Energikrisens Implikationer for en Dansk Plastvirksomhed

En analyse af priselasticitet og forecasting

Forfattere

Christian Øst Cloos

Denesan Shanmuganathan

Mathias Krogsgaard

Et speciale udarbejdet for kandidatuddannelsen

Cand.Oecon

Aalborg University Business School

Denmark

3. Juni 2022

Projektinformation

Titel:

Energikrisens Implikationer for en Dansk Plastvirksomhed

English Title:

The Implications of the Energy Crisis for a Danish Plastic Company

Forfattere:

Christian Øst Cloos

Denesan Shanmuganathan

Mathias Krogsgaard

Projektperiode:

Kandidatafhandling, 10. semester

Forårssemester 2022

Omfang:

Antal Sider: 79

Total: 114

Vejleder:

Daniel S. Hain

Forord:

Afhandlingen er udført af de tre forfattere i kollaboration med vejleder og samarbejdsvirksomhed. Processen har været afhængig af disse samarbejder, hvorfor forfatterne retter en stor tak til Daniel S. Hain og specielt to medarbejdere hos samarbejdsvirksomheden. Specialet ville ikke have samme kvalitet uden denne støtte og sparring.

Abstract

This paper seeks to examine how the energy crisis impacts a Danish company in the plastic molding industry through a cooperation agreement. The problem area is approached using a variety of econometric and statistic modeling techniques to analyze the implications for the company's revenue and construct forecasts hereof.

Based on the fact that the company can change its product pool between every time step, the paper has utilized proxies for its sales amount and price due to the heterogeneous structure of the company.

A short- and long-term price elasticity is estimated to study the implications of the energy crisis on the company's revenue. The SVAR-model yields that a 1% increase in the firm's sales price causes the sales of plastic products to fall by 0,15% and 0,17%, respectively. The analytical results turn out to be consistent with the hypothesis, which states the appearance of a relatively more elastic estimate in the long run. Based on the inelastic results, the energy crisis has a positive impact on the company's revenue.

By considering the combination of RMSE and graphical illustration, the VAR-model provides the best forecast of the company's revenue. The forecast obtains a total average deviation of 9,7% from the observed values, which is 0,5% point lower than the simpler univariate Prophet-model with XGBoost Errors. Therefore, it can be discussed whether the company needs such complex models for forecasting. Overall, the neural network models perform the worst, which might be due to the simplistic nature of regression and the shorter length of the dataset.

Indhold

1 Indledning **1**

1.1 Problemformulering 2

1.2 Afgrænsning 3

2 Redegørelse **5**

2.1 Energikrisen 5

2.2 Makroøkonomiske Implikationer 9

2.3 Operationalisering af Omsætningen 12

2.3.1 Samarbejdsvirksomheden 12

2.3.2 Priselasticitet 13

2.4 Delkonklusion 15

3 Metode **16**

3.1 Metodiske Refleksioner 16

3.2 ARMA 17

3.2.1 Stationaritet 17

3.2.2 ARIMA 18

3.2.3 Informationskriterier 19

3.3 (S)VAR 20

3.3.1 Modelegenskaber 21

3.3.2 Den Reducerede VAR 23

3.3.3 Stationaritetsbetingelser for en VAR-model 24

3.3.4 Antallet af Lags i en VAR-model 26

3.3.5 Forecasting i en VAR-model . . . 26
3.3.6 Den Strukturelle VAR . . . 27
3.3.7 Tilgang til Identifikation af den Strukturelle VAR . . . 29
3.3.8 Impulse Response Funktioner . . . 31
3.3.9 Estimering af Egenpriselasticiteter . . . 32
3.4 Maskinlæring . . . 33
3.5 Neurale Netværk . . . 34
3.5.1 Modelarkitektur . . . 34
3.5.2 Recurrent Neural Network . . . 36
3.5.3 Rolling Forecast . . . 38
3.6 Præstationsmål . . . 39
3.7 Delkonklusion . . . 40

4 Datastruktur **42**
4.1 Udvælgelse af Variable . . . 42
4.1.1 Proxy Variable . . . 45
4.2 Datamanipulation . . . 46
4.3 Delkonklusion . . . 47

5 Analyse Del 1 **49**
5.1 Egenpriselasticiteter i en SVAR-model . . . 49
5.1.1 Estimat- og Hypotesebeskrivelse . . . 49
5.1.2 Modelspecifikationer . . . 50
5.1.3 Analyseresultater . . . 51

6 Analyse Del 2 **57**
6.1 Baseline ARIMA . . . 57
6.1.1 Modelspecifikationer . . . 57

6.1.2 Forecasting og Præstationsevne . . . 58
6.2 VAR Forecast . . . 59
6.2.1 Modelspecifikationer . . . 60
6.2.2 Forecasting og Præstationsevne . . . 61
6.3 Maskinlæring Forecast . . . 62
6.3.1 Modelspecifikationer . . . 62
6.3.2 Forecasting og Præstationsevne . . . 63
6.4 Neurale Netværk Forecast . . . 64
6.4.1 Modelspecifikationer . . . 65
6.4.2 Forecasting og Præstationsevne . . . 65
6.5 Resultater . . . 67
6.6 Delkonklusion . . . 70

7 Diskussion **71**
7.1 Refleksion over de Analytiske Resultater . . . 71
7.2 Ruslands Invasion af Ukraine . . . 72
7.3 Begrænsninger . . . 73
7.3.1 Fremtidig Forskning og Udvikling . . . 75

8 Konklusion **77**
8.1 Vidensbidrag . . . 78
8.1.1 Industriel indvirkning . . . 79
8.1.2 Politiske overvejelser . . . 79
8.1.3 Teoretiske bidrag . . . 79

9 Litteratur **80**

A Bilag **84**
A.1 Sæson-Trend-Loess Decomposition (STL) . . . 84

A.2 Sæsonmæssige Variationer og Unit Root 88
A.2.1 Undersøgelse af Sæsonmæssige Variationer 88
A.2.2 Undersøgelse af Unit Root 89
A.3 Opsætning af SVAR-model 91
A.4 Opsætningen af ARIMA-model 97
A.4.1 Opsætning af VAR-model 105
A.5 Inflation i Importeret Plast 107

1 | Indledning

I år 2021 blev Danmark og resten af Europa ramt af markante energiprisstigninger, som er forårsaget af faktorer såsom vejrforhold og coronakrisen. Den aktuelle energikrise har haft implikationer for husholdninger såvel som virksomheder i form af højere el- og varmeregninger. En kombination af et lavere rådighedsbeløb blandt forbrugere samt højere salgspriser, pålagt af den industrielle sektor, har manifesteret sig i den europæiske økonomi. I denne forbindelse tillægges landenes indbyrdes konkurrenceevne en øget betydning, som derved i større grad er bestemt af deres interne energiafhængighed i den pågældende krise.

De højere salgspriser har implikationer for efterspørgsel og derigennem omsætning for den industrielle sektor. Herunder opstår der ydermere en øget usikkerhed i forhold til virksomhedens forventningsdannelse, hvor den fremtidige planlægning forstyrres. Et værktøj i forventningsdannelsen kan således være fordelagtigt i form af økonometriske og statistiske modelopsætninger.

I afhandlingen foretages der en undersøgelse af energikrisen og dets implikationer for en industriel virksomhed ved at indgå en samarbejdsaftale med en dansk virksomhed indenfor plastindustrien[1]. De opsatte problemstillinger analyseres ved brug af økonometriske og statistiske modelopsætninger, hvor det førstenævnte undersøges ved at tage udgangspunkt i et nyere studie af Yilmazkuday, 2019. I dette studie anvendes en Structural Vector Autoregressive model (SVAR) til at estimere handelselasticiteter over flere tidshorisonter. I kandidatafhandlingen appliceres den samme tilgang til at undersøge, hvor følsom samarbejdsvirksomhedens efterspørgsel er overfor en prisændring for derved at identificere implikationerne af energikrisen på virksomhedens omsætning. Dernæst behandles problemstillingen omkring forventningsdannelsen ved at benytte en variation af forecast-modeller for at skabe en grad af diversitet.

[1] Betegnes 'Samarbejdsvirksomheden' i kandidatafhandlingen.

I denne forbindelse er det nødvendigt at indsamle data, der opfanger sammenhængen mellem energipriser og økonomiske variable. I Narayan et al., 2014 findes der belæg for en sammenhæng mellem oliepris og den økonomiske vækst i udviklede lande, hvilket understøtter undersøgelsen af afhandlingens problemstilling. Derudover er der, ifølge litteraturreviewet Ghoddusi et al., 2019, manglende forskning indenfor brugen og sammenligning af komplicerede modeller indenfor energiøkonomi, herunder neurale netværk.

1.1 Problemformulering

På baggrund af den indledende del søger kandidatafhandlingen at besvare følgende problemstilling:

Hvordan påvirkes samarbejdsvirksomheden af den aktuelle energikrise?

Besvarelsen knyttes op til de følgende underspørgsmål for at skabe en mere fyldestgørende besvarelse.

Disse tre underspørgsmål vil blive besvaret igennem afhandlingen:

- Hvad har forårsaget den aktuelle energikrise?
- Hvilke implikationer har energikrisen for samarbejdsvirksomhedens omsætning?
- Hvilken statistisk model præsterer bedst til forecast af omsætningen i kriseperioden?

I den indledende del redegøres der for den aktuelle energikrise, hvor sammenhængen i energimarkedet beskrives. I denne forbindelse belyses såvel de makroøkonomiske implikationer af energikrisen for den danske økonomi ved at betragte husholdnings- og virksomhedssektoren. Derudover introduceres samarbejdsvirksomheden efterfulgt af det mikroøkonomiske begreb, priselasticitet, som senere appliceres i den analytiske sektion.

Herefter inddrages der et metodologisk kapitel, hvor relevante økonometriske og statistiske metoder inddrages i overensstemmelse med de etablerede metodiske refleksioner. I det efterfølgende kapitel præsenteres og udvælges dé variable, der anses som mest velegnede til operationalisering af problemstillingen. Sammenspillet mellem disse kapitler danner således fundamentet for den analytiske

besvarelse.

Den analytiske sektion er todelt, hvor den første del analyserer implikationerne af energikrisen for samarbejdsvirksomhedens omsætning. I denne forbindelse estimeres det, hvor følsom virksomhedens efterspørgsel er overfor en prisændring på kort og lang sigt igennem anvendelsen af en SVAR-model. I den anden halvdel af analysen anvendes økonometriske og statistiske modeller til forecasting af omsætningen, som efterfølgende evalueres ud fra 'Root Mean Square Error' (RMSE) samt grafisk præstation.

I den afsluttende del diskuteres de analytiske resultater, der er opnået igennem udarbejdelsen af afhandlingen. Derudover inddrages Ruslands invasion af Ukraine og dets indvirkning på energikrisen og deraf samarbejdsvirksomheden. Herefter beskrives de begrænsninger, der er erfaret, og herunder hvad fremtidig forskning og udvikling, kan bidrage med. Afslutningsvis besvarer konklusionen afhandlingens problemstilling ved at tage udgangspunkt i de opsatte delkonklusioner. I denne sammenhæng beskrives de industrielle, politiske og teoretiske indvirkninger, som afhandlingen giver anledning til.

I det følgende afsnit gøres det eksplicit, hvilke valg og fravalg, der foretages med formålet om at indsnævre det valgte problemfelt. Derudover fastlægges der definitioner af begreber, som anvendes igennem afhandlingen.

1.2 Afgrænsning

I kandidatafhandlingen tages der udgangspunkt i data fra en dansk virksomhed indenfor plaststøberi. Afhandlingen afgrænser sig da fra at behandle plastmarkedet som helhed, da hensigten er at udarbejde mere specifikke og dermed brugbare analyseresultater for den enkelte virksomhed. Dette udelukker dog ikke, at andre virksomheder indenfor plastindustrien eller lignende energiintense industrier kan anvende afhandlingens vidensbidrag. For at opfange energikrisen er det nødvendigt at erhverve data, som er opgjort på månedlig basis, da kriseperioden er relativ kort på

nuværende tidspunkt, og forespørgslen imødekommes af samarbejdsvirksomheden.

Produktionen i virksomheden er ordrebestemt, hvilket medfører et bredt heterogent produktsortiment. I denne forbindelse kan hele produktsortimentet i princippet blive udskiftet mellem hver periode, hvilket foranlediger strukturelle ændringer over tid. Derfor er det nødvendigt, i udførelsen af en troværdig og repræsentativ analyse, at betragte homogene mål for salgsmængde og pris, som ikke ændrer struktur over tid. I afhandlingen anvendes der således proxy'er, som formår at homogenisere sammenhængene imellem produkter på tværs af perioder.

Energiforbruget anvendes som proxy for salget af plastprodukter. Samarbejdsvirksomheden har, igennem de seneste år, foretaget energieffektiviseringer i samarbejde med Energistyrelsen (Plast Forum, 2020). Derudover har virksomheden oplyst, at der kun foretages maskinopgraderinger, når det er nødvendigt, hvilket gør, at der ikke forekommer større uventede udskiftninger af maskiner, som deraf optræder som brud i tidsserien. Det oplyses samtidigt, at virksomheden har haft stor fokus på, at udstyr og maskiner ikke er i unødvendig brug. I denne forbindelse antages det, at teknologiniveauet følger en relativ konstant udvikling, hvilket tilnærmelsesvis isolerer de produktionsmæssige udsving i tidsserien.

I den første del af afhandlingens analyseafsnit undersøges det, hvor følsom efterspørgslen efter plastprodukter er ved en ændring i salgsprisen, da målet er at analysere implikationerne af energikrisen på samarbejdsvirksomhedens omsætning. Efterspørgslen skal således forstås som virksomhedens salg af støbte plastprodukter, og det antages dermed, at al efterspørgsel imødekommes af en tilsvarende produktion. Der afgrænses fra andre typer af elasticiteter end priselasticitet, da denne mikroøkonomiske størrelse er tilstrækkelig for at tolke på implikationerne af energikrisen på samarbejdsvirksomhedens omsætning. Herunder afgrænses der ligeledes fra en undersøgelse af markedsforhold og andre tilhørende mikroøkonomiske aspekter.

Der afgrænses fra en betragtning af samarbejdsvirksomhedens profit, da den ønskede besvarelse udelukkende omfatter energikrisens påvirkningen på salget af plastprodukter og derigennem omsætningen i virksomheden.

2 | Redegørelse

I dette kapitel redegøres der for, hvilke faktorer, der har været årsag til den aktuelle energikrise. I denne forbindelse præsenteres der deskriptiv statistik for en række energipriser i perioden 2011-2022 for at sammenholde faktorer med den observerede udvikling. Derudover betragtes krisen fra et makroøkonomisk perspektiv, hvor implikationerne for den danske økonomi ekspliciteres. Dernæst beskrives samarbejdsvirksomheden og det mikroøkonomiske begreb, priselasticitet, som benyttes i analyseafsnittet.

2.1 Energikrisen

Kort forklaret omfatter energikrisen høje energipriser, som primært skyldes en mangel på vedvarende energi, hvilket har skabt en større afhængighed af naturgas. Naturgas er et fossilt brændstof, som er den tredjestørste energikilde på verdensplan. Denne findes naturligt i undergrunden og anvendes til produktion af varme og strøm (DR1, 2022).

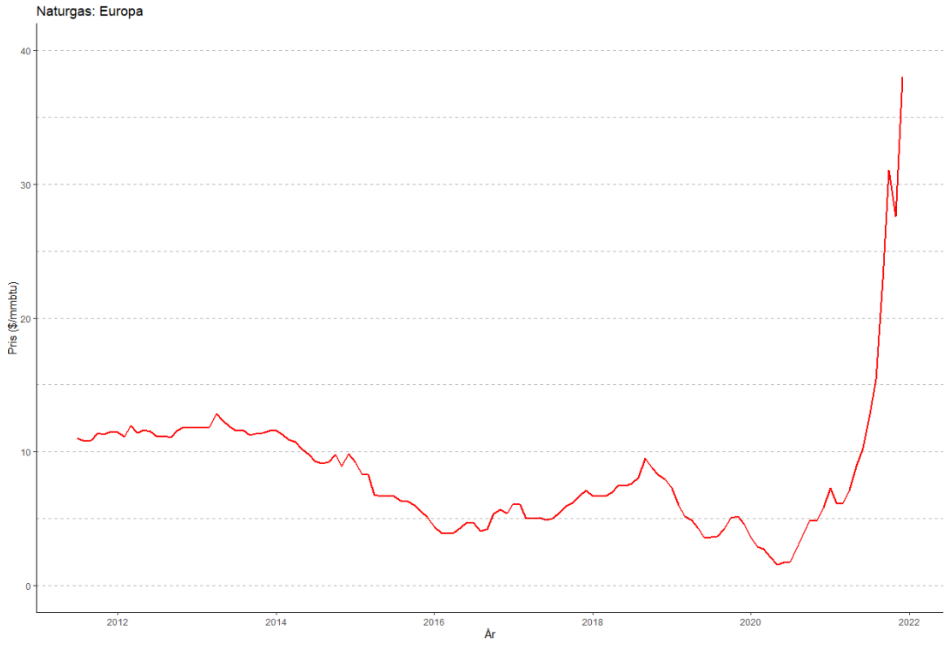
Den danske energiproduktion består af både fossile brændstoffer (olie, kul og naturgas) samt vedvarende energikilder (vind, solenergi og biomasse). I Danmark er gasnettet konstrueret af et transmissions- samt distributionsnet, som fordeler gassen til slutbrugerne. Transmissionsgasnettet forbinder landsdelene sammen og skaber en forbindelse til produktionen af gas i Nordsøen og til det europæiske gasnet igennem Tyskland og Sverige (Dansk Gasteknisk Center, n.d.). Derudover importerer Danmark energi fra private producenter og industrielle værker fra ind- og udland, herunder lande såsom Norge, Sverige og Tyskland, som producerer vha. vandkraft (EnergiFyn, 2021).

Energikrisen, der ramte i år 2021, opstod på baggrund af uhensigtsmæssige vejrforhold. I Danmark

stammer 80% af energiforbruget fra vedvarende energikilder. En blanding af mindre blæst og regn i år 2021 betød, at produktionen af energi fra de danske vindmøller faldt samtidigt med, at billig strøm fra vandkraftværkerne i Norge og Sverige ikke kunne erhverves i samme grad som tidligere. Et begrænset udbud af energi har således forårsaget en stigning i prisen. Når vedvarende energikilder ikke producerer nok strøm, anvender Danmark fossile brændstoffer som et alternativ. Problemet opstår hermed, idet Europas gaslagre har været forholdsvist lave. En kold vinter i år 2020 efterfulgt af en varm sommer i 2021 har skabt en øget efterspørgsel efter naturgas, hvilket har lagt pres på naturgaslagrene (DR1, 2022).

En anden årsag til energikrisen og de pressede gaslagre har været coronapandemien. I år 2020 lukkede Danmark og resten af verden ned med en forventning om en længerevarende økonomisk krise. I denne forbindelse valgte mange af gasproducenterne at lade ikke-kritiske funktioner arbejde hjemmefra og herigennem skrue ned for produktionen, selvom der var et eksisterende behov for naturgas. Flere energiproducerende virksomheder var samtidigt tvunget til det samme eller helt at stoppe produktion af energi pga. coronasmittetilfælde i virksomheden og dertil isolation af medarbejdere (EnergiWatch, 2022). I år 2021, hvor store dele af verdensøkonomien kom i gang, var udbuddet af naturgas betydeligt lavere end efterspørgslen, hvilket pressede prisen på naturgas yderligere opad (DR1, 2022).

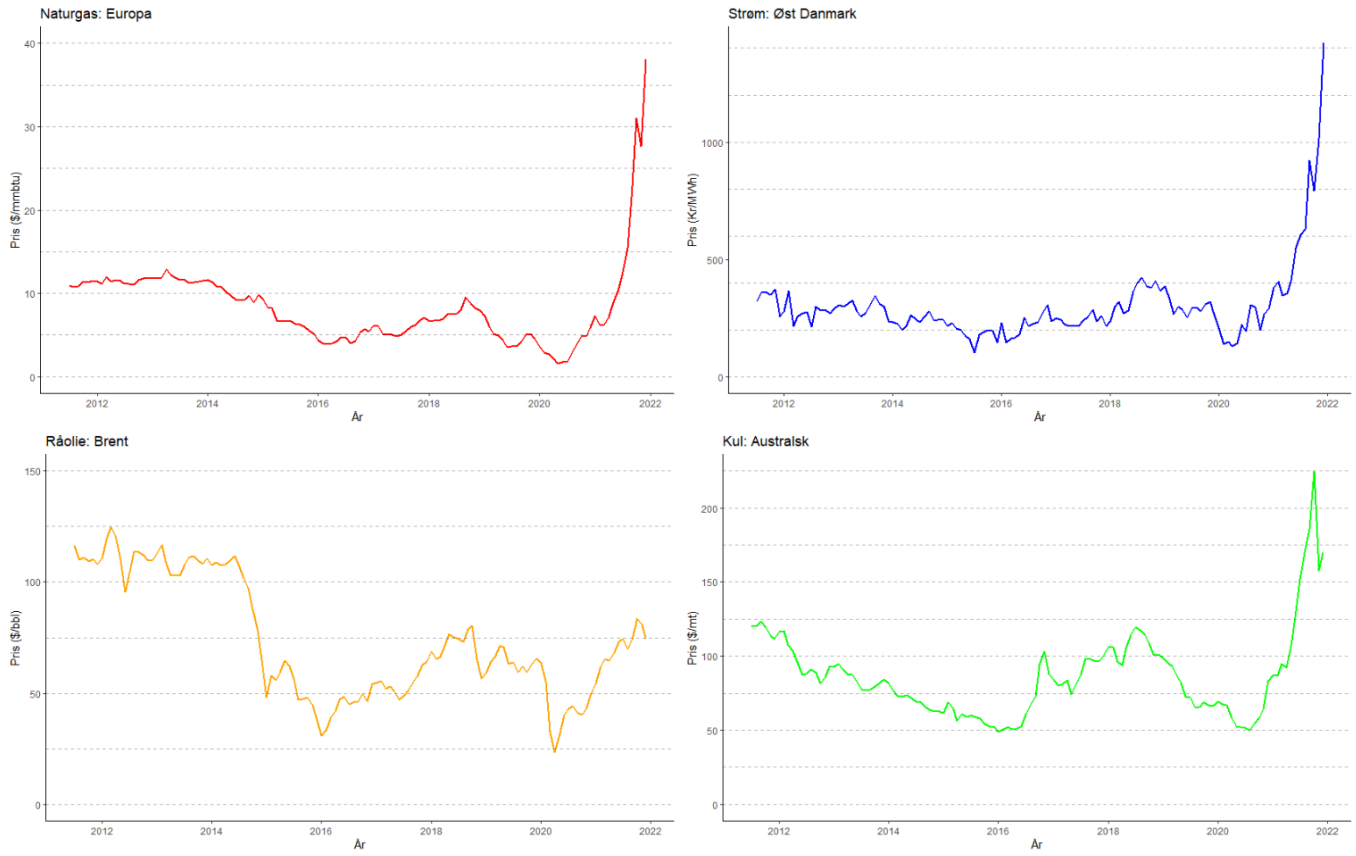
De nævnte årsager kan forklare den observerede udvikling i prisen på naturgas på det europæiske energimarked i år 2021, som illustreres i Figur 2.1. Forud for prisstigningen forekom der et lavpunkt i 2020, hvilket har styrket den generelle oplevelse af prisen på naturgas, som er mere end tidoblet fra lavpunktet.



Figur 2.1: Naturgas prisudvikling.

Kilde: Energistyrelsen, 2022

I Figur 2.2 sammenholdes Figur 2.1 med prisen på el, råolie og kul for at beskrive sammenhængen mellem priserne på energimarkedet. Herudfra illustreres det, at den stigende prisudvikling i naturgas og el også afspejles i mindre grad i kul. Årsagen til denne sammenhæng er, at kul, ligesom naturgas, kan benyttes til at producere strøm. Afhængigheden af kul er dog faldet over tid, og der er implementeres i større grad vedvarende energikilder. Den svagere sammenhæng mellem el og kul kan observeres i figuren, hvor kulprisen ikke er steget så kraftigt som for naturgas (Energi Danmark, 2018). En alternativ måde at producere strøm på er igennem benyttelsen af olie. Afhængigheden af olie er faldet og endda mere end kul, hvilket afspejles i prisudviklingen på råolie, som viser sig ikke at have en synlig sammenhæng med elprisen. Hertil nævnes det, at lande som f.eks. Danmark har stoppet al produktion af el baseret på olie (NRGI, 2022).



Figur 2.2: Prisudvikling for energi.

