller udvælges med udgangspunkt i Akaike Information Criteria, AIC. AIC beregnes for en række modeller, udvalgt med afsæt i ACF og PACF. Disse benyttes som et pejlemærke til at bestemme antallet af parametre p og q, der forekommer signifikante, og klarlægger herved den øvre grænse for, hvor mange lags der testes for i specialet (Newbold, 1975). Det fremgår i bilag B3, at op til 10 lags er signifikante på et 10% signifikansniveau. Herved testes alle kombinationer af p og q op til 10 lags, med og uden drift, hvilket medfører test af 240 individuelle modeller for hver aktie. Herunder udvælges den ARIMA model, som har den laveste AIC værdi.

Figur 7.2 illustrerer et udpluk af specialets estimerede AIC værdier for Visa, med henblik på at visualisere, hvorledes AIC ændres i takt med kompleksiteten i modellen.

Figur: 7.2 Illustration af AIC for Visa Inc.



Figuren er udarbejdet i Excel, hvor x-aksen angiver modellen og y-aksen AIC. Y-aksen er omvendt, så de øverste værdier tilsvarer de laveste AIC. AIC illustreres for alle modeller mellem ARIMA(2,1,0) og (5,1,10) med og uden drift for Visa Inc. Den orange søjle angiver den bedste ARIMA model givet ved (3,1,1) med drift.

Den optimale ARIMA model illustreres ved den orange søjle og er givet ved ARIMA(3,1,1) med drift. Det fremgår, at AIC indledningsvist er faldende, hvorved kompleksitet i modellen medvirker til en øget forklaringsgrad. Øges kompleksiteten yderligere, vil AIC dog

stige i takt med kompleksiteten, illustreret ved trendlinjen. Dette er i overensstemmelse med, at eksisterende litteratur finder, at mindre komplekse modeller ofte vil opnå en bedre performance. Endvidere har modellerne uden drift en højere AIC end modellerne med drift, hvilket implicerer, at Visa er præget af trend. Modeller, der inddrager denne komponent, vil herved bedre estimere det faktiske data og opnå en lavere AIC. AIC værdierne for alle estimerede modeller angives i bilag B9, hvortil de udvalgte modeller angives i tabel 7.1[15].

Tabel: 7.1 ARIMA - Modeludvægelse

Aktie	ARIMA	Aktie	ARIMA
Alphabet Inc. A	ARIMA(6,1,7), drift	Alphabet Inc. C	ARIMA(4,1,8), drift
Amazon.com Inc.	ARIMA(3,1,3), drift	Tesla Inc	ARIMA(4,1,2)
Home Depot Inc.	ARIMA(4,1,3), drift	B. Hathaway	ARIMA(9,1,5), drift
JP Morgan	ARIMA(2,1,4), drift	Bank of America	ARIMA(2,1,9), drift
U. Health Group	ARIMA(6,1,3), drift	Merck & Co.	ARIMA(6,1,4), drift
Abbvie Inc.	ARIMA(9,1,7), drift	Eli Lilly	ARIMA(5,1,4), drift
U. Parcel Service	ARIMA(8,1,7)	Caterpillar Inc.	ARIMA(9,1,8), drift
Apple Inc.	ARIMA(7,1,7), drift	Microsoft Corp	ARIMA(4,1,4), drift
Nvidia Corp	ARIMA(5,1,7), drift	Broadcom Inc.	ARIMA(3,1,7), drift
Visa Inc. A	ARIMA(3,1,1), drift	Mastercard Inc.	ARIMA(1,1,1), drift
Udarbejdet i RStudio			

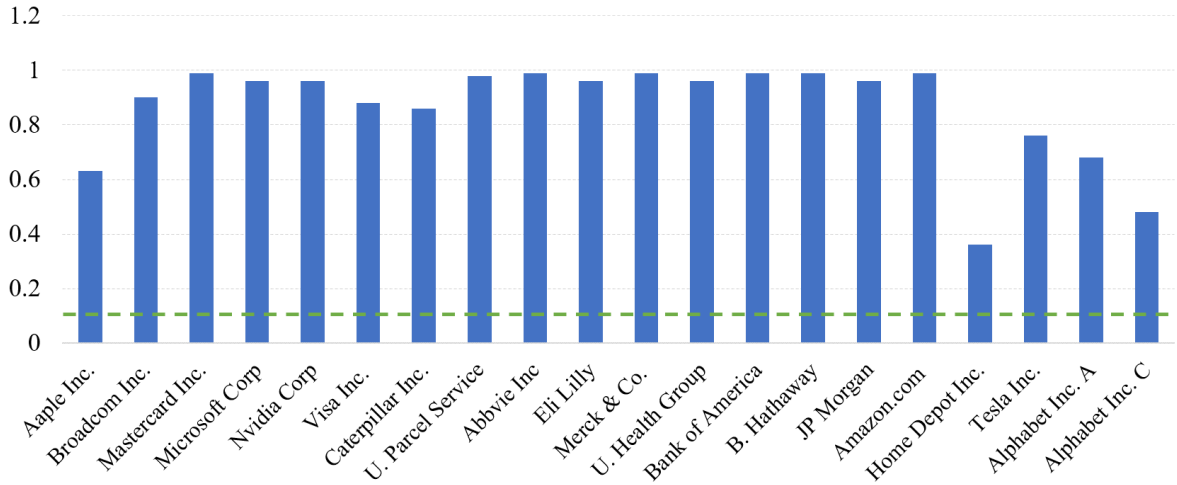
Tabellen angiver den bedste ARIMA model for hver aktie. De tilhørende AIC værdier illustreres i bilag B3.

I forbindelse med modeludvælgelsen er en random walk, ARIMA(0,1,0), også testet for hver aktie. Som angivet i tabel 7.1 kan resultaterne ikke understøtte random walk hypotesen, eftersom én ARIMA(0,1,0) ikke forekommer bedst for nogen af specialets aktier. Alle

[15] Ved estimering af ARIMA modeller anvender specialet ARIMA() funktionen fra stats pakken i RStudio. Herunder anvendes maximum likelihood estimation til estimering af modellens parametre, hvorfor der testes og udarbejdes AIC for alle modeller med parametre, som resulterer i en stationær ARIMA model.

ARIMA modellerne benytter en Box Cox transformation til at reducere anomalier, herunder manglende homoskedasticitet, additivitet og normalitet. For at undersøge validiteten af de udvalgte modeller, udføres diagnostisk. Herunder testes modellerne for unit root, hvor testen og resultaterne fremgår i bilag B3. Det kan heraf konkluderes, at ingen af modellerne indeholder unit root, da alle inverse rødder ligger indenfor enhedscirklen. Modellerne testes endvidere for autokorrelation i en Ljung Box, hvor p-værdierne illustreres i figur 7.3[16].

Figur: 7.3 Ljung Box p-værdier



Figuren illustrerer p-værdierne ved test af autokorrelation i en Ljung box. Den grønne linje angiver hertil et signifikansniveau på 10%. Figuren angiver, at det på et 10% signifikansniveau kan konkluderes, at der ikke foreligger autokorrelation i nogle af specialets modeller.

Ud fra ovenstående kan det på et 10% signifikansniveau konkluderes, at der ikke foreligger autokorrelation i nogle af specialets modeller. Hertil undersøges det, hvorvidt de residuale følger en normalfordeling i bilag B3. Det kan ikke med udgangspunkt i Jarque-Bera testen endegyldigt vurderes, at de residuale er normalfordelte, hvor en visualisering af de residuale dog antyder, at de tilnærmelsesvist følger en normalfordeling. Fordelingen kan herved stadig påvirke de endelige resultater i forbindelse med forecasting. Ovenstående resultater tilsidesættes midlertidigt, med henblik på en gennemgang af specialets LSTM modeller. De udarbejdede forecasts sammenfattes efterfølgende i en analyse af ARIMA og LSTM.

[16]Det er essentielt, at modellerne ikke indeholder autokorrelation, da dette medfører, at OLS ofte vil underestimere variansen af parametrene, og herved øge sandsynligheden for at vurdere insignifikante parametre signifikante.

7.2 Udvælgelse af LSTM

Specialets LSTM modeller udvælges med udgangspunkt i mean squared error, MSE[17]. Specialet tager afsæt i én LSTM model for hver aktie, hvor der forecastes daglige ændringer i kurspoint ved et one-step-ahead forecast[18]. Design af et deep neural network kan anses som værende en arbitrær og iterativ disciplin. Træning af LSTM modeller kan være særligt krævende, grundet den ofte store mængde data, omfattende beregninger og parameterjusteringer. Modellens netværksstruktur skal hertil justeres og optimeres for at opnå det bedste resultat, hvilket kræver både tid og omfattende beregningsressourcer. Herved fremgår LSTM modeller ofte mere krævende end ARIMA modeller, grundet den øgede kompleksitet. Dette er i forbindelse med estimering af lang hukommelseshistorie, flere justerbare elementer og tusindvis af parameterjusteringer, som kræver omfattende matriceoperationer og beregninger (Moolayil, 2019). Dette medvirker til, at der for de udarbejdede LSTM modeller må tages forbehold for en begrænset computerkraft og tid. Herved kan en test af alle mulige kombinationer for modellens arkitektur udelukkes. Modellen opstilles i stedet på baggrund af en balance mellem performance og evnen til hurtigt at lære ved justering af parametre. Specialet tager afsæt i den generaliserede guideline fra *"Learn Keras for deep neural Networks - A fast-track approach to modern deep learning with python"* af Jojo Moolayil i 2019, til udvælgelse og design af forskellige netværk (Moolayil, 2019).

Med henblik på at opnå den bedste forecastperformance for specialets LSTM modeller, optimeres modellernes hyperparametre. Hyperparametre angiver de egenskaber, som navigerer processen under træning af modellen og angiver herved strukturen for netværket. Disse forekommer derfor essentielle i design af en LSTM model, da de direkte kontrollerer

[17] Ved estimering af LSTM modeller anvender specialet keras pakken i RStudio.

[18] LSTM forecaster daglige ændringer, for at tage udgangspunkt i stationært data. Principielt er stationaritet ikke nødvendigt, men kan medvirke til en øget forecast performance, og gør det samtidig lettere at træne det neurale netværk (Li, 2019).

adfærden i træningsalgoritmen og dermed modellens performance. Hyperparamentre adskiller sig fra modellens parametre, der beskriver de egenskaber, som justeres og optimeres gennem træningen af modellen, herunder vægte og bias. I kontrast hertil kan modellens hyperparametre ikke tillæres, men defineres i stedet under konstruktionen af modellen og holdes konstant under træningen. Hyperparametre kan hertil grupperes baseret på deres funktion og, hvilke aspekter i modellen som de påvirker. Arkitekturparametre omfatter de parametre, som er afgørende for kompleksiteten i modellen, herunder antal layers og neuroner i hvert layer. Træningsparametre angiver hertil de parametre, der er afgørende for træningsprocessen, herunder læringsrate, epochs og batches (Moolayil, 2019).

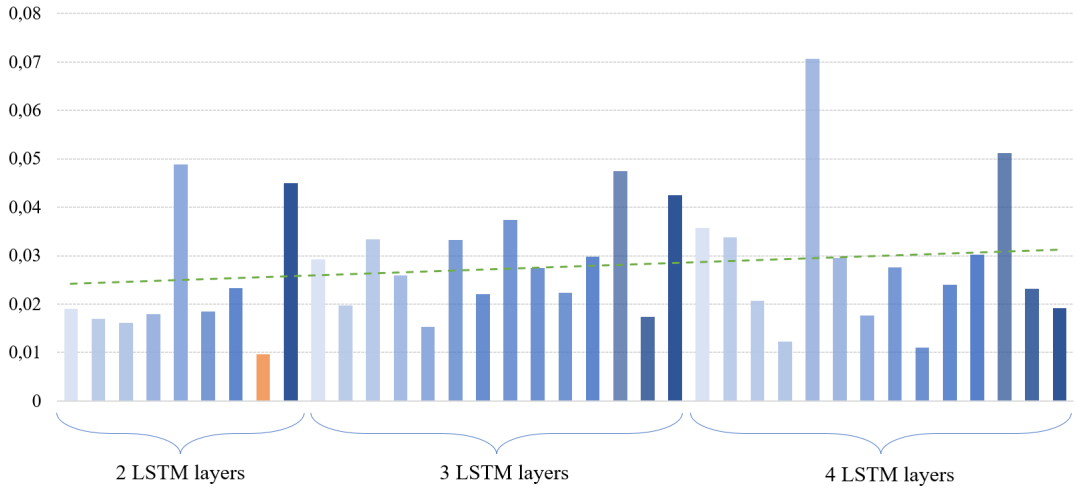
7.2.1 Neuroner og layers

Antallet af neuroner i modellens hidden layers tager i specialet altid en værdi, som er i potens af to, da træningsalgoritmen lettere kan bearbejde data i denne form[19]. Herved testes forskellige kombinationer af antal neuroner for værdier i potens af to i intervallet [8;128], hvor antallet i efterfølgende layers holdes konstant eller mindskes (Moolayil, 2019).
Antallet af layers i en LSTM model har signifikant betydning for modellens forecastperformance, hvor tilføjelsen af flere layers i et netværk generelt medvirker til en bedre performance. Ved inddragelse af flere layers, øges kompleksiteten og træningstiden dog markant, hvortil mange layers vil nødvendiggøre et højt antal epochs for at opnå lovende resultater. Det fremgår i litteraturen, at den øgede kompleksitet kan medføre, at marginaleffekten falder eller bliver negativ for modeller med mere end tre layers. Specialet tester derfor modeller med henholdsvis to, tre og fire layers. Hertil inkluderes et dense layer mellem, hvert LSTM layer, som har til formål at forbinde alle relevante informationer i modellen, hvor units i specialet er fastsat til 32 for alle modeller. For at visualisere udviklingen i MSE, ved

[19]Det forekommer mere beregningsmæssigt effektivt at anvende antal neuroner i potens af to, grundet måden, computerhukommelse er organiseret på. Mange moderne computeres hardware bruger binær repræsentation og allokerer hukommelsesressourcer i blokke, der er baseret på potenser af to.

ændringer i antal layers og units, illustrerer figur 7.4 MSE for en række udvalgte modeller estimeret for Visa. MSE angives for henholdsvis to, tre og fire layers, hvortil units varierer mellem 8 og 128. En mørkere søjle indikerer flere inkluderede units i hvert LSTM layer.

Figur. 7.4 MSE ved forskellige antal layers og units



Figuren er udarbejdet i Excel. Modellen har for alle layers og units en drop-out rate på 0,2 og batch size på 16. Dette er for at visualisere betydningen af en ændring i layers og units. Det angives i bilag B10, at resultaterne vil være lignende ved enhver anden værdi for modellens resterende hyperparametre. Den orange søjle angiver den model med den laveste MSE.

Ovenstående trendlinje angiver, at modeller med 2 LSTM layers generelt opnår de laveste værdier af MSE, hvor modellen med den laveste MSE er givet ved den orange søjle. Det fremgår endvidere med udgangspunkt i de enkelte modeller, at der forekommer stor variation i fejlene, når antal layers holdes konstant. Dette kan skyldes en tilfældighed i LSTM, som medfører, at modeller, der har tilnærmelsesvist samme arkitektur kan opnå meget forskellige resultater. Dette fremgår ved 4 LSTM layers, hvor flere modeller har de største fejl, imens nogle modeller opnår fejl, som er mindre relativt til en række modeller med færre layers. Ovenstående tilfældighed vil i LSTM medvirke til, at selvom kompleksitet ikke generelt medvirker til en bedre performance, kan det ikke udelukkes, at én mere kompleks model og én given træning, i højere grad estimerer det faktiske data. Der bør derfor tages højde for dette ved en eventuel udelukkelse af test af diverse hyperparametre.

Der anvendes en ReLU aktiveringsfunktion i modellens hidden layers, hvor der i output

layer anvendes en lineær aktiveringsfunktion. Herved er det som tidligere nævnt muligt for modellen at bearbejde komplekse ikke-lineære mønstre, hvortil det endelige output kan tage enhver værdi (Ghatak, 2019, kap. 1 s. 10).

7.2.2 Epochs

Det optimale antal epochs vælges med udgangspunkt i tabsfunktionen. Det er gældende, at antal epochs øges op til punktet, hvor tabsfunktionen ikke længere er faldende. I tilfælde af, at der anvendes for få epochs, forekommer det muligt at opnå en bedre model ved at øge antallet. For mange epochs vil modsat kunne medføre overfitting. En trial and error tilgang kan være særligt krævende, hvortil det optimale antal epochs varierer for enhver given model. Specialet anvender derfor callbacks, til at stoppe træningen af netværket på et passende tidspunkt relativt til tabsfunktionen (Rai, 2019). Callbacks stopper modellen ved det punkt, hvor tabsfunktionen ikke har været faldene for en række epochs. I specialet anvendes en 'patience' på 10 epochs, hvilket angiver, at modellen ikke kører flere epochs, såfremt tabsfunktionen ikke har været faldende i de sidste 10 epochs[20].