Kilde: Energistyrelsen, 2022

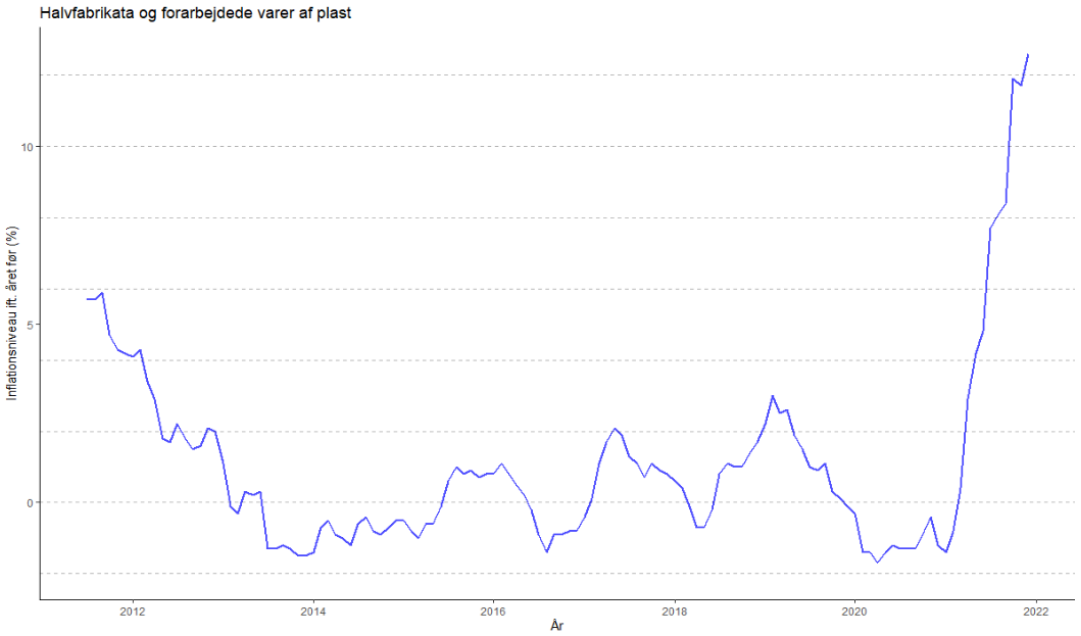
Ud fra disse sammenhænge kan det derfor vises, at elprisen i Danmark, først og fremmest, er påvirket af produktionen fra de vedvarende energikilder i Norden. Dernæst påvirkes elprisen af presset på naturgaslagrene, der benyttes som et alternativ til strømproduktion (DR1, 2022). Hvis disse to produktioner ikke er tilstrækkelige, importeres der europæisk naturgas og endnu et alternativ er da kul. Der bliver ikke i større grad benyttet olie til produktion af el.

Priserne på alternative energikilder afhænger dermed af deres nyttighed i produktionen af el og påvirker derigennem husholdnings- og virksomhedssektoren igennem energinødvendigheden. De makroøkonomiske implikationer for den danske økonomi beskrives i det følgende afsnit.

2.2 Makroøkonomiske Implikationer

I det forrige afsnit blev det beskrevet, hvordan en blanding af mangel på vedvarende energi fra ind- og udland samt et begrænset udbud af naturgas i Europa har været nogle af de primære årsager til den aktuelle energikrise.

Højere energipriser har negative effekter på både husholdnings- og virksomhedssektoren igennem flere kanaler, som derigennem har implikationer for den danske økonomi. Den første negative effekt indebærer højere el- og varmeregninger (DR1, 2022), som, alt andet lige, påvirker den enkelte husholdning negativt igennem et fald i rådighedsbeløbet. En lavere indkomst medfører en lavere efterspørgsel i økonomien, hvilket foranlediger et fald i produktionen og derigennem indkomsten på baggrund af en lavere beskæftigelse. Den anden effekt optræder igennem en øget inflation i den danske økonomi, som skyldes prisstigninger på brændstof, elektricitet og gas (Danmark & Bernhard, 2022; Danmarks Statistik, 2022). De stigende energipriser kan ikke kun mærkes hos de danske forbrugere, men såvel de danske virksomheder, som ligeledes oplever øgede omkostninger til el og varme såvel som transport af varer. Omkostningspresset resulterer i et prispres, som i sidste ende rammer den enkelte husholdning (Jersild & Munch, 2022). Som et eksempel afspejler Figur 2.3 i sammenhæng med Figur 2.1 empirisk, at de stigende energipriser i høj grad er blevet pålagt slutbrugeren, når der tages udgangspunkt i markedet for plast. En øget inflation i økonomien forårsager, at husholdningens købekraft, alt andet lige, falder. Som forklaret ovenfor, vil dette bevirke til et fald i efterspørgslen efter varer og dermed et fald i den danske BNP-vækst.

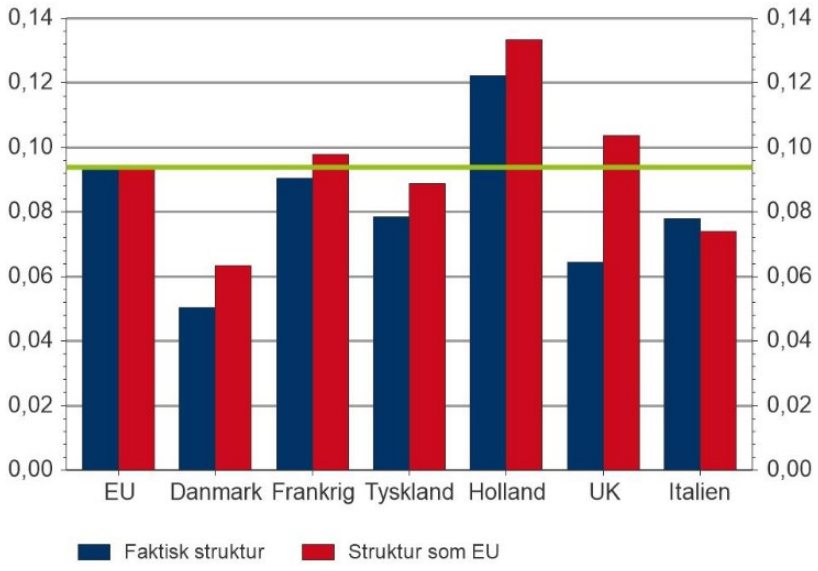


Figur 2.3: Udvikling i inflation for plastmarkedet.

Kilde: Danmarks Statistik, 2015

Energikrisen har medført højere priser til de danske forbrugere igennem øgede produktions- og transportomkostninger. Hvor meget prisen stiger for den enkelte husholdning afhænger af, hvor sensitiv den danske virksomhedssektor er overfor prisstigninger i energi. Et mål for dette er energiintensiteten, som angiver hvor effektivt energi anvendes og defineres som forholdet mellem energiforbrug og output. En virksomheds eller branches energiintensitet kan være en måde at måle, hvor energieffektiv branchen eller virksomheden er set i forhold til lignende virksomheder og brancher. Jo mere effektiv en branche er, desto mindre sensitiv vil den være overfor ændringer i energipriserne. Energiintensitet kan således være et mål for den indenlandske såvel som udenlandske priskonkurrenceevne. Hvis den danske industri er forholdsvist mere energieffektiv og dermed mindre energiintensiv i forhold til udlandet, så vil danske virksomheder være mindre følsom overfor pludselige ændringer som f.eks. energikriser kan forårsage. Den danske industri vil dermed stå stærkere i forhold til udlandet i form af en bedre priskonkurrenceevne (Energistyrelsen, n.d.)

I det følgende anvendes EU Kommissionens definition af energiintensitet fra det fælles europæiske projekt, Odyssee-Mure, for at bidrage med en fælles indikator for energieffektivitet. Denne fælles indikator er funderet på fælles retningslinjer således, at en retvisende sammenligning mellem deltagerlandene kan udarbejdes. Odyssee-Mure definerer energiintensiteten som energiforbruget målt i kilo olieækvivalenter pr. bruttoværditilvækst i euro i faste priser dvs. mængden af energi, som virksomhederne bruger, når de skal skabe en euro i værdi (Energistyrelsen, n.d.). Figur 2.4 viser en sammenligning af energiintensiteten mellem EU som helhed og specifikke lande som Danmark, Frankrig, Tyskland, Holland, UK og Italien. Ud fra de blå søjler, som måler graden af energiintensitet med den faktiske struktur i det pågældende land eller område, kan det aflæses, at Danmark udnytter dets industrisammensætning. Et eksempel, der kan nævnes herved, er medicinalsektoren, som skaber meget værdi ved et forholdsvist beskedent forbrug af energi. Ved at korrigere for forskelle i branchestruktur, dvs. ved aflæsning af den røde søjle, dannes et mere retvisende sammenligningsgrundlag. Ud fra dette mål observeres det, at danske virksomheder har en lavere energiintensitet relativt til de andre udvalgte lande (Kristensen, 2021).



Figur 2.4: Sammenligning af energiintensitiet med og uden strukturkorrektion for 2019.
Kilde: Kristensen, 2021

Opsummerende er de danske virksomheder bedre til at ekstrahere værdi ud af energien og dermed mindre afhængige af energien relativt til andre lande, hvilket giver en priskonkurrencefordel i energikrisen. Når de danske virksomheder har en relativt højere energieffektivitet og i denne sammenhæng en lavere energiintensitet, medvirker det til, at virksomhedernes efterspørgsel efter energi ikke er så høj som i andre lande. Dette betyder, at virksomhederne er forholdsvist mindre sensitive overfor prisændringer i energi, og inflationen i den danske økonomi stiger dermed mindre. De danske virksomheder står altså stærkere i energikrisen internationalt set, da den danske priskonkurrence er relativt bedre. Dette er gunstigt for den danske eksport, som, alt andet lige, vil falde mindre end andre lande, som ligeledes implicerer en mindre negativ indvirkning på den danske produktion såvel som indkomst (Kristensen, 2021). Energikrisen påvirker den danske økonomi negativt igennem højere priser, men i mindre grad end andre lande i Europa.

2.3 Operationalisering af Omsætningen

I dette afsnit redegøres der for afhandlingens mikroøkonomiske fundament. Indledningsvist redegøres der kort for samarbejdsvirksomheden og den tilhørende værdikæde. Derudover redegøres der for priselasticitet, som nærmere kan specificere energikrisens effekt på virksomhedens omsætning.

2.3.1 Samarbejdsvirksomheden

Samarbejdsvirksomheden er afhængig af deres omkostninger til bl.a. input og energi. Eftersom virksomheden støber plastelementer i forskellige størrelser, medfører det, at omkostningerne, der er forbundet med fremstillingen af disse varer, vil have en stor variation, da der skal bruges forskellige mængder plast og energi til produktionen. Denne produktion er baseret på maskiner, der støber deres plastelementer. Virksomheden producerer selv deres støbeforme, hvilket resulterer i en mere omstillingsparat produktion af plastprodukter og dermed kortere leveringstid. Disse egenskaber øger virksomhedens konkurrenceevne i markedet.

Derudover er virksomheden ordrebestemt, hvilket skaber muligheden for at indgå aftaler om produktion af større partier. Dette betyder således, at der er mulighed for at sænke inputprisen såvel som de marginale omkostninger ved produktionen, når der indgås en større produktionsordre. En gensidig fordelagtig situation opstår således for begge parter, da salgsprisen sænkes for køberen, og produktionen sikres for virksomheden (Qin et al., 2006). Hertil forbeholder virksomheden sig retten til at kunne genkalkulere salgsprisen på baggrund af væsentlige omkostningsændringer.

Den gensidig fordelagtige situation er gældende, så længe virksomheden kan rumme den ønskede produktionskapacitet, og de marginale omkostninger ikke bliver større ved at producere en ekstra enhed. Salgsprisen afspejler derfor en kombination af direkte og indirekte omkostninger som bl.a. produktionsomkostninger og lønninger, hvortil der pålægges en mark-up. Der vil således være en specifik salgspris for hver ordre, hvor den samlede ordrepris er et resultat af salgsprisen og outputmængden (Bhimani, 2019). Virksomhedens operationelle omsætning vil blive givet af de akkumulerede ordresalgspriser. Hvordan denne påvirkes af stigende omkostningen til energi kan forklares igennem priselasticitet som gennemgås i følgende afsnit.

2.3.2 Priselasticitet

Priselasticitet er et procentuelt mål for, hvor meget efterspørgslen efter en given vare ændrer sig ved en ændring i prisen på varen. Begrebet er derfor et mål for, hvor sensitiv forbrugsmarkedet er overfor en given prisændring. Størrelsen på priselasticiteten er bestemmende for, hvordan en virksomheds omsætning påvirkes (Mankiw & Taylor, 2011).

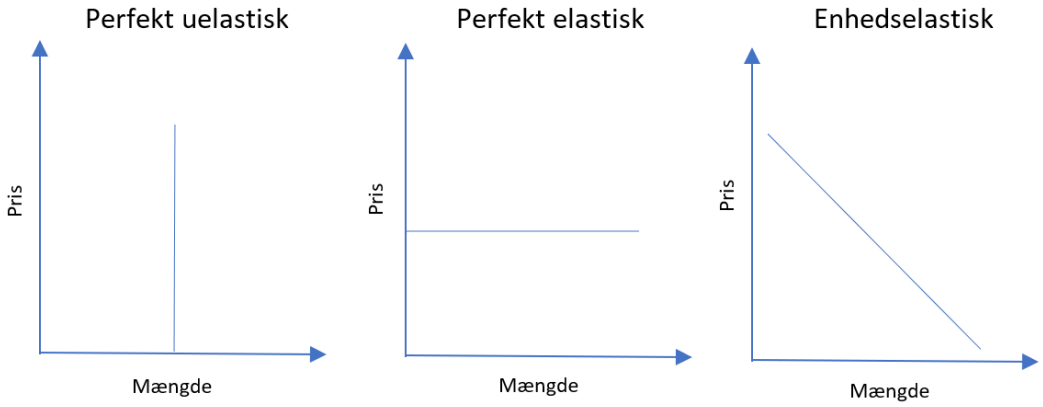
Efterspørgslen efter et givet produkt kan karakteriseres som værende af uelastisk, enhedselastisk eller elastisk karakter j.f. Figur 2.5, og hver af disse påvirker omsætningen på forskellig vis. Hvis efterspørgslen efter en given vare er uelastisk, hvilket typisk er gældende for nødvendighedsvarer, implicerer det, at den procentvise ændring i efterspørgslen er mindre end den procentvise ændring i prisen. I dette tilfælde ligger priselasticiteten mellem 0 og -1[1]. En uelastisk efterspørgsel vil således

[1]Der kan ses bort fra minustegnet, som blot udtrykker den negative procentuelle ændring i efterspørgslen.

påvirke en given virksomheds omsætning positivt. I tilfælde af perfekt uelasticitet, som angiver et ekstremt scenarie, forekommer der ikke en procentuel ændring i efterspørgslen ved en prisstigning.

Hvis efterspørgslen er elastisk, hvilket oftest er forbundet med luksusgoder, betyder det, at den procentvise ændring i efterspørgslen er større end den procentvise ændring i prisen. Priselasticiteten påtager sig en værdi, der er større end 1. En elastisk efterspørgsel har en negativ indvirkning på en given virksomheds omsætning. I tilfælde af perfekt elasticitet forekommer der en uendelig stor procentuel ændring i efterspørgslen på baggrund af en prisstigning.

Hvis efterspørgslen efter en given vare derimod er enhedselastisk dvs. priselasticiteten er lig 1, vil den procentuelle ændring i efterspørgslen være proportional med prisændringen. Implikationen af dette er, at en given virksomheds omsætning vil være uændret (Varian, 2019). I denne afhandling ses der bort fra perfekt priselasticiteter, da disse blot er ekstreme og dermed ikke-realistiske scenarier (Nechyba, 2011).



Figur 2.5: Scenarier for priselasticitet.

Kilde: Egen skitsering

Teoretisk set vil elasticiteten være relativt mere elastisk på længere sigt, da køberen vil tilpasse sit forbrug eller indkøb over tid ved f.eks. substitution. Efterspørgslen bliver således mere følsom overfor prisændringen. I forbindelse med priselasticitet kan der udregnes en såkaldt egenpriselasticitet,

hvilket blot er et procentuelt mål for, hvor meget en 1% stigning i prisen på en given vare påvirker efterspørgslen efter varen, hvilket er typen, der appliceres i afhandlingens analysedel (Nechyba, 2011).

2.4 Delkonklusion

Dette redegørende kapitel introducerede tre dele, hvor der indledningsvis blev redegjort for den aktuelle energikrise, og hvordan den opstod på baggrund af coronapandemiens nedlukning og manglende vedvarende energi. Dette inkluderede en empirisk forklaring af udviklingen i prisen på strøm og dets sammenhæng med alternative energikilder. Herunder blev det illustreret, at naturgas og kul benyttes til strømproduktion i Danmark, og at olie ikke anvendes.

Sektionen efter præsenterede de makroøkonomiske implikationer for den danske økonomi, hvor sammenhængen mellem højere energipriser og lavere BNP, alt andet lige, blev fastlagt. Dernæst blev begrebet energiintensitet introduceret for at beskrive en virksomheds energieffektivitet. I denne forbindelse bidrog Odyssee-Mure projektet, udarbejdet af EU-kommissionen, med en fælles indikator for energieffektivitet. Herigennem blev det illustreret, at Danmark står stærkere i energikrisen relativt set på baggrund af en lavere energiintensitet. Danmark opnår således en konkurrencemæssig fordel overfor andre lande.

Afslutningsvis blev samarbejdsvirksomheden og dens tilhørende værdikæde beskrevet. Dernæst blev begrebet, priselasticitet, gennemgået, hvor der blev opsat tre scenarier. Disse scenarier havde forskellige implikationer for en given virksomheds omsætning.

3 | Metode

Dette kapitel gennemgår det metodologiske grundlag for kandidatafhandlingens analyseafsnit. Herunder inddrages ARIMA-modellen, som bidrager med et forecast, der betragtes som baseline. Dernæst redegøres der for et højere økonometrisk niveau igennem anvendelsen af en (S)VAR-model. Ydermere indebærer afsnittet en forklaring af impulse response funktioner, og hvordan disse appliceres som et analytisk redskab i afhandlingen. Afslutningsvis præsenteres maskinlæring og herigennem neurale netværk for at udnytte disse databehandlingsegenskaber som en alternativ forecast-metode. Hvorfor netop disse metoder er udvalgt er uddybet i følgende segment.

3.1 Metodiske Refleksioner

ARIMA-modellen er en del af metodeafsnittet, da det er fordelagtigt at sammenligne andre modeller med en simpel univariate baseline model. Modellen er valgt, da den anvender både tidligere værdier samt et bevægende gennemsnit i forklaringen af samarbejdsvirksomhedens omsætning. En algoritme gør det bekvemt at definere den bedste ARIMA-model, hvor der såvel tages højde for stationaritet. Et alternativ til kombinationen af disse led er typisk at betragte en AR(1)-proces, hvilket anses som værende for simpel en model. (S)VAR-modellen inddrages som udvidelse til ARIMA-modellen, da inkluderingen af flere variable kan benyttes til at øge forklaringsgraden af udviklingen i samarbejdsvirksomhedens omsætning. Et alternativ til dette er en Autoregressive Distributed Lag (ARDL)-model, hvor det ikke er nødvendigt for inkluderede variable at være endogene. Da der arbejdes med økonomiske variable i afhandlingen, er det problematisk at retfærdiggøre eksistensen af asymmetri, dvs. at variablene ikke påvirker hinanden gensidigt, hvilket fører til anvendelsen af (S)VAR-modellen. SVAR-modellen bidrager med en undersøgelse af energikrisens virkning på omsætningen ved brug af impulse response funktioner og i denne sammenhæng priselasticiteter.

Et alternativ hertil vil være en Error-Correction Model (ECM), hvor både kort- og langsigtede effekter kan findes. Fremkomsten af de langsigtede effekter kræver, at der eksisterer et langsigtet forhold mellem ikke-stationære variable (cointegration) (Enders, 2014), hvilket ikke altid er tilfældet i praksis. Neurale Netværk implementeres for at bidrage med en alternativ vinkel på kandidatafhandlingens problemstilling. Derudover, jf. Afsnit 1, er der et uudnyttet potentiale i opsætningen af energiøkonomiske modeller. Der findes herigennem flere metodiske muligheder, i denne afhandling er der fokus på netværk, der excellerer ved sekventiel data.

Det teoretiske grundlag for hver af de nævnte metoder beskrives igennem kapitlet, hvor den modelmæssige rækkefølge følger samme opsætning som i dette segment.

3.2 ARMA

En Autoregressive Moving Average (ARMA) er en univariate tidsserie model, som benytter tidligere værdier (lags) af den afhængige variabel (AR-proces) samt det bevægende gennemsnit (MA-proces) til at prædiktere fremtidige værdier. I denne forbindelse kræves der en tidsinvariant dataserie før, at modellen kan anvendes til forecasting, hvilket betegnes som stationaritet.

3.2.1 Stationaritet

For en stationær tidsserie gælder det, at gennemsnit og varians er uafhængig af tid, hvilket implicerer, at der ikke eksisterer en trend eller sæsonfluktuering i datasættet. Stationaritetsbetingelserne formaliseres herunder (Wooldridge, 2009):

- Forventet y_t værdi er konstant og uafhængig af tiden: $E[y_t] = \mu$
- Variansen til y_t skal være konstant og uafhængig af tiden: $var(y_t) = \sigma^2$
- Kovariansen mellem y_t og y_{t+k} er konstant, da $Cov(y_t, y_{t+k}) = f(k)$.

Tidsserier er ofte ikke stationære, hvilket betegnes som eksistensen af unit root. Derfor er det

nødvendigt at transformere værdierne i tidsserierne ved at differentiere dem, som vist i Ligning (3.1). Tidsseriens middelværdi stabiliseres derved, og der korrigeres dermed for trends. I tilfælde af sæsoneffekter indsættes der dummy's for at udbedre disse.

$$\Delta Y_t = Y_t - Y_{t-1} \tag{3.1}$$

Det kan variere, hvorvidt der skal én eller flere differentieringer til for at opnå stationaritet, men én differentiering er ofte tilstrækkeligt. Som følge af én differentiering reduceres antallet af observationer med én (Wooldridge, 2009).

Når der tages den første differens af tidsserien, arbejdes der med en stationær model, hvilket udtrykkes igennem tilføjelsen af en yderligere integrationskomponent (I) til ARMA-modellen. Der forekommer derved en transformering af ARMA til ARIMA.

3.2.2 ARIMA

Autoregressive Integrated Moving Averag (ARIMA) er en forecast-model, der består af tre komponenter. Den første komponent er den autoregressive del, $AR(p)$, som inkluderer lags af den afhængige variable som uafhængige variable i modellen.

Den autoregressive proces kan opskrives på formen:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2) \tag{3.2}$$

I Ligning (3.2) beskriver y_t den afhængige variabel, som bliver forecastet baseret på de laggede værdier, hvor antallet af lags udtrykkes igennem p. Hertil findes parameterværdierne, α_i, hvor α_0 er en konstant. Derudover indgår et fejlled, som antages at være white noise (Enders, 2014).

Moving Average processen, $MA(q)$, er den del af ARIMA-modellen, som benytter fejlleddet i tidligere perioder til at danne et forecast. MA-processen formaliseres herunder:

$$\chi_t = \sum_{i=0}^{q} \beta_i \epsilon_{t-i} \qquad \epsilon_t \sim WN(0, \sigma^2) \tag{3.3}$$

Hvor ϵ_t angiver white noise komponenten med inkluderingen af q lags. I denne sammenhæng indgår der en parameter, β_i. Ved at sammensætte $AR(p)$ og $MA(q)$ fås $ARMA(p,q)$ modellen. Ligningen for $ARMA(p,q)$ opskrives:

$$y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=0}^{q} \beta_i \epsilon_{t-i} + \epsilon_t \qquad \epsilon_t \sim WN(0, \sigma^2) \tag{3.4}$$

I det tilfælde, hvor $q = 0$, eller $p = 0$, vil dette være hhv. en ren AR- eller MA-proces.

Ved brugen af en ARMA-model skal tidsserien, der benyttes, være stationær før, at den kan benyttes til forecasting. Det observeres empirisk, at tidsserier typisk ikke er stationære, dvs. der er unit root j.f. Afsnit 3.2.1, og der tages dermed den første differens for at overkomme dette problem. Dette gøres automatisk i ARIMA-modellen igennem integrationskomponenten, $I(d)$, som påtager værdien 1. Herefter kan stationaritet kontrolleres i ARIMA-modellen ved at betragte de inverse AR- og MA-rødder, som skal ligge indenfor enhedscirklen.

Opsummerende angiver denne sammensætning af de tre komponenter, $ARIMA(p,d,q)$. Ved betingelsen $d = 0$, vil $ARIMA(p,0,q)$ svare til en $ARMA(p,q)$ model (Enders, 2014). ARIMA opsætningen kan indeholde sæsonkomponenter, og når disse inkluderes betegnes modellen i stedet SARIMA.

SARIMA er en udbygning af $ARIMA(p,d,q)$, hvor der pålægges yderligere tre led (P,D,Q), som repræsenterer sæsonkomponenter. Hele modellen står dermed på formen $SARIMA(p,d,q)(P,D,Q)$, hvor det efterfølgende (P,D,Q) angiver et AR, I, og MA led for sæsonkomponenterne. Derudover angives en værdi m der bestemmer cyklussen af sæson, dvs. ved månedlig data er $m = 12$ (Brownlee, 2019). Sammensætningen af $(p,d,q)(P,D,Q)$ bestemmes på baggrund af det relevante informationskriterie, som inddrages i det følgende.