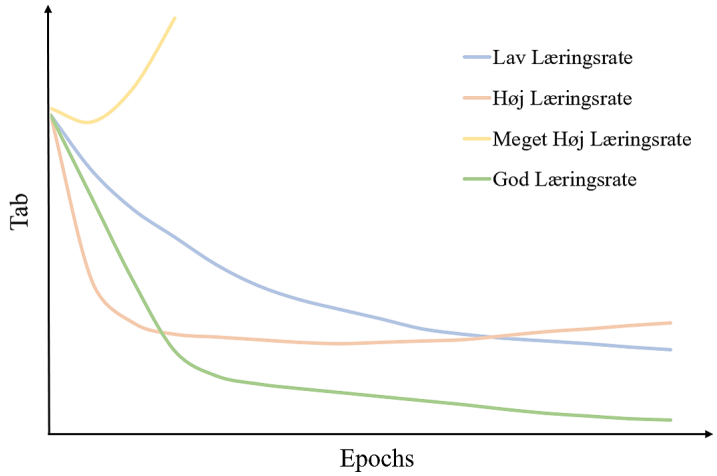
7.2.3 Læringsrate og regulering

Ovenstående hyperparametre tilpasses og justeres med henblik på at forbedre de udvalgte modellers performance og skaber herved strukturen for læringsprocessen. I denne proces, indsamler modellen information og lærer systematiske mønstre i det udvalgte data. Herunder opdateres modellens vægte og struktur løbende for bedst muligt at tilpasse modellen. Læringsraten angiver herunder en hyperparameter, som kontrollerer vægtene i netværket med hensyn til tabsgradienten. Den bestemmer, hvor hurtigt eller langsomt, modellen bevæger sig mod de optimale vægte. Figur 7.5 illustrerer en sammenligning af ændringen i

[20]Det er ikke altid tilfældet, at en model ikke vil forbedres, fordi tabsfunktionen for én epochs ikke mindskes, hvorfor specialet sætter patience til 10. Dette er udvalgt med udgangspunkt i en balance i at opnå den bedst mulige model og undgå overfitting.

tabsfunktionen i forbindelse med ændringer i antal epochs ved forskellige læringsrater.

Figur: 7.5 Læringsrater (Rakhecha, 2010)



Figuren er udarbejdet i Excel og tager afsæt i understanding learning rate (Rakhecha, 2010). Herunder illustreres ændringen i tabsraten i takt med ændringen i epochs for forskellige læringsrater.

Det fremgår, at en lav læringsrate medfører, at modellen ved et lokalt minimum vil tage længere tid om at konvergere. Det ønskes derfor at opnå en læringsrate, som både lærer relativt hurtigt og samtidig konvergerer mod et optimalt minimum. Dette er ensbetydende med, at tabsfunktionen bør mindskes i passende rate, mens antallet af epochs stiger. Det fremgår af bilag B4, at alle specialets LSTM modeller har en høj- til optimal læringskurve, da tabsfunktionerne forekommer meget faldende ved de første epochs, for efterfølgende at stabiliseres ved et forskelligt antal epochs.

Modellen kan ligeledes træne dataet for godt, også kendt som overfitting. Dette forekommer når detaljer og støj i data fejlagtigt bliver inkorporeret som mønstre og bekendte komponenter i modellen. Støj refererer til mønstre i data, som opstår grundet tilfældighed, og derved ikke er repræsentativt for den reelle udvikling. Dette er problematisk i tilfælde af, at modellen ikke er i stand til at skelne mellem reelle indsigter og støj, og derfor tilpasser vægte for at imødekomme det nye mønster, som i virkeligheden er støj. Dette vil have en negativ effekt på modellens evne til at generalisere data. For at reducere og undgå overfitting anvender specialet regulering, ved implementering af en drop-out mekanisme. Spe-

cialet anvender derfor et drop-out layer ved hvert LSTM layer, for at forhindre overfitting og mindske beregningskrav ved deaktivering af tilfældigt udvalgte neuroner under træning af netværket. Drop-out rate inkluderes på henholdsvis 20% og 30%, som er et anerkendt kompromis mellem at opnå en høj nøjagtighed og undgå overfitting (Chowdhury, 2021).

7.2.4 Batch Size

Specialets batch size er på henholdsvis 16, 32 og 64. Disse er bredt accepteret og anvendt, da de ofte vil medvirke til en bedre læringsproces og herved mere glat læringskurve for netværket (Moolayil, 2019). Ovenstående tuning af modellens hyperparamentre, resulterer i 516 forskellige LSTM modeller, for hver af specialets bæredygtige aktier. Modellerne sammenlignes med udgangspunkt i mean squared error for validerings- og træningsdata, hvor den bedste model for hver aktie udvælges. Figur 7.6 angiver den udvalgte LSTM model for Visa Inc. og illustrerer herved opsætningen og flowet for det udarbejdede netværk.

Figur. 7.6 Opbygning af LSTM Visa Inc.

Figuren illustrerer opbygningen af den estimerede optimale LSTM model for Visa Inc. fra input til output. Modellen består af to LSTM layers, med 128 units i det første layer og 64 i det andet. Hertil inddrages en drop-out rate på 0,2 og et dense layer mellem de to LSTM layers.

Det fremgår af ovenstående, at modellen tager afsæt i to LSTM layers, hvor antal units er givet ved 128 og 64. Hertil inddrages en drop-out rate på 0,2, som anvendes mellem hvert layer i modellen. Herved bearbejder modellen relevant input til ét output, som i specialet for hver periode tilsvarer det daglige afkast for Visa Inc.

Specialets udvalgte LSTM modeller angives i tabel 7.2. Størstedelen af modellerne tager afsæt i to eller tre layers, hvortil der forekommer en del variation i units og batch size for hver model. Det fremgår for alle modeller optimalt at anvende en drop-out rate på 0,2%, og at inkludere ét timestep til forecasting af den efterfølgende periode. Herved er specialets resultater overordnet i overensstemmelse med eksisterende litteratur. Dette i forbindelse

med, at det kun for Microsoft er fordelagtigt at inkludere fire layers, hvortil drop-out raten tilsvarer den drop-out rate, som størstedelen af eksisterende litteratur ligeledes anvender.

Tabel: 7.2 LSTM- Modeludvægelse

Aktie	Layers	Units	Batch size	Drop-out
Alphabet Inc. A	3	64, 64, 32	64	0,2
Alphabet Inc. C	3	64, 32, 32	32	0,2
Amazon.com	2	128, 128	64	0,2
Home Deposit Inc.	3	128, 128, 64	16	0,2
Tesla Inc.	3	64, 32, 32	64	0,2
Bank of America	3	32, 32, 16	64	0,2
B. Hathaway	3	128, 128, 64	16	0,2
JP Morgan	3	16, 16, 8	64	0,2
Abbvie Inc.	2	128, 128	16	0,2
Eli Lilly	3	16, 16, 16	64	0,2
Merck & Co.	2	64, 32	32	0,2
U. Health Group	4	128, 128, 128, 64	32	0,2
Caterpillar Inc	3	128, 64, 64	64	0,2
U. Parcel Service	2	32, 16	64	0,2
Broadcom Inc.	2	64, 64	32	0,2
Mastercard Inc.	2	64, 64	64	0,2
Microsoft Corp	4	16, 16, 16, 16	32	0,2
Nvidia Corp	2	64, 64	32	0,2
Visa Inc.	2	128, 64	16	0,2
Apple Inc.	2	128, 64	16	0,2
Udarbejdet i RStudio				

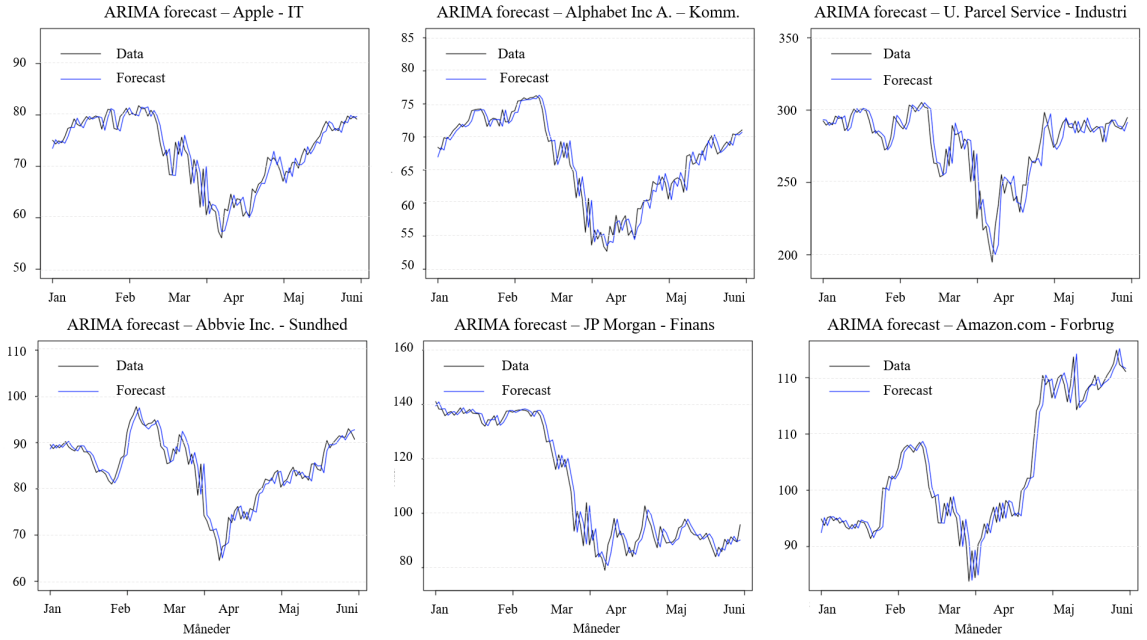
Tabellen angiver specialets optimale LSTM modeller. Herunder angives den enkelte models opbygning i forbindelse med antal layers, units, batch size og drop out rate.

8 Forecast og evaluering af ARIMA og LSTM

Følgende afsnit har til formål at sammenligne og evaluere ARIMA og LSTM modellernes forecastperformance, hvortil en random walk inddrages for at vurdere, hvorvidt modellerne er i stand til at forecaste specialets aktier. Indledningsvist illustreres de udarbejdede forecasts for at klarlægge generelle tendenser, som skaber kontekst i forbindelse med tilhørende forecastsfejl. Endvidere anvendes evalueringsmetoderne beskrevet i *4.4 Evalueringsmetoder - Forecast* til at vurdere ARIMA og LSTM modellernes forecastperformance.

Figur 8.1 illustrerer et ARIMA forecast, for hver af de inkluderede sektorer, op imod de faktiske kurser i perioden primo til medio 2020[21].

Figur 8.1 Forecast - ARIMA



Figuren er udarbejdet i RStudio og illustrerer aktieafkast og ARIMA forecasts fra primo til medio 2020. Her angiver x-aksen tiden og y-aksen afkastet, som varierer for de individuelle forecasts.

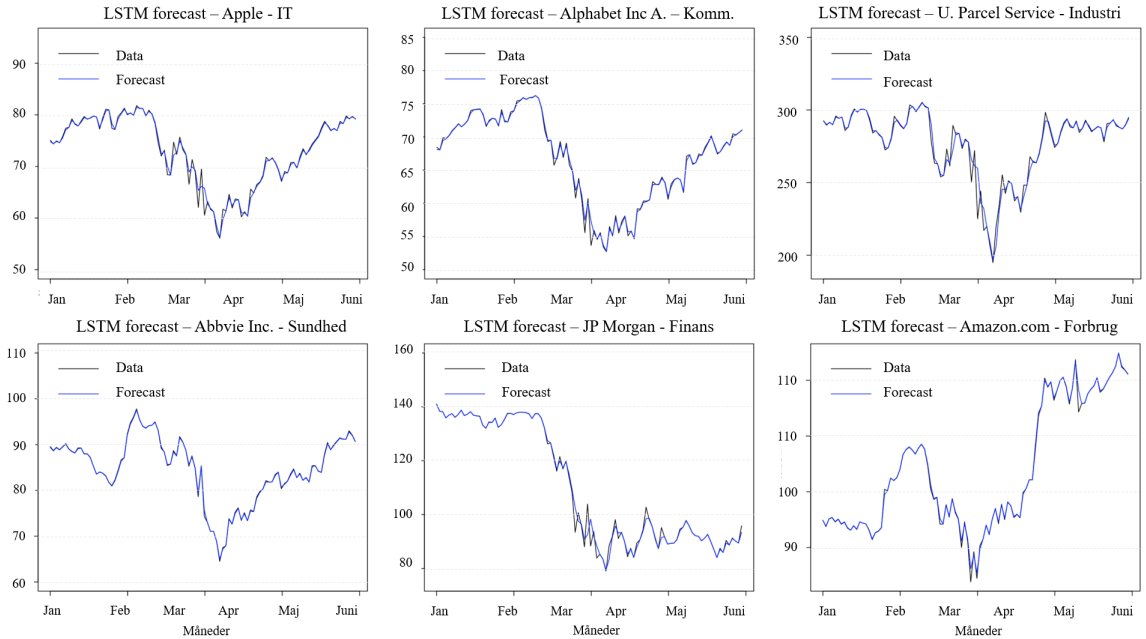
Det fremgår af ovenstående, at valget om at benytte one-step ahead forecast har stor betydning for præcisionen af de udarbejde forecasts. Dette er eftersom udsvingene umiddelbart

[21] Figur 8.1 illustrerer ikke hele den forcastede periode fra primo 2020 til ultimo 2022, da det for perioden er svært at vurdere forecast og faktisk data op mod hinanden. Derfor er en kortere periode valgt som visualisering.

fanges en periode forsinket, hvilket ikke er overraskende da modellen udelukkende tager afsæt i tidligere information samt kender den faktiske kurs i sidste periode, y_{t-1}. Forsinkelsen i de forecastede udsving kan være problematisk i meget volatile perioder eller ved forecasting af volatile aktier. Dette kan medvirke til, at forecast og faktisk data bevæger sig modsatrettet, hvoraf forventningerne til ARIMA i en mean-variance nedsættes.

For at sammenligne ARIMA og LSTM illustrerer figur 8.2 de udvalgte LSTM forecasts, for perioden og aktierne, tilsvarende figur 8.1. I forbindelse med sammenligningen af ARIMA og LSTM´s fejl og forecast konverteres de forecastede værdier i LSTM til kurser. Det fremgår, at alle specialets modeller i høj grad er i stand til at fange udsvingene i de bæredygtige aktier, hvor LSTM modellerne, relativt til ARIMA, er i stand til at fange udsvingene i de faktiske frem for forsinkede perioder.

Figur: 8.2 Forecast - LSTM

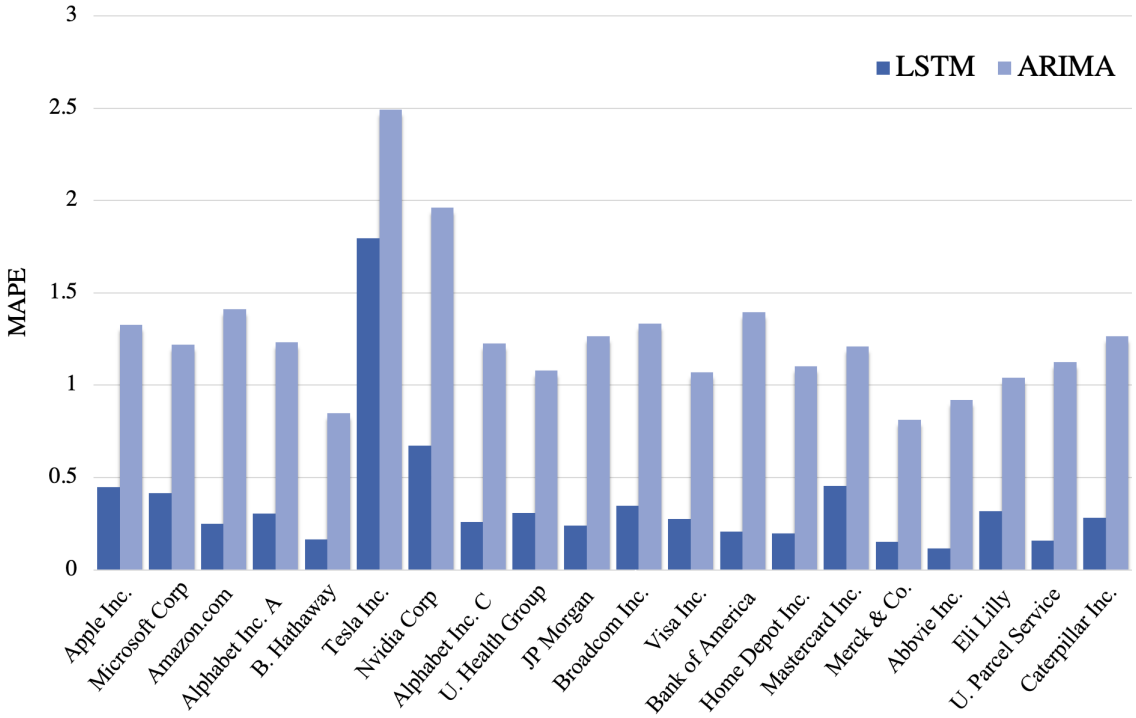


Figuren er udarbejdet i RStudio og illustrerer aktieafkast og LSTM forecasts fra primo til medio 2020. Her angiver x-aksen tiden og y-aksen afkastet, som varierer for de individuelle forecasts.

Det kan ikke med udgangspunkt i figur 8.1 og 8.2 endegyldigt vurderes, hvorvidt ARIMA eller LSTM har den relativt bedste forecastperformance. Det vurderes derfor relevant at inkludere en række evalueringsmetoder for at præcisere forecastnøjagtigheden.

Figur 8.3 illustrerer MAPE for de enkelte aktier for ARIMA og LSTM. MAPE angiver gennemsnittet af de absolutte forskelle mellem faktiske og forecastede værdier i procent af de faktiske værdier. Det fremgår herunder i eksisterende litteratur, at en MAPE på under 5% implicerer en høj og acceptabel forecastperformance (J. Green og Newman, 2017).

Figur: 8.3 MAPE - ARIMA & LSTM



Figuren er udarbejdet i Excel, hvor x-aksen angiver den enkelte aktie og y-aksen MAPE. De mørke søjler angiver ARIMA og de lyse LSTM. LSTM opnår for alle aktier lavere procentvise fejl end ARIMA.

Det fremgår af ovenstående, at alle specialets forecastmodeller opnår en MAPE på under 5% og herved vurderes acceptable i henhold til forecastperformance. Dette forekommer ikke overraskende, da specialet udelukkende anvender one-step-ahead forecast. Forecastet for enhver periode er herved opdateret i henhold til den forsinkede periode y_{t-1}, hvortil forecasthorisonten altid er givet ved én periode. Dette vil alt andet lige ofte resultere i re-

lativt præcise forecasts. Det angives endvidere, at LSTM for alle specialets aktier opnår en lavere MAPE end ARIMA. Figuren illustrerer afslutningsvist, at fejlene i høj grad varierer mellem de enkelte aktier, hvortil der forekommer en tendens til, at både ARIMA og LSTM opnår henholdsvis høje eller lave fejl ved de samme aktier. En årsag til, at ARIMA og LSTM opnår markant højere forecastfejl ved eksempelvis Tesla kan være, at den historiske udvikling ikke er i overensstemmelse med udviklingen i specialets udvalgte test periode, som angivet i bilag B2. Dette er gældende for størstedelen af specialets aktier, hvor den historiske afvigelse er større for de aktier, som har relativt større forecastfejl.
Tabel 8.1 angiver hertil et generelt overblik af forecastfejlene, givet ved RMSE, MAE og MAPE. Herunder sammenfattes fejlene i henhold til middelværdien, medianen, minimum- samt maksimumsværdien for ARIMA, LSTM og differencen herimellem.

Tabel: 8.1 Forecastfejl - ARIMA og LSTM

Model	Middel	Median	Minimum	Maksimum
ARIMA RMSE	3,61	3,24	0,66	8,26
LSTM RMSE	1,60	1,00	0,19	7,01
Difference RMSE	**2,01**	**2,24**	**0,47**	**1,25**
ARIMA MAE	2,28	2,06	0,42	5,46
LSTM MAE	0,75	0,51	0,06	3,87
Difference MAE	**1,52**	**1,55**	**0,36**	**1,60**
ARIMA MAPE	1,27	1,22	0,81	2,49
LSTM MAPE	0,37	0,28	0,12	1,80
Difference MAPE	**0,90**	**0,94**	**0,70**	**0,70**
Udarbejdet i R-Studio				

Tabellen angiver middelværdien, medianen, maksimum og minimum af forecastfejlene for ARIMA og LSTM. LSTM opnår generelt de laveste forecastfejl ved både RMSE, MAE og MAPE.

Der fremgår en tendens til, at ARIMA har større forecastfejl end LSTM, da alle ovenstående værdier er mindre for LSTM. Det fremgår for begge modeller, at middelværdien er højere end medianen, hvilket implicerer, at få modeller opnår væsentlig større forecastfejl relativt til de resterende modeller. Dette illustreres i figur 8.2 ved, at eksempelvis Tesla opnår væsentlig større forecastfejl end de resterende modeller. Endvidere er forholdet mellem MAE og RMSE ligeledes højere for ARIMA, hvilket implicerer en større varians i fejlene. Med henblik på at relatere ovenstående til bæredygtighedselementet, angives de gennemsnitlige MAPE værdier for hver ESG-rating i tabel 8.2.

Tabel: 8.2 Gennemsnitlig MAPE - ESG-rating

ESG-rating	"A"rating	"AA"rating	"AAA"rating
LSTM	0,43	0,26	0,29
ARIMA	1,32	1,16	1,24
Udarbejdet i RStudio			

Tabellen angiver den gennemsnitlige MAPE for alle ARIMA- og LSTM modeller opdelt for hver ESG-rating.

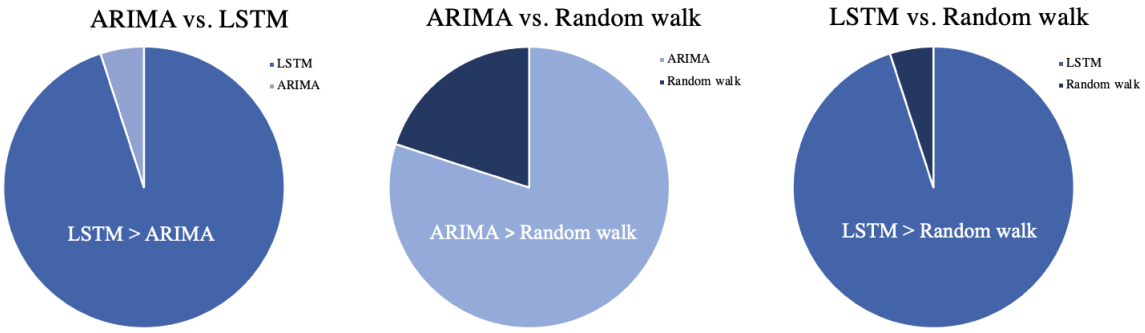
Det fremgår af ovenstående, at specialets aktier med ESG-ratingen "AA"opnår de laveste gennemsnitlige MAPE værdier for både ARIMA og LSTM. Dette implicerer, at modellerne opnår den bedste forecastperformance for aktierne herunder. Hertil opnår "AAA"aktierne lavere gennemsnitlige fejl end "A"ratede aktier. Det kan herved med afsæt i forecastfejlene anskues mere sikkert for en investor at anvende forecast af mere bæredygtige aktier til investeringsbeslutninger. LSTM opnår hertil de laveste MAPE for alle ESG-ratings.

Ovenstående resultater implicerer hermed, at LSTM er mere fordelagtig til forecasting af de bæredygtige aktier. Dette kan være grundet, at ARIMA antager, at forholdet mellem den afhængige og uafhængige variabel er lineært, samt at standardafvigelsen i fejlleddet er konstant. Forholdet i data er dog i praksis mere komplekst end teorien antager. Herunder

kan ikke-lineære mønstre i data, samt den øgede volatilitet i forecastperioden, som illustreret i *6.2 Deskriptiv statistik* i figur 6.4, have betydning for ARIMA modellernes mindre nøjagtighed. Modsat tager LSTM højde for afhængighedsproblemer på tværs af tidsseriens observationer. Dette er eftersom modellen indeholder en læringsalgoritme, benævnt *stokastisk gradient descent*, som tidligere klarlagt i *4.3 Teoretisk gennemgang af LSTM*. Dette betyder, at LSTM kan selektere i tilfælde af støjende inputsekvenser, som ofte præger aktiekurser og samtidig bearbejder ikke-lineære mønstre. Det kan hermed tyde på, at bæredygtige aktier indeholder ikke-lineære mønstre, som LSTM i højere grad formår at fange.

For at klarlægge, hvorvidt forskellen mellem modellernes forecastperformance forekommer signifikant, inddrager specialet en Diebold-Mariano test. Denne anvendes for at sammenligne ARIMA og LSTM, hvortil modellerne testes op mod en random walk. Resultaterne er angivet i bilag B5, hvortil figur 8.4 illustrerer resultaternes overordnede tendenser.

Figur: 8.4 Diebold-Mariano test



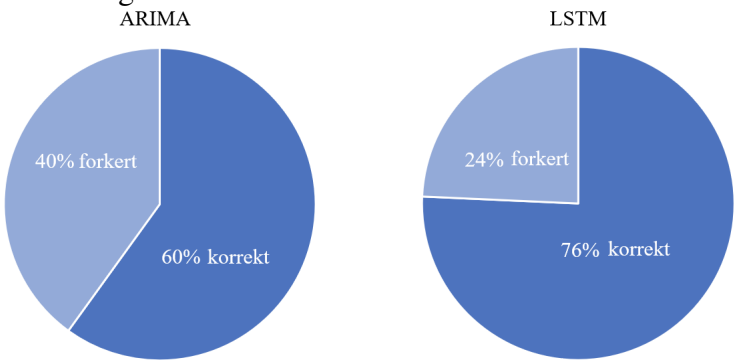
Figuren illustrerer tendensen i resultaterne af Mariano-Diebold testen. Med afsæt i forskellige signifikansniveauer, angivet i bilag B5, fremgår det, at LSTM for størstedelen af specialets aktier er signifikant mere nøjagtig end ARIMA. Begge modeller er hertil generelt mere nøjagtige end en random walk.

Det fremgår i figuren, at alle modellernes nøjagtighed er signifikant forskellige, hvortil det på forskellige signifikansniveauer fremgår, at den ene model forekommer mere nøjagtig. Som angivet i figuren til venstre, er der en tendens til, at LSTM er signifikant mere nøjagtig end ARIMA. Figuren i midten og til højre illustrerer hertil, at både ARIMA og LSTM, for

størstedelen af specialets aktier, er overlegen i henhold til en random walk.

Ovenstående klarlægger ARIMA og LSTM modellernes forecastperformance med afsæt i en række evalueringsmetoder, som angiver forskellene i de forecastede og faktiske værdier. Det fremgår, at LSTM for alle specialets aktier har en højere forecastperformance, og herved i højere grad tilnærmes det faktiske data. Forecastfejlene tager ikke forbehold for forskellen i den forecastede og faktiske retning af afkastet. Dette er af betydning, da specialet anvender de udarbejdede forecasts med henblik på at maksimere profitten i mean-variance optimerede porteføljer. I tilfælde af, at retningen af afkastet forecastes forkert kan der forekomme negative afkast. Dette kan eksempelvis komme til udtryk, hvis det forecastede afkast er positivt, imens det faktiske afkast er negativt. I dette tilfælde, vil aktien vægtes positivt i den samlede portefølje, trods et faktisk fald i kursen, hvilket vil medføre et fald i porteføljeværdien. Det vurderes derfor relevant at undersøge, hvorvidt de udarbejdede forecasts tenderer til at følge retningen af de faktiske værdier. Herved illustreres succesraten for korrekt forecastede retninger i figur 8.5.

Figur: 8.5 Succesrate - ARIMA & LSTM

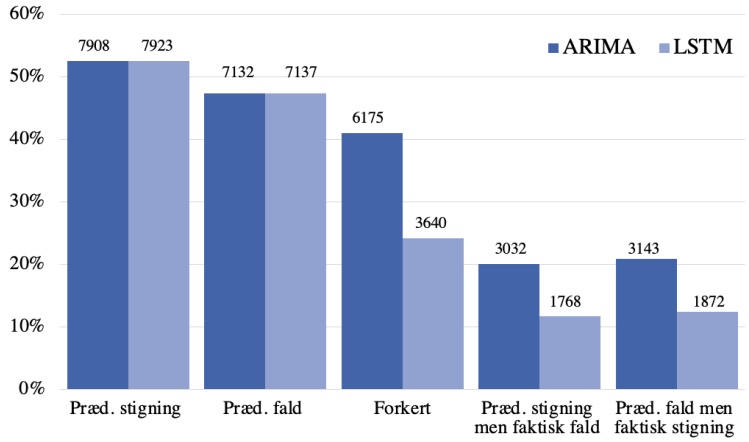


Figuren illustrerer den procentvise succesrate, hvor modellen formår at prædiktere retningen i afkastet korrekt. Den mørkeblå angiver andel korrekte prædikteringer, mens den lyse angiver forkerte.

Det fremgår af ovenstående, at både ARIMA og LSTM formår at forecaste retningen af afkastet korrekt for mere end halvdelen af dagene. Herunder fremgår det, at ARIMA har en succesrate på 60%, hvortil LSTM opnår en succesrate på 76%. Herved fremstår begge mo-

deller bedre i henhold til, hvis specialet havde anvendt en såkaldt "blindfolded monkey", til at angive, hvorvidt kurserne vil stige eller falde. Ovenstående resultater, i forbindelse med, at ARIMA har en lavere succesrate, er i overensstemmelse med figur 8.1, hvor det fremgår at ARIMA, vil tendere til at være én periode forsinket relativt til de faktiske værdier.
Med henblik på at undersøge, hvorvidt modellerne performer bedst på et bear eller bull marked, illustreres de forecastede fald og stigninger i figur 8.6. Forecastede stigninger, ved faktiske fald og forecastede fald ved faktiske stigninger fremgår hertil til højre i figuren.

Figur: 8.6 Forecastede retninger



Figuren er udarbejdet i Excel og angiver forecastede stigninger og fald. Hertil illustreres antal forkerte prædikteringer af retningen, samt prædikterede stigninger ved faktiske fald og vice versa. Der forekommer i investeringsperioden 8017 faktiske stigninger og 7025 faktiske fald.

Ovenstående figur angiver, at ARIMA og LSTM, for alle specialets aktier, forecaster 7908 og 7923 stigninger, samt 7132 og 7137 fald, hvorved modellerne tilnærmelsesvist forecaster lige mange stigninger og fald. Det fremgår yderligere, i overensstemmelse med figur 8.4, at ARIMA generelt har flere fejl i forecast af retningen. Der forekommer hertil ingen tydelige tendenser til, at nogle af modellerne er bedre til at forecaste fald eller stigninger, da fejlene er tilnærmelsesvist ligeligt fordelt. Det kan derfor ikke vurderes, hvorvidt modellerne performer bedst i perioder med vækst eller nedgang. Dette understøttes i en Pesaran Timmermann test angivet i bilag B8. Det fremgår, at alle modellerne har p-værdier tilnærmelsesvis lig nul, hvorved nullhypotesen om, at data og forecast er uafhængigt fordelt, kan

forkastes på et 5% signifikansniveau. Herved antages alle specialets udarbejdede modeller, i nogen grad at være i stand til at forecaste retningen af de faktiske afkast.
Hvorvidt ARIMA eller LSTM bedst fanger udsvingene i data, kan ligeledes anskues i variansen af det faktiske data relativt til forecastene. For at sammenligne afvigelsen på tværs af aktier og sektorer angiver tabel 8.3 den procentvise afvigelse for ARIMA og LSTM.

Tabel: 8.3 Afvigelse i faktisk og forcastede varians

Aktie	Sektor	Afvigelse ARIMA	Avigelse LSTM
Alphabet Inc. A	Kommunikation	4,5%	4,2%
Alphabet Inc. C	Kommunikation	4,6%	4,1%
Amazon.com	Dis. forbrug	-5,9%	-7,3%
Home Deposit Inc.	Dis. forbrug	2,8%	3,1%
Tesla Inc.	Dis. forbrug	3,5%	3,5%
Bank of America	Finans	4,8%	4,8%
B. Hathaway	Finans	-0,2%	-0,7%
JP Morgan	Finans	4,3%	4,2%
Abbvie Inc.	Sundhed	-2,9%	-3,5%
Eli Lilly	Sundhed	-6,5%	-7,4%
Merck & Co.	Sundhed	-27,1%	-28,5%
U. Health Group	Sundhed	1,4%	0,8%
Caterpillar Inc	Industri	-7,0%	-7,6%
U. Parcel Service	Industri	4,2%	3,9%
Broadcom Inc.	IT	-2,2%	-2,3%
Mastercard Inc.	IT	-0,8%	-0,9%
Microsoft Corp	IT	4,8%	4,4%
Nvidia Corp	IT	2,7%	2,2%
Visa Inc.	IT	-0,4%	-0,1%
Udarbejdet i Excel			

Tabel: 8.3 Afvigelse i faktisk og forcastede varians

Aktie	Sektor	Afvigelse ARIMA	Avigelse LSTM
Apple Inc.	IT	4,0%	3,5%
Total	-	94,7%	97,0%
Udarbejdet i Excel			

Tabellen angiver den procentvise afvigelse i den faktiske og forecastede varians for ARIMA og LSTM. Summen af de absolutte forskelle er angivet i total, hvor ARIMA i højere grad tilnærmes den faktiske varians.