3.2.3 Informationskriterier

Antallet af lags i ARIMA-modellen bestemmes ud fra en vægtning af datasættets fit samt enkeltheden af modellen. Hertil kan der benyttes Akaike's (AIC) eller Bayesian's (BIC) informationskriterie til at finde det optimale antal af lags. AIC og BIC opstilles på følgende vis:

$$AIC = 2 \cdot k + N \cdot ln(SSR) \tag{3.5}$$

$$BIC = N \cdot ln(SSR) + k \cdot ln(N) \tag{3.6}$$

Hvor N angiver det samlede antal observationer, mens k beskriver antallet af parametre i modellen. Afhængigt af størrelsen på datasættet benyttes der enten AIC eller BIC. AIC appliceres typisk, når der arbejdes med et mindre datasæt, og BIC, når datasættet er større. Gældende for begge kriterier er at lave værdier er bedre. Valget af lags sker således på baggrund af det laveste estimat af hhv. AIC eller BIC (Wang et al., 2021).

Som beskrevet kan det være svært at vælge den optimale kombination af (p, d, q) i ARIMA-modellen. I denne forbindelse tages der udgangspunkt i en *auto.arima* funktion, som anvender en algoritme til at bestemme det optimale antal lags og differentieringer, hvorved det undgås at fitte modellen manuelt ud fra autokorrelationsfunktioner, herunder ACF og PACF (Al-Chalabi et al., 2018). Algoritmen opsætter ARIMA-modeller med forskellige komponentværdier og rapporterer de tilhørende informationskriterier. Herudfra udvælger funktionen automatisk den bedste model baseret på det valgte informationskriterie (RDocumentation, n.d.).

Den estimerede ARIMA-model benyttes som baseline i forbindelse med forecasting. For at øge det økonometriske kompleksitetsniveau introduceres (S)VAR-modellen i den kommende sektion.

3.3 (S)VAR

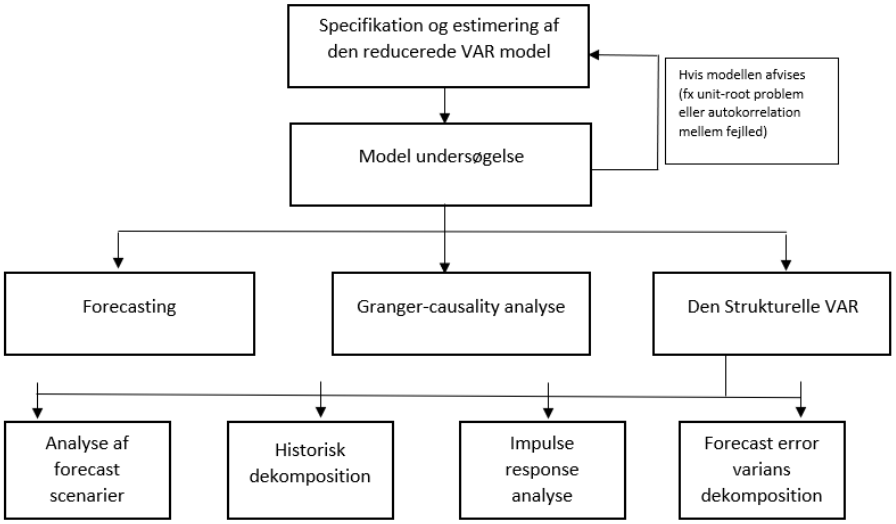
I dette afsnit præsenteres to modeller af højere økonometrisk niveau i form af et multivariate system i modsætning til ARIMA-modellen. I den første del af afsnittet redegøres der for VAR-modellen som et værktøj til at danne et in-sample forecast af samarbejdsvirksomhedens omsætning. Dernæst introduceres en SVAR-model, som tillader operationaliseringen af samtidige effekter. I denne forbindelse inddrages metoden tilhørende impulse response funktioner, der åbner op for undersøgelsen af energikrisens effekt på virksomhedens omsætning over tid.

Kort forklaret foretages der, i afhandlingens Afsnit 5, et stød til virksomhedens salgspris for at betragte virkningen på efterspørgslen efter plastprodukter. Et kort- og langsigtet estimat for egenpriselasticiteten fremkommer herigennem, og størrelsen på estimaterne kan tolkes som energikrisens effekt på omsætningen over tid. I denne forbindelse tages der et metodologisk udgangspunkt i et nyere studie af Yilmazkuday, 2019, hvori SVAR-opsætningen benyttes til at estimere egenpriselasticiteten, som defineret i Afsnit 2.3.2. Opsætningen af afsnittet forekommer i en ikke-kronologisk rækkefølge i forhold til besvarelsen af problemstillingen, da estimeringen af VAR-modellen er krævet for at kunne identificere SVAR-modellen.

3.3.1 Modelegenskaber

En Vector Autoregression Model (VAR) kan introduceres som en analogi til den simple AR(1), hvor der nu inkluderes mere end én endogen variabel. Der oprettes således et system bestående af m-ligninger og m-variable, hvor hver variabel forklares af egne lags samt lags af de resterende m-1-variable, som indgår i det pågældende system. Alle variable i en given VAR-model betegnes dermed oftest som endogene, og modeltypen er særligt brugbar, da den åbner op for en interaktion mellem flere endogene variable over tid.

Der sondres typisk mellem to typer af VAR-modeller; den reducerede VAR og den strukturelle VAR. I denne forbindelse skal det påpeges, at det altid er nødvendigt at estimere den reducerede VAR uafhængigt af den valgte problemstilling. Årsagen til dette er, at den strukturelle VAR ikke kan estimeres direkte. I Figur 3.1 forekommer der en grov skitsering af VAR-modelleringsprocessen og dertil, hvilke problemstillinger, som de to modeltyper kan anvendes til at undersøge nærmere.



Figur 3.1: Skitsering af VAR-modelleringsprocessen.
Kilde: Lütkepohl, 2013

Den valgte problemstilling er således bestemmende for, hvorvidt der eksisterer et behov for at identificere den strukturelle fra den reducerede VAR. Som vist i Figur 3.1 er målet med kandidatafhandlingen at anvende begge modeltyper, da der udarbejdes et forecast samt en impulse response analyse, som beskrevet i den indledende del af Afsnit 3.3.

Forskellen mellem de to modeltyper optræder ved, at den strukturelle VAR ikke kun lader variablene interagere med hinanden igennem deres lags, men denne tillader såvel en eksistens af samtidige effekter mellem de endogene variable i systemet. Dette uddybes i de efterfølgende afsnit. Derudover tager de to modeltyper udgangspunkt i to forskellige typer af kausalitet. Den reducerede VAR beskæftiger sig med den prædiktive kausalitet, 'Granger Causality', hvor kausaliteten i den strukturelle VAR er økonomisk funderet (Enders, 2014).

3.3.2 Den Reducerede VAR

I den reducerede VAR-model interagerer de endogene variable med hinanden igennem deres lags, hvilket betyder, at det betingede gennemsnit i modeltypen udelukkende er bestemt af fortiden. Formidlingen, samt den anvendte notation i dette afsnit, er baseret på artiklen af Lütkepohl, 2013, hvor der tages udgangspunkt i et eksempel med to økonomiske variable indkomst, y_t, og realrenten, i_t.

Der foreligger et stærkt økonomisk teoretisk grundlag for, at en stigning i realrenten, alt andet lige, vil foranledige et fald i indkomsten, og derfor vil der ikke være en grund til at tro, at hverken y_t eller i_t optræder som eksogent bestemte. På baggrund af dette teoretiske grundlag kan den reducerede VAR anvendes, som tillader, at ligningen for indkomsten, y_t, i denne periode, påvirkes af tidligere værdier af y_t selv, hvilket kan formaliseres som $(y_{t-1}, y_{t-2}, ..., y_{t-q})$. Ligningen tilhørende realrenten, i_t, opskrives ved at tage udgangspunkt i samme opsætning. En reduceret VAR-model med to variable og q lags betegnes en 2VAR(q).

I den næste del formaliseres en 2VAR(1)-model, hvor der tages udgangspunkt i eksemplet med y_t og i_t. Derudover inkluderes der ét lag dvs. $q = 1$.

Ligningen for indkomsten, y_t, kan formaliseres til:

$$y_t = v_y + \phi_{yy} y_{t-1} + \phi_{yi} i_{t-1} + \varepsilon_{y,t} \tag{3.7}$$

Og ligningen for realrenten, i_t, er:

$$i_t = v_i + \phi_{iy} y_{t-1} + \phi_{ii} i_{t-1} + \varepsilon_{i,t} \tag{3.8}$$

Hvor v indgår som en konstant, ϕ, er parameterværdierne, og ε repræsenterer fejlleddet.

Ligningssystemet kan således opskrives på matrice form:

$$\underbrace{\begin{bmatrix} y_t \\ i_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} v_y \\ v_i \end{bmatrix}}_{v} + \underbrace{\begin{bmatrix} \phi_{yy} & \phi_{yi} \\ \phi_{iy} & \phi_{ii} \end{bmatrix}}_{\phi_1} \underbrace{\begin{bmatrix} y_{t-1} \\ i_{t-1} \end{bmatrix}}_{Y_{t-1}} + \underbrace{\begin{bmatrix} \varepsilon_{y,t} \\ \varepsilon_{i,t} \end{bmatrix}}_{\varepsilon_t} \tag{3.9}$$

Hvilket er ækvivalent med:

$$Y_t = v_0 + \phi_1 Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim IID(0, \Sigma_\varepsilon) \tag{3.10}$$

I den indledende del af afsnittet blev det beskrevet, hvordan det betingede gennemsnit udelukkende var bestemt af fortiden i den reducerede VAR. Dette skyldes, at de samtidige effekter mellem variablene y_t og i_t ikke indgår heri, men disse opfanges i stedet i fejlleddets varians-kovarians matrice, som betegnes en 'full matrix' i den reducerede VAR form.

Udtrykket for fejlleddet i Ligning (3.10) kan opskrives på matrice form:

$$\begin{pmatrix} \varepsilon_{y,t} \\ \varepsilon_{i,t} \end{pmatrix} \qquad \sim IID \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{yy} & \sigma_{yi} \\ \sigma_{iy} & \sigma^2_{ii} \end{pmatrix} \right) \tag{3.11}$$

I det følgende afsnit opstilles stationaritetsbetingelserne for en VAR-model.

3.3.3 Stationaritetsbetingelser for en VAR-model

På samme måde som ARIMA-modellen kan den reducerede VAR karakteriseres som værende et stationært system, hvis en række betingelser overholdes:

- Vektoren for gennemsnittet skal være konstant og endelig: $E[Y_t] = \mu$
- Kovarians-matricen skal være konstant og endelig: $E[(Y_t - \mu)(Y_t - \mu)^T)$
- Autokovariansen afhænger af mængden af lags og ikke af tiden, t

En VAR-model betegnes som stationær hvis:

$$det(\phi - \lambda I) = 0 \tag{3.12}$$

Løsningerne til Ligning (3.12), som optræder i form af egenværdier, er mindre end 1 i absolut værdi, hvilket kan formaliseres til $(|\lambda| < 1)$[1]. Egenværdierne, tilhørende companion-matricen, ϕ, skal med andre ord ligge indenfor enhedscirklen. Hvis dette ikke er tilfældet, er systemet ikke stationært, og VAR-modellen kan således ikke anvendes. I denne forbindelse tages der typisk differensen af de inkluderede variable i systemet, indtil betingelsen opfyldes.

En opfyldelse af stationaritetskriteriet implicerer, at VAR-modellen konvergerer mod det ubetingede gennemsnit over tid. I det følgende illustreres dette ved at tage udgangspunkt i en VAR(1)-model:

$$Y_t = v_0 + \phi_1 Y_{t-1} + \varepsilon_t \tag{3.13}$$

$Y_{t-1} = v_0 + \phi_1 Y_{t-2} + \varepsilon_{t-1}$ substitueres ind på Y_{t-1}'s plads i Ligning (3.13):

$$Y_t = (I + \phi_1)v_0 + \phi_1^2 Y_{t-2} + (I + \phi_1)\varepsilon_{t-1} \tag{3.14}$$

Ved gentagelse af substitution over tid fremkommer Ligning (3.15):

$$Y_t = \phi^{t-1} Y_1 + \sum_{j=0}^{t-2} \phi_1^j v_0 + \sum_{j=0}^{t-2} \phi_1^j \varepsilon_{t-j} \tag{3.15}$$

Ud fra Ligning (3.15) bemærkes det, at effekten af initialværdierne for Y_1, og effekten af forstyrrelsen, som optræder igennem fejlleddet, ε_{t-j}, kun vil forsvinde, hvis $\phi^j \to 0$, når $j \to \infty$. Det forrige er gældende, hvis processen er stationær.

$\phi^j \to 0$ når $j \to \infty$ er et udtryk for, at momenterne er endelige. Med andre ord, når tiden går mod uendelig, $j \to \infty$, vil processen bevæge sig mod dets ubetingede gennemsnit som vist i Ligning (3.16):

$$E[Y_t] = \sum_{i=0}^{\infty} \phi_1^j v_0 = (I - \phi)^{-1} v_0 = \mu \tag{3.16}$$

[1]N.B. Dette er en generel betingelse, da enhver VAR(q) kan opskrives som en VAR(1).

Den reducerede VAR-model estimeres typisk ved brug af Ordinary Least Squares (OLS), hvilket anvendes som en standard estimeringsmetode.

3.3.4 Antallet af Lags i en VAR-model

Valget af lags, der inkluderes i en VAR-model, er en essentiel proces, og det kan have betydelige implikationer for modellens styrke, hvis der træffes et forkert valg på dette punkt. I denne forbindelse eksisterer der et trade-off mellem tilføjelsen af for mange lags og for få lags. Et misspecifikationsproblem kan opstå, hvis der inkluderes for få lags, og resultatet er 'omitted variable bias'. Dette medfører en fremkomst af uoptimale estimatorer. På den anden side medfører en tilføjelse af for mange lags en øget varians i estimatorerne, og modellen bliver således mindre efficient. Hver gang der tilføjes et lag i modellen, så bliver der opbrugt én frihedsgrad, hvilket har implikationer ved en lille stikprøve. Ydermere bestemmes antallet af lags i modellen på baggrund af en mulig eksistens af seriel korrelation, som betyder, at fejleddene er korreleret med tidligere tidsperioder. Eksistensen af dette kan resultere i en øget varians og derigennem en mindre efficient model. Ved tilstedeværelsen af seriel korrelation inkluderes der lags, indtil korrelationen er statistisk insignifikant. Antallet af lags bestemmes ved brug af informationskriteriet, AIC, som måler informationstabet i en model, og den foretrukne model vælges ud fra den laveste AIC-værdi jf. Afsnit 3.2.3. Herefter undersøges det, hvorvidt der er seriel korrelation tilstede ved brug af den såkaldte Portmanteu-test.

3.3.5 Forecasting i en VAR-model

I dette underafsnit formaliseres processen bag forecasting i en VAR-model. Processen forklares ved at tage udgangspunkt i variablen, y_t fra det forrige eksempel i Afsnit 3.3.2.

Betragt en VAR(1)-model på formen:

$$Y_t = v_0 + \phi_1 Y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim IID(0, \Sigma_\varepsilon) \tag{3.17}$$

Et forecast af én periode frem, dvs. $t+1$, kan formaliseres på følgende måde, hvor der tages det

betingede gennemsnit på begge sider af lighedstegnet:

$$\begin{aligned} E_T[Y_{T+1}] &= v_0 + \phi_1 E_T[Y_T] + E_T[\varepsilon_T] \\ &= v_0 + \phi_1 Y_T \end{aligned} \tag{3.18}$$

Hvor et forecast af anden periode, dvs. $t+2$, fremkommer ved rekursivt at tage udgangspunkt i Ligning (3.18):

$$\begin{aligned} E_T[Y_{T+2}] &= v_0 + \phi_1 E_T[Y_{T+1}] \\ &= v_0 + \phi_1(v_0 + \phi_1 Y_T) \\ &= v_0 + \phi_1 v_0 + \phi_1^2 Y_T \end{aligned} \tag{3.19}$$

Processen bag forecasting kan beskrives som værende dynamisk, idet forecasts bliver dannet på baggrund af tidligere forecasts, og der tages således udgangspunkt i en rekursiv tilgang. Når der dannes et forecast på længere sigt, dvs. $t+j$, og j øges over tid, konvergerer forecastet mod processens ubetingede gennemsnit, hvilket skyldes VAR-modellens stationære egenskaber, som beskrevet i Afsnit 3.3.3.

3.3.6 Den Strukturelle VAR

I den strukturelle VAR (SVAR) foretages der ikke en statistisk undersøgelse af de inkluderede variable i systemet, men der kobles i stedet økonomisk teori på det pågældende data. Dette medfører således, at der er fokus på økonomisk fundering i SVAR-modellen til forskel fra den reducerede VAR, som udelukkende bearbejder dataet fra en statistisk vinkel. En SVAR-model identificeres ved at pålægge et givet antal af restriktioner, så der opstår en overensstemmelse mellem antallet af parametre i de to modeltyper. I denne forbindelse dannes der en kausal rangorden med udgangspunkt i økonomisk teori, hvor rangordenen af variablene bestemmes ud fra, hvilken variabel, der *antages* at være den mest eksogene til den mest endogene variabel. En overensstemmelse mellem parametre i de to modeltyper medfører, at SVAR-modellen er præcist identificeret, og der kan således udarbejdes en impulse response analyse. Analysen indebærer en undersøgelse af, hvordan et stød til

én variabel i systemet, alt andet lige, påvirker de andre variable. En bemærkelse kan inddrages her, hvor forskellige kausale rangordener skaber forskellige SVAR-modeller. Med andre ord kan én reduceret VAR, som er estimeret fra ét datasæt, danne mange forskellige SVAR-modeller (Enders, 2014).

I det følgende formaliseres SVAR-modellen med udgangspunkt i det tidligere eksempel i Afsnit 3.3.2. En 2SVAR(1)-model opsættes således med hhv. y_t og i_t, hvor der inkluderes ét lag i modellen.

Ligningen for indkomsten, y_t, opskrives:

$$y_t = \omega_y - b_{0yi} i_t + b_{1yi} i_{t-1} + b_{1yy} y_{t-1} + u_{yt} \tag{3.20}$$

hvor ligningen for realrenten, i_t, er:

$$i_t = \omega_i - b_{0iy} y_t + b_{1ii} i_{t-1} + b_{1iy} y_{t-1} + u_{it} \tag{3.21}$$

Ligningssystemet kan ligeledes omskrives på matrice form:

$$\underbrace{\begin{bmatrix} 1 & b_{0yi} \\ b_{0iy} & 1 \end{bmatrix}}_{B_0} \underbrace{\begin{bmatrix} y_t \\ i_t \end{bmatrix}}_{Y_t} = \underbrace{\begin{bmatrix} \omega_y \\ \omega_i \end{bmatrix}}_{w} + \underbrace{\begin{bmatrix} b_{1yy} & b_{1yi} \\ b_{1iy} & b_{1ii} \end{bmatrix}}_{B_1} \underbrace{\begin{bmatrix} y_{t-1} \\ i_{t-1} \end{bmatrix}}_{Y_{t-1}} + \underbrace{\begin{bmatrix} u_{yt} \\ u_{it} \end{bmatrix}}_{u_t} \tag{3.22}$$

hvilket er ækvivalent med:

$$B_0 Y_t = w + B_1 Y_{t-1} + u_t, \qquad u_t \sim IDD(0, \Sigma_u) \tag{3.23}$$

I Ligning (3.20) og (3.21) bemærkes det, hvordan de samtidige effekter mellem de inkluderede variable samt lags indgår i det betingede gennemsnit i SVAR-modellen. Dette betyder såvel, at varians-kovarians matricen, tilhørende fejlleddet i SVAR-modellen, er en 'diagonal matrix', da denne ikke længere opfanger den samtidige effekt mellem variablene som illustreret i Ligning (3.24).

Fejlleddet fra Ligning (3.23) opskrives på matrice form:

$$\begin{pmatrix} u_{y,t} \\ u_{i,t} \end{pmatrix} \qquad \sim IID \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{yy} & 0 \\ 0 & \sigma^2_{ii} \end{pmatrix} \right) \tag{3.24}$$

3.3.7 Tilgang til Identifikation af den Strukturelle VAR

Dette afsnit benyttes til at beskrive identifikationsproblemet samt, hvilken tilgang, som anvendes til at overkomme problemet således, at SVAR-modellen kan identificeres fra den reducerede VAR. I den følgende gennemgang inddrages en formaliseret overgang fra SVAR-modellen til den reducerede VAR, da denne anvendes som et led i identifikationen senere.

Ved at tage udgangspunkt i Ligning (3.23), som er SVAR-modellen opskrevet på matrice form, kan den reducerede VAR findes ved at multiplicere med den inverse B_0-matrice, B_0^{-1}, på begge sider af lighedstegnet:

$$\begin{bmatrix} y_t \\ i_t \end{bmatrix} = \overbrace{\begin{bmatrix} 1 & b_{0yi} \\ b_{0iy} & 1 \end{bmatrix}^{-1} \begin{bmatrix} \omega_y \\ \omega_i \end{bmatrix}}^{v} + \overbrace{\begin{bmatrix} 1 & b_{0yi} \\ b_{0iy} & 1 \end{bmatrix}^{-1} \begin{bmatrix} b_{1yy} & b_{1yi} \\ b_{1iy} & b_{1ii} \end{bmatrix}}^{\phi_1} \begin{bmatrix} y_{t-1} \\ i_{t-1} \end{bmatrix} + \overbrace{\begin{bmatrix} 1 & b_{0yi} \\ b_{0iy} & 1 \end{bmatrix}^{-1} \begin{bmatrix} u_{yt} \\ u_{it} \end{bmatrix}}^{\epsilon_t} \tag{3.25}$$

Den reducerede VAR optræder derved i Ligning (3.26):

$$Y_t = v + \phi_1 Y_{t-1} + \varepsilon_t \tag{3.26}$$

Ydermere kan varians-kovarians matricen, tilhørende fejlleddet for den reducerede VAR, Σ_ε, findes ved at bruge formlen herunder:

$$\Sigma_\varepsilon = B_0^{-1} \Sigma_u (B_0^{-1})^T \tag{3.27}$$

I Afsnit 3.3.6 pointeres det, at den strukturelle VAR ikke kan identificeres direkte fra den reducerede VAR, hvilket betegnes identifikationsproblemet. Med andre ord eksisterer der ikke en 1:1 overensstemmelse mellem antallet af parametre i de to modeltyper, og SVAR-modellen betegnes derfor som ikke-identificeret. I denne forbindelse er det nødvendigt at pålægge et givet antal af restriktioner, før identifikationen af SVAR-modellen kan forekomme. Den traditionelle fremgangsmåde er anvendelsen af den såkaldte 'Cholesky Decomposition', som er en standard identifikationsordning. Ved brug af denne tages der udgangspunkt i B_0-matricen i Ligning (3.23) på venstre side, som indeholder de samtidige effekter mellem de inkluderede variable. Matricen transformeres til en

'lower-triangular form', når ovenstående matrice decompositionsmetode bruges. Der dannes således en kausal orden, hvor de økonomiske variable i systemet rangordnes fra den mest eksogene til den mest endogene variabel funderet i økonomisk teori.

Identifikationsproblemet kan forklares ud fra udtrykket for varians-kovarians matricen tilhørende fejlleddet for den reducerede VAR i Ligning (3.27). I en 2VAR(1) består venstre siden, Σ_ε, af tre parametre, hvor højre side består af fire parametre. I B_0^{-1}-matricen indgår der to parametre, hvilket ligeledes er gældende for, Σ_u. Der er altså flere parametre på højre side i forhold til venstre side. Denne uoverensstemmelse mellem parametre i de to modeltyper problematiserer identifikationen af den strukturelle fra den reducerede VAR.

De omtalte restriktioner, som skal pålægges SVAR-modellen før, at en identifikation kan finde sted, udregnes ud fra nedenstående formel:

$$\frac{m(m-1)}{2} \tag{3.28}$$

hvor m angiver antallet af variable, som er medtaget i systemet.

Cholesky Decomposition som Identifikationsordning

I Ligning (3.25) illustreres den formaliserede overgang fra den strukturelle til den reducerede VAR, hvor der multipliceres med B_0^{-1} på begge sider af lighedstegnet. Når der er pålagt de nødvendige restriktioner på SVAR-modellen, jf. Ligning (3.28), forekommer der en overensstemmelse mellem antallet af parametre på højre og venstre side af Ligning (3.27). I denne forbindelse inddrages den allerede nævnte Cholesky Decomposition, som anvendes til at løse denne uoverensstemmelse. Et valg, der foretages for at gøre anvendelsen af teknikken nemmere, er en bytning af normaliseringen i SVAR-modellen. Normaliseringen byttes fra B_0-matricen, hvor diagonalen ikke længere er lig 1, over til en normalisering af Σ_u, som nu bliver en identitetsmatrice.

Normaliseringsskiftet medfører, at udtrykket for varians-kovarians matricen, tilhørende fejlleddet for den reducerede VAR, ændrer sig til følgende:

$$\Sigma_\varepsilon = B_0^{-1}(B_0^{-1})^T \tag{3.29}$$

Brugen af en Cholesky Decomposition på en positivt defineret matrice, som Σ_ε i Ligning (3.29), resulterer i en lower-triangular matrice, P, og P', i Ligning (3.30):

$$\Sigma_\varepsilon = B_0^{-1}(B_0^{-1})^T = PP' \tag{3.30}$$

I Ligning (3.30) vises det, at B_0^{-1} er lig P, som betegnes en Cholesky faktor af varians-kovarians matricen tilhørende fejlleddet for den reducerede VAR. Det må dermed betyde, at den inverse Cholesky faktor, P^{-1}, er lig B_0. På dette tidspunkt kan den strukturelle VAR identificeres fra den reducerede VAR ved at tage udgangspunkt i Ligning (3.25), hvor der foretages en multiplikation af den inverse Cholesky faktor på begge sider af lighedstegnet. Dette medfører en annullering af B_0^{-1}, hvilket illustreres herunder:

$$P^{-1}Y_t = P^{-1}B_0^{-1}w + P^{-1}B_0^{-1}B_1Y_{t-1} + P^{-1}B_0^{-1}u_t \tag{3.31}$$

Hvilket resulterer i fremkomsten af den strukturelle VAR i Ligning (3.32):

$$P^{-1}Y_t = w + B_1Y_{t-1} + u_t \tag{3.32}$$

3.3.8 Impulse Response Funktioner

I dette afsnit introduceres impulse response funktioner, som kort blev berørt i Afsnit 3.3. Som beskrevet, tillader identifikationen af SVAR-modellen brugen af impulse response funktioner. Disse implementeres i analyseafsnittet for at undersøge effekten af energikrisen på samarbejdsvirksomhedens omsætning over tid.

I det følgende benyttes en egenskab, som indebærer, at enhver stationær VAR(q) kan repræsenteres som en Vector Moving Average(∞) (VMA) til at identificere impulse response funktionerne. En VAR(1) og dets tilhørende VMA-repræsentation opskrives herunder:

$$Y_t = \mu + \sum_{i=0}^{\infty} \Phi_1^i \epsilon_{t-i} \tag{3.33}$$

Hvor Φ angiver parameterværdierne. Ved at tage udgangspunkt i Ligning (3.33) er næste trin at finde en strukturel VMA (SVMA) repræsentation. Omskrivningsprocessen forekommer herunder:

$$Y_t = \mu + \sum_{i=0}^{\infty} \Phi_1^i \epsilon_{t-i} = \mu + \sum_{i=0}^{\infty} C_i \epsilon_{t-i} \tag{3.34}$$

Der henvises til Afsnit 3.3.7, hvori det illustreres, at $\epsilon_t = B_0^{-1} u_t = P u_t$. Dette anvendes i det følgende, og resultatet bliver en SVMA(1), som indebærer impulse response funktionerne, Θ:

$$Y_t = \mu + \sum_{i=0}^{\infty} C_i B_0^{-1} u_{t-i} = \mu + \sum_{i=0}^{\infty} \Theta_i u_{t-i} \tag{3.35}$$

I Ligning (3.35) kan det ses, at impulse response funktionerne afhænger af C_i'erne, som er selve MA-processen, og den estimerede Cholesky factor. En multiplicering af matricerne resulterer i impulse response funktionerne svarende til Θ_i. Da den estimerede Cholesky factor indgår heri, implicerer det, at funktionerne tager udgangspunkt i den valgte kausale rangorden af variablene.

Anvendelsen af Cholesky Decomposition og den tilhørende normalisering, jf. Afsnit 3.3.7, medfører, at impulse response funktionerne måles i form af standardafvigelser. Derudover, når der betragtes log-differens variable, som multipliceres med 100, betyder det, at impulse response funktionerne skal tolkes som procentpoint stød.

3.3.9 Estimering af Egenpriselasticiteter

Estimationen af egenpriselasticiteter tager sit udgangspunkt i definitionen af priselasticitet. I dette tilfælde er priselasticiteten givet ved den totale procentvise mængdeændring i salget af plastprodukter divideret med den totale procentvise prisændring på plastprodukter (Mankiw & Taylor, 2017), som illustreret i Ligning (3.36)[2]:

$$\theta = -\frac{\Delta \text{Solgte mængde af plastprodukter}}{\Delta \text{Prisen på plastprodukter}} = -\frac{\Delta lnsales_{i,t}}{\Delta lnprice_{i,t}} \tag{3.36}$$

[2]Elasticiteter er angivet med et negativt fortegn

hvor Δ angiver den procentvise ændring i en variabel. Når disse procentvise ændringer implementeres i den strukturelle VAR defineres egenpriselasticiteten, θ, som den kumulative impulse response af $\Delta lnsales_{i,t}$ divideret med den kumulative impulse response af $\Delta lnprice_{i,t}$ begge som følge af et engangsstød til salgsprisen (Yilmazkuday, 2019). Ligningen opskrives herunder:

$$\theta = -\frac{\text{Kumulative response af } \Delta lnsales_{i,t}}{\text{Kumulative response af } \Delta lnprice_{i,t}} \tag{3.37}$$

I kandidatafhandlingen betragtes der egenpriselasticiteter, θ, af samarbejdsvirksomhedens salg af plastprodukter på baggrund af energikrisen både på kort sigt og lang sigt. Tidshorisonterne defineres senere i analyseafsnittet. I analyseafsnittet undersøges det således, hvordan en stigning i salgsprisen påvirker samarbejdsvirksomhedens omsætning over tid. Som det næste gennemgås modelopsætningen, der benyttes under maskinlæring.

3.4 Maskinlæring

I sektionen gennemgås, hvilke modeller, der anvendes under maskinlæring. Modelopsætningen tager udgangspunkt i 'Modeltime'-pakken i R, som indeholder de ønskede modeller. Herudfra benyttes Generalized Linear Model (GLMNET) og Prophet samt ARIMA og Prophet med Extreme Gradient Boosted (XGBoosted) residualer.

GLMNET er en lineær regressionsmodel, som estimeres vha. 'maximum likelihood', hvilket resulterer i en fleksibel og solid maskinlæringsmodel. Modellen tager udgangspunkt i en standard 'gaussian' fordeling, hvor der opstilles et minimeringsproblem. Derudover standardiseres variansen i dataserien, hvilket afhjælper potentielle outliers (Hastie et al., 2022).

Prophet-modellen er udarbejdet af Facebook, og den ligger tilgængeligt som open source. Modellen er udviklet med formålet om at øge tilgængeligheden af en modelopsætning indenfor forecasting. Værktøjet er målrettet virksomheder, som derigennem kan håndtere diverse problematikker såsom sæson, irregulære stød og manglende observationer. Prophet-modellen er opstillet som en additiv regressionsmodel, og den består af fire primær komponenter; lineær eller logistisk vækst trend, årlig

sæson, ugentlig sæson og en liste af irregulære forventede stød såsom helligdage. Modellen benytter således det leverede data til at bestemme en trend, og den justerer derefter udviklingen baseret på dummy variable. (Letham & Taylor, 2017).

Ud over GLMNET og Prophet implementeres der XGBoost til at modellere residualerne i ARIMA- og Prophet-modellen. XGBoost påtager en trælignende form, hvor hver gren angiver mulige inputs i modellen. Forskellen mellem XGBoost og Random Forest opstår i selve læringsprocessen, hvor XGBoost optimerer hvert træ baseret på de tidligere dannede træer. Der opnås herigennem en forbedret model ved benyttelse af dette maskinlæringselement (XGBoost Developers, 2021). En yderligere øget kompleksitet, under maskinlæring, inddrages ved brug af neurale netværk i den følgende sektion.

3.5 Neurale Netværk

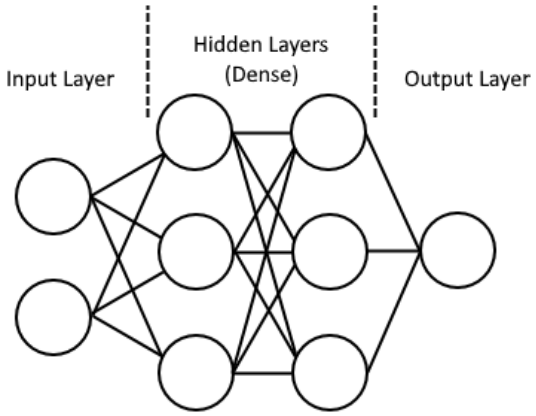
Dette afsnit introducerer neurale netværk, og hvordan disse kan bidrage til et højt niveau af forecasting. Afsnittet fokuserer på Recurrent Neural Networks (RNN), da disse netværk excellerer i forecasting med tidsserier. I denne forbindelse inddrages tre forskellige netværk, den basale RNN, Long-Short-Term-Memory (LSTM) og Gated-Recurrent-Unit (GRU).

3.5.1 Modelarkitektur

Neurale netværk bliver genereret ved at sammensætte layers af databehandling. Dette resulterer i en dybere læringsproces end traditionel maskinlæring. Sammensætningen af de valgte layers betegner arkitekturen i modellen, som kan tilpasses efter formålet ved at inkludere layers, der excellerer i f.eks. numerisk-, tekst- eller grafisk data (Hvitfeldt & Silge, 2021). I denne kandidatafhandling benyttes layers, der er specialiseret i tidsserier med fokus på numerisk data og regression.

Neurale netværk lærer og tilpasser sig efter de input, den modtager i form af modelarkitektur og dataserie. Et eksempel på en modelarkitektur er illustreret i Figur 3.2. Konstruktionen af netværket

er baseret på ét input layer, to hidden layers og ét output layer. Et input layer er defineret af den forinden datamanipulerede vektor. Hvordan dette data yderligere behandles, bestemmes dernæst af de inkluderede hidden layers og til slut af output layer. I hvert layer defineres der et antal hyperparametre, som definerer behandlingsstrukturen i et layer. Disse parametre inkluderer typisk antallet af units, som er antallet af nodes i et layer samt aktiveringsfunktionen, der benyttes til databehandlingen før næste layer (Hvitfeldt & Silge, 2021). Dette producerer en vægt i hver unit og bestemmer, hvordan informationen bliver allokeret imellem dem.



Figur 3.2: Eksempel på et simpelt dense neuralt netværk.
Kilde: Sharma, 2022

Et dense layer er det mest almindelige layer, som inkluderes i størstedelen af de neurale netværk i form af et output layer, da det tvinger informationen til den ønskede dimension (Sharma, 2022). I modsætning til input- og hidden layers indeholder et output layer en simpel lineær aktiveringsfunktion ved regressionsmodeller, da informationen skal udleveres som én værdi, og denne aktiveringsfunktion determinerer denne værdi (Brownlee, 2021). Figur 3.2 illustrerer en dense 'feed-forward' arkitektur, som danner bindinger mellem alle units, dvs. information kan allokeres frit (Sharma, 2022). Udover et dense output layer er det nødvendigt med et layer, som kan udnytte tidsseriedata. Dette tillades ved brug af enten RNN, GRU eller LSTM.

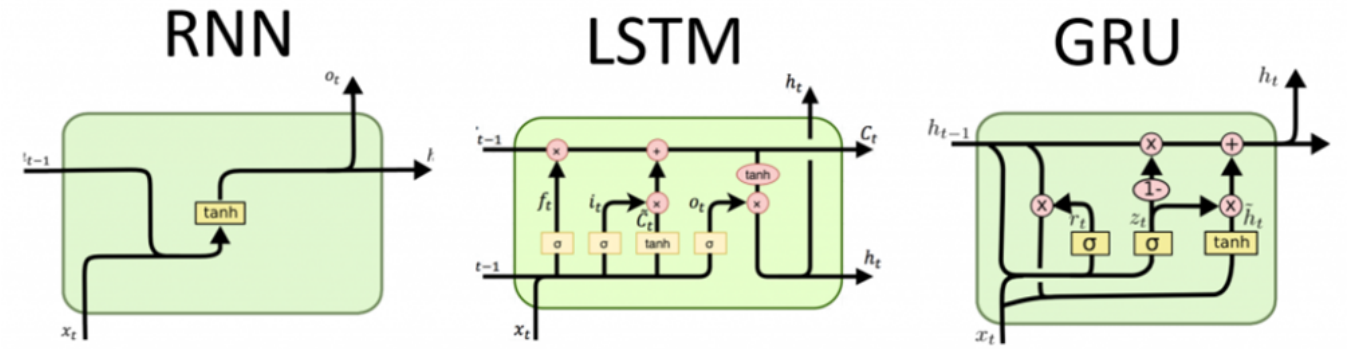
3.5.2 Recurrent Neural Network

GRU og LSTM tager udgangspunkt i den simple RNN-struktur. Strukturen muliggør en behandling af tidligere information, hvorfor det egner sig til tidsserier. Figur 3.3 illustrerer de tre modellers indre opbygning, og her ses det, at RNN har en simplere struktur end LSTM og GRU. I modsætning til RNN indeholder LSTM og GRU gates, der tillader dem at huske og glemme information alt efter, hvad der er relevant. LSTM er den mest komplicerede af de tre modeller, da den er indeholder tre gates; input, forget og output gate. Input og forget gaten samarbejder om at bestemme, hvilken information, der er nyttig at beholde i modellen. Output gaten producerer den vektor, der skal videreføres til næste layer i netværket (Petneházi, 2018). GRU forekommer simplere, da den benytter to gates; en update gate og en reset gate. Update gaten kontrollerer informationen fra tidligere, og hvilken information der skal lægges til. Reset gaten bestemmer aktiveringen af data output til næste layer (Petneházi, 2018).

RNN, GRU og LSTM har et matematisk fundament, som er baseret på feed-forward og 'backpropagation'. Dette fundament bestemmer, hvordan inputvektoren behandles og videreføres til det næste layer i netværksarkitekturen (dProgrammer lopez, 2019). Når backpropagation benyttes, under træningen af modellen, kan der opstå problemer med 'exploding' eller 'vanishing gradients', som forekommer når en værdi multipliceres gentagende gange med en vægt $\neq 1$. Gradients indeholder information omkring læringsprocessen for modellen, som bliver insignifikant, hvis de er vanishing, hvilket forårsager, at modellen ikke længere lærer af informationen. Ved exploding gradients forekommer det modsatte, hvor informationen får for meget vægt (Pascanu et al., 2013).

Det kan observeres i Figur 3.3, at de tre metoder stabiliseres af en 'tanh-funktion'. Funktionen anvendes for at stabilisere inputtet og til trods for dette indeholder RNN-cellen stadig problemerne. Problemerne udbedres i nogen grad i GRU og LSTM, hvor benyttelse af gates kan fjerne information, der enten har for lidt forklaring eller for meget (Ardel, 2018). Problematikken vokser ved flere gentagelser af træning på samme information. Udover tahn-stabiliseringsfunktionen suppleres LSTM og GRU af en sigmoid aktiveringsfunktion. Disse funktioner benyttes som standard, men

kan ændres alt efter formålet (Chollet et al., 2015).



Figur 3.3: RNN, LSTM & GRU Workings.
Kilde: dProgrammer lopez, 2019

En fejlkilde, der kan forekomme, ved at benytte neurale netværk, er overfitting, som er et større problem ved RNNs. Overfitting betyder, at modellen lærer af sammenhænge, der kun forekommer i træningsdatasættet, hvilket resulterer i, at modellen præsterer mindre godt på andet data. Disse netværk leder således efter sammenhænge i dataserien og lærer deraf for at skabe den bedst præsterende model. Fejlkilden kan delvist bekæmpes igennem en modificeret netværksarkitektur, hvor der indføres et 'dropout layer', der optræder som et ekstra hidden layer i modellen. Dette layer medfører, at informationen, der videreføres imellem layers i arkitekturen, deles på tilfældig vis, hvorved kun den ene del trænes på. Størrelsen på den ikke-trænede del er defineret i procent i et dropout layer (Chollet et al., 2015). Dette resulterer i, at modellen ikke nødvendigvis opfanger disse lokale sammenhænge i træningsdatasættet med samme hastighed som før.

Når modelarkitekturen er bestemt, dvs. hvilke layers, der skal indgå, kan modellen sammensættes. Sammensætningen kræver 'compiling', hvor en 'loss function' defineres. Den valgte loss function er typisk 'Mean Squared Error' (MSE). Denne beregner et tal, der beskriver distancen mellem de observerede og prædikterede værdi i gennemsnit (Chollet et al., 2015). Modellen justerer derfor de indledende arbitrære vægte mht. den valgte loss function, som minimeres under træningen.

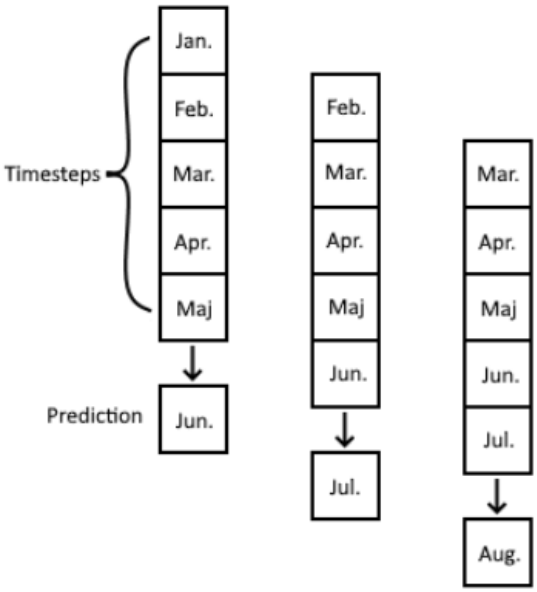
Træningsdatasættet er en forudbestemt andel af hele dataserien, som typisk udgør størstedelen af dataserien. Når modellen fittes herpå, udvælges der et antal epochs, som bestemmer, hvor mange gange, modellen skal træne på det samme informationsgrundlag. I denne forbindelse behandler modellen informationen én gang og gemmer resultater samt vægte herfra. Dernæst benyttes dette i en gentagende proces, indtil antallet af epochs er opnået (Chollet et al., 2015). Den resterende andel af dataserien angiver testdatamængden, som modellen testes på efterfølgende for at kontrollere præstationen af modellen udenfor træningsserien. Denne præstation angiver typisk, hvorvidt der eksisterer et overfittingsproblem.

3.5.3 Rolling Forecast

I sektionen redegøres der kort for 'rolling forecast', som er fremgangsmåden, der benyttes i de neurale netværk til forecasting i afhandlingen. I denne forbindelse tages der udgangspunkt i et eksempel fra Brownlee, 2018, hvortil der inddrages en visualisering af den bagvedliggende proces.

I litteraturen konverteres dataet på matrice form, hvorefter der foretages en normalisering af datapunkterne, da aktiveringsfunktionerne i de neurale netværk er sensitive overfor forskellige størrelsesordener. Dernæst opdeles datasættet i to, herunder et trænings- og testdatasæt. Herudfra appliceres der en 'TimeSeriesGenerator', som opsætter anvendelsen af 'many-to-one'-teknikken, hvor antallet af timesteps defineres.

I Figur 3.4 illustreres den anvendte many-to-one rolling forecast fremgangsmåde, hvilket indebærer, at der bruges flere tidligere perioder, dvs. antallet af timesteps, til at prædiktere én periode frem i tiden.



Figur 3.4: Visualisering af Rolling Forecast.

Kilde: Egen Skitsering

I det næste trin opsættes arkitekturen i det neurale netværk, hvorefter modellen trænes på det tilhørende datasæt. Afslutningsvis benyttes den trænede model på testdatasættet og evalueres ud fra præstationsmålet, som redegøres for i næste sektion.

3.6 Præstationsmål

Dette afsnit gennemgår et præstationsmål, der anvendes til at evaluere og sammenligne forecast-modellerne i kandidatafhandlingen. Evalueringen forekommer på baggrund af Root Mean Squared Error (RMSE) samt grafisk kontrol.

RMSE angiver, hvor langt de prædikterede værdier ligger fra de observerede i gennemsnit, hvilket kan beregnes ved brug af Ligning (3.38). I de neurale netværk tages der udgangspunkt i MSE som loss function. Forskellen på RMSE og MSE er inkluderingen af kvadratroden, som påføres udtrykket for at kunne tolke på det endelige tal. Med andre ord sørger kvadratroden for, at RMSE angiver

en værdi, der er, i samme størrelsesorden, som tidsserien, der forecastes. Værdien, for RMSE, angives fremadrettet i afhandlingen som en gennemsnitlig procentuel afvigelse for at imødekomme samarbejdsvirksomhedens anonymitetsønske.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(z_f - z_o)^2}{n}}$$
$$\% = \frac{RMSE}{Mean} \cdot 100 \tag{3.38}$$

Hvor z_f og z_o betegner hhv. den prædikterede og observerede værdi, og n angiver det samlede antal af forecastede værdier (Chai & Draxler, 2014).

3.7 Delkonklusion

Dette kapitel dannede rammerne for afhandlingens anvendte metode til besvarelse af problemstillingen. Indledningsvist blev der redegjort for ARIMA-modellen, som angiver en baseline i forbindelse med forecasting. Den optimale ARIMA-model var fundet ved at tage udgangspunkt i *auto.arima*-algoritmen, som rapporterer informationskriterierne, AIC og BIC. Hvorvidt AIC eller BIC benyttes, afhænger af stikprøvestørrelsen, hvor den laveste værdi repræsenterer den bedste model.

Dernæst blev der introduceret mere avancerede økonometriske modeller i form af en (S)VAR. VAR-modellen blev præsenteret for at danne et forecast med udgangspunkt i et multivariate system til forskel fra ARIMA-modellen, hvor der blot er én endogen variabel. Herudfra blev den strukturelle VAR identificeret for at betragte impulse response funktioner, og herunder egenpriselasticiteter, som anvendes til at undersøge implikationerne af energikrisen for samarbejdsvirksomhedens omsætning.

I den afsluttende del blev der redegjort for, hvordan maskinlæring og neurale netværk kan benyttes som forecast-metode. I denne forbindelse blev GLMNET, Prophet og XGBoost Errors præsenteret.

Herunder forklares det, hvorfor LSTM er et solidt redskab til at behandle tidsseriedata samt, hvilke modelmæssige forskelle, der eksisterer mellem RNN, GRU og LSTM. Derudover blev forskellige problematikker ved neurale netværk pointeret såsom overfitting, exploding og vanishing gradients, og hvordan disse kan udbedres. I afsnittet om præstationsmål blev RMSE introduceret som et sammenligningsgrundlag blandt forecast-modellerne i sammenspil med grafisk kontrol.

4 | Datastruktur

I dette kapitel præsenteres data fra samarbejdsvirksomheden og eksterne kilder, samt hvilken datastruktur, der danner udgangspunktet for kandidatafhandlingens analyse. Herunder beskrives det, hvorfor de udvalgte variable er anvendelige som proxy'er. Afslutningsvis redegøres der for, hvordan dataet klargøres til undersøgelse af problemstillingen.

4.1 Udvælgelse af Variable

I afsnittet introduceres det tilgængelige data, og hvilke variable, der herfra videreføres til databehandlingen. Som beskrevet i Afsnit 3.3.4, er det fordelagtigt med en simpel model for at mindske tabet af frihedsgrader, hvilket især gør sig gældende ved færre observationer. På baggrund af dette tages der udgangspunkt i en beregning af omsætningen ved brug af salgspris og den solgte mængde. Da der, jf. Afsnit 1.2, er heterogene strukturer i virksomheden, er det således nødvendigt at anvende proxy'er for efterspørgslen efter samarbejdsvirksomhedens plastprodukter og deres salgspris. Der findes flere variable, som kan bruges som proxy'er for disse, hvilket er vist i Tabel 4.1.

Proxy	Virksomhedsspecifik		Industrispecifik	
Salgspris	Plastindkøbspris pr. Enhed	(Månedsfrekvens)	Inflation i Plast	(Månedsfrekvens)
	Plastindkøbspris Total	(Årsfrekvens)		
Efterspørgsel	Energiforbrug i Produktion	(Månedsfrekvens)	Industri Energiforbrug	(Månedsfrekvens)
	Plastindkøbsmængde	(Årsfrekvens)	Industri Plastindkøb/salg	(Månedsfrekvens)

Tabel 4.1: Variabelmuligheder.