Det fremgår, at ARIMA og LSTM i høj grad fanger størrelsen af udsvingene i det faktiske data. Merck og Co. har dog en stor afvigelse på -27% og -28%, hvilket implicerer, at modellerne i høj grad undervurderer udsvingenes størrelse. Summen af de absolutte forskelle i variansen fremgår under *total*, hvor ARIMA i begrænset omfang er LSTM overlegen.

Ovenstående afsnit fremsætter, at både ARIMA og LSTM fremgår signifikant mere nøjagtige relativt til en random walk. Begge modeller opnår hertil en succesrate for forecast af retningen på over 50%, og præsterer herved bedre end en såkaldt “blindfolded monkey". Endvidere vurderes det, at LSTM opnår en bedre forecastperformance, hvorved de opnåede resultater er i overensstemmelse med eksisterende litteratur.
Der skal herunder tages forbehold for, at et mere nøjagtigt forecast ikke nødvendigvis bevirker til, at modellen vil opnå en højere profit ved handel på de finansielle markeder. Dette er essentielt i forbindelse med praktisk anvendelse af modellerne til at optimere afkastet over en given periode. Herved tilsidesættes ovenstående resultater midlertidigt, mens specialet undersøger, hvorvidt det forekommer økonomisk fordelagtigt, at implementere de forecastede værdier i en mean-variance analyse.

9 Porteføljestrategier

9.1 Antagelser og begrænsninger

Følgende afsnit har til formål at klarlægge en række antagelser og begrænsninger, som anvendes i specialets mean-variance analyser. For at undersøge anvendeligheden af ARIMA og LSTM i optimal porteføljesammensætning, fremsættes en række forskellige investeringstrategier, som udarbejdes med udgangspunkt i empirisk data samt forecastede værdier. Specialet udarbejder herved følgende fem strategier:

Strategi 1: Simpel strategi, hvor alle aktier indledningsvist vægtes ligeligt i porteføljen.

Strategi 2: Mean-variance analyse med buy and hold, baseret på empirisk data.

Strategi 3: Mean-variance analyse med daglig rebalancering baseret på empirisk data.

Strategi 4: Mean-variance analyse med daglig rebalancering, baseret på ARIMA forecasts.

Strategi 5: Mean-variance analyse med daglig rebalancering, baseret på LSTM forecasts.

Alle strategierne holdes herunder op mod S&P 500. Dette er for at vurdere, hvorvidt det forekommer fordelagtigt at anvende en mean-variance analyse til at opnå et merafkast relativt til markedet samt, hvorledes det forekommer profitabelt at investere bæredygtigt.
Strategi 1 og 2 anvendes endvidere som en baseline for, hvorvidt det forekommer profitabelt, at anvende daglig rebelancering af porteføljen. Strategi 4 og 5 tager hertil afsæt i de forecastede værdier af ARIMA og LSTM. En daglig rebelancering af porteføljen angiver, at der hver dag i investeringsperioden, udarbejdes en ny tangentportefølje, hvilket alt andet lige medfører, at vægtene i porteføljen ændres dagligt. Vægtene kan hertil både forekomme positive eller negative, afhængigt af, hvorvidt porteføljen går kort eller lang i den enkelte aktie. Shorting kan medføre mere profitable afkast, men forekommer også risikabelt, efter-

som investoren kan tabe flere penge end investeret (Hillier m.fl., 2011, kap. 4 s. 84). Dette forventes, at kunne være økonomisk fordelagtigt, da alle aktier korrelerer positivt og herved generelt følger udviklingen på det generelle marked og med hinanden.
Der pålægges endvidere en begrænsning for ekstreme vægte. Denne begrænsning tager afsæt i antagelsen om, at det er muligt at diversificere risikoen i en portefølje ved inkludering af flere forskellige aktier, som ikke korrelerer perfekt. Herunder fastsættes en øvre grænse for tangentporteføljens vægtning på +20% og -10% for hver aktie i hver periode[22].

Analysen tager yderligere afsæt i et fiktivt investeringsbeløb på 1 million US dollars. Investeringsbeløbet antages herunder at kunne erstattes med et alternativt beløb og replicerer de opnåede resultater. Aktiernes lukkekurser angiver endvidere købs- og salgskursen, hvortil der pålægges en transaktionsomkostning på 1% for hvert beløb, der handles hver periode. Dette er, som tidligere angivet, i modstrid med antagelsen om et friktionsløst marked. Specialet antager det herunder profitabelt at handle alle dage, hvor det forventede afkast fratrukket transaktionsomkostninger er positivt.
Alle investeringsstrategier, som tager afsæt i en mean-variance analyse, anvender en varians-kovariansmatrice, der beregnes med udgangspunkt i empirisk data for primo 2017 til ultimo 2019. Den beregnes herved for en 3-årig periode tilsvarende specialets udvalgte investeringsperiode, hvortil der kun inddrages det nyeste tilgængelige data, hvor dette antages at være det mest repræsentative for nuværende sammenhænge i data. For porteføljer, der rebalanceres dagligt, antages det risikofrie aktiv at have et procentvist afkast lig nul, hvortil buy and hold strategierne antager et afkast på 1%[23].

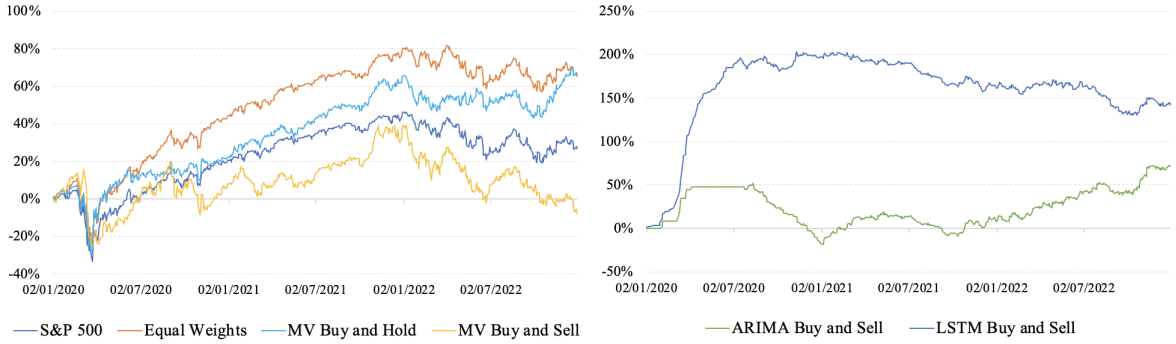
[22]Porteføljesammensætningen tager ikke højde for prisen på hver aktie, hvorved vægtene ikke altid resulterer i køb og salg af hele aktier. Dette kan medvirke til et mindre tab, som ikke implementeres i analysen.

[23]Det risikofrie afkast for investeringsperioden angives som den gennemsnitlige rente på amerikanske statsobligationer med udløb om 10 år.

9.2 Resultater - mean-variance

Følgende afsnit har til formål at evaluere og klarlægge resultaterne for specialets udvalgte porteføljer, som opstilles med udgangspunkt i ovenstående antagelser. For at sammenligne de udarbejdede porteføljer, tages der afsæt i porteføljernes afkast, merafkast, gennemsnitlige daglige afkast, standardafvigelse, Sharpe ratio, information ratio og maximum drawdown, beskrevet i *5.1 Evalueringsmetoder - Mean-variance*. Hertil relateres alle resultaterne til S&P 500, som er specialets benchmark og ligeledes en baseline til, hvorvidt det forekommer profitabelt at investere i mere bæredygtige porteføljer. De akkumulerede procentvise afkast for de seks investeringsstrategier illustreres i figur 9.1.

Figur: 9.1 Kummulative procentvise afkast



Beregningerne er udarbejdet i Excel. Figuren illustrerer de realiserede kummulative procentvise afkast for investeringsperioden, hvor x-aksen angiver tid og y-aksen de procentvise afkast. De to grafer har forskellige y-akser, hvor den venstre er i intervallet [-40%; 100%] og den højre [-50%; 250%].

Ovenstående angiver det akkumulerede daglige procentvise afkast for porteføljen, og afspejler herved den daglige procentvise ændring i porteføljeværdien, frem for den procentvise stigning i porteføljeværdien relativt til det oprindelige investeringsbeløb.

Det fremgår, at alle strategierne, undtagen mean-variance buy & sell, opnår et positivt akkumuleret procentvist afkast for den udvalgte investeringsperiode. Alle strategierne, undtagen de forecastede porteføljer, opnår hertil negative procentvise afkast i primo 2020. Dette kan være grundet det store fald på det amerikanske aktiemarked, som angivet i 6.2 *Deskriptiv statistik*, hvorfor dette ikke anses kritisk. Det gælder hertil for alle strategier uden

anvendelse af forecast, at 2021 er præget af vækst, hvortil afkastene efterfølgende tilnærmelsesvist stabiliseres, eller opnår beskedne fald. Herved følger disse strategier i høj grad den generelle udvikling på det amerikanske aktiemarked. ARIMA opnår hertil positive afkast for primo 2020, hvorefter der forekommer en overvægt negative afkast indtil primo 2022. LSTM opnår hertil den største stigning i starten af perioden, for efterfølgende at stabiliseres. Ovenstående grafer angiver herved, at LSTM genererer den største vækst i porteføljeværdien for alle specialets udvalgte investeringsstrategier.

Med henblik på at vurdere en investeringsstrategi, skal der hertil tages højde for en række andre faktorer, herunder det givne risikoniveau for porteføljen. Resultaterne for specialets porteføljer og S&P 500 illustreres derfor yderligere i tabel 9.1.

Tabel: 9.1 Porteføljeresultater

Mål	S&P 500	Simpel	MV B & H	MV B & S	ARIMA	LSTM
Totalt afkast i %	17,9%	66,5%	66,1%	-21,0%	87,4%	271,4%
Merafkast i %	-	48,6%	48,3%	-38,8%	69,5%	253,5%
Dagligt afkast %	0,04%	0,09%	0,09%	-0,01%	0,09%	0,19 %
Min. afkast %	-12,0%	-12,9%	-16,8%	-15,1%	-6,8%	-5,5%
Maks. afkast %	9,4%	10,0%	11,6%	11,5%	14,2%	14,2%
Daglig Std. afv.	1,6%	1,7%	1,9%	1,9%	1,4%	1,7%
Indbringende salg i %	-	-	-	52,6%	61,0%	85,8%
Antal handler	2	2	2	753	467	749
Udarbejdet i RStudio						

Tabellen angiver resultaterne for de seks investeringsstrategier i henhold til en række mål for afkast og risiko herunder det totale afkast, daglige afkast, samt spredningen heraf, angivet ved standardafvigelsen og største daglige fald og stigning i porteføljen. Afslutningsvist fremgår antal handler og indbringende salg. Alle mål fremgår både for buy and hold og buy and sell for at sammenligne afkast og risikoniveauet. Alle værdier er afrundet til en decimal, hvor det daglige afkast er afrundet til to.

Ovenstående angiver det totale afkast, daglige afkast, samt spredningen heraf, angivet ved standardafvigelsen og største daglige fald og stigning i porteføljen. Afslutningsvist fremgår antal handler og indbringende salg. Alle mål fremgår både for buy and hold og buy and sell for at sammenligne afkast og risiko ved de forskellige investeringsstrategier. Herunder inddrages det daglige aspekt for at vurdere risikoeksponeringen for en investor over en treårig investeringsperiode. Dette er i henhold til, hvis investoren en given dag skulle ønske at stoppe sin investering. Endvidere angives det totale afkast i % af det oprindelige investeringsbeløb på 1 mio. US dollars. Det fremgår af tabellen, at S&P 500, opnår et samlet afkast på 17,9% og en standardafvigelse i de daglige procentvise afkast på 1,6%. Herunder opnår den simple og mean-variance buy and hold et afkast på 66,5% og 66,1%, hvortil mean-variance med daglig rebalancering opnår et negativt afkast på -21%. Det fremgår hertil, at de tre strategier opnår en beskedent højere standardafvigelse end S&P 500.

Anvendelse af de forecastede ARIMA værdier, angiver et merafkast på 69,5% og en standardafvigelse på 1,4%, som er den laveste for de udvalgte investeringsstrategier. LSTM opnår hertil et merafkast på 253,5%, hvorved denne strategi performer det absolut højeste afkast for investeringsperioden. Dette understøttes i tabel 9.1, hvor det fremgår, at LSTM angiver den strategi med flest indbringende salg relativt til antal handler. Herunder fremgår 85,8% af alle handler profitable. Relativt hertil opnår ARIMA 61,0% og mean-variance buy and sell 52,6% profitable handler. Dette er i overensstemmelse med tidligere resultater i figur 8.5, hvor det fremgår, at LSTM har en succesrate på 76% i henhold til korrekt prædiktering af retningsskift i kursen. Dertil har ARIMA en succesrate på 60% korrekt prædiktion. Det fremgår endvidere, at ARIMA har en standardafvigelse på 1,4%, hvortil standardafvigelsen for alle specialets andre porteføljer er højere relativt til S&P 500. Dette er i overensstemmelse med, at specialets porteføljer alle består af færre aktier end S&P 500, hvorfor disse er mindre diversificerede og herved generelt er præget af en højere risiko.

Med henblik at sammenfatte ovenstående mål for afkast og risiko, angiver tabel 9.2 annualiserede mål for Sharpe ratio, information ratio, og maximum drawdown.

Tabel: 9.2 Annualiserede porteføljeresultater

Mål	S&P 500	Simpel	MV B&H	MV B&S	ARIMA	LSTM
Ann. Sharpe ratio	0,2	0,7	0,7	-0,2	1,0	2,0
Ann. information ratio	-	2,6	1,1	-0,6	0,5	1,3
Max. drawdown %	33,9%	31,9%	33,3%	40,6%	51,1%	54,2%
Udarbejdet i RStudio						

Tabellen angiver den annualiserede Sharpe ratio og information ratio samt maximum drawdown for alle specialets investeringsstrategier.

Det fremgår, at LSTM opnår den højeste Sharpe ratio. Hertil opnår ARIMA den næsthøjeste værdi, hvorved det kan anses økonomisk fordelagtigt at anvende forecast i porteføljeteori. Mean-variance buy and sell opnår hertil negative værdier, da porteføljen opnår et negativt afkast. Endvidere opnår alle specialets porteføljer en positiv information ratio på nær mean-variance buy and sell, hvilket implicerer et højere afkast givet risiko relativt til S&P 500. Anskues de seks investeringsstrategier i henhold til maximum drawdown, fremgår det, at S&P 500, den simple strategi og mean-variance både med og uden handel, opnår tilnærmelsesvist samme værdi. Herunder angives det største fald i porteføljeværdien, fra seneste top til bund, for strategierne at være mellem 30% og 40%. LSTM opnår den højeste drawdown på 54,2%, hvor porteføljeværdien, grundet den store vækst i primo 2020, ikke falder under det oprindelige investeringsbeløb på 1 mio. US dollars. Endvidere opnår ARIMA en drawdown på 51,1%, hvor porteføljeværdien i ultimo 2020 og ultimo 2021 falder under det oprindelige investeringsbeløb. Begge perioder kendetegnes ved nedgang på det generelle aktiemarkedet, hvor ARIMA ikke formår at kapitalisere på dette.

Ovenstående angiver, at specialets udarbejdede porteføljer er i stand til at præstere et overnormalt afkast relativt til markedet, men er ligeledes forbundet med en højere risiko. Relateres dette til den efficiente markedshypotese implicerer ovenstående, at det er muligt at opnå systematiske højere afkast relativt til markedet, men at dette samtidig er forbundet med en højere risiko. Grundet, at der forekommer en højere risiko i specialets porteføljer i forbindelse med højere standardafvigelse eller maximum drawdown, er resultaterne ikke nødvendigvis i modstrid med semi-stærk efficiens. Specialets resultaterne er hermed ikke i modstrid med den efficiente markedshypotese.

10 Diskussion

Følgende afsnit tager indledningsvist afsæt i en diskussion af, hvorvidt det historiske data kan anses som repræsentativt for fremtidigt data. Endvidere diskuteres mean-variance frameworket og kritik heraf. Specialets tidligere udarbejdede forecast tilkobles, hvortil fordele og ulemper tydeliggøres. Afslutningsvist diskuteres det bæredygtige element, herunder den bæredygtige omstilling og betydning heraf for specialets opnåede resultater.

10.1 Anvendelse af forecast i mean-variance

9.2 Resultater af mean-variance angiver, at den simple strategi opnår et højere afkast givet risiko, relativt til de empiriske mean-variance porteføljer. Dette skaber en undren i henhold til, at mean-variance udarbejdes med et formål om at optimere Sharpe ratio. Det opnåede resultat kan være grundet, at mean-variance tager afsæt i den historiske varians, kovarians samt afkast, som ikke nødvendigvis er repræsentativ for fremtidige værdier, da udviklingen i data ikke er identisk over tid. Dette fremsættes i tabel 10.1, som angiver den historiske daglige varians samt afkast, beregnet for perioden primo 2017 til ultimo 2019. De realiserede værdier i investeringsperioden beregnes hertil for primo 2020 til ultimo 2022.