Kilde: Egen opsætning

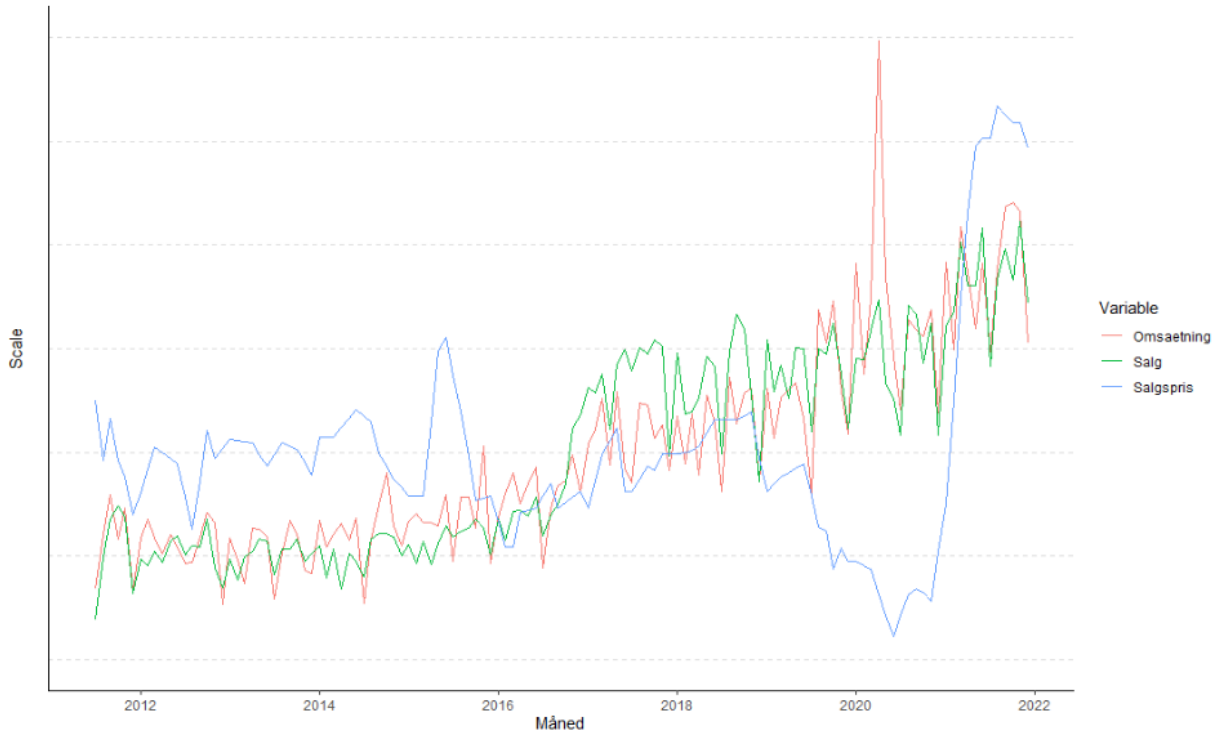
Samarbejdsvirksomhedens omsætning er opgivet på månedsbasis, hvilket nødvendiggør, at de andre

tidsserier er opgjort i samme frekvens. Dette medfører, at de to virksomhedsspecifikke variable, *Plastindkøbspris total* og *Plastindkøbsmængde*, fravælges. Hvis disse inkluderes i modellen, vil værdierne fremstå konstante igennem året med én ændring, da årsværdien opdeles i de 12 måneder for det gældende regnskabsår. De tilbageværende muligheder for proxy'er for efterspørgslen står mellem *Energiforbrug i Produktion*, *Industri Energiforbrug* og *Industri Plastindkøb/salg*, hvor mulighederne for salgsprisen er *Plastindkøbspris pr. Enhed* og *Inflation i Plast*.

De industrispecifikke variable fravælges, da der i afhandlingen benyttes en virksomhedspecifik omsætning. Sammensætningen af industri- og virksomhedsspecifikke variable vil typisk medføre en problematik i form af en upræcis forklaring af omsætningen. Et eksempel kan herved gives, når der betragtes en situation, hvor samarbejdsvirksomheden oplever et positivt eller negativt stød. Dette kan være et lokalt stød for virksomheden, som således ikke afspejles i industridata.

Opsummerende fravælges de industrispecifikke variable, hvilket resulterer i én mulig proxykombination for efterspørgsel og salgspris; *Energiforbrug i Produktion* og *Plastindkøbspris pr. Enhed*, hvor enheden måles i kg. Disse variable indgår dermed i kandidatafhandlingens analyseafsnit.

I Figur 4.1 illustreres tidsserien for de udvalgte variable samt samarbejdsvirksomhedens omsætning. Værdierne er normaliseret i overensstemmelse med virksomhedens ønskede anonymitet.



Figur 4.1: Normaliseret udvikling i de udvalgte variable.

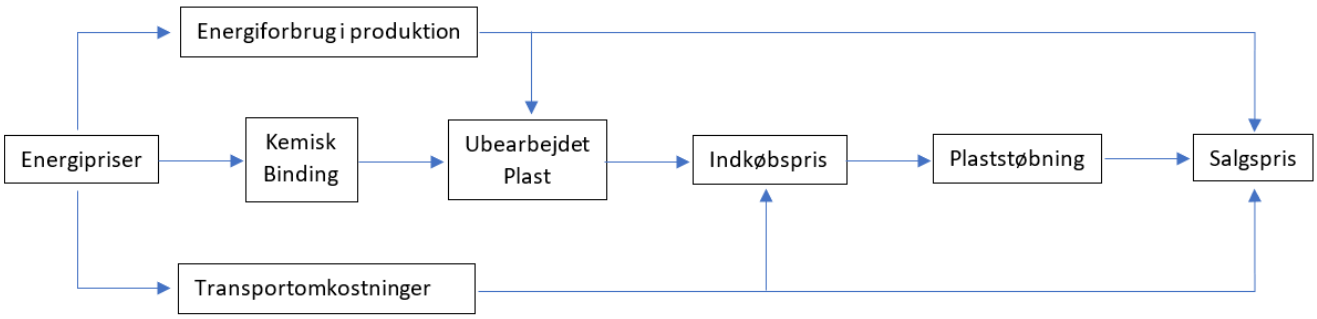
Kilde: Plot fra R

I Figur 4.1 observeres det, at omsætningen ligger over salget i størstedelen af perioden fra år 2011 til år 2016, hvilket kan forklares ud fra udviklingen i salgsprisen. Fra år 2016 til år 2019 observeres der et relativt lavere niveau for salgsprisen, hvilket stemmer overens med, at omsætningen ligger under salgsmængden. I år 2020 rammer salgsprisen sit laveste punkt, hvilket står i kontrast til, at omsætningen ligger over salgsmængden. Der er således andre faktorer, som har spillet ind på de stigninger, som har fundet sted i den reelle salgspris i dette år. Fra år 2021 observeres der en stigning i salgsprisen, hvilket er forårsaget af den aktuelle energikrise. Dette stemmer overens med, at omsætningen ligger højere end salgsmængden i størstedelen af perioden, hvilket viser, at omsætningen i større grad er præget af stigninger i salgsprisen, som således er forekommet igennem øgede inputpriser. I det kommende afsnit gøres det eksplicit, hvorfor de to udvalgte variable kan benyttes som proxy'er for salgsmængde og -pris.

4.1.1 Proxy Variable

Det teoretiske belæg for at kunne benytte samarbejdsvirksomhedens energiforbrug, som proxy for salgsmængden, kan beskrives ved at tage udgangspunkt i det følgende scenarie. Når efterspørgslen efter virksomhedens plastvarer øges, medfører dette en stigning i produktionen. Denne stigning opfanges, alt andet lige, igennem et øget energiforbrug og vice versa. I denne forbindelse antages det, at teknologiniveauet følger en forholdsvist konstant udvikling, hvilket medfører en tilnærmelsesvis isolering af de produktionsmæssige svingninger i tidsserien. Ud fra de teoretiske og empiriske overvejelser, som yderligere beskrevet i Afsnit 1.2, er det således muligt at anvende energiforbruget som proxy variabel.

Indkøbsprisen på plast anvendes som proxy for samarbejdsvirksomhedens salgspris, da en øget inputpris på plast, alt andet lige, påføres salgsprisen på virksomhedens plastvarer og vice versa. Derudover angiver proxy'en prisudviklingen i en af de plastkomponenter, der indgår i størstedelen af samarbejdsvirksomhedens produktion. Virksomheden benytter mange forskellige plasttyper med varierende mængde af data. Den valgte plasttype afspejler den generelle prisudvikling i plastmarkedet. Dette ses bl.a. ved, at plasttyperne generelt, jf. Figur A.12, har oplevet en markant prisstigning i 2021, hvilket opfanges af den udvalgte plasttype. Stigningen i år 2021 er forårsaget af den aktuelle energikrise, hvor både øgede omkostninger til el- og varme samt produktionsomkostninger i plastproduktionen afspejles igennem en øget inputpris for samarbejdsvirksomheden. Der forekommer da en dobbelteffekt, idet der indgår energi i den kemiske binding i plast (Plastindustrien, n.d.) samt i den grundlæggende produktion af plast. Derudover importerer Danmark alt ubearbejdet plast, hvilket yderligere bliver påvirket af prisen på brændstof j.f. Afsnit 2.2. I det henviste afsnit er det vist empirisk, at en øget inputpris påføres slutbrugeren igennem en højere salgspris. Sammenhængen mellem energipris og salgspris er illustreret i Figur 4.2 herunder.



Figur 4.2: Energikrisens indvirkning på salgspris.

Kilde: Egen skitsering

4.2 Datamanipulation

I dette afsnit beskrives den nødvendige datamanipulation, der foretages for at klargøre de udvalgte tidsserier fra Afsnit 4.1 til videre brug i afhandlingens analyse.

Den ene variabel, salgspris, indeholder manglende datapunkter ved tilfældige positioner i dataserien. I få tilfælde optræder disse manglende observationer i forlængelse af hinanden og opleves fra 1 til 3 perioder. Værdierne er indført ved at anvende lineær interpolation, som angiver en lineær sammenhæng mellem hver observation. Denne teknik tillader en udregning af alle mellemliggende punkter ved benyttelse af de kendte værdier på hver side af den manglende værdi. Teknikken er opstillet i Ligning (4.1).

$$p = p_{max} + \left(\frac{p_{min} - p_{max}}{s_{max} - s_{min}} \right) (s - s_{min}) \tag{4.1}$$

hvor p angiver salgsprisens værdi ved max og min dvs. øvre og nedre kendte værdi. Hertil angiver s observationsnummerering for den ønskede værdi og ligeledes min og max for de kendte observationer (Andersen, 2019). Anvendelsen af linear interpolation kan påføre bias til salgspris-tidsserien, idet den ikke indeholder alle de korrekte observerede værdier. Til trods for dette er inputprisen på plast

pr. enhed stadig den bedste mulighed for at operationalisere samarbejdsvirksomhedens salgspris. De manglende værdier er tilfældige og forekommer som en mindre del af dataserien. Den påførte bias vil dermed være ubetydelig, da de ikke opstår som større sammenhængende serier.

Der er observeret og testet for strukturelle brud i dataserierne. Samarbejdsvirksomhedens salg oplever et statistisk signifikant brud i slutningen af år 2016, hvilket illustreres i Figur A.2. I denne forbindelse er der opstået et 'regimeskift' i samarbejdsvirksomheden, hvilket påvirker de multivariate modeller. Derfor er det nødvendigt for disse modeller at benytte en afskåret dataserie fra 2017-01 til 2021-12. Inddragelsen af tidsperioden, hvor det gamle regime er tilstede, vil, alt andet lige, ikke være repræsentativt for, hvad der sker i et nyt regime. Ydermere vil inkluderingen af det forhenværende regime medføre inkonsistens i modellernes parametre, og de opstår derigennem som ustabile og foranderlige. Dette brud optræder ikke som statistisk signifikant i tidsserien for virksomhedens omsætning, hvilket betyder, at de univariate modeller kan benytte hele dataserien tilbage fra 2011-07.

I SVAR-modellen tages der udgangspunkt i data fra år 2017-01 til 2020-12 for samarbejdsvirksomhedens salgsmængde og -pris. Formålet er at analysere, hvordan energikrisen påvirker virksomhedens omsætning. Afskærringen af data fra 2011-2016 forekommer på baggrund af ovenstående argumentation omkring regimeskifte. Derudover er år 2021 afskåret, da energikrisen rammer i denne periode, og formålet er netop at modellere implikationerne af denne krise ved at betragte en normal periode uden stød. Estimaterne for en hhv. kort- og langsigtet egenpriselasticitet vil afspejle en periode inden energikrisen, og det antages således, at den adfærdsmæssige struktur er uændret ift. substitution. Egenpriselasticiteterne estimeres i det kommende kapitel, hvor SVAR-opsætningen appliceres til dette.

4.3 Delkonklusion

Dette kapitel introducerede datastrukturen i kandidatafhandlingen, og hvordan denne danner fundamentet for analysen. Indledningsvist forekom der en udvælgelse af de benyttede variable samt

en refleksion over nogle af de muligheder, der har været til rådighed. I denne forbindelse blev det bestemt, at de følgende to variable; *Energiforbrug i Produktion* og *Plastindkøbspris pr. Enhed*, optræder som proxy'er for hhv. samarbejdsvirksomhedens salgsmængde og -pris. Dernæst blev variablenes udvikling illustreret i Figur 4.1 og en problematik, ift. proxy'en for salgspris, blev understreget. Herunder blev det forklaret, hvorfor de udvalgte variable kan benyttes som proxy'er for salgsmængde og -pris. Afslutningsvist blev den nødvendige datamanipulation gennemgået, hvor det blev beskrevet, at indkøbsprisen indeholdte manglende værdier, og at dette blev udbedret vha. lineær interpolation. Derudover blev implikationerne af et større strukturelt regimeskift forklaret, hvilket har medført, at multivariate modellerne tager udgangspunkt i en afskåret dataserie. Herefter blev det forklaret, hvilken dataperiode SVAR-modellen anvender. Den valgte periode er i overensstemmelse med det bagvedliggende formål om at undersøge implikationerne af energikrisen for omsætningen under normale forhold.

5 | Analyse Del 1

I dette kapitel analyseres det, hvordan den aktuelle energikrise påvirker samarbejdsvirksomhedens omsætning ved at opsætte en SVAR-model. Opsætningen af modeltypen tillader undersøgelsen af, hvor følsom virksomhedens efterspørgsel efter plastprodukter er overfor en ændring i salgsprisen igennem anvendelsen af impulse response funktioner.

5.1 Egenpriselasticiteter i en SVAR-model

I denne sektion undersøges det, hvordan en stigning i salgsprisen påvirker samarbejdsvirksomhedens omsætning ved brug af en SVAR-opsætning. Modelvalget tillader en operationalisering af krisens implikationer for omsætningen igennem anvendelsen af impulse response funktioner, som appliceres i forbindelse med estimeringen af en kort- og langsigtet egenpriselasticitet. Indledningsvist opstilles kandidatafhandlingens hypotese omkring størrelsen på disse estimater, hvorefter de underliggende specifikationer bag den anvendte SVAR-model beskrives. Afslutningsvist præsenteres og tolkes der på analyseresultaterne fra modellen.

5.1.1 Estimat- og Hypotesebeskrivelse

Estimaterne for en kort- og langsigtet egenpriselasticitet tolkes i overensstemmelse med Afsnit 2.3.2, hvori det såvel står beskrevet, hvordan estimaternes størrelse angiver, hvorvidt virksomhedens omsætning påvirkes af energikrisen. Da samarbejdsvirksomheden støber forskellige produkter af plast, hvilket betyder, at virksomheden har et bredt ordrebestemt produktsortiment, kan godetype ikke benyttes som hypotesegrundlag for estimaternes størrelse. Der opstilles i stedet en hypotese for egenpriselasticiterne på kort og lang sigt ud fra virksomhedens ordrebestemte salg. På kort sigt er der en forventning om, at egenpriselasticiteten er af uelastisk karakter, da kunderne er bundet

til købsordren med faste betingelser, og de er i nogen grad nødt til at acceptere prisændringer på daværende tidspunkt. I denne forbindelse kan det inddrages, at virksomheden har sikret sig ret til at genkalkulere priserne på deres salgsordre, såfremt der forekommer væsentlige omkostningsstigninger jf. Afsnit 2.3.1. Det forventes derimod, at egenpriselasticiteten er relativt mere elastisk på længere sigt, da samarbejdsvirksomhedens kunder i større grad har mulighed for at substituere over til andre producenter på denne tidshorisont.

5.1.2 Modelspecifikationer

I det følgende opsættes SVAR-modellen med udgangspunkt i Ligning (3.20) og Ligning (3.21), som omskrives til et system bestående af to ligninger herunder:

$$p_t = \omega_p - b_{1pp}p_{t-1} + b_{1ps}s_{t-1} + u_{pt} \tag{5.1}$$

$$s_t = \omega_s - b_{0sp}p_t + b_{1sp}p_{t-1} + b_{1ss}s_{t-1} + u_{st} \tag{5.2}$$

SVAR-modellen består af variablene: p_t ($lnprice_t$), som betegner samarbejdsvirksomhedens salgspris, samt s_t ($lnsales_t$), som angiver virksomhedens salg af plastprodukter. På baggrund af heterogene strukturer i virksomheden anvendes plastindkøbsprisen pr. enhed og energiforbruget i produktionen på månedlig frekvens som proxy'er for hhv. salgspris og salg jf. Afsnit 4.1. Ud fra AIC-informationskriteriet inkluderes der ét lag af hver endogen variabel i modellen jf. Bilag A.3, og ln angiver, at der er taget logaritmen til variablene. Modellen tager således udgangspunkt i et datasæt bestående af 46 observationer.

De inkluderede variable i systemet testes for stationaritet, og det viser sig, at p_t-tidsserien har en unit root. Stationaritet i tidsserien opfyldes efter at have taget den første differens, og operationen foretages på begge af de ovenstående variable, da dette er nødvendigt for at skabe en meningsfuld model jf. Bilag A.2.2. Der er ydermere udført en diagnostik af modellen, hvilket omfatter en undersøgelse af stationaritet, da dette er et essentielt krav til brugbarheden af denne. Det næste

trin er at kontrollere modellen for seriel korrelation, da det skal sikres, at modellen har fejlled, som opfører sig som en white noise proces. SVAR-modellen overholder ovennævnte tests jf. Bilag A.3.

Frihedsgraderne i SVAR-modellen kan udregnes ved at tage udgangspunkt i følgende formel: $n - (k+1)$, hvor k angiver antallet af variable, og n er antal observationer. Efter at have taget den første differens og log af variablene er antallet af observationer reduceret til 46, og antallet af frihedsgrader i modellen kan dermed udregnes til $46 - (2 + 1) = 43$.

SVAR-modellen identificeres ved brug af Cholesky Decomposition, hvor der er pålagt én restriktion i overensstemmelse med Ligning (3.28). Resultatet er således en lower-triangular form, som danner en økonomisk funderet kausalorden gående fra den mest eksogene til den mest endogene variabel i systemet, som yderligere beskrevet i Afsnit 3.3.6. I SVAR-modellen er, p_t, valgt som den mest eksogene variabel i systemet, og s_t, som den mest endogene. Implikationen af den valgte kausalorden er, at et stød til salgsprisen i denne periode påvirker salget af plastprodukter i samme periode. Efter den første periode igangsættes MA-processen, og variablene interagerer med hinanden i systemet over tid, som forklaret i Afsnit 3.3.8.

5.1.3 Analyseresultater

I dette afsnit beskrives det indledningsvist, hvordan den kort- og langsigtede egenpriselasticitet identificeres, hvorefter beregningen af disse forekommer ved at anvende den præsenterede fremgangsmåde i Afsnit 3.3.9.

Den valgte modelspecifikation for SVAR-modellen medfører, at resultaterne fra impulse response analysen ikke kan tolkes som elasticiteter. Der tages med andre ord udgangspunkt i en logaritmisk-differens model, og de herfra resulterende estimater forekommer på baggrund af ét procentpoint stød, hvilket ikke kan opgøres i elasticiteter. En ren logaritmisk SVAR-model er derimod nødvendig for, at estimaterne kan tolkes direkte som elasticiteter, men da der eksisterer unit root i én af tidsserierne jf. Bilag A.2.2, er denne modelspecifikation ikke en mulighed. Egenpriselasticiterne identificeres i stedet ved at betragte en ratio mellem de kumulative impulse response funktioner som beskrevet i

Afsnit 3.3.9. Ligningen gengives herunder:

$$\theta = -\frac{\text{Kumulative response af } \Delta lnsales_{i,t}}{\text{Kumulative response af } \Delta lnprice_{i,t}} \tag{5.3}$$

Ligning (5.3) kan udregnes til enhver tidshorisont efter et éngangsstød til salgsprisen, p_t, i den første måned. I kandidatafhandlingen er det valgt, at estimeringen af den kortsigtede egenpriselasticitet, θ_{SR}, forekommer tre måneder efter stødet til p_t, hvor den langsigtede egenpriselasticitet, θ_{LR}, estimeres tolv måneder efter stødet. I den valgte periode optræder de kumulative impulse response funktioner som værende statistisk signifikante. Den langsigtede tidshorisont er bestemt ud fra konvergensen af disse funktioner.

I det følgende præsenteres analyseresultaterne fra SVAR-modellen, hvor formålet er at undersøge, hvordan energikrisen påvirker samarbejdsvirksomhedens omsætning. Den aktuelle energikrise imiteres ved at foretage et stød til virksomhedens inputpris, der benyttes som proxy for salgsprisen. Ræsonnementet bag dette forklares i Afsnit 4.1.1, hvor variabelsammenhæng såvel illustreres i Figur 4.2.

Den kortsigtede egenpriselasticitet for SVAR-modellen, θ_{SR}, findes ved først at foretage et éngangsstød til samarbejdsvirksomhedens salgspris, p_t, på ét procentpoint i den første måned. Virksomhedens salg af plastprodukter, s_t, responderer med et samlet fald på 0,66 procentpoint tre måneder efter stødet. Det næste trin er at lave et stød til salgsprisen af samme størrelsesorden i den første måned og betragte den samlede effekt af stødet på salgsprisen selv efter tre måneder, hvilket resulterer i en samlet stigning på 4,32 procentpoint. Disse ovenstående estimater kan aflæses i Tabel 5.1 nedenfor, og tabellerne for de kumulative impulse response estimater, for hele perioden, kan findes i Bilag A.3.

Impulse:	lnprice	lnprice
Response:	lnsales	lnprice
SR: 3. måned	-0,006612753	0,04320193
LR: 12. måned	-0,007215645	0,04336999

Tabel 5.1: Kumulative impulse response funktioner.

Kilde: Output fra R

Den kortsigtede egenpriselasticitet, θ_{SR}, estimeres tre måneder efter stødet:

$$\theta_{SR} = \frac{-0,66}{4,32} = -0,15 \tag{5.4}$$

En kortsigtet egenpriselasticitet på -0,15 implicerer således, at en 1% stigning i samarbejdsvirksomhedens salgspris forårsager et fald i virksomhedens salgsmængde på 0,15% efter tre måneder.

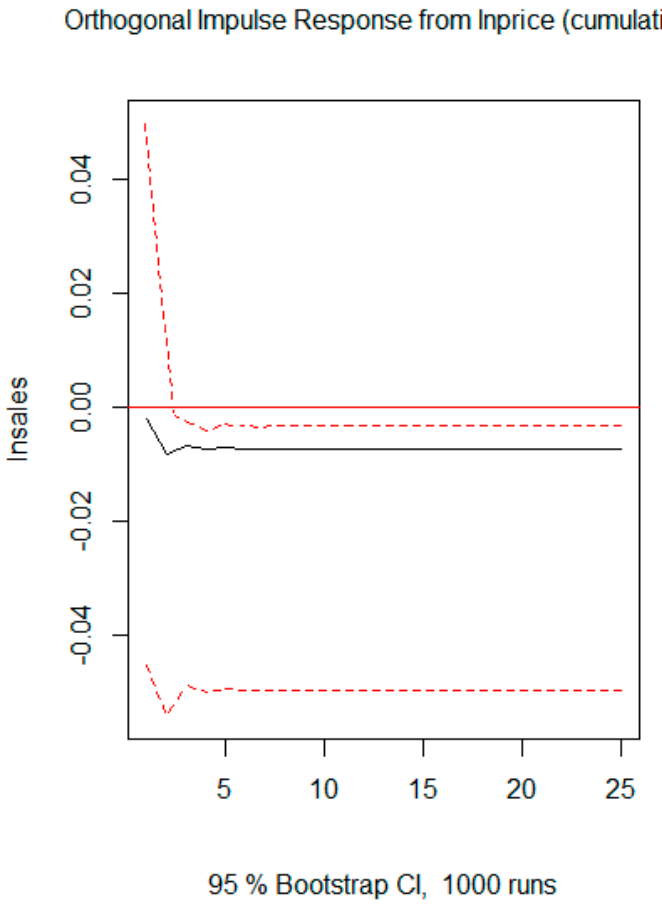
En langsigtet egenpriselasticitet for SVAR-modellen, θ_{LR}, udregnes ved brug af samme tilgang som ovenfor. Der foretages indledningsvist et éngangsstød til samarbejdsvirksomhedens salgspris, p_t, på ét procentpoint i den første måned. Virksomhedens salg af plastprodukter responderer med et samlet fald på 0,72 procentpoint tolv måneder efter stødet. Herefter foretages der et stød til salgsprisen af samme størrelsesorden i den første måned, hvor den samlede effekt af stødet på salgsprisen findes. Resultatet er således en samlet stigning på 4,33 procentpoint tolv måneder efter stødet.

En langsigtet egenpriselasticitet, θ_{LR}, estimeres tolv måneder efter stødet:

$$\theta_{LR} = \frac{-0,72}{4,33} = -0,17 \tag{5.5}$$

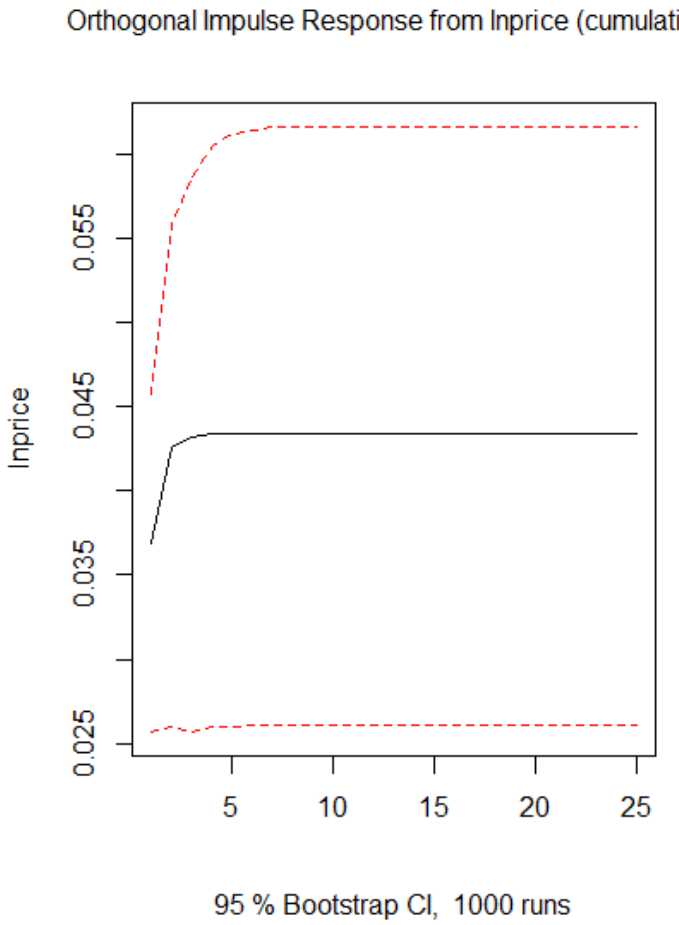
En langsigtet egenpriselasticitet på -0,17 betyder, at en 1% stigning i samarbejdsvirksomhedens salgspris foranlediger et fald i virksomhedens salg af plastprodukter på 0,17% efter tolv måneder.

De rapporterede estimater for en kort- og langsigtet egenpriselasticitet er statistisk signifikante, da SVAR-modellens kumulative impulse response funktioner er statistisk signifikante for den valgte tidshorisont, der tages udgangspunkt i. Dette illustreres i Figur 5.1 og 5.2 herunder, hvor det observeres, at konfidensintervallerne ligger under 0-linjen, og disse krydser ikke denne linje fra tredje måned af.



Figur 5.1: Kumulative impulse responses af salgspris på salg.

Kilde: Plot fra R



Figur 5.2: Kumulative impulse responses af salgspris på salgspris.

Kilde: Plot fra R

De ovenstående estimater tilfredsstiller således den opstillede hypotese. Egenpriselasticiteten er uelastisk på kort sigt, da det fundne estimat ligger mellem 0 og -1. Opsummerende forventes det, at samarbejdsvirksomhedens nuværende kunder er bundet op til eksisterende ordre på kort sigt, og der vil, alt andet lige, være en omkostning i form af tid og tabt salg for deres kunder, hvis de springer fra på nuværende tidspunkt. Det forrige står i kontrast til en langsigtet egenpriselasticitet,

der er relativt mere elastisk. Dette understøtter afhandlingens hypotese om, at kunderne har haft længere tid til at overveje andre producenter, når der betragtes en længere tidshorisont.

Ud fra de resulterende statistisk signifikante estimater for en kort- og langsigtet egenpriselasticitet påpeges det, at den aktuelle energikrise påvirker samarbejdsvirksomhedens omsætning positivt under den analyserede tidshorisont. Forskellen mellem den procentuelle stigning i samarbejdsvirksomheden salgspris på 1% og faldet i salgsmængden bliver mindre over tid. På kort sigt øges virksomhedens omsætning med $1 - 0,15 = 0,85\%$, hvor omsætningen øges med $1 - 0,17 = 0,83\%$ på længere sigt. Den øgede omsætning er 0,02 procentpoint mindre efter ét år, og det kan dermed ses, at størstedelen af omsætningsstigningen vedholdes over tid.

En anden faktor, som supplerer forklaringen om, at energikrisen påvirker samarbejdsvirksomhedens omsætning positivt, er, at plastindustrien er én af de brancher, som bruger mest energi. Dette skaber et generelt incitament til at foretage energieffektiviseringer og derigennem opnå en lavere energiintensitet relativt til mindre energitunge industrier. Derudover besidder Danmark en lav energiintensitet, når der tages højde for branchestruktur, som illustreret i Figur 2.4. Virksomheder i Danmark, og især de mere energitunge industrier, opnår således priskonkurrencemæssige fordele overfor udlandet, hvilket er gunstigt for efterspørgslen og derigennem omsætningen.

Analyseresultaterne, som afspejler en generel uelastisk egenpriselasticitet, understøttes af den observerede udvikling i omsætningen i år 2021, hvor den aktuelle energikrise optræder. I Figur 4.1 observeres det, at der forekommer en gennemsnitlig stigning i omsætningen på baggrund af coronakrisen fra år 2019 til år 2020. Fra år 2020 til år 2021 forekommer der endnu en gennemsnitlig stigning i omsætningen, hvilket skyldes en højere salgsmængde og især en højere salgspris, som i større grad må være drevet af øgede omkostninger jf. Figur 2.3. Den øgede omsætning i år 2021 må i større grad være drevet af en øget salgspris på baggrund af det observerede forhold mellem omsætning og salget. I den næste analysedel anvendes mulige forecast-metoder, der kan opstilles i kriseperioden, hvor resultaterne sammenfattes med den første del af analysen.

6 | Analyse Del 2

I dette kapitel udføres den anden del af analysen, hvor der dannes forskellige in-sample forecasts af samarbejdsvirksomhedens omsætning baseret på ARIMA, VAR, maskinlæring og neurale netværk. I modellerne benyttes der en forecast-horisont på fire perioder, da modellerne tager udgangspunkt i mest muligt modellerings- og træningsdata fra energikrisen.

6.1 Baseline ARIMA

I sektionen opsættes den første univariate forecast-model, som er en ARIMA-model. Indledningsvist introduceres den valgte modelopsætning, hvorefter der dannes et in-sample forecast. Dernæst rapporteres præstationsmålet, RSME, som er det anvendte sammenligningsgrundlag blandt forecast-modellerne.

6.1.1 Modelspecifikationer

ARIMA-modellen opsættes med udgangspunkt i Ligning (3.4):

$$r_t = \alpha_0 + \alpha_1 r_{t-1} + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + sar1 + sma1 + sma2 + e_1 d_1 + e_2 d_2 + \epsilon_t \tag{6.1}$$

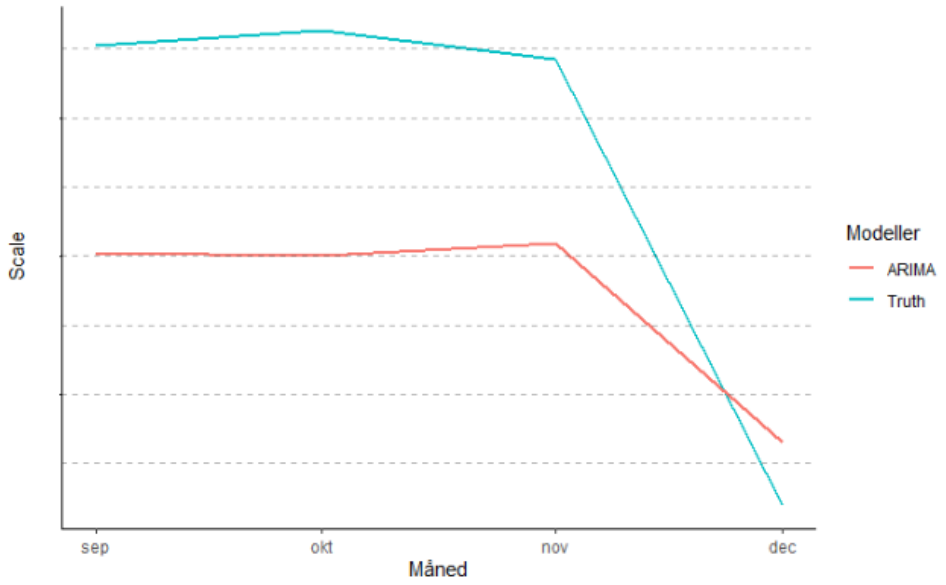
Ligning (6.1) afspejler en modellering af samarbejdsvirksomhedens omsætning, r_t, ved brug af en ARIMA-model, med ét AR-led ($p = 1$) og to MA-led ($q = 2$), integreret af første orden ($d = 1$). Derudover inkluderes der ét sæsonkomponent til AR-leddet og to komponenter til MA-leddet, hvilket implicerer, at der afslutningsvist er tale om en $SARIMA(1,1,2)(1,0,2)[12]$, hvor $m = 12$, da der tages udgangspunkt i data på månedlig basis. Modelopsætningen er fundet ved at anvende både *auto.arima* algoritmen i samspil med nødvendige manuelle tilføjelser jf. Bilag A.4. Herudover medtages der to dummy variable i SARIMA-modellen, d_1 og d_2, som kontrollerer for effekterne af

coronakrisen i år 2020, hvilket illustreres i Figur 4.1. Modellen anvender data fra år 2011-07 til 2021-08, dvs. 122 observationer, til forecasting af de resterende fire måneder af tidsserien til og med 2021-12.

Samarbejdsvirksomhedens omsætning testes for stationaritet, hvilket opfyldes jf. Bilag A.2.2. Ifølge *auto.arima* algoritmen opnås den bedste model ved at betragte SARIMA-modellen af første integrationsorden, $d = 1$ til trods for denne opfyldelse. Ydermere er der foretaget en diagnostik af modellen, hvilket indebærer en undersøgelse af stationaritet, seriel korrelation, ARCH og normalitet af residualerne. Disse elementer er essentielle for modellens anvendelighed. Modellen overholder ovenstående tests foruden seriel korrelation, og løsningen på dette problem er den manuelle tilføjelse af ekstra led til modellen. Eksistensen af seriel korrelation medfører en øget varians og derigennem en mindre efficient model.

6.1.2 Forecasting og Præstationsevne

I det følgende konstrueres der et in-sample forecast af samarbejdsvirksomhedens omsætning ud fra en univariate modelopsætning. I Figur 6.1 plottes de prædikterede værdier, som angiver den røde kurve, op imod de observerede værdier, som angiver den lyseblå kurve. Ud fra den grafiske illustration kan det ses, at de tre første prædikterede værdier ligger i et markant lavere niveau i forhold til de observerede værdier. Den fjerde prædikterede værdi ligger betydeligt tættere på den observerede værdi, hvilket betyder, at afstanden mellem kurverne altså mindskes over tid. Et mål for den gennemsnitlige afstand mellem kurverne udregnes med udgangspunkt i målet fra Afsnit 3.6.



Figur 6.1: ARIMA Forecast.

Kilde: Plot fra R

SARIMA-modellen formår at danne et forecast, som præsterer med en RMSE på $11,2\%$. De prædikterede værdier afviger således med $11,2\%$ fra de observerede værdier i gennemsnit. Denne baseline-præstationen sammenlignes med de mere komplekse modeller i de kommende afsnit.

6.2 VAR Forecast

I denne sektion øges det økonometriske niveau ved at anvende en VAR-model til forecasting af omsætningen. Der indledes med den valgte modelopsætning, hvor der afslutningsvist dannes et forecast som vurderes ud fra det anvendte præstationsmål.

6.2.1 Modelspecifikationer

VAR-modellen opsættes med udgangspunkt i Ligning (3.10), som kan omskrives til et ligningssystem bestående af tre ligninger herunder:

$$r_t = v_r + \phi_{rr} r_{t-1} ... \phi_{rr} r_{t-4} + \phi_{rs} s_{t-1} ... \phi_{rs} s_{t-4} + \phi_{rp} p_{t-1} ... \phi_{rp} p_{t-4} + e_1 d_1 + e_2 d_2 + \varepsilon_{r,t} \tag{6.2}$$

$$s_t = v_s + \phi_{sr} r_{t-1} ... \phi_{sr} r_{t-4} + \phi_{ss} s_{t-1} ... \phi_{ss} s_{t-4} + \phi_{sp} p_{t-1} ... \phi_{sp} p_{t-4} + e_1 d_1 + e_2 d_2 + \varepsilon_{s,t} \tag{6.3}$$

$$p_t = v_p + \phi_{pr} r_{t-1} ... \phi_{pr} r_{t-4} + \phi_{ps} s_{t-1} ... \phi_{ps} s_{t-4} + \phi_{pp} p_{t-1} ... \phi_{pp} p_{t-4} + e_1 d_1 + e_2 d_2 + \varepsilon_{p,t} \tag{6.4}$$

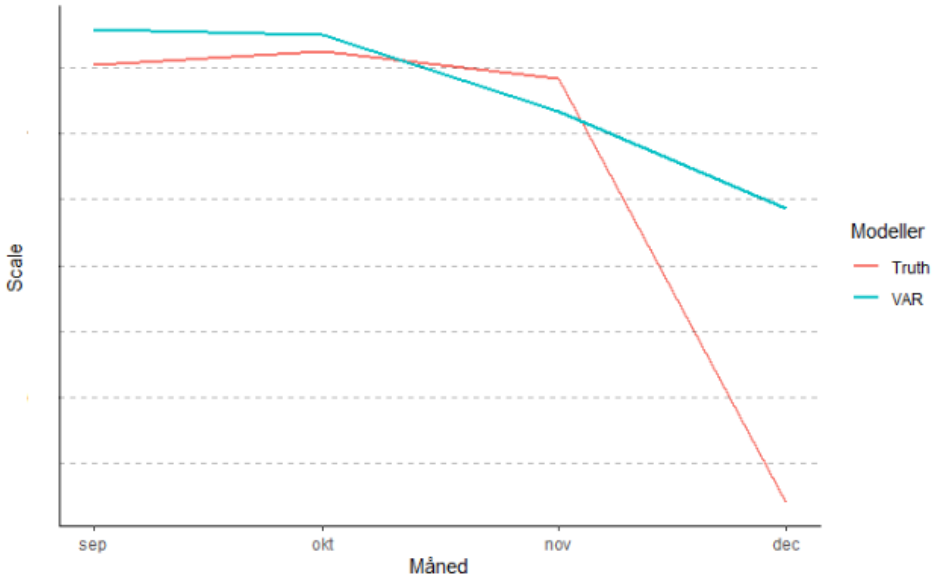
VAR-modellen indeholder tre variable: samarbejdsvirksomhedens omsætning, r_t ($lnrevenue_t$), virksomhedens salg af plastprodukter, s_t ($lnsales_t$) og salgsprisen, p_t ($lnprice_t$). Ud fra AIC informationskriteriet er det mest optimalt at inkludere fire lags af hver endogen variabel i modellen jf. Bilag A.4.1. Derudover tilføjes der to dummy variable, som er identiske med dem, der blev medtaget i SARIMA-modellen i forrige sektion for netop at kontrollere for coronakrisen i år 2020. VAR-modellen anvender data fra år 2017-01 til 2021-08, hvilket er 56 observationer, til forecasting af de resterende fire måneder af tidsserien til og med 2021-12.

De tre variable, som er medtaget i VAR-modellen, testes for stationaritet. Som beskrevet i Bilag A.2.2, er der unit root i samarbejdsvirksomhedens salgspris, p_t. Stationaritet opfyldes efter at have taget den første differens, og dette foretages på alle de inkluderede variable. Herudover er der foretaget en diagnostik af modellen i forhold til stationaritet og seriel korrelation. VAR-modellen overholder ovenstående tests jf. Bilag A.4.1.

Frihedsgraderne i VAR-modellen udregnes ved at tage udgangspunkt i formlen fra tidligere: $n - (k + 1)$. Efter at have taget den første differens og log af variablene bliver antallet af observationer reduceret til 56, og antallet af frihedsgrader i modellen udregnes til $56 - (18 + 1) = 37$.

6.2.2 Forecasting og Præstationsevne

I det følgende konstrueres der et in-sample forecast af samarbejdsvirksomhedens omsætning, r_t, i en multivariate opsætning. Implikationen af dette er, at forrige perioder af salg og salgspris såvel benyttes i forecastet af omsætningen til forskel fra SARIMA-modellen. I Figur 6.2 plottes de prædikterede værdier, som angiver den røde kurve, op imod de observerede værdier, som angiver den lyseblå kurve. I det grafiske plot følger de første tre prædikterede værdier tilnærmelsesvis de observerede værdier, hvor den fjerde prædikterede værdi afviger i større grad fra den observerede værdi. Et mål for den gennemsnitlige afstand mellem kurverne udregnes på samme måde som tidligere ud fra præstationsmålet i Afsnit 3.6.



Figur 6.2: VAR Forecast.

Kilde: Plot fra R

VAR-modellen opnår en RMSE på $9,7\%$, hvilket implicerer, at de prædikterede værdier afviger med $9,7\%$ fra de observerede værdier i gennemsnit.

6.3 Maskinlæring Forecast

I den følgende sektion dannes der forecasts ved brug af maskinlæringsmetoder. Herunder anvendes der fire forskellige modeller, som er knyttet til modeltime pakken i R. Disse modeller er; ARIMA m. XGBoost Errors, Prophet, Prophet m. XGBoost Errors og GLMNET. Yderligere forklaringer af disse modeller findes i Afsnit 3.4, hvor XGBoost bruges til modellering af residualerne i ARIMA- og Prophet-modellen.

6.3.1 Modelspecifikationer

De fire modeller er opsat som univariate modeller, hvilket implicerer, at de udelukkende tager udgangspunkt i omsætningen. Derudover er tre af modellerne opsat som multivariate igennem inkluderingen af de to ekstra variable fra Afsnit 4.1. Univariate modellerne konstrueres på baggrund af det fulde datasæt fra år 2011-07 til år 2021-12, da der ikke forekommer statistisk signifikante brud i denne dataserie. Ved multivariate modellerne er tidsserien afskåret til år 2017-01 til 2021-12. Dette er foretaget, da salget oplever et strukturelt brud jf. Afsnit 4.2.

Opsætningen af ARIMA-modellen m. XGBoost Errors sker ved at bruge *auto.arima* algoritmen. Resultatet heraf er en $SARIMA(0,1,1)(2,0,0)[12]$ model. I Ligning (6.5) er modellen vist, og opsætningen følges i overensstemmelse med Afsnit 3.2.2.

$$r_t = \alpha_0 + \beta_1 \epsilon_{t-1} + sar1 + sar2 + \epsilon_t \tag{6.5}$$

Prophet-modellerne er opstillet efter standard opsætning med automatisk registrering af sæsonelementer og uafhængige dummy variable. Ydermere forbedres modellen ved inkludering af en lineær trend, når der bruges XGBoost Errors, hvilket ligeledes er gældende for GLMNET-modellen.

6.3.2 Forecasting og Præstationsevne

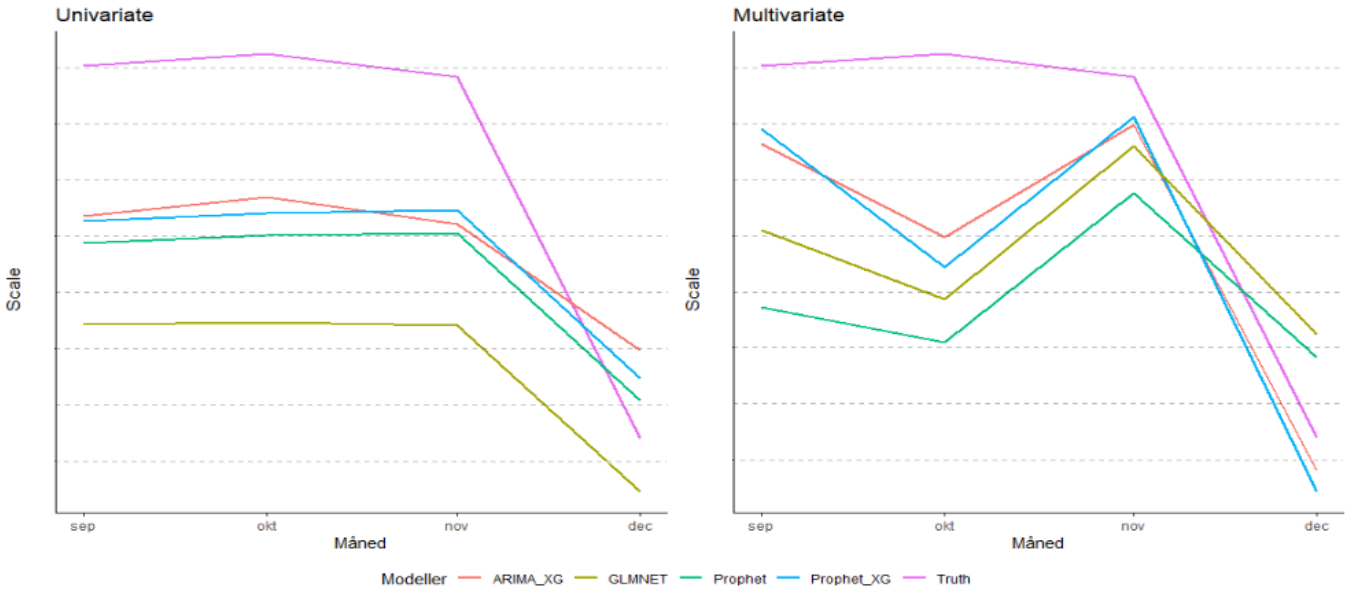
Præstationsmålet, RMSE, er rapporteret for disse modeller i Tabel 6.1 herunder.

Model	*Univariate*	*Multivariate*
ARIMA XGBoost Errors	10,3 %	-
Prophet	11,4 %	15,4 %
Prophet XGBoost Errors	10,2 %	**8,9** %
GLMNET	17,2 %	12,2 %

Tabel 6.1: Maskinlæring RMSE for Univariate og Multivariate.

Kilde: Output fra R

Tabellen viser, at multivariate Prophet-modellen m. XGBoost Errors præsterer bedst ud fra en evaluering af RMSE efterfulgt af dens univariate version. Multivariate modellen opnår en gennemsnitlig afvigelse på $8,9\%$ fra de observerede værdier. Modellerne forbedres markant ved inkludering af XGBoost Errors. Dette tyder på, at der er en sammenhæng i residualerne, som kan modelleres, hvilket XGBoost-tilføjelsen formår at opfange. I denne forbindelse kan det ydermere vises, at ARIMA-modellen m. XGBoost fremstår markant bedre end ARIMA fra det tidligere Afsnit 6.1. De ovenstående resultater skal sammenlignes med deres grafiske præstation ud fra Figur 6.3. Univariate modellerne har generelt den samme udvikling, hvilket såvel gør sig gældende for multivariate modellerne. Den største forskel mellem de to typer, er stødet, som multivariate modellerne oplever i oktober måned, hvilket påvirker deres præstation negativt.



Figur 6.3: Forecast-resultater fra Maskinlæringsmodellerne.

Kilde: Plot fra R

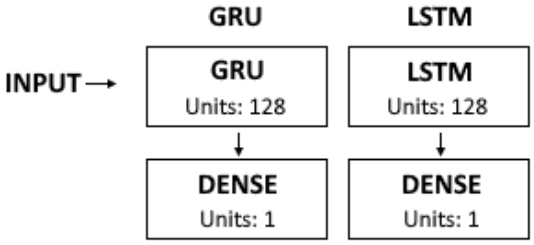
En sammenfattet evaluering viser således, at univariate modellerne præsterer bedst i dette forecast-scenarie. Herunder angives univariate Prophet m. XGBoost Errors for at være den bedst præsterende maskinlæringsmodel ud fra både RMSE og grafisk kontrol. Multivariate modellerne skaber generelt en usikkerhed igennem det negative stød.

6.4 Neurale Netværk Forecast

I denne sektion opstilles der flere forskellige uni- og multivariate neurale netværk. Den generelle fremgangsmåde, som der tages udgangspunkt i, er rolling forecast-teknikken, som er redegjort for i Afsnit 3.5.3.

6.4.1 Modelspecifikationer

De neurale netværk opsættes manuelt igennem diverse arkitekturer og ændringer af parametre. En simpel arkitektur præsterer bedst for disse neurale netværk målt på RMSE. En mulig forklaring kan være et lavt kompleksitetsniveau samt relativt få observationer. De manuelle ændringer i arkitekturer indebærer en inkludering af dropout layers, yderligere LSTM eller GRU layers og dense layers. Disse mere komplekse arkitekturer præsterer værre, og derfor tages der udgangspunkt i den simplere arkitektur under opsætningen af modellerne. Netværksarkitekturen illustreres i Figur 6.4 herunder.



Figur 6.4: Modelarkitektur på forecast-modellerne.

Kilde: Output fra R

6.4.2 Forecasting og Præstationsevne

De neurale netværk, uni- og multivariate, er opsat på forskellige typer af dataformer (level / procentændring), datalængder og timesteps. Datasættet tager først udgangspunkt i perioden 2011-07 til 2021-12, hvorefter der forekommer en afskæring fra 2017-01 til 2021-12 jf. Afsnit 4.2. Modellerne er yderligere testet ved at ændre antallet af timesteps fra 3 til 6. Modellerne testes på disse forskellige elementer for at identificere den bedste opsætning, og Tabel 6.2 viser resultaterne heraf.