Tabel: 10.1 Daglig varians & afkast

Aktie	Sektor	Hist. var.	Real. var.	Hist. afkast	Real. afkast
Alphabet Inc. A	Komm.	0,08%	0,12%	0,08%	0,06%
Alphabet Inc. C	Komm.	0,02%	0,05%	0,08%	0,06%
Amazon.com Inc.	Dis. forbrug	0,02%	0,05%	0,13%	0,02%
Home Deposit Inc.	Dis. forbrug	0,02%	0,04%	0,07%	0,07%
Tesla Inc.	Dis. forbrug	0,01%	0,03%	0,13%	0,30%
Bank of America	Finans	0,02%	0,05%	0,07%	0,03%
Udarbejdet i RStudio					

Tabel: 10.1 Daglig varians & afkast

Aktie	Sektor	Hist. var.	Real. var.	Hist. afkast	Real. afkast
B. Hathaway	Finans	0,01%	0,03%	0,05%	0,05%
JP Morgan	Finans	0,01%	0,06%	0,07%	0,02%
Abbvie Inc.	Sundhed	0,02%	0,05%	0,06%	0,09%
Eli Lilly	Sundhed	0,04%	0,06%	0,08%	0,16%
Merck & CO Inc.	Sundhed	0,02%	0,07%	0,06%	0,05%
U. Health Group	Sundhed	0,03%	0,03%	0,09%	0,10%
Caterpillar Inc.	Industri	0,09%	0,21%	0,07%	0,09%
U. Parcel Service	Industri	0,02%	0,05%	0,01%	0,07%
Broadcom Inc.	IT	0,03%	0,06%	0,09%	0,11%
Mastercard Inc.	IT	0,01%	0,04%	0,15%	0,05%
Microsoft Corp	IT	0,01%	0,04%	0,13%	0,08%
Nvidia Corp	IT	0,02%	0,06%	0,15%	0,18%
Visa Inc.	IT	0,02%	0,04%	0,12%	0,04%
Apple Inc.	IT	0,03%	0,05%	0,14%	0,10%
Udarbejdet i RStudio					

Tabellen angiver den historiske og realiserede varians og middelværdi af afkastet, som er relevant ved anvendelse af mean-variance. Den historiske daglige varians samt afkast, er beregnet for perioden primo 2017 til ultimo 2019. De realiserede værdier i investeringsperioden beregnes hertil for primo 2020 til ultimo 2022.

Ovenstående angiver, at den realiserede varians forekommer højere end den historiske for alle specialets aktier. Dette implicerer, at det realiserede risikoniveau er højere end det historiske, hvor den daglige varians for hovedparten af specialets aktier er mere end fordoblet. Hermed vurderes den historiske varians ikke at være repræsentativ for den fremtidige varians. Det fremgår yderligere, at den historiske og realiserede middelværdi for det procentvise afkast, varierer for størstedelen af de bæredygtige aktier. Afkastene både under- og overestimeres herunder med udgangspunkt i det historiske data. Afslutningsvist fremgår

det, at forholdet mellem afkast og risiko ændres for trænings- og testdata. Dette forekommer problematisk i forbindelse med konstruktionen af porteføljevægtene. Herunder vil aktier, som har en øget varians og et mindre afkast, under porteføljekonstruktionen fremstå langt mere fordelagtige end tilfældet. Et eksempel herpå er Mastercard Inc., som er en af de aktier med den historiske laveste varians og højeste afkast. Herved vægtes aktien højt i porteføljen, hvor bilag B7 angiver, at aktien pålægges en vægt på 20% og herved har stor indflydelse på porteføljens performance. I forbindelse med, at variansen i investeringsperioden er firedoblet og det procentvise afkast i flere perioder er negativt, vil dette medvirke til økonomiske tab. Hermed formår teorien ikke at tage højde for nye omstændigheder, hvilket i praksis understøtter, at teoriens grundlag ikke altid forekommer holdbart til beregning af optimale porteføljer. En yderligere problematik ved forskelle i de forventede og faktiske værdier er i forbindelse med, at porteføljevægtene har tendens til at være ekstremt følsomme, hvor en lille ændring i det forventede afkast, af kun én aktie, kan ændre den optimale porteføljesammensætning betydeligt. Denne følsomhed resulterer i flere perioder i en stærkt koncentreret portefølje, der ikke tilbyder tilstrækkelige diversificeringsfordele.

Estimeringen af parametrene fremgår hertil kun valide såfremt de procentvise afkast er normalfordelte, hvilket i praksis ofte ikke er tilfældet. Da alle procentvise afkast for specialets aktier har en kurtosis over tre, vil anvendelsen af mean-variance, resulterer i mere risikable allokeringer end beregnet. Dette er grundet, at tunge haler i fordelingen kan medføre betydelige fejl i estimatet af kovariansmatricen. Dette udtrykkes i specialet ved, at mean-variance opnår en højere standardafvigelse end forventet, hvorved teorien ikke opnår den forventede optimerende diversificering. Denne kritik menes at kunne mindskes, hvis teorien inkluderer flere faktorer end udelukkende det forventede afkast, variansen samt kovariansen. Dette kan være inddragelse af proxyer, for ny information om makroøkonomiske variable, såsom inflation, renter, energipriser eller volatilitet i kurserne. Udfordringen ved at

inkorporere flere faktorer er, at en analytisk løsning til, hvordan en investor optimerer sine investeringsbeslutninger er kompleks og svær at definere. Dette er dog udenfor specialets omfang at undersøge, men vurderes som en mulighed for at imødegå denne problematik.

Anvendelsen af historisk data kan ligeledes anses som problematisk i henhold til tidshorisonten. Dette relateres til buy and hold strategierne, hvor målene for afkast og risiko holdes konstant for hele investeringsperioden. En buy and hold strategi uden justeringer, tager herved ikke forbehold for ændringer i strukturen og udviklingen i både de individuelle aktier, samt forholdet herimellem. Det vurderes derfor relevant at inkludere strategier, som tager afsæt i løbende rebalancering. Dette er i forbindelse med, at der for hver periode justeres i henhold til ny information, hvorfor disse strategier i højere grad tager højde for ændringer i porteføljens risikoprofil. Det fremgår i tabel 9.1, at buy and sell strategien genererer et negativt afkast og en højere standardafvigelse relativt til buy and hold. Dette skyldes, at strategien tager afsæt i et glidende gennemsnit af det historiske data for tre år, der som tidligere angivet ikke er repræsentativt for investeringsperioden. Herved formår strategien, på trods af det opdaterede data og rebalancering, ikke at udnytte mulighederne ved buy and sell, da det historiske gennemsnit ikke fanger udsvingene i kurserne. Det vurderes herved relevant at implementere specialets forecasts i analysen. Dette tager afsæt i en antagelse om, at ARIMA og LSTM i højere grad formår at forecaste disse udsving. En ulempe heraf er givet ved, at det forventede afkast ændres markant for hver periode, hvilket kan resultere i ekstreme og vekslende vægte. Et forkert forventet fortegn, kan medvirke til en porteføljesammensætning, hvor de enkelte aktier genererer et negativt afkast. Ændringen i porteføljens værdi er i nogle perioder positiv, men beskedent relativt til det mulige afkast, imens det i andre perioder medfører negative afkast. Ved en strategi med handel, vil afkastene for porteføljen akkumuleres i hver periode, hvilket kan medføre enten større tab eller profit. Dermed kan forecast i mean-variance anses fordelagtig, men risikabel.

Specialets ARIMA forecast er underlagt en kritik i forbindelse med at forecaste udsvingene én periode forsinket, hvilket er særligt problematisk i volatile perioder. Dette illustreres eksempelvis for Alphabet Inc. i figur 8.1, hvor det fremgår, at kursen er særlig volatil, hvorfor forecast og data bevæger sig modsatrettet. De koncentrerede vægte i mean-variance kombineret med forkerte forecastede retninger, medvirker til flere perioder med høje negative afkast. Hermed kan de forecastede udsving, alt andet lige ikke vurderes som værende en fordel, såfremt forecastet ikke er i stand til at fange udsvingene i den faktiske periode. Herunder opnår ARIMA stadig et højere afkast for hele perioden relativt til strategierne, som anvender historisk data. Endvidere formår LSTM at forecaste 76% af retningsskiftene korrekt og opnår hermed en højere succesrate end ARIMA. Dette medfører, at LSTM størstedelen af perioderne er præget af positive afkast, som akkumuleres over tid og herved opnår det højeste totale afkast. Dette illustreres i figur 9.1 ved, at LSTM opnår kontinuerlige positive afkast for hele perioden. Det er herunder af relevant betydning, at LSTM ikke formår at forecaste alle perioder korrekt, hvor ekstreme vægte i mean-variance kan medvirke til høje negative afkast for disse perioder. Dette begrænses i specialet i forbindelse med de fremsatte begrænsninger for de procentvise vægte i porteføljen givet ved +20% og -10%. Hvis ikke en investor ønsker at sætte lignede begrænsninger, vil en tilsvarende analyse, kunne medvirke til ekstreme stigninger eller fald i porteføljeværdien.

Ovenstående implicerer herved, at historisk data ikke kan vurderes at være repræsentativt for fremtidig data. Det kan derfor anses fordelagtigt at anvende forecast i en porteføljekonstruktion, såfremt modellen er i stand til korrekt at fange udsving i kurserne. Det er derfor relevant at inkludere resultaterne fra *8 Forecast og evaluering af ARIMA og LSTM* for at vurdere, hvorvidt modellerne er i stand til at forecaste bæredygtige aktier.

10.2 Forecasting af bæredygtige aktier

Med henblik på at vurdere, hvorvidt specialets ARIMA og LSTM modeller er i stand til at forecaste bæredygtige amerikanske aktier, tages der afsæt i Gilliland, 2013 mål for forecast performance. Det er her gældende, at en model bør opnå en bedre performance end en random walk for at kunne vurderes at være i stand til at forecaste en given tidsserie. Det kan med udgangspunkt i forecastfejl og Diebold-Mariano testen, angivet i bilag B3, vurderes, at både LSTM og ARIMA er signifikant mere nøjagtige end en random walk og herved er i stand til at forecaste udsving i kurserne. Det fremgår endvidere, at faktisk data og forecast for begge modeller er afhængigt fordelt, hvilket implicerer, at modellerne er i stand til at fange retningen af afkastet. Specialet kan med udgangspunkt i resultaterne herved ikke understøtte random walk hypotesen, hvorfor det antages, at der fremgår systematiske mønstre i de amerikanske bæredygtige aktiekurser, som specialets modeller er i stand til at forecaste. Det fremgår yderligere, at LSTM er ARIMA overlegen til forecasting af alle specialets bæredygtige aktier. Dette vurderes i henhold til, at LSTM opnår en højere succesrate ved forecast af retningen i afkastet og lavere forecastfejl.

Specialet angiver hertil en række problematikker i forbindelse med forecasting af aktiemarkedet. Herunder kan jagten på den bedste forecast model forekomme kompleks og præget af en naturlig begrænsning i henhold til tid og computerkraft. Endvidere vil der over tid fremgå nye relevante informationer og mønstre i data, som bør inkorporeres i modellen. Komplekse modellers anvendelighed, vil derfor reduceres over tid, hvis dataet ændres og modellen ikke løbende justeres. I relation til nye informationer og mønstre i data, antages det at være afgørende for modellernes forecastperformance, om udviklingen i data er tilnærmelsesvist ens for trænings- og testdata. Dette er da modellerne trænes og estimeres med afsæt i det udvalgte træningsdata. Herved vil parametrene være bestemt ud fra den historiske udvikling, hvorfor modellerne udarbejder forecasts med afsæt i disse mønstre.

Ændres dette for testdata vil modellerne alt andet lige opnå en lavere forecastperformance. Det illustreres i *6 Databehandling*, at udviklingen og volatiliteten i aktierne er væsentlig forskelligt for perioden primo 2013 til ultimo 2019 relativt til primo 2020 til ultimo 2022. Særligt for ARIMA er det gældende, at modellen ikke er i stand til at forecaste skiftende volatilitet, hvor LSTM i højere grad kan tilpasse og justere parametrene løbende. Specialet søger at imødekomme denne problematik i forbindelse med one step ahead forecasts med rolling window. Herved vil modellerne løbende opdateres i henhold til den forsinkede værdi, hvorfor størrelsen på fejlene mindskes. På trods af, at det vurderes, at modellerne er i stand til at forecaste de bæredygtige aktier forventes det, at modellerne vil opnå en bedre performance, hvis udviklingen i testdata i højere grad tilnærmes træningsdata.

10.2.1 LSTM - Netværksarkitektur, reproducerbarhed og fortolkning

Konstruktionen af arkitekturen i et neuralt netværk sammenlignes ofte med en arbitrær kunst frem for en eksakt videnskab. Inden for deep learning og neurale netværk findes der mange forskellige typer af modeller, hvor specialet udelukkende anvender LSTM, som er egnet til tidsserier. Blandt disse modeller er der utallige måder at strukturere et netværk på.

Konstruktionen af LSTM modeller tager afsæt i en omfangsrig tuning af modellens hyperparametre, som angiver netværkets struktur. Valget om antal layers, antal units, aktiveringsfunktioner og andre hyperparametre, er ofte baseret på en trial and error tilgang. Dette skyldes, at neurale netværk kan betragtes som en black box, da det ikke forekommer muligt at fortolke, hvad netværket præcist lærer under træningen. Det er herunder ekstremt vanskeligt at få indsigt i og fortolke parametrene og vægtene, som netværket anvender. Komplekse neurale netværk kan have titusindvis af parametre eller endnu flere. Interaktionen mellem disse parametre og kompleksiteten af netværket er ikke klart defineret, herunder om der er ikke-lineære sammenhænge mellem variablerne. Da alle LSTM modeller ikke kan testes,

har specialet ikke nødvendigvis opnået den optimale LSTM model. Dette kan anses som en generel kritik af LSTM, hvor specialet har valgt den bedste af alle estimerede modeller, som stadig opnår en høj forecastperformance. Hertil er en ofte kritiseret egenskab ved neurale netværk, at det er udfordrende at reproducere de samme resultater. Dette skyldes, at der er en vis grad af tilfældighed involveret i træningen. De tusindvis af parametre, som netværket justerer under træningen, kan variere betydeligt fra træning til træning, hvorved to netværk, med præcis samme arkitektur, kan opnå helt forskellige resultater. De tilfældige forhold og startværdier under træningen af et neuralt netværk er ikke en fejl, men en bevidst og væsentlig egenskab, der bidrager til netværkets evne til at lære.

På trods af, at LSTM modeller er bredt anerkendte og ligeledes i specialet opnår den signifikant højeste forecastperformance, er modellen herved ikke uden kritik. For mennesker kan det være umuligt at fortolke modellen, hvilket kan forekomme særligt problematisk ved anvendelse af modellen i praksis. Hvis modellen ikke kan fortolkes og forstås, vil det være umuligt at estimere risikoen bag de opnåede forecasts. Som tidligere angivet, er de finansielle markeder under konstant udvikling, hvorfor udviklingen i data sjældent forekommer identisk over tid. En model, som tidligere har opnået en høj forecastperformance, vil ikke nødvendigvis fortsætte med at gøre det, såfremt markedet ændres. Det er derfor essentielt for en investor at forstå, hvilke elementer en model vurderer essentielle eller af mindre betydning. Dette er med henblik på at vurdere, om nye relationer, som kan blive gældende på markedet, også fanges af modellen. På trods af, at forecasts og tests er nyttige til at reducere usikkerheden for en forecastmodel, bør resultaterne af enhver model herved udelukkende betragtes som en retningslinje.

10.2.2 ARIMA - Udvælgelseskriterier og kompleksitet

Det fremgår af ovenstående, at LSTM er ARIMA overlegen til forecasting af alle specialets aktier. En kritik af LSTM fremsættes ved, at det ikke er muligt for en investor at tolke på modellen. I kontrast hertil har ARIMA et større teoretisk grundlag, hvor alle elementer i modellen kan fortolkes. Derfor er det lettere at justere modellen, da metoden og fremgangsmåden er mere udspecificeret i litteraturen. Det kan derfor være fordelagtigt for en investor at arbejde videre på at opnå en bedre forecastperformance for ARIMA. Der kan være flere måder til at justere en ARIMA model, herunder ved at inddrage sæsonkompomenter, andre transformationer af data eller flere lags. Specialet tager afsæt i relativt simple ARIMA modeller, ved at begrænse mængden af komponenter og lags i modellen. Dette er grundet, at flere studier finder, at kompleksitet i ARIMA modeller reducerer forecastperformance (K. C. Green og Armstrong, 2015). Kompleksiteten i specialets modeller, kan mindskes i forbindelse med valget af udvælgelseskriteriet. Herunder anvendes AIC, som tidligere angivet inkorporerer en strafparameter for hvert lag i modellen. Havde specialet valgt et kriterie, som straffer inkluderingen af lags hårdere, vil dette have resulteret i mere simple modeller. Det kan derfor ikke udelukkes, at det er muligt at opnå en bedre ARIMA model ved enten at inkludere flere komponenter, eller reducerer kompleksiteten i modellen.

Det teoretiske grundlag i ARIMA kan ligeledes være en udfordring i forbindelse med, at der sættes flere antagelser op for modellen. Er disse ikke opfyldt, vil resultaterne være invalide, hvorfor modellen ikke kan anvendes til forecasting. Der skal hertil tages forbehold for, at finansielt data generelt vil være præget af ikke-lineære mønstre, hvorved LSTM ofte vil være ARIMA overlegen. Ovenstående angiver, at LSTM ofte vil opnå den bedste forecastperformance, men samtidig være mere risikofyldt i forbindelse med tolkning af modellen. Det vil derfor være essentielt for en investor at tage stilling til, hvorvidt det mere nøjagtige forecast opvejer risikoen ved ikke at kunne tolke på den udarbejdede model.

10.3 Inkorporering af økonomisk teori

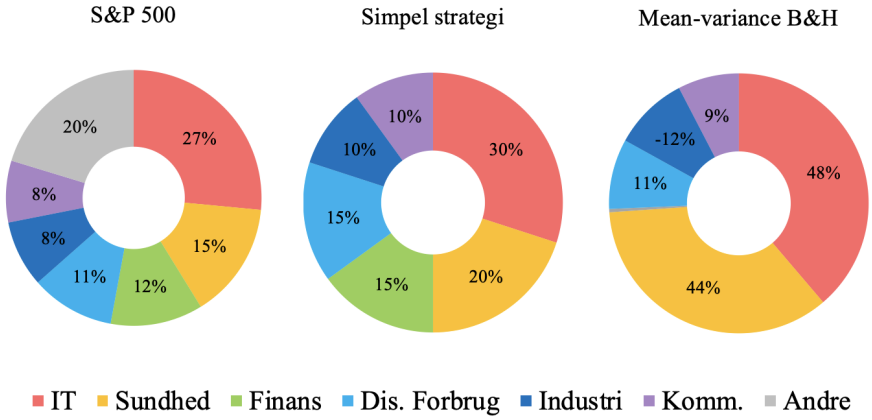
Et af specialets primære valg, som er af stor betydning for de endelige resultater, er anvendelsen af univariate modeller. Dette er grundet et ønske om at sammenligne ARIMA og LSTM. ARIMA er per definition en univariate model, hvorved LSTM ligeledes inddrages uden inkluderingen af yderligere variable. Brugen af endogene variable, vil kunne medvirke til at fange dynamiske interaktioner og feedback-effekter. Dette vil derfor kunne bidrage til mere nøjagtige forecasts i forbindelse med at imødegå udviklingen i de individuelle sektorer og andre makroøkonomiske faktorer (Singh, 2023).

Denne metode vil også være relevant i forbindelse med, at specialet har til formål at forecaste bæredygtige aktier. En undersøgelse fra 2017 af Kodra & Partalidou finder en positiv sammenhæng mellem selskabernes økonomiske og miljømæssige præstationer (Tran m.fl., 2020). Hermed kan det argumenteres for, at bæredygtighedselementet har en relation til selskabets økonomiske præstation. En forklarende variabel, givet ved eksempelvis ESG-rating, vil herved kunne bidrage til mere nøjagtige forecasts. Endvidere er der, grundet det store fokus på bæredygtighed, opstået en række politiske begivenheder i forbindelse med ny EU lovgivning, som har medført en stigning i efterspørgslen. Hertil kan der eventuelt klarlægges årsagssammenhænge mellem lovgivning og bæredygtige aktiers kursændring. Eksisterende litteratur og forskning af bæredygtighed og dens påvirkning på udviklingen i kurserne er herunder af begrænset omfang. Det anses herved vanskeligt at udvælge endogene variable, som vides at have signifikant effekt på bæredygtige aktier og som hermed bidrager til mere nøjagtige forecasts.

10.4 Investering med et bæredygtigt formål

Specialets bæredygtige porteføljer tager afsæt i udvalgte selskaber i S&P 500, som har en stærk miljømæssig, social og ledelsesmæssig praksis. Disse er sammensat for at undersøge, hvorvidt der forekommer økonomisk incitament til at investere mere bæredygtigt. Tabel 9.1 angiver, at størstedelen af de bæredygtige porteføljer, medfører et merafkast relativt til S&P 500. Det anses derfor relevant at anskue dette i henhold til sektordiversificeringen af porteføljerne herunder, hvorvidt de tilsvarer S&P 500. Dette er for at vurdere om merafkastet skyldes, at de bæredygtige aktier generelt performer bedre end S&P 500 eller er et resultat af en fundamental anden opsætning i porteføljen. Figur 10.1 illustrerer vægtningen i S&P 500, den simple strategi samt mean-variance buy and hold.

Figur: 10.1 S&P 500, Simpel strategi & Mean-variance - Buy and Hold



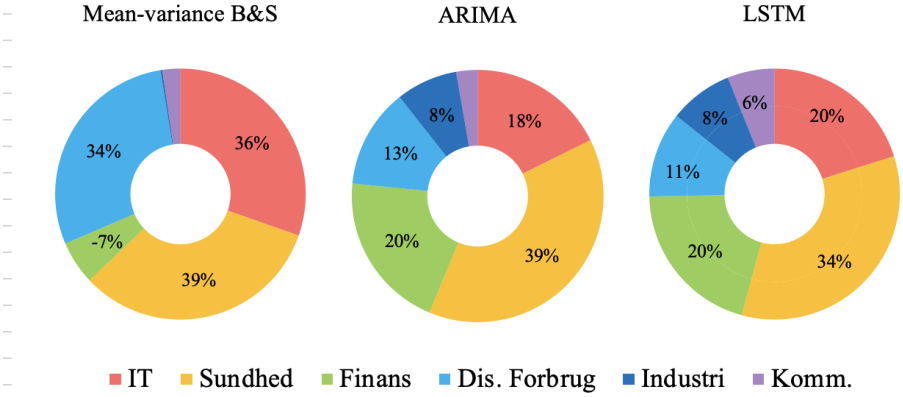
Figuren illustrerer den procentvise værdivægtede vægtning af aktierne i forskellige sektorer for S&P 500, mean-variance buy and hold og simpel strategi. Alle procentvise vægte rundes op til nærmeste hele værdi.

Den simple strategi afspejler bedst sektorfordelingen i S&P 500, hvilket er i overensstemmelse med, at aktierne er udvalgt med henblik på at afspejle denne. Det fremgår i tabel 9.1, at den simple strategi opnår et afkast på 66,5% og performer markant bedre relativt til S&P 500, som opnår et samlet afkast på 17,9%. Herved opnår bæredygtige aktier et merafkast uden nogen betydelig stigning i risikoen. Mean-variance buy and hold opnår resultater tilsvarende den simple strategi, hvor fordelingen adskiller sig mere fra S&P 500. Aktierne i sundhedssektoren vægtes herunder væsentlig højere, imens der er negative porteføljevægte

for aktierne i industrisektoren. Porteføljernes merafkast implicerer hermed, at investering i bæredygtige aktier er overlegne i henhold til investering i S&P 500 i investeringsperioden. En årsag hertil kan være givet ved kursudviklingen for specialets investeringsperiode. Herunder var aktiemarkedet ekstraordinært turbulent og stærkt præget af covid-19. Mange af specialets aktier, herunder United Postal Service, Mastercard og Nvidia Corp, har i perioden oplevet stor vækst og volatilitet. I kontrast hertil har flere selskaber inden for fossile brændstoffer oplevet mindsket efterspørgsel og vækst. Dette kan have påvirket væksten i S&P 500 negativt, uden betydning for specialets bæredygtige porteføljer.

Med henblik på at sammenligne ovenstående resultater med bæredygtige porteføljer med rebalancering, illustreres vægtene for alle specialets buy and sell strategier i figur 10.2.

Figur: 10.2 Mean-variance, ARIMA & LSTM - Buy and Sell



Figuren illustrerer den procentvise værdivægtede vægtning af sektorfordeling i mean-variance buy and sell, ARIMA og LSTM. Alle procentvise vægte rundes op til nærmeste hele værdi..

Det fremgår, at vægtene for ARIMA og LSTM tilnærmelsesvist følger hinanden og S&P 500. Begge porteføljer går i gennemsnit lang i alle sektorerne, hvortil der investeres i alle aktier i hver periode. Herved opnår både ARIMA og LSTM en gennemsnitlig sektordiversificering, som det er ønsket for en investor. Dette kan givet merafkastet understøtte, at forecast i mean-variance er økonomisk fordelagtigt for bæredygtige investeringer. Dette kan være grundet anvendelsen af ESG som udvælgelseskriterie, da selskaber med høj ESG-rating, ofte tager forbehold for en række risici, som kan være gavnlige for investorer.

En yderligere faktor af relevant betydning er specialets definition af bæredygtighed. Herunder udelukkes aktier med en ESG-rating under "A". Såfremt grænsen for ekskludering havde været højere, eksempelvis "AAA", havde resultaterne formentlig været anderledes. Det vil herunder ikke være muligt at opnå en sektordiversificering, som skildrer S&P 500. Dette vil medvirke til mere risikoprægede porteføljer og mindre afkastmuligheder. Anskues dette i henhold til specialets porteføljekonstruktion, forekommer det relevant at inddrage specialets forecasts i relation til forskellige ESG-ratings. Det fremgår i *8 Forecast og evaluering af ARIMA og LSTM*, at forecastnøjagtigheden øges ved forecasting af aktier med ESG-ratingen "AA"og "AAA". Dette er i overensstemmelse med, at eksisterende litteratur finder, at forecastmodeller ofte performer bedre på mindre volatile tidsserier, som også kendetegner bæredygtige aktier (Silvola og Landau, 2021, kap. 8 s. 164). Herved vil ekskluderingen af alle aktier under "AA", medvirke til en fordel ved anvendelse af forecast til porteføljesammensætning, i forbindelse med en højere forecastingnøjagtighed.

11 Konklusion

Specialets resultater, analyser og diskussioner er udarbejdet med et formål om at besvare følgende problemformulering:

Formår ARIMA og LSTM modeller at forecaste bæredygtige aktier, og herunder at præstere et overnormalt afkast? Hvorledes kan indsigter herfra anvendes til at skabe incitament til investering med et bæredygtigt formål?

For at undersøge, hvorvidt modellerne formår at forecaste bæredygtige aktier, estimeres og sammenlignes random walk, ARIMA og LSTM. For at vurdere, om modellerne er i stand til at præstere et overnormalt afkast ved bæredygtig investering, inddrages udelukkende bæredygtige aktier, der sammenlignes med et benchmark givet ved S&P 500. Hertil inddrages en mean-variance analyse, imens resultaterne relateres til den efficiente markedshypotese. De udvalgte ARIMA og LSTM modeller evalueres med afsæt i specialets evalueringsmetoder, hvor det fremgår, at både ARIMA og LSTM er signifikant mere nøjagtige end en random walk. Modellerne opnår hertil en succesrate for forecastet retning på over 50% og præsterer herved bedre end en såkaldt “blindfolded monkey". Random walk hypotesen kan hermed ikke understøttes, hvorfor det antages, at der forekommer systematiske mønstre i de bæredygtige kurser, som ARIMA og LSTM er i stand til at forecaste. Endvidere kan det konkluderes, at LSTM opnår en bedre forecastperformance relativt til ARIMA. Resultaterne er herved i overensstemmelse med eksisterende litteratur.

Modellernes forecasts anvendes efterfølgende i en mean-variance analyse. Resultaterne angiver, at LSTM opnår den højeste Sharpe ratio. Hertil opnår ARIMA de næsthøjeste værdier, hvorved det kan anses økonomisk fordelagtigt at anvende forecast i porteføljeteori. Hertil fremgår en positiv information ratio for tilnærmelsesvis alle udarbejdede porteføljer,

hvilket implicerer et højere afkast givet risiko relativt til S&P 500. Anskues investeringsstrategierne i henhold til maximum drawdown, fremgår det, at S&P 500, den simple strategi og mean-variance med og uden handel, opnår værdier mellem 30% og 40%. ARIMA og LSTM opnår de højeste drawdown på 51,1% og 54,2%. For LSTM falder porteføljeværdien aldrig under det oprindelige investeringsbeløb, imens ARIMA i flere perioder opnår økonomiske tab i porteføljen. Relateres dette til den efficiente markedshypotese implicerer ovenstående, at det er muligt at opnå systematiske højere afkast relativt til markedet, men at det samtidig er forbundet med en højere risiko. Da der forekommer en højere risiko i specialets porteføljer i forbindelse med højere standardafvigelse eller maximum drawdown, er resultaterne ikke nødvendigvis i modstrid med semi-stærk efficiens. Specialets resultaterne er hermed ikke i modstrid med den efficiente markedshypotese.

Ud fra ovenstående, kan det konkluderes, at bæredygtige porteføljer, både baseret på forecastede værdier og simple strategier udkonkurrer det traditionelle marked, givet ved S&P 500. Herved opnår bæredygtige aktier et væsentligt merafkast uden nogen betydelig stigning i risikoen, hvorved der skabes et økonomisk incitament til at investere mere bæredygtigt.

12 Litteraturliste

Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock price prediction using the ARIMA model. *International Conference on Computer Modelling and Simulation*. Hentet 3. maj 2023, fra https://www.researchgate.net/publication/261179224%20_Stock%20_price%20_prediction%20_using%20_the%20_ARIMA%20_model

AnalystPrep. (2019). Kurtosis and Skewness. Hentet 25. marts 2023, fra https://analystprep.com/cfa-level-1-exam/quantitative-methods/kurtosis-and-skewness-types-of-distributions/

AnalystPrep. (2020). Skewness. Hentet 25. marts 2023, fra https://analystprep.com/study-notes/frm/part-1/quantitative-analysis/probabilities/attachment/page-33a/

Ayodele, T. O. (2010). Types of machine learning algorithms. *New Advances in Machine Learning*. Hentet 4. maj 2023, fra https://www.researchgate.net/publication/221907660_Types_of_Machine_Learning%20_Algorithms

Baheti, P. (2022). Types of Neural Network Activation Functions: How to Choose. *V7 Labs*. Hentet 29. maj 2023, fra https://www.v7labs.com/blog/neural-networks-activation-functions

Balsara, N. J., Chen, G., & Zheng, L. (2007). The Chinese stock market: An examination of the random walk model and technical trading rules. *Quarterly Journal of Business and Economics*. Hentet 17. maj 2023, fra https://www.jstor.org/stable/40473435

Barnes, R. (2022). Benchmark Your Returns With Indexes. *Investopedia*. Hentet 4. maj 2023, fra https://www.investopedia.com/articles/basics/06/benchmark.asp

BlackRock. (2023). iShares ESG Screened S&P 500 ETF. Hentet 9. marts 2023, fra https://www.ishares.com/us/literature/fact-sheet/xvv-ishares-esg-screened-s-p-500-etf-fund-fact-sheet-en-us.pdf

Blackrock. (2023). iShares ESG Screened S&P 500 ETF. Hentet 9. marts 2023, fra https://www.ishares.com/us/products/315917/ishares-esg-screened-s-p-500-etf#esgAnalytics

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Broadridge. (2021). ESG: Transforming asset management and fund distribution. Hentet 14. marts 2023, fra https://www.broadridge.com/_assets/pdf/broadridge_esg_white_paper_march_2021.pdf

Brodtman, Z. (2021). The Importance and Reasoning behind Activation Functions. *Towards Data Science*. Hentet 3. maj 2023, fra https://towardsdatascience.com/the-importance-and-reasoning-behind-activation-functions-4dc00e74db41

Business for Good. (n.d). SUSTAINABLE DEVELOPMENT GOALS. Hentet 3. maj 2023, fra https://b1g1.com/sustainable-development-goals?gclid=CjwKCAjwhNWZBhB_EiwAPzlhNpgl

Campbell, J. Y., & Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*. Hentet 3. maj 2023, fra https://www.jstor.org/stable/40056860

Chen, J. (2023). Skewness: Positively and Negatively Skewed Defined With Formula. *Investopedia*. Hentet 21. maj 2023, fra https://www.investopedia.com/terms/s/skewness.asp

Chen, Y.-L., Shen, S.-L., & Zhou, A. (2022). Assessment of red tide risk by integrating CRITIC weight method, TOPSIS-ASSETS method, and Monte Carlo simulation. *Environmental Pollution*. Hentet 29. maj 2023, fra https://www.sciencedirect.com/science/article/pii/S0269749122014683

Chowdhury, K. (2021). 10 Hyperparameters to keep an eye on for your LSTM model — and other tips. *Medium*. Hentet 17. maj 2023, fra https://medium.com/geekculture/

10-hyperparameters-to-keep-an-eye-on-for-your-lstm-model-and-other-tips-f0ff5b63fcd4

DeepAI. (n.d). What is Bias Vector? Hentet 4. maj 2023, fra https://deepai.org/machine-learning-glossary-and-terms/bias-vector?