Den markerede værdi i Tabel 6.2 angiver den laveste RMSE fra de neurale netværk, som er opnået i det afskårede datasæt med 5 timesteps. Herunder opnår en multivariate LSTM-model i levels en RMSE på $12,6\%$ efterfulgt af en multivariate GRU i levels, som præsterer næstbedst med en RMSE

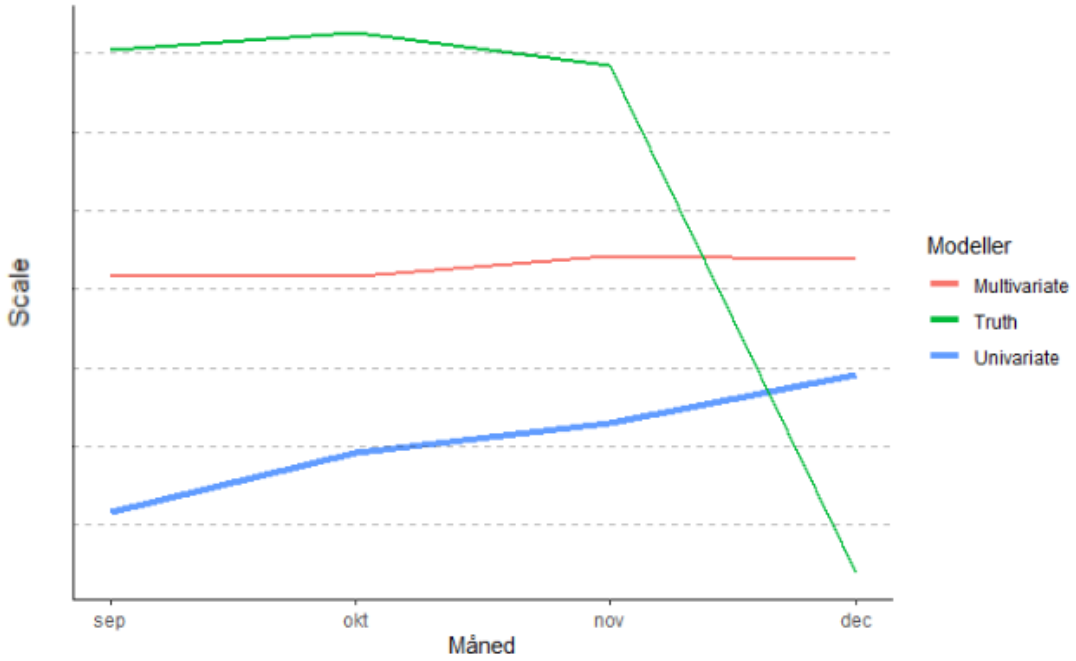
på 15, 2%. De neurale netværk præsterer generelt bedre ved brug af det afskårede datasæt, hvilket kan skyldes, at de ikke kan håndtere det strukturelle brud, der forekommer i salgs-tidsserien i det fulde datasæt. Derudover observeres der en forbedring ved brug af mere end 3 timesteps og mindre end 6.

		Fuld Data 2011-2021				*Skåret Data 2017-2021*			
Layer	*Antal Timesteps til Forecast:*	3	4	5	6	3	4	5	6
LSTM	Multivariate Procent	22,6 %	19,0 %	23,3 %	26,5 %	27,1 %	29,3 %	25,3 %	24,0 %
	Multivariate Level	40,1 %	26,4 %	28,0 %	28,2 %	19,7 %	16,6 %	**12,6** %	17,4 %
	Univariate Procent	20,8 %	20,1 %	21,9 %	28,1 %	25,8 %	24,6 %	19,2 %	20,3 %
	Univariate Level	36,9 %	32,7 %	27,0 %	28,7 %	23,1 %	23,2 %	22,5 %	21,5 %
GRU	Multivariate Procent	23,1 %	19,5 %	24,5 %	29,9 %	32,4 %	28,7 %	23,8 %	19,8 %
	Multivariate Level	27,4 %	17,1 %	27,9 %	21,3 %	18,8 %	19,6 %	15,2 %	15,9 %
	Univariate Procent	22,6 %	21,9 %	22,6 %	28,7 %	28,2 %	25,4 %	18,7 %	19,6 %
	Univariate Level	22,1 %	26,2 %	22,4 %	27,0 %	23,4 %	23,2 %	21,5 %	22,4 %

Tabel 6.2: RMSE Resultater for Neurale Netværks Modeller.
Kilde: Output fra R

Forecastet for multivariate LSTM-modellen i levels og den bedst præsterende univariate GRU-model i procent, baseret på det afskårede data med 5 timesteps, illustreres i Figur 6.5. Det observeres, at begge modeller har forecastet en næsten konstant værdi, hvilket kan antyde, at de neurale netværk ikke formår at opfange udviklingen i den forholdsvist korte dataserie.

Det er valgt kun at fremhæve den bedst præsterende model indenfor hver variate-kategori, da det ikke anses som nødvendigt at grafisk vise alle 64 modeller. Ud fra de to fremhævede modeller evalueres multivariate LSTM-modellen i levels, som den bedst præsterende på baggrund af RMSE og grafisk kontrol.



Figur 6.5: Forecastresultater fra Neurale Netværk.

Kilde: Plot fra R

I den følgende sektion forekommer der en sammenfatning af de forecastede resultater for de bedst præsterende modeller målt på RMSE og grafisk kontrol.

6.5 Resultater

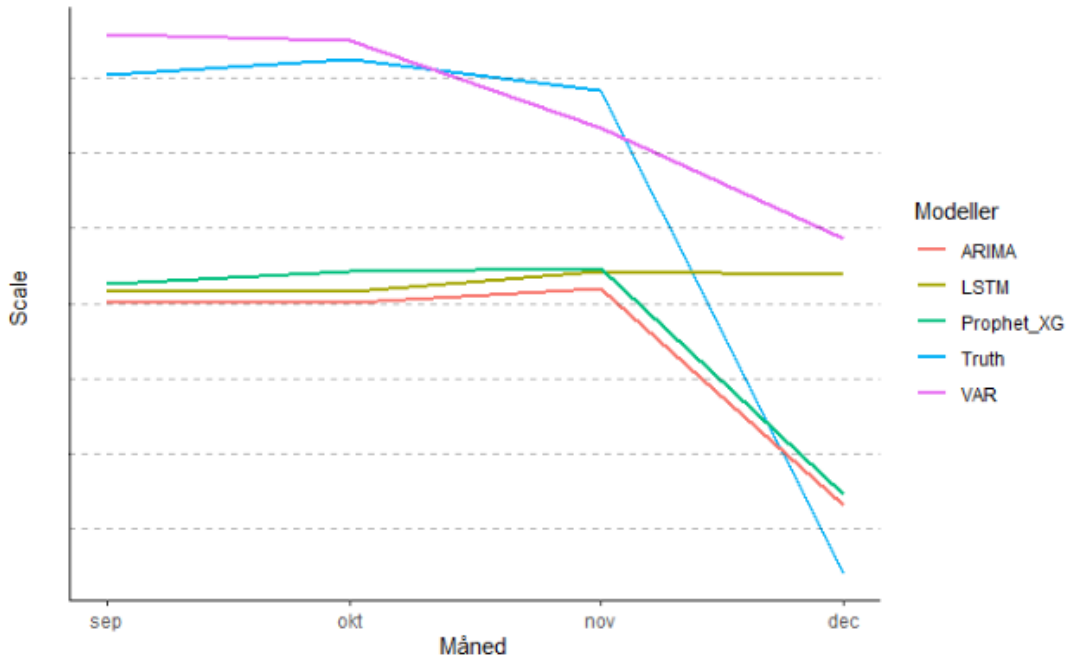
I Tabel 6.3 opstilles RMSE for både uni- og multivariate modeller. Den bedst præsterende model bestemmes ud fra en evaluering af RMSE og den grafiske illustration. I Tabel 6.3 aflæses det, at VAR-modellen er den bedste model, da modellen opnår den laveste RMSE. VAR-modellen konkurrerer tæt med Prophet-modellen m. XGBoost Errors, som i gennemsnit afviger 0,5 procentpoint mere fra de observerede værdier. Den dårligst præsterende model, LSTM-modellen, opnår en procentuel gennemsnitlig afvigelse fra de observerede værdier på 12,6 %, hvilket er 2,9 procentpoint mere i forhold til VAR-modellen.

	RMSE	
Model	*Univariate*	*Multivariate*
ARIMA - Manuel	11,2 %	-
Prophet m. XGBoost Errors	10,2 %	-
VAR	-	**9,7** %
LSTM	-	12,6 %

Tabel 6.3: RMSE og tilhørende procentuel gennemsnitlig afvigelse fra observerede værdier.
Kilde: Output fra R

Den grafiske evaluering af modellernes præstation i Figur 6.6 understøtter det ovenstående resultat. VAR-modellen opfanger i større grad niveau- og udviklingsmæssige ændringer relativt til de andre modeller. På den anden side er den grafiske vurdering, at LSTM-modellen er den svageste model, da de prædikterede værdier er tæt på at være konstante i forecast-perioden. Dette kan skyldes, at datamængden er begrænset. LSTM-modellen er oftest stærkest, når der betragtes data af større størrelse, da den således, i større grad, kan udnytte sine hukommelsesmæssige evner. Prophet- og ARIMA-modellen prædikterer tilnærmelsesvis den samme udvikling, hvor der observeres en lille niveaumæssig forskel mellem de to modeller. I denne forbindelse kan det observeres, at ARIMA-modellen er kompetetiv med de mere komplicerede modeller, da den præsterer bedre end LSTM-modellen og relativt tæt på VAR-modellen målt på RMSE.

VAR-modellen formår at danne et in-sample forecast, som opfanger den generelle uelasiske egenpriselasticitet bedst, da de prædikterede værdier ligger på niveau med den højere omsætning, der observeres i år 2021, hvor energikrisen er tilstede. På baggrund af RMSE og den grafiske præstation evalueres det, at VAR-modellen opnår det bedste resultat.



Figur 6.6: Forecast-resultater.

Kilde: Plot fra R

Præstationen af VAR-modellen i sammenspil med egenpriselasticiteten, beregnet i første del af analysen, kan udtrykke kvaliteten af en fremtidig forecast, da der er en lille procentpoint forskel mellem estimaterne for kort- og langsigt priselasticitet. Dette medvirker til en relativ konstant struktur over tid ift. ændringen i salgsmængden på baggrund af en stigning i salgsprisen. En forecast-model, der anvender strukturen fra tidligere perioder, vil således være mere præcis på længere sigt. Herudfra kan VAR-modellen, for samarbejdsvirksomheden, benyttes til at danne forecast på længere sigt under antagelsen om, at der ikke forekommer ændringer i substitutionsadfærd samt andre stød i dataserien.

6.6 Delkonklusion

I den følgende sektion sammenfattes resultaterne fra analyse del 1 og 2. I den første del blev implikationerne af energikrisen for samarbejdsvirksomhedens omsætning undersøgt ved at applicere en SVAR-model. I denne forbindelse blev der estimeret en kort- og langsigtet egenpriselasticitet på hhv. -0,15 og -0,17, hvilket stemmer overens med den opsatte hypotese i analyseafsnittet. Den uelastiske egenpriselasticitet implicerer, at energikrisen påvirker samarbejdsvirksomhedens omsætning positivt.

I den anden del af analysen anvendes flere typer af modeller til forecasting af omsætningen i energikrisen. De økonometriske modelopsætninger indebar anvendelsen af en ARIMA-baseline samt en VAR-model. Andre statistiske modeller, indenfor maskinlæring og neurale netværk, blev ligeledes opsat såsom en Prophet- og LSTM-model. En evaluering af modellerne på baggrund af RMSE og grafisk præstation viste, at VAR-modellen præsterede bedst. Modellen opfanger den uelastiske egenpriselasticitet fra analyse del 1 bedst i form af en prædiktion af det højere observerede omsætningsniveau i kriseperioden. Derudover blev det illustreret, at de neurale netværk præsterede værst blandt alle modellerne. Årsagen til dette kunne højst sandsynligt tillægges den begrænsede datamængde, som anvendes i afhandlingen. VAR-modellen og maskinlæringsmodellerne optrådte derfor som de bedste forecast-modeller.

7 | Diskussion

I kapitlet reflekteres der over de analytiske resultater, som er opnået i afhandlingen. Dernæst diskuteres det, hvilke implikationer Ruslands invasion af Ukraine har og kan have for energikrisens varighed og i denne sammenhæng samarbejdsvirksomheden. Afslutningsvist inddrages de erfarede begrænsninger, som har haft en indflydelse på besvarelsen af afhandlingens problemstilling. I denne forbindelse præsenteres der ligeledes løsningsforslag til, hvad der kunne have forbedret besvarelsen samt yderligere tilføjelser.

7.1 Refleksion over de Analytiske Resultater

En refleksion over de opnåede resultater i de to analysedele forekommer udelukkende på baggrund af egne forventninger, da der ikke eksisterer andet litteratur på den aktuelle energikrise i forhold til implikationer (estimater) og forecasting for den danske plastindustri.

I den første analysedel blev energikrisens implikationer for samarbejdsvirksomhedens omsætning undersøgt ved brug af en SVAR-model, hvor der blev estimeret en statistisk signifikant kort- og langsigtet egenpriselasticitet. Resultaterne heraf opfyldte den opsatte hypotese i Afsnit 5.1.3. Til trods for at hypotesen opfyldes, var forventningen, at der ville have været en større forskel mellem det kort- og langsigtet estimat for egenpriselasticiteten i form af en relativt mere elastisk egenpriselasticitet på længere sigt end, hvad der var tilfældet. Dette skyldes umiddelbart, at den kortsigtede tidshorisont er defineret for sent, da den største ændring i salgsmængden forekommer efter den første måned på baggrund af et éngangsstød til salgsprisen. I denne forbindelse kan det nævnes, at den kortsigtede tidshorisont oprindeligt var defineret til én måned, men da den kumulative impulse response funktion i Figur 5.1 ikke er statistisk signifikant i denne måned, var dette ikke en mulighed.

I den anden analysedel var forventningen, at multivariate modellerne kunne udkonkurrere univariate versionerne, da disse ville kunne forklare omsætningen i større grad i form af en inkludering af både salgsmængde og -pris. Resultatet var heraf, at den økonometriske multivariate VAR-model klarede sig bedst med en lille præstationsafvigelse fra maskinlæringens univariate Prophet-model med XGBoost Errors. Forventningen opfyldes således delvist, da Prophet- og ARIMA-modellen formår at udkonkurrere LSTM-modellen. Afvigelsen mellem de forskellige modelversioner er ikke markant, hvorfor det kan diskuteres, hvorvidt det er nødvendigt at opsætte mere komplekse modeller såsom en VAR-model, som kræver undersøgelser for stationaritet og andre diagnostikker.

7.2 Ruslands Invasion af Ukraine

Usikkerheden omkring energikrisens varighed øges, og hvilken andel af energiprisstigningerne, der skyldes den aktuelle energikrise sløres af optrapningerne til Ruslands invasion af Ukraine. I det seneste år er den russiske præsident Vladimir Putin blevet beskyldt for at have nedjusteret forsyningerne af gas til Europa for at lægge et pres på godkendelsen af Nord Stream 2-gasledningen. En instrumentalisering af naturgas er således forekommet, hvilket har skabt færre leverancer af gas fra Rusland end hidtil. Det europæiske energimarked påvirkes derved, da Rusland dækker mere end 40% af Europas samlede forbrug af naturgas. Derudover har det øgede konfliktniveau på baggrund af Ruslands invasion af Ukraine ligeledes haft implikationer for energimarkedet, hvor der allerede er iværksat nedjusteringer overfor Polen og Bulgarien i forhold til udbuddet af gas. De to ovenstående faktorer resulterer i højere priser på energi, herunder naturgas, som pålægges allerede stigende priser forårsaget af den eksisterende energikrise (DR1, 2022).

Ruslands invasion af Ukraine og dets implikationer for samarbejdsvirksomheden i form af højere og vedvarende energipriser diskuteres i det følgende. I Afsnit 5 opnås der en generel uelastisk egenpriselasticitet for efterspørgslen efter virksomhedens plastprodukter, hvor analyseresultatet er, at egenpriselasticiteten er relativt mere elastisk på længere sigt. Forventningen om, at invasionen af Ukraine vil øge varigheden af energikrisen, åbner op for en diskussion i forhold til analyseresultatet

ud over den analyserede periode. Hvis der ses bort fra konvergens i SVAR-modellen, og egenpriselasticiteten fortsat bliver relativt mere elastisk på endnu længere sigt, vil samarbejdsvirksomheden blive yderligere påvirket igennem et fald i den vedholdte omsætningsstigning.

Hvordan samarbejdsvirksomhedens omsætning påvirkes af en længerevarende energikrise afhænger såvel af Danmarks priskonkurrenceevne overfor andre lande, hvilket påvirkes af forskellen i energiintensiteten mellem de pågældende lande over tid. I Figur 2.4 ses det, at Danmark besidder den laveste energiintensitet relativt til andre lande, når der tages højde for branchestruktur. I denne forbindelse kan det diskuteres, hvorvidt dette er gældende på endnu længere sigt, når andre lande ligeledes arbejder imod øget uafhængighed af Rusland og derigennem en ændring i deres energistruktur. Et eksempel på energiuafhængighed afspejles f.eks. igennem Danmarks grundlæggelse af en energiø med målet om at kunne centralisere dansk energi (Regeringen, 2021). Hvis andre lande, alt andet lige, opnår en lavere energiintensitet over tid, falder Danmarks priskonkurrenceevne. En kobling kan laves til egenpriselasticiteten, som fortsat forventes at være relativt mere elastisk på endnu længere sigt, når invasionen af Ukraine tages i betragtning. En relativt forværret priskonkurrenceevne medfører således på sigt, at samarbejdsvirksomhedens kunder ikke kun har mere tid til at substituere, men de har ligeledes et større udvalg af internationale producenter, som nu er blevet mere favorable prismæssigt.

Længden på energikrisen og herigennem et højere prisniveau på energi er således ukendt, og hvorledes Putin fortsætter med at skrue ned for energieksporten til andre lande er uvist på nuværende tidspunkt, da det afhænger af fremtidige begivenheder mellem de europæiske lande og Rusland.

7.3 Begrænsninger

Operationaliseringen af kandidatafhandlingens problemstilling er forekommet på baggrund af anvendelsen af proxy variable, da virksomhedens salg er ordrebestemt, hvilket indebærer et bredt heterogent produktsortiment samt priser, som defineres pr. ordre. I denne forbindelse opstår der

en problematik i at finde de mest optimale proxy variable for virksomhedens salgsmængde samt salgspris ud fra den begrænsede mængde af variable, der er stillet til rådighed. Inputprisen på plast, der er udvalgt som proxy i afhandlingen, anses som værende det mest homogene og dermed pålidelige mål, der er til rådighed. I SVAR-modellen tages der udgangspunkt i perioden fra år 2017 til år 2020. Antagelsen om, at inputprisen er styrende for virksomhedens salgspris er korrekt i det meste af perioden, hvilket kan observeres i Figur 4.1. Fra år 2017 til 2019 ligger salgsmængden over omsætningen i forskellige grader, hvilket stemmer overens med udviklingen i inputprisen, der angiver virksomhedens salgspris. I år 2020 holder antagelsen ikke, da omsætningen ligger betydeligt over salgsmængden, hvilket betyder, at forskellen i disse angiver, at der er en øget salgspris. Dette kan ikke observeres ud fra den anvendte proxy, og stigningen i den reelle salgspris er således forårsaget af andre faktorer såsom en øget mark-up. Da den anvendte proxy ikke opfanger stigningen i den reelle salgspris, forekommer der et bias i analyseresultaterne. Et eksempel på en bedre proxy for salgsprisen kunne derfor være en variabel, som kombinerer inputpriser og mark-up.

En identisk problematik opstår i forbindelse med virksomhedens salgsmængde, hvor energiforbruget anvendes som et homogent mål, da virksomhedens produktion er fordelt ud på mange forskellige varer. Som beskrevet i Afsnit 1.2, er hovedargumentet for at kunne anvende energiforbruget, som et mål for den solgte mængde, at samarbejdsvirksomheden har haft stor fokus på, at maskiner ikke er i unødvendig drift. I denne forbindelse kan det diskuteres, hvorvidt det er værd at søge efter de bedste proxy'er, når mindre komplicerede univariate modeller såsom ARIMA og Prophet kan benyttes til at opnå tilnærmelsesvis samme resultater, som diskuteret i forrige sektion.

En begrænsning i forbindelse med undersøgelsen af problemstillingen er, at datamængden ligger i underkanten af, hvad der generelt er fordelagtigt for økonometriske og statistiske modeller. Et eksempel kan gives i forbindelse med en (S)VAR-model, som kræver et større antal observationer, da modellen hurtigt kan konsumere mange frihedsgrader ved tilføjelse af lags i ligningssystemet. En implikation af dette er en generelt mindre efficient model. Derudover er færre observationer forekommet på baggrund af et strukturelt brud i den betragtede periode. Der er således valgt en tidsperiode

uden statistisk signifikante brud fremfor flere observationer, da det øger repræsentativitet samt stabiliserer estimaterne i modellerne. Implikationerne af den begrænsede tidsserie kommer ydermere til udtryk i LSTM-modellen, som typisk præsterer bedre ved større kompleksitetsgrad og datamængde, hvilket illustreres i Figur 6.6.

Definitionen af den langsigtede tidshorisont i forbindelse med estimeringen af egenpriselasticiteten er valgt på baggrund af SVAR-modellens præmisser. En langsigtet tidshorisont defineres typisk til at være længere end ét år, men da modellen er konvergeret på dette tidspunkt, kan der ikke udvindes yderligere information for en længere tidshorisont. Ydermere findes der ikke andre studier på området, hvilket resulterer i, at den langsigtede tidshorisont endeligt defineres på baggrund af denne konvergens.

7.3.1 Fremtidig Forskning og Udvikling

Den datamæssige problematik kan udbedres, givet mere tid og ressourcer, ved benyttelse af en anden proxy for samarbejdvirksomhedens salgsmængde, som ikke indeholder et statistisk signifikant brud. Inkluderingen af den nuværende salgsmængde begrænser brugen af en længere tidsserie i modellerne, hvilket resulterer i en mindre efficient model igennem færre frihedsgrader. Derudover, da der eksisterer bias i den udvalgte proxy for salgsprisen, findes der ligeledes en mulig forbedring i denne sammenhæng. Samarbejdsvirksomhedens udvalg af data er ukendt, hvilket har problematiseret en præcis forespørgsel efter alternative homogene proxy'er.

I forbindelse med det fremtidige arbejde kan der inddrages en undersøgelse af, hvilken variabel, som forklarer samarbejdsvirksomhedens omsætning mest fra måned til måned. Lloyd Shapley har, igennem sit fokus på spilteori, givet en matematisk beskrivelse af, hvilke spillere, der forklarer mest i et kooperativt spil. Shapleys idé er udvidet og kan anvendes i forbindelse med neurale netværk og forecasting, hvilket gennemgås i Ferreira, 2019. Dette har især en øget relevans, når der betragtes modeller med flere variable relativt til, hvad der er inkluderet i denne afhandling. Implementeringen af Shapley værdier vil således kræve en udvidelse af de nuværende forecast-modeller med yderligere

variable i det fremtidige arbejde. I denne forbindelse kunne det være yderst interessant at vide, hvilken variabel, som forklarer omsætningen mest i perioder, hvor der forekommer stød, da det kan give en indikator for, hvilken type af stød, der fremtidigt kan forventes.

En fremtidig udvidelse af afhandlingen kan ydermere forekomme ved at betragte en alternativ tidsperiode for at undersøge om forecast-resultaterne er konsistente udenfor den anvendte periode. Da denne afhandling fokuserer på en kriseperiode, vil dette inddrage usikkerhed i analyseresultaterne, og det forventes, at modellerne ville præstere bedre udenfor denne periode. En undersøgelse af denne hypotese kunne være interessant for at afklare, hvorledes den medtagne usikkerhed er af større betydning for modellernes forecast.

8 | Konklusion

I den følgende sektion besvares hovedproblemstillingen ved at tage udgangspunkt i de tre underspørgsmål i Afsnit 1.1, som betragtes i kronologisk rækkefølge.

Den aktuelle energikrise opstod i år 2021 på baggrund af coronapandamiens nedlukning og en manglende vedvarende energi. En blanding af mindre blæst og regn har medført, at produktionen af energi fra de danske vindmøller faldt samtidigt med, at billig strøm fra Norge og Sveriges vandkraftværker ikke kunne erhverves i samme grad som før. En manglende vedvarende energi har således forårsaget, at Danmark har været nødsaget til at anvende bl.a. naturgas som et alternativ til den traditionelle strømkilde. I denne forbindelse opstår problemet, da Europas gaslagre har været relativt lave, hvilket er forekommet på baggrund af en kold vinter og varm sommer i år 2020 og 2021. Herudover har coronakrisen været årsag til, at produktionen af gas har været nedjusteret, da der har eksisteret en forventning om en længerevarende økonomisk krise. Disse ovenstående faktorer og begivenheder har været hovedårsagerne til den aktuelle energikrise.