Delua, J. (2021). Supervised vs. unsupervised learning: What's the difference. *IBM.com*. Hentet 24. maj 2023, fra https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning

Difference Between. (n.d). Difference Between AIC and BIC. Hentet 4. maj 2023, fra http://www.differencebetween.net/miscellaneous/difference-between-aic-and-bic/

Dowell, G., Hart, S., & Yeung, B. (2000). Do Corporate Global Environmental Standards Create or Destroy Market Value? *Management Science*. Hentet 17. maj 2023, fra https://www.jstor.org/stable/2661584

Dwight, C. A. (2019). Jarque-Bera Test of Normality. *Medium*. Hentet 20. april 2023, fra https://collinsdwight.medium.com/jarque-bera-test-of-normality-a108a1515b22

Enders, W. (2014). *Applied econometric time series*. Wiley.

Erhvervstyrrelsen. (2021). Ny vejledning om ESG-rapportering. Hentet 4. maj 2023, fra https://erhvervsstyrelsen.dk/ny-vejledning-om-esg-rapportering

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The journal of Finance*. Hentet 3. maj 2023, fra http://www.jstor.com/stable/2325486

Filbeck, G., & Gorman, R. F. (2004). The Relationship between the Environmental and Financial Performance of Public Utilities. *Environmental and Resource Economics*. Hentet 17. maj 2023, fra https://link.springer.com/article/10.1023/B:EARE.0000044602.86367.ff

Finans Danmark. (n.d). Bæredygtige investeringer. Hentet 3. maj 2023, fra finansdanmark.dk/aktuelle-emner/baeredygtig-finans/baeredygtige-investeringer/

FN‘s Verdensmål. (2016). VERDENSMÅLENE. Hentet 4. maj 2023, fra https://www.verdensmaalene.dk/fakta/verdensmaalene

Galili, T. (2016). Lambda (Box-Cox transformation parameter) value for forecasting using ARIMA). *R-bloggers*. Hentet 4. maj 2023, fra https://www.r-bloggers.com/2016/02/lambda-box-cox-transformation-parameter-value-for-forecasting-using-arima/

Ghatak, A. (2019). *Deep learning with R*. Springer.

Gilliland, M. (2013). FVA: A reality check on forecasting practices. *The International Journal of Applied Forecasting*. Hentet 29. maj 2023, fra https://forecasters.org/wp-content/uploads/FVA_A-Reality-Check_Foresight29.pdf

Green, J., & Newman, P. (2017). Disruptive innovation, stranded assets and forecasting: the rise and rise of renewable energy. *Review of Economics and Finance*. Hentet 5. marts 2023, fra https://www.tandfonline.com/doi/full/10.1080/20430795.2016.1265410

Green, K. C., & Armstrong, J. S. (2015). Simple versus complex forecasting: The evidence. *Journal of Business Research*. Hentet 17. maj 2023, fra https://www.sciencedirect.com/science/article/pii/S014829631500140X

Gudimalla, V. (2021). Concept of Gradient Descent in Machine Learning. *Medium*. Hentet 4. maj 2023, fra https://varshithagudimalla.medium.com/concept-of-gradient-descent-algorithm-in-machine-learning-44f587ac16ac

Gupta, F., Prajogi, R., & Stubbs, E. (1999). The information ratio and performance. *Journal of Portfolio Management*. Hentet 24. maj 2023, fra https://www.pm-research.com/content/iijpormgmt/26/1/33

Han, R., Wang, W., & MA, Y. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*. Hentet 12. marts 2023, fra https://doi.org/10.1016/j.eswa.2020.113973

Hart, S. L., & Ahuja, G. (1996). Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance. *Business strategy and the Environment*. Hentet 21. maj 2023, fra https://doi.org/10.1002/(SICI)1099-0836(199603)5:1%3C30::AID-BSE38%3E3.0.CO;2-Q

Hayes, A. (2022). Maximum Drawdown (MDD) Defined, With Formula for Calculation. *Investopedia*. Hentet 4. maj 2023, fra https://www.investopedia.com/terms/m/maximum-drawdown-mdd.asp

Hillier, D., Grinblatt, M., & Titman, S. (2011). *Financial Markets and Corporate Strategy European*. McGraw Hill.

IBM. (2023a). Neutral Network. Hentet 24. marts 2023, fra https://www.ibm.com/cloud/learn/neural-network

IBM. (2023b). Recurrent Neural Networks. Hentet 4. maj 2023, fra https://www.ibm.com/topics/recurrent-neural-networks

Kampakis, S. (2020). Performance measures: RMSE and MAE. *The Data Scientist*. Hentet 9. maj 2023, fra https://thedatascientist.com/performance-measures-rmse-mae/

Kavlakoglu, E. (2020). AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference? *IBM*. Hentet 29. maj 2023, fra https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks

KEFM. (n.d). Klima, Energi og Forsyningsministeriet. Hentet 3. maj 2023, fra kefm.dk

Kenton, W. (2023). Beta: Definition, Calculation, and Explanation for Investors. *Investopedia*. Hentet 17. maj 2023, fra https://www.investopedia.com/terms/b/beta.asp

King, R. G., Plosser, C. I., & Rebelo, S. T. (2002). Production, Growth and Business Cycles: Technical Appendix. *Computational Economics*. Hentet 17. maj 2023, fra https://doi.org/10.1023/A:1020529028761

Kotze, K. (n.d.). Forecasting. Hentet 5. marts 2023, fra https://kevin-kotze.gitlab.io/tsm/ts-4-note/

Kvilhaug, S. (2023). A breakdown of the fiscal and monetary responses to the pandemic. *Investopedia*. Hentet 21. maj 2023, fra https://www.investopedia.com/government-stimulus-efforts-to-fight-the-covid-19-crisis-4799723

Lee, S. I., & Yoo, S. J. (2018). Threshold-based Portfolio: The Role of the Threshold and its applications. *The Journal of Supercomputing*. Hentet 24. maj 2023, fra https://link.springer.com/article/10.1007/s11227-018-2577-1

Li, S. (2019). Time series analysis, visualization & forecasting with LSTM. *Towards Data Science*. Hentet 12. marts 2023, fra https://towardsdatascience.com/time-series-analysis-visualization-forecasting-with-lstm-77a905180eba

Ma, Q. (2020). Comparison of ARIMA, ANN and LSTM for stock price prediction. *E3S Web of Conferences*. Hentet 3. maj 2023, fra https://www.e3s-conferences.org/articles/e3sconf/abs/2020/78/e3sconf_iseese2020_01026/e3sconf_iseese2020_01026.html

Madsen, J. M. (2022). Vil krisen skabe et nyt bear-marked? *Nordea Invest*. Hentet 23. maj 2023, fra https://www.nordeafunds.com/da/artikler/vil-krisen-skabe-et-nyt-bear-marked

Maheswari, J. P. (2019). Breaking the curse of small data sets in Machine Learning. *Towards Data Science*. Hentet 21. maj 2023, fra https://towardsdatascience.com/breaking-the-curse-of-small-data-sets-in-machine-learning-part-2-894aa45277f4

Mallikarjuna, M., & Rao, R. P. (2019). Evaluation of forecasting methods from selected stock market returns. *Financial Innovation*. Hentet 17. maj 2023, fra https://jfin-swufe.springeropen.com/articles/10.1186/s40854-019-0157-x

Maverick, J. (2023). S&P 500 Average Return. *Investopedia*. Hentet 27. marts 2023, fra https://www.investopedia.com/ask/answers/042415/what-average-annual-return-sp-500.asp

McKinsey. (n.d). The net-zero transition: What it would cost, what it could bring. Hentet 3. maj 2023, fra https://www.mckinsey.com/capabilities/sustainability/our-insights/the-net-zero-transition-what-it-would-cost-what-it-could-bring

McLeman-Hasselgaard, E., & Nielsen, R. L. (2017). Forecast Analyse. *Aalborg Universitet*. Hentet 4. maj 2023, fra https://projekter.aau.dk/projekter/files/259875842/Projekt_Forecast_analyse.pdf

Milonas, N., Rompotis, G., & Moutzouris, C. (2022). The performance of ESG funds vis-à-vis non-ESG funds. *The Journal of Impact and ESG Investing*. Hentet 29. maj 2023, fra https://www.pm-research.com/content/pmrjesg/early/2022/03/19/jesg20221041

Molnar, C. (2023). Interpretable Machine Learning. *Lulu.Com.* Hentet 4. maj 2023, fra https://christophm.github.io/interpretable-ml-book/

Moolayil, J. (2019). *Learn Keras for Deep Neural Networks*. Springer. Hentet 14. marts 2023, fra https://link.springer.com/book/10.1007/978-1-4842-4240-7

MSCI. (2023). ESG Ratings & Climate. Hentet 14. marts 2023, fra https://www.msci.com/

Newbold, P. (1975). The Principles of the Box-Jenkins Approach. *Palgrave Macmillan Journals.* Hentet 4. maj 2023, fra https://www.jstor.org/stable/3007750

Nielsen, M. (2015). *Neutral Network and Deep Learing*. Determination Press. Hentet 24. maj 2023, fra http://neuralnetworksanddeeplearning.com/chap2.html,

Olah, C. (2015). Understanding LSTM Networks. *Colah's blog*. Hentet 24. maj 2023, fra http://colah.github.io/posts/2015-08-Understanding-LSTMs/?fbclid=IwAR176XT5

Otexts. (n.d). ARIMA modelling in R. Hentet 4. maj 2023, fra https://otexts.com/fpp2/arima-r.html

Pesaran, M. H., & Timmermann, A. (1992). A Simple Nonparametric Test of Predictive Performance. *Journal of Business & Economic Statistics*. Hentet 25. maj 2023, fra https://www.jstor.org/stable/1391822

Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time Series with R*. Springer Science & Business Media.

Prabhakaran, S. (2021). ARIMA Model-Complete Guide to Time Series Forecasting in Python. *Machine Learning+*. Hentet 4. maj 2023, fra https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/

Rai, B. (2019). *Advanced Deep Learning with R*. Packt Publishing Ltd. Hentet 21. maj 2023, fra https://ebookcentral.proquest.com/lib/aalborguniv-ebooks/detail.action?pq-origsite=primodocID=6005546

Rakhecha, A. (2010). Understanding Learning Rate. Hentet 24. marts 2023, fra https://towardsdatascience.com/https-medium-com-dashingaditya-rakhecha-understanding-learning-rate-dd5da26bb6de

Saeed, M. (2022). An Introduction to Recurrent Neural Networks and the Math That Powers Them. *Machine Learning Mastery*. Hentet 24. maj 2023, fra https://llk.dk/0ifvzr

Sakia, R. M. (1992). Box-Cox Transformation Technique: A Review. *Journal of the Royal Statistical Society: Series D (The Statistician)*. Hentet 4. maj 2023, fra https://www.jstor.org/stable/2348250

Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. *Time Series Analysis and Its Applications*. Hentet 12. marts 2023, fra https://link.springer.com/chapter/10.1007/978-3-319-52452-8_3

Siami-Namini, S., & Namin, A. S. (2018). Forecasting Economics and Financial Time Series: ARIMA vs. LSTM. *Cornell University*. Hentet 23. maj 2023, fra https://arxiv.org/abs/1803.06386

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). A Comparative Analysis of Forecasting Financial Time Series Using ARIMA, LSTM, and BiLSTM. *Cornell University*. Hentet 23. maj 2023, fra https://arxiv.org/abs/1911.09512

Silvola, H., & Landau, T. (2021). *Sustainable Investing: Beating the Market with ESG*. Springer.

Singh, A. (2023). Multivariate Time Series Analysis With Python for Forecasting and Modeling. *Analytics Vidhya*. Hentet 17. maj 2023, fra https://www.analyticsvidhya.com/blog/2018/09/multivariate-time-series-guide-forecasting-modeling-python-codes/

Smith, T. (2023). Hvad er sektorer? *Investopedia*. Hentet 11. maj 2023, fra https://www.investopedia.com/terms/r/randomwalktheory.asp

Su, J., & Yang, W. (2022). Artificial intelligence in early childhood education: A scoping review. *Computers and Education: Artificial Intelligence*. Hentet 17. maj 2023, fra https://www.sciencedirect.com/science/article/pii/S2666920X22000042?via%5C%3Dihub

Telle, K. (2006). "It Pays to be Green" – A Premature Conclusion? *Environmental and Resource Economics*. Hentet 29. maj 2023, fra https://doi.org/10.1007/s10640-006-9013-3

Thombre, A. M. (2018). Using Machine Learning to find out When to use Box-Cox Transformation on Time Series Data. *IEEE*. Hentet 24. maj 2023, fra https://ieeexplore.ieee.org/document/8745372

Thune, K. (2021). Efficient Market Hypothesis (EMH). *The balance*. Hentet 4. maj 2023, fra https://www.thebalancemoney.com/efficient-markets-hypothesis-emh-2466619

Tran, T., Do, H., Vu, T., & Do, N. (2020). The factors affecting green investment for sustainable development. *Decision Science Letters*. Hentet 4. maj 2023, fra https://www.researchgate.net/publication/341750564_The_factors_affecting_green_investment_for_sustainable_development

Vorgaard, P. (2020). Hvad er sektorer? *Euroinvester*. Hentet 24. maj 2023, fra https://www.euroinvestor.dk/nyheder/hvad-er-sektorer

World Economic Forum. (2022). The net-zero transition: What it would cost, what it could bring. Hentet 3. maj 2023, fra https://www3.weforum.org/docs/WEF_Winning_the_Race_to_Net_Zero_2022.pdf

Wrg, A. (2019). From Econometrics to Machine Learning. *Medium*. Hentet 15. maj 2023, fra https://towardsdatascience.com/from-econometrics-to-machine-learning-ee182f3a45d7

Xiao, R., Feng, Y., Yan, L., & Ma, Y. (2022). Predict stock prices with ARIMA and LSTM. *Cornell University*. Hentet 3. maj 2023, fra https://arxiv.org/abs/2209.02407

Yamak, P. T., Yujian, L., & Gadosey, P. K. (2019). A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting. *Association for Computing Machinery*. Hentet 23. maj 2023, fra https://dl.acm.org/doi/10.1145/3377713.3377722

Your Green Wealth. (2023). MSCI ESG Fund Rating. Hentet 21. april 2023, fra https://yourgreenwealth.com/database/

Zaiontz, C. (2021). Diebold-Mariano Test. *Real Statistics Resources*. Hentet 15. februar 2023, fra https://real-statistics.com/time-series-analysis/forecasting-accuracy/diebold-mariano-test/

Ziegler, A., & Nogareda, J. S. (2009). Environmental management systems and technological environmental innovations: Exploring the causal relationship. *Research Policy*. Hentet 29. maj 2023, fra https://www.sciencedirect.com/science/article/pii/S0048733309000122

13 Bilagsoversigt

Bilag B1: Specialets data for 20 udvalgte aktier og S&P 500.

https://github.com/JakobTheSlott/Kandidat/blob/main/Assets.csv

https://github.com/JakobTheSlott/Kandidat/blob/main/SP%20500.csv

Bilag B2: Beregninger af deskriptiv Statistik.

https://github.com/JakobTheSlott/Kandidat/blob/main/LSTM%20folder/Bilag%20B2%20-%20Deskriptiv-Statistik.rmd

Bilag B3: ARIMA udvælgelse for alle specialets aktier.

https://github.com/JakobTheSlott/Kandidat/tree/main/ARIMA%20Folder

Bilag B4: LSTM udvælgelse for alle specialets aktier.

https://github.com/JakobTheSlott/Kandidat/tree/main/LSTM%20folder

Bilag B5: Forecast og evaluering af ARIMA og LSTM.

https://github.com/JakobTheSlott/Kandidat/tree/main/Forecast%20og%20Evaluering

Bilag B6: Mean-variance analyse.

https://github.com/JakobTheSlott/Kandidat/blob/main/Bilag%20B6%20-%20Mean-variance-analyse.xlsm

Bilag B7: Mean-variance vægte.

https://github.com/JakobTheSlott/Kandidat/tree/main/Weights

Bilag B8: Pesaran-Timmermann test.

https://github.com/JakobTheSlott/Kandidat/blob/main/Forecast%20og%20Evaluering/Bilag-B8-Pesaran-Timmermann-test.xlsx

Bilag B9: AIC værdier for ARIMA.

https://github.com/JakobTheSlott/Kandidat/blob/main/ARIMA%20Folder/AICValues.csv

Bilag B10: MSE værdier for LSTM.

https://github.com/JakobTheSlott/Kandidat/blob/main/LSTM%20folder/MSE-Assets.csv

14 Appendiks A

Følgende appendiks har til formål at klarlægge og uddybe resultaterne fremsat i specialet. Herunder illustreres fejl og resultater af udførte tests for alle udvalgte bæredygtige aktie.

Tabel A.1 - Test for autokorrelation - Ljung Box

Tabel: A.1 Test for autokorrelation - Ljung Box

Aktie	P-værdi	Aktie	P-værdi
Alphabet Inc. A	0,68	Alphabet Inc. C	0,48
Amazon.com	0,64	Tesla Inc	0,76
Home Depot Inc.	0,36	B. Hathaway	0,99
JP Morgan	0,96	Bank of America	0,99
U. Health Group	0,96	Merck & Co Inc.	0,99
Abbvie Inc.	0,99	Eli Lilly	0,96
U. Parcel Service	0,98	Caterpillar Inc.	0,86
Apple Inc.	0,63	Microsoft Corp	0,99
Nvidia Corp	0,96	Broadcom Inc.	0,90
Visa Inc. A	0,88	Mastercard Inc.	0,96
Udarbejdet i RStudio			

Tabel A.2 - Fejlprædikteringer for ARIMA

Tabel: A.2 Fejlprædikteringer for ARIMA

Aktie	Model	RMSE	MAE	MAPE
Alphabet Inc. A	**ARIMA**	1,86	1,11	1,19
Alphabet Inc. C	**ARIMA**	1,87	1,11	1,18
Amazon.com Inc.	**ARIMA**	2,93	1,76	1,38
Tesla Inc.	**ARIMA**	6,92	3,90	2,38
Home Depot Inc.	**ARIMA**	4,87	2,79	1,05
Bershire Hathawa	**ARIMA**	3,46	2,03	0,83
JP Morgan	**ARIMA**	2,40	1,39	1,25
Bank Of America	**ARIMA**	0,67	0,39	1,34
U. Health Group	**ARIMA**	6,49	3,77	1,02
Merck & Co Inc.	**ARIMA**	1,08	0,62	0,77
Abbvie Inc.	**ARIMA**	1,63	0,96	0,85
Eli Lilly	**ARIMA**	3,74	2,15	0,99
U. Parcel Service	**ARIMA**	2,81	1,58	1,04
Caterpillar Inc.	**ARIMA**	3,19	1,91	1,18
Apple Inc.	**ARIMA**	2,48	1,50	1,26
Microsoft Corp	**ARIMA**	4,39	2,67	1,20
Nvidia Corp	**ARIMA**	4,78	2,78	1,93
Broadcom Inc.	**ARIMA**	8,85	5,37	1,32
Visa Inc. A	**ARIMA**	3,56	2,07	1,07
Mastercard Inc.	**ARIMA**	6,32	3,73	1,19
Udarbejdet i RStudio				

Tabel A.3 - Fejlprædikteringer for LSTM

Tabel: A.3 Fejlprædikteringer for LSTM

Aktie	Model	RMSE	MAE	MAPE
Alphabet Inc. A	**LSTM**	0,73	0,35	0,35
Alphabet Inc. C	**LSTM**	0,69	0,38	0,38
Amazon.com Inc.	**LSTM**	1,06	0,58	0,42
Tesla Inc.	**LSTM**	7,42	4,28	2,02
Home Depot Inc.	**LSTM**	1,77	0,60	0,23
Bershire Hathawa	**LSTM**	1,04	0,60	0,25
JP Morgan	**LSTM**	1,04	0,47	0,40
Bank Of America	**LSTM**	0,26	0,14	0,44
U. Health Group	**LSTM**	2,96	1,27	0,35
Merck & Co Inc.	**LSTM**	0,31	0,20	0,24
Abbvie Inc.	**LSTM**	0,47	0,34	0,30
Eli Lilly	**LSTM**	2,11	0,89	0,41
U. Parcel Service	**LSTM**	2,95	1,27	0,35
Caterpillar Inc.	**LSTM**	1,03	0,62	0,36
Apple Inc.	**LSTM**	1,23	0,69	0,55
Microsoft Corp	**LSTM**	2,26	1,10	0,48
Nvidia Corp	**LSTM**	3,09	1,53	0,87
Broadcom Inc.	**LSTM**	2,35	1,63	0,38
Visa Inc. A	**LSTM**	1,27	0,62	0,32
Mastercard Inc.	**LSTM**	3,49	1,74	0,55
Udarbejdet i RStudio				

Tabel A.4 - Test af bias for ARIMA

Der testes for, hvorvidt der forekommer bias i specialets udvalgte ARIMA og LSTM modeller. Hertil angiver nulhypotesen, at modellen ikke er biased, givet ved $H_0 : \alpha = 0$ & $\beta = 0$. Her angiver $\alpha = 0$, at der ikke forekommer en niveauforskel mellem forecast og faktiske observationer, imens $\beta = 0$ betyder, at modellen ikke indeholder diskrepans, som betyder, at forecast og faktiske observationer ikke følges ad.

Tabel: A.4 Test for bias

Aktie	P-værdi	Aktie	P-værdi
Alphabet Inc. A	0,20	Alphabet Inc. C	0,29
Amazon.com Inc.	0,01	Tesla Inc	0,19
Home Depot Inc.	0,63	Berkshire Hathaway	0,06
JP Morgan	0,19	Bank Of America	0,50
UnitedHealth Group	0,07	Merck & Co Inc.	0,00
Abbvie Inc.	0,01	Eli Lilly	0,00
U. Parcel Service Inc.	0,20	Caterpillar Inc.	0,14
Apple Inc.	0,18	Microsoft Corp	0,07
Nvidia Corp	0,44	Broadcom Inc.	0,53
Visa Inc. A	0,04	Mastercard Inc. A	0,03
Udarbejdet i RStudio			

Tabel A.5 - Diebold-Mariano test - LSTM & ARIMA

Tabel: A.5 Diebold-Mariano test - LSTM & ARIMA

Aktie	M1$\neq M2$	M1 < M2	M1 > M2
Alphabet Inc. A	2.2e-16	2.2e-16	1,00
Alphabet Inc. C	2.2e-16	2.2e-16	1,00
Amazon.com Inc.	2.2e-16	2.2e-16	1,00
Tesla Inc.	0,91	0,45	0,55
Home Depot Inc.	7.42e-16	3.71e-16	1,00
Berkshire Hathaway	2.2e-16	2.2e-16	1,00
JP Morgan	2.2e-16	2.2e-16	1,00
Bank of America Corp	2.2e-16	2.2e-16	1,00
U. Health Group	2.2e-16	2.2e-16	1,00
Merck & CO Inc.	2.2e-16	2.2e-16	1,00
Abbive Inc.	2.2e-16	2.2e-16	1,00
Eli Lilly	2.41e-10	1.21e-10	1,00
U. Parcel Service	6.06e-08	3.03e-08	1,00
Caterpillar Inc.	2.2e-16	2.2e-16	1,00
Apple Inc.	2.2e-16	2.2e-16	1,00
Microsoft Corp	2.2e-16	2.2e-16	1,00
NVIDIA Corp	3.37e-09	1.68e-09	1,00
Broadcom Inc.	2.2e-16	2.2e-16	1,00
VISA Inc. A	2.2e-16	2.2e-16	1,00
Mastercard Inc	2.2e-16	2.2e-16	1,00
Udarbejdet i R studio			
M1: ARIMA & M2: LSTM			

Tabel A.6 - Diebold-Mariano test - Random Walk

Tabel: A.6 Diebold-Mariano test - Random Walk

Aktie & Model	M1$\neq M2$	M1$<$ M2	M1$>$ M2
Alphabet Inc. A ARIMA	0,02	0,99	0,01
Alphabet Inc. A LSTM	2.2e-16	1,00	2.2e-16
Alphabet Inc. C ARIMA	0,03	0,98	0,02
Alphabet Inc. C LSTM	2.2e-16	1,00	2.2e-16
Amazon.com Inc. ARIMA	0,00	1,00	0,00
Amazon.com Inc. LSTM	2.22e-16	1,00	2.2e-16
Tesla Inc. ARIMA	2.039e-05	1.02e-05	1,00
Tesla Inc. LSTM	0,86	0,43	0,57
Home Depot Inc. ARIMA	0,26	0,87	0,13
Home Depot Inc. LSTM	2.22e-16	1,00	2.2e-16
Berkshire Hathawa ARIMA	5.752e-05	1,00	2.876e-05
Berkshire Hathawa LSTM	2.2e-16	1,00	2.2e-16
JP Morgan ARIMA	0,00	1,00	0,00
JP Morgan LSTM	2.2e-16	1,00	2.2e-16
Bank Of America ARIMA	0,02	0,99	0,01
Bank Of America LSTM	2.2e-16	1,00	2.2e-16
UnitedHealth Group ARIMA	0,17	0,91	0,09
UnitedHealth Group LSTM	2.2e-16	1,00	2.2e-16
Merck & Co Inc. ARIMA	0,01	1,00	0,00
Merck & Co Inc. LSTM	2.2e-16	1,00	2.2e-16
Abbvie Inc. ARIMA	0,16	0,08	0,92
Abbvie Inc. LSTM	2.2e-16	1,00	2.2e-16
Eli Lilly ARIMA	0,00	1,00	8.197e-05
Eli Lilly LSTM	6.914e-11	1,00	3.457e-11

Tabel: A.6 Diebold-Mariano test - Random Walk

Aktie & Model	M1$\neq M2$	M1$<$ M2	M1$>$ M2
U. Parcel Service Inc. ARIMA	0,50	0,75	0,25
U. Parcel Service Inc. LSTM	1.759e-07	1,00	8.796e-08
Caterpillar Inc. ARIMA	0,04	0,02	0,98
Caterpillar Inc. LSTM	2.2e-16	1,00	2.2e-16
Apple Inc. ARIMA	0,18	0,91	0,09
Apple Inc. LSTM	2.2e-16	1,00	2.2e-16
Microsoft Corp ARIMA	3.042e-07	1,00	1.521e-07
Microsoft Corp LSTM	2.2e-16	1,00	2.2e-16
Nvidia Corp ARIMA	3.881e-07	1,00	1.941e-07
Nvidia Corp LSTM	3.666e-10	1,00	1.833e-10
Broadcom Inc. ARIMA	3.01e-08	1,00	1.505e-08
Broadcom Inc. LSTM	2.2e-16	1,00	2.2e-16
Visa Inc. A ARIMA	0,00	1,00	6.412e-05
Visa Inc. A LSTM	2.2e-16	1,00	2.2e-16
Mastercard Inc. A ARIMA	0,00	1,00	0,00
Mastercard Inc. A LSTM	2.2e-16	1,00	2.2e-16
Udarbejdet i R studio			
M1: Model angivet i tabel			
M2: Random Walk			

Tabel A.7 - Pesaran-Timmermann test

Tabel: A.7 Pesaran-Timmermann test

Aktie	Antal	Korrekte	Succesrate	P-værdi
Alphabet Inc. A ARIMA	753	441	58,57%	2.25e-06
Alphabet Inc. A LSTM	753	568	75,43%	0.00e+00
Alphabet Inc. C ARIMA	753	442	58,70%	1.69e-06
Alphabet Inc. C LSTM	753	581	77,16%	0.00e+00
Amazon.com Inc. ARIMA	753	438	58,17%	3.12e-06
Amazon.com Inc. LSTM	753	578	76,76%	0.00e+00
Tesla Inc. ARIMA	753	428	56,84%	1.80e-03
Tesla Inc. LSTM	753	587	77,95%	0.00e+00
Home Depot Inc. ARIMA	753	451	59,89%	8.21e-08
Home Depot Inc. LSTM	753	570	75,70%	0.00e+00
Berkshire Hathaway ARIMA	753	455	60,42%	5.24e-09
Berkshire Hathaway LSTM	753	577	76,63%	0.00e+00
JP Morgan ARIMA	753	445	59,10%	2.98e-07
JP Morgan LSTM	753	561	74,50%	0.00e+00
Bank Of America ARIMA	753	460	61,09%	5.94e-10
Bank Of America LSTM	753	572	75,96%	0.00e+00
UnitedHealth Group ARIMA	753	431	57,24%	4.60e-05
UnitedHealth Group LSTM	753	571	75,83%	0.00e+00
Merck & Co Inc. ARIMA	753	462	61,35%	2.33e-10
Merck & Co Inc. LSTM	753	577	76,63%	0.00e+00
Abbvie Inc. ARIMA	753	437	58,03%	1.33e-05
Abbvie Inc. LSTM	753	557	73,97%	0.00e+00
Eli Lilly ARIMA	753	461	61,22%	4.48e-10
Udarbejdet i Excel				

Tabel: A.7 Pesaran-Timmermann test

Aktie	Antal	Korrekte	Succesrate	P-værdi
Eli Lilly LSTM	753	575	76,36%	0.00e+00
U. Parcel Service Inc. ARIMA	753	443	58,83%	6.31e-07
U. Parcel Service Inc. LSTM	753	563	74,77%	0.00e+00
Caterpillar Inc. ARIMA	753	449	59,63%	8.80e-08
Caterpillar Inc. LSTM	753	571	75,83%	0.00e+00
Apple Inc. ARIMA	753	428	56,84%	1.23e-04
Apple Inc. LSTM	753	576	76,49%	0.00e+00
Microsoft Corp ARIMA	753	448	59,50%	1.41e-07
Microsoft Corp LSTM	753	568	75,43%	0.00e+00
Nvidia Corp ARIMA	753	419	55,64%	1.80e-03
Nvidia Corp LSTM	753	584	77,56%	0.00e+00
Broadcom Inc. ARIMA	753	420	55,78%	1.44e-03
Broadcom Inc. LSTM	753	556	73,84%	0.00e+00
Visa Inc. A ARIMA	753	451	59,89%	4.65e-08
Visa Inc. A LSTM	753	548	72,78%	0.00e+00
Mastercard Inc. A ARIMA	753	438	58,17%	4.71e-06
Mastercard Inc. A LSTM	753	562	74,63%	0.00e+00
Total ARIMA	15.060	8.847	58,75%	-
Total LSTM	15.060	11402	75,71%	-
Udarbejdet i RStudio				

Tabel A.8 - Porteføljе gennemsnitlige vægt

Tabel: A.8 Porteføljе gennemsnitlige vægt

Aktie	Sektor	Simpel	MV KH	MV KS	ARIMA	LSTM
Alphabet Inc. A	Komm.	5%	20%	-4%	4%	2%
Alphabet Inc. C	Komm.	5%	-10%	2%	4%	0%
Amazon.com	Dis. forbrug	5%	-7%	-2 %	1 %	0%
Home Depot Inc.	Dis. forbrug	5%	19 %	18 %	4 %	11 %
Tesla Inc.	Dis. Forbrug	5%	-2%	18 %	3 %	0 %
Bank of America	Finans	5%	-10%	-8%	16%	0%
B. Hathaway	Finans	5%	-10 %	-2%	16%	9%
JP Morgan	Finans	5%	19%	4%	4%	8%
Abbvie Inc.	Sundhed	5%	-10%	-4%	16%	6%
Eli Lilly	Sundhed	5%	15%	20%	4%	12%
Merck & Co.	Sundhed	5%	19%	5%	4%	15%
U. Health Group	Sundhed	5%	19%	18%	16%	10%
Caterpillar Inc	Industri	5%	-2%	-9%	2%	3%
U. Parcel Service	Industri	5%	-10%	10%	-4%	6%
Apple Inc.	IT	5%	-10%	19%	4%	3%
Broadcom Inc.	IT	5%	19%	-3%	1%	6%
Mastercard Inc.	IT	5%	19%	2%	-4%	4%
Microsoft Corp	IT	5%	19%	18%	3%	2%
Nvidia Corp	IT	5%	6%	4%	1%	-2%
Visa Inc.	IT	5%	-7%	-6%	4%	6%
Udarbejdet i RStudio						

15 Appendiks B

Følgende appendiks har til formål at klarlægge og uddybe resultaterne fremsat i specialet. Herunder illustreres alle formler, som anvendes i bilag B8 for mean-variance og ikke forklares i specialet.

Porteføljeudregninger

Det totalte afkast i % er udregnet ved start værdien fratrukket slutværdien inklusiv restriktioner, som divideres med slutværdien for at give det procentvise totale afkast:

$$(X_{slutvrdi} - X_{initialvrdi})/X_{initialvrdi}$$

Merafkast i % er udregnet ved start værdien fratrukket slutværdien inklusiv restriktioner, som divideres med slutværdien for at give det procentvise totale afkast. Til sidst fratrækkes benchmark for at give merafkast i %:

$$((X_{slutvrdi} - X_{initialvrdi})/X_{initialvrdi}) - X_{benchmark}$$

Dagligt afkast er udregnet som gennemsnittet af de daglige afkast inklusiv omkostninger, ved porteføljerne som handler:

$$(X_1 + X_2..X_n)/n$$

Min. afkast er det laveste daglige afkast og *maks. afkast* er det højeste daglige afkast.

Den daglige standardafvigelse beregnes som kvadratroden af den daglige varians, σ^2.

$$\sigma = \sqrt{\frac{\sum(X_i - \mu)^2}{N}}$$