I kandidatafhandlingen opstilles der en SVAR-model, som tager udgangspunkt i data fra år 2017-2020, for at undersøge implikationerne af energikrisen på samarbejdsvirksomhedens omsætning. Analysen foretages ved at betragte, hvor følsom samarbejdsvirksomhedens efterspørgsel er overfor en ændring i salgsprisen. I denne forbindelse viser estimationen af en kort- og langsigtet egenpriselasticitet, at en forøgelse af salgsprisen på 1% foranlediger et fald i samarbejdsvirksomhedens salg med 0,15% på kort sigt og 0,17% på længere sigt. En relativt mere elastisk egenpriselasticitet på sigt stemmer overens med de opsatte forventninger i analyseafsnittet, da der opstår en øget mulighed for at substituere over til en anden producent. Den generelt uelastiske egenpriselasticitet implicerer således, at energikrisen påvirker samarbejdsvirksomhedens omsætning positivt.

Der anvendes flere typer af modeller til forecasting af samarbejdsvirksomhedens omsætning i krisepe-

rioden. Heriblandt konkurrerer ARIMA-modellen, som baseline, imod de bedst præsterende modeller indenfor hver type, hvor Prophet, VAR og LSTM inkluderes. Det er tydeliggjort, at LSTM-modellen muligvis begrænses af en lille datamængde, som er yderligere afskåret på baggrund af et strukturelt brud i samarbejdsvirksomhedens salg. Det kan dermed konkluderes, at VAR-modellen præsterer bedst efterfulgt af maskinlæringsmodellen Prophet m. XGBoost Errors baseret på RMSE og grafisk præstation.

I diskussionen inddrages Ruslands invasion af Ukraine, da dette er aktuelt for år 2022, og det har implikationer for effekt og varighed af energikrisen. Invasionen forlænger energikrisen på ukendt tid og forårsager højere og vedvarende energipriser, hvilket påvirker samarbejdsvirksomheden. I analyseafsnittet findes der estimater for en generel uelastisk egenpriselasticitet for efterspørgslen efter virksomhedens plastprodukter, hvor egenpriselasticiteten bliver relativt mere elastisk på længere sigt. Da invasionen af Ukraine forventes at øge varigheden af energikrisen, kan analyseresultaterne betragtes ud over den analyserede periode. Hvis egenpriselasticiteten fortsat bliver mere elastisk på endnu længere sigt, medvirker det til, at den vedholdte omsætningsstigning bliver mindre. Derudover kan det diskuteres, hvorledes egenpriselasticiteten på sigt ændrer sig på baggrund af tilpasninger i priskonkurrenceevnen, da andre lande stræber efter øget uafhængighed af russisk energi.

8.1 Vidensbidrag

Afhandlingen har anvendt artiklerne Ghoddusi et al., 2019; Narayan et al., 2014 som belæg for undersøgelsen af problemstillingen i Afsnit 1. Afhandlingen har fundet belæg for at anvende en indirekte energivariable til beskrivelse af en ikke-energivariable for en energiintensiv virksomhed. Derudover har afhandlingen sammenlignet en bred variation af forecast-modeller, herunder neurale netværk, som foreslået i litteraturen. Det blev således konkluderet, at de komplekse neurale netværk ikke præsterede, i denne afhandling, som forventet i den inddragede litteratur. Herefter præsenteres den nyerhvervede viden til hhv. de industrielle, politiske og teoretisk indvirkninger.

8.1.1 Industriel indvirkning

Afhandlingens brugbarhed for industrien forekommer ved en øget indsigt i den eksisterende heterogenitet i virksomheder, og hvilke proxy'er, der potentielt kan benyttes i en modelopsætning til at undersøge en lignende problemstilling. Som beskrevet i Afsnit 7.3, er det uvist, om der anvendes de mest fordelagtige proxy variable, da disse homogene variable kan kritiseres på andre punkter, hvor det er nødvendigt med antagelser jf. Afsnit 1.2. Den enkelte virksomhed kan i sidste ende vurdere, om det er nødvendigt at anvende de mere komplekse modeller såsom VAR og neurale netværk, da de simplere modeller er konkurrencedygtige i denne afhandling.

8.1.2 Politiske overvejelser

I tilfælde af at andre lande øger deres energiuafhængighed og fokuserer på at blive mindre energiintensive på baggrund af Ruslands instrumentalisering af naturgas kan det overvejes, hvorledes kompensation for energiintense industrier bør indføres. Kompensationen skal indføres for at fastholde den danske priskonkurrenceevne overfor udlandet og derigennem for at undgå en relativt mere elastisk egenpriselasticitet på længere sigt. Derudover bibeholdes skatteindtægten i Danmark. Afhandlingen bidrager således til en generel forståelse af den potentielle indvirkning, som Ruslands invasion af Ukraine kan have på Danmarks internationale konkurrenceevne i fremtiden.

8.1.3 Teoretiske bidrag

Datagrundlaget er fortsat bestemmende for, hvilken modelopstilling, der bør anvendes. Kandidatafhandlingen understreger det allerede eksisterende metodiske fundament omkring brugen af neurale netværk. I denne forbindelse har de neurale netværk behov for en højere kompleksitetsgrad ved en større observationsmængde eller flere uafhængige variable for at excellere. Herunder er det erfaret, at en modelopsætning baseret på VAR eller simplere maskinlæring såsom Prophet med XGBoost Errors kan producere bedre forecast-resultater.

9 | Litteratur

Al-Chalabi, H., Al-Douri, Y. K., & Lundberg, J. (2018). Time series forecasting using arima model.

Andersen, M. M. (2019). Lineær interpolation i pointgivning af budgetter [First used on: 06-05-2022]. https://mikl.dk/da/post/2019-budget-point-linear-interpolation/

Ardel, N. (2018). How lstm networks solve the problem of vanishing gradients [First used on: 01-04-2022]. https://medium.datadriveninvestor.com/how-do-lstm-networks-solve-the-problem-of-vanishing-gradients-a6784971a577

Bhimani, A. (2019). Management and cost accounting (7th ed.).

Brownlee, J. (2018). How to use the timeseriesgenerator for time series forecasting in keras [First used on: 19-05-2022]. https://machinelearningmastery.com/how-to-use-the-timeseriesgenerator-for-time-series-forecasting-in-keras/

Brownlee, J. (2019). A gentle introduction to sarima for time series forecasting in python [First used on: 02-05-2022]. https://machinelearningmastery.com/sarima-for-time-series-forecasting-in-python/

Brownlee, J. (2021). How to choose an activation function for deep learning [First used on: 31-03-2022]. https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/

Chai, T., & Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (mae)?

Chollet, F. et al. (2015). Keras [First used on: 31-03-2022]. https://keras.io

Danmark, F., & Bernhard, M. (2022). Betaler syv millioner kroner mere om måneden - høje energipriser ender hos forbrugerne [First used on: 17-03-2022]. https://www.tv2east.dk/guldborgsund/betaler-syv-millioner-kroner-mere-om-maaneden-hoeje-energipriser-ender-hos-forbrugerne

Danmarks Statistik. (2015). Pris1115: Prisindeks for indenlandsk vareforsyning (2015=100) efter varegruppe og enhed [First used on: 15-03-2022]. https://www.statbank.dk/statbank5a/SelectVarVal/Define.asp?Maintable=PRIS1115&PLanguage=0

Danmarks Statistik. (2022). Største stigning i forbrugerpriserne i over 13 år [First used on: 15-03-2022]. https://www.dst.dk/da/Statistik/nyheder-analyser-publ/nyt/NytHtml?cid=32942

Dansk Gasteknisk Center. (n.d.). Gassystemet [First used on: 09-03-2022]. https://www.gasfakta.dk/gas-i-danmark/gassystemet

dProgrammer lopez. (2019). Rnn, lstm & gru [First used on: 28-03-2022]. http://dprogrammer.org/rnn-lstm-gru

DR1. (2022). Explainer: Hvorfor bliver det dyrere for dig? [First used on: 9-03-2022]. https://smltar.com

Enders, W. (2014). Applied econometric time series [First used on: 31-03-2022].

Energi Danmark. (2018). Danmarks kulforbrug falder fortsat [First used on: 15-03-2022]. https://www.energidanmark.dk/markedsinfo/nyheder/nyheder/nyheder-2018/11/danmarks-kulforbrug-falder-fortsat/

EnergiFyn. (2021). Hvor kommer strømmen fra? [First used on: 09-03-2022]. https://www.energifyn.dk/privat/el/vaerd-at-vide-om-el/hvor-kommer-strommen-fra/

Energistyrelsen. (n.d.). Fakta om energiintensitet [First used on: 16-03-2022]. https://ens.dk/ansvarsomraader/energi-klimapolitik/erhvervslivets-energiforhold/fakta-om-energiintensitet

Energistyrelsen. (2022). Energipriser og afgifter [First used on: 10-03-2022]. https://ens.dk/service/statistik-data-noegletal-og-kort/energipriser-og-afgifter

EnergiWatch. (2022). Overblik: Corona i energibranchen [First used on: 14-03-2022]. https://energiwatch.dk/Energinyt/Politik___Markeder/article12002452.ece

Ferreira, A. (2019). Interpreting recurrent neural networks on multivariate time series [First used on: 16-05-2022]. https://towardsdatascience.com/interpreting-recurrent-neural-networks-on-multivariate-time-series-ebec0edb8f5a

Ghoddusi, H., Creamer, G. G., & Rafizadeh, N. (2019). Machine learning in energy economics and finance: A review [First used on: 17-05-2022]. https://www.sciencedirect.com/science/article/pii/S0140988319301513

Hastie, T., Qian, J., & Tay, K. (2022). An introduction to glmnet.

Hvitfeldt, E., & Silge, J. (2021). Supervised machine learning for text analysis in r [First used on: 31-03-2021]. https://smltar.com

Jersild, A. O., & Munch, J. (2022). En el-regning fra 20 til 100 millioner på et år: Høje energipriser presser virksomhederne [First used on: 17-03-2022]. https://www.dr.dk/nyheder/penge/en-el-regning-fra-20-til-100-millioner-paa-et-aar-hoeje-energipriser-presser

Kristensen, S. (2021). Høje energipriser giver en konkurrencefordel til danske virksomheder [First used on: 16-03-2022]. https://www.sydbank.dk/privat/artikler/artikel/hoeje-energipriser-giver-en-konkurrencefordel-til-danske-virksomheder

Letham, B., & Taylor, S. J. (2017). Prophet: Forecasting at scale [First used on: 02-05-2022]. https://research.facebook.com/blog/2017/02/prophet-forecasting-at-scale/

Lütkepohl, H. (2013). Vector autoregressive models [First used on: 31-03-2022]. https://www.elgaronline.com/view/edcoll/%209780857931016/9780857931016.00012.xml

Mankiw, N. G., & Taylor, M. P. (2011). Microenonomics (2nd ed.).

Mankiw, N. G., & Taylor, M. P. (2017). Microeconomics (4th).

Narayan, P. K., Sharma, S., Poon, W. C., & Westerlund, J. (2014). Do oil prices predict economic growth? new global evidence [First used on: 17-05-2022]. https://www.sciencedirect.com/science/article/pii/S0140988313002478

Nechyba, T. J. (2011). Microeconomics: An intuitive approach with calculus (International Edition).

NRGI. (2022). Fakta om olieenergi [First used on: 15-03-2022]. https://nrgi.dk/privat/elforbrug/energiformer/olie/

Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks [First used on: 05-04-2022]. https://arxiv.org/pdf/1211.5063

Petneházi, G. (2018). Recurrent neural networks for time series forecasting [First used on: 28-03-2022]. https://www.researchgate.net/publication/330102696_Recurrent_Neural_Networks_for_Time_Series_Forecasting

Plast Forum. (2020). 16 plastvirksomheder indgår ny aftale med energistyrelsen [First used on: 15-03-2022]. https://www.plastforum.dk/article/view/699681/16_plastvirksomheder_indgar_ny_aftale_med_energistyrelsen

Plastindustrien. (n.d.). Hvordan laver man plast? fremstilling af plastik er en kompliceret proces [First used on: 15-03-2022]. https://plast.dk/hvordan-laver-man-plast/

Qin, Y., Tang, H., & Guo, C. (2006). Channel coordination and volume discounts with price-sentive demand.

RDocumentation. (n.d.). Rdocumentation [First used on: 07-04-2022]. https://www.rdocumentation.org/

Regeringen. (2021). Tillæg til klimaaftale om energi og industri af 22. juni 2020 vedr. ejerskab og konstruktion af energiøer mv. [First used on: 12-05-2022]. https://kefm.dk/Media/5/E/Aftaletekst%5C%20-%5C%20Energi%5C%C3%5C%B8er%5C%20-%5C%20Ejerskab%5C%20og%5C%20konstruktion%5C%20af%5C%20energi%5C%C3%5C%B8er%5C%20mv.pdf

Sharma, P. (2022). Keras dense layer explained for beginners [First used on: 31-03-2022]. https://machinelearningknowledge.ai/keras-dense-layer-explained-for-beginners/

Varian, H. R. (2019). Intermediate microeconomics (9th ed.).

Wang, C.-C., Chien, C.-H., & Trappey, A. J. C. (2021). On the application of arima and lstm to predict order demand based on short lead time and on-time delivery requirements.

Wooldridge, J. M. (2009). Introductory econometrics - a modern approach (4th ed.).

XGBoost Developers. (2021). Xgboost documentation [First used on: 02-05-2022]. https://xgboost.readthedocs.io/en/stable/tutorials/model.html

Yilmazkuday, H. (2019). Estimating the trade elasticity over time.

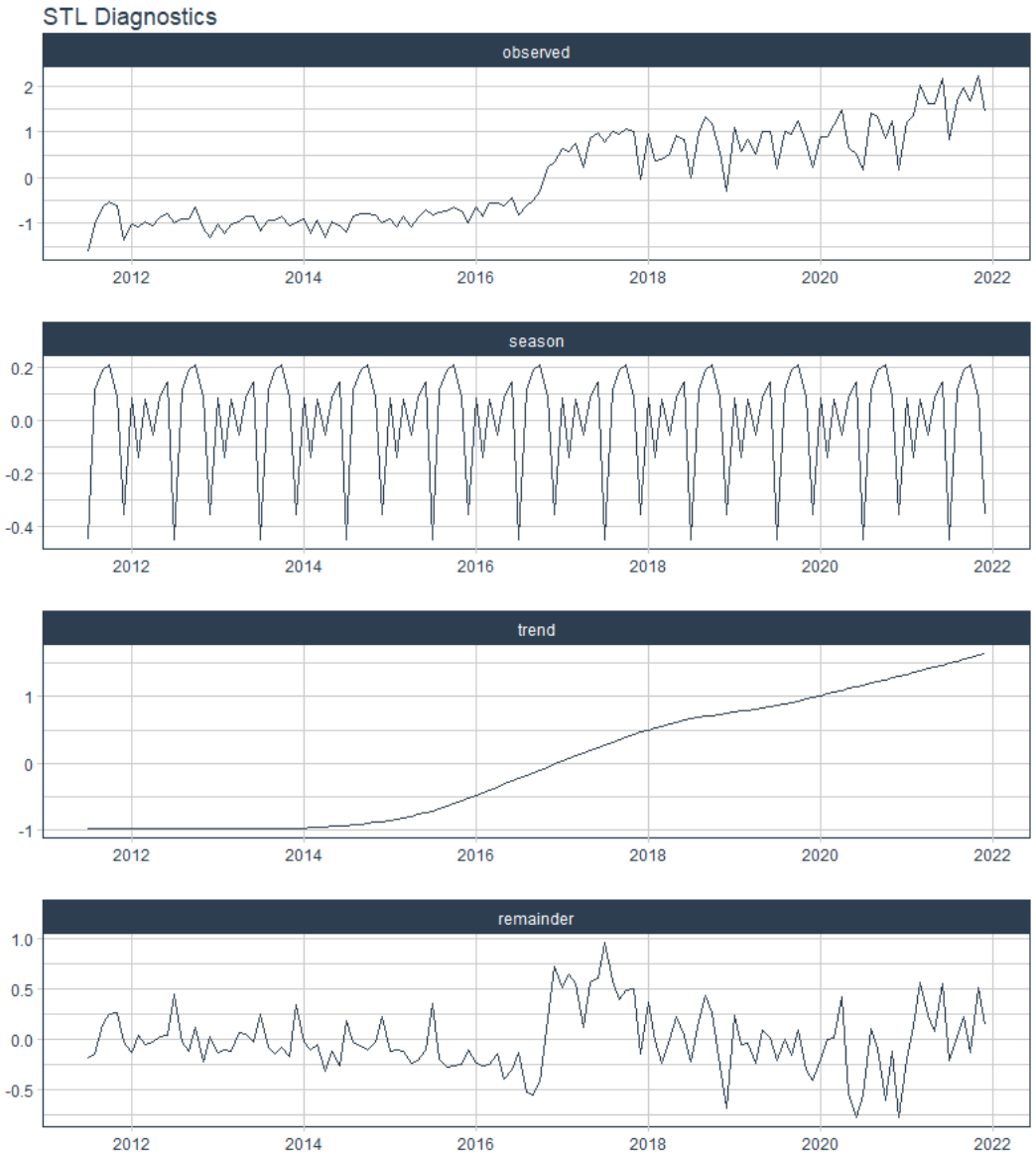
A | Bilag

A.1 Sæson-Trend-Loess Decomposition (STL)

I den indledende sektion foretages der en grafisk undersøgelse af de udvalgte tidsserier ved brug af STL Decomposition. Decompositionsprocessen kan opstilles på baggrund af følgende ligning: $X_t = seasonal_t + trend_t + cyclical + \varepsilon_t$, hvor X_t angiver den udvalgte tidsserie, som består af de efterfølgende komponenter. De følgende figurer viser således fire plots for hver af tidsserierne, hvor det første plot illustrerer alle komponenter (observed). Dernæst isoleres sæson- og trendkomponenterne, hvor det fjerde plot udelukkende illustrerer det cykliske komponent og fejlleddet (remainder).

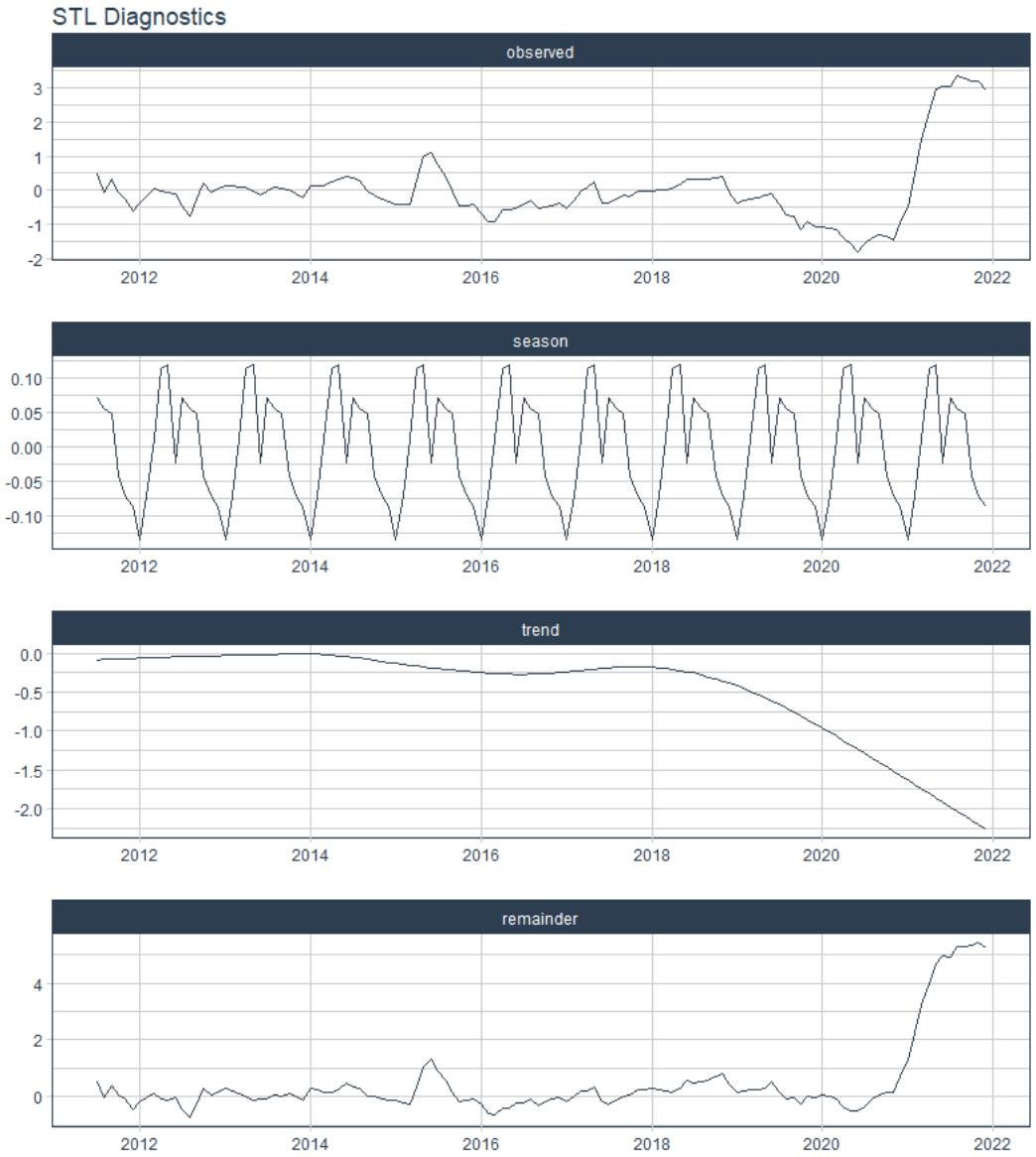
Figur A.1: Omsætning Decomposition.

Kilde: Plot fra R



Figur A.2: Salgsmængde Decomposition.

Kilde: Plot fra R



Figur A.3: Salgspris Decomposition.

Kilde: Plot fra R

A.2 Sæsonmæssige Variationer og Unit Root

I denne sektion testes de udvalgte tidsserier for eksistensen af statistiske signifikante sæsonudsving, da det er det cykliske komponent, der søges i disse tidsserier. Derudover foretages der tests for unit root, da stationaritet er en essentiel betingelse for, at de økonometriske modeller kan anvendes til at besvare kandidatafhandlingens problemstilling.

A.2.1 Undersøgelse af Sæsonmæssige Variationer

Undersøgelsen af sæsonmæssige fluktuationer i tidsserierne foretages ved at tage udgangspunkt i dummy variabel metoden, som benytter regressionen herunder:

$$\omega_t = \alpha + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_3 + \beta_4 D_4 ... + \beta_{11} D_{11} \tag{A.1}$$

Den lineære regression indeholder den afhængige variabel, dvs. tidsserien, som testes for sæsonudsving, ω_t, og denne bliver således forklaret af 11 dummy variable, D_i, til 12 kategorier svarende til den månedlige frekvens. I denne forbindelse foretages der en F-test, hvor nulhypotesen, H_0, udtrykker, at disse dummy variable ikke har en fælles statistisk signifikant effekt på den afhængige variabel, ω_t, hvilket betyder, at der statistisk evidens for, at der ikke er sæsonudsving i tidsserien. Hvis, H_0, derimod kan forkastes til fordel for alternativhypotesen, H_1, implicerer det, at der er statistisk evidens for sæsonmæssige fluktuationer i tidsserien.

Sæson Test	*Test-statistik:*	*P-værdi:*
Omsætning	1.06	0.41
Salg	2.11	0.06
Salgspris	0.13	0.99

Tabel A.1: Test af sæsonkomponent i tidsserie.

Kilde: Output fra R

F-statistikkerne tilhørende omsætning, salg samt salgspris sammenholdes med deres respektive p-værdier i Tabel A.1, og det ses, at disse ligger over et 5% signifikansniveau. Nulhypotesen, H_0, kan ikke forkastes til fordel for alternativhypotesen, H_1, hvilket implicerer, at der er statistisk evidens for, at der ikke er sæsonudsving i de udvalgte tidsserier.

A.2.2 Undersøgelse af Unit Root

I dette underafsnit undersøges de udvalgte variable for unit root, da stationaritet er et krav for at kunne opstille og derved anvende de inkluderede økonometriske modeller i kandidatafhandlingen.

En regulær Dickey-Fuller Test (DF) antager på forhånd, at fejlleddet er white noise, hvilket betyder, at der ikke eksisterer seriel korrelation (ε_t og ε_{t-k} er serielt ukorreleret for alle k > 0). Dette er en stærk antagelse i praksis, og derfor appliceres Augmented Dickey-Fuller Test (ADF), som udvider den traditionelle test med tilføjelsen af lags af den differencerede afhængige variabel som uafhængig variabel for at imødekomme problemet (Wooldridge, 2009).

Augmented Dickey-Fuller testen (ADF) opskrives på følgende regression:

$$\Delta\omega_t = \alpha + \beta t + \delta\omega_{t-1} + \sum_{i=1}^{k} \gamma_i \Delta\omega_{t-1} + \varepsilon_t \tag{A.2}$$

hvor ω betegner den afhængige variabel, som testes for unit root.

I det følgende opstilles nulhypotesen, $H_0 : \delta = 0$ mod alternativhypotesen, $H_1 : \delta < 0$. Hvis nulhypotesen, H_0, ikke kan forkastes til fordel for alternativhypotesen, H_1, implicerer det, at der er unit root i den pågældende tidsserie, ω_t, dvs. gennemsnit og varians er ikke invariant over tid. På den anden side, hvis nulhypotesen, H_0 kan forkastes til fordel for alternativhypotesen, H_1, så er der ikke unit root i tidsserien.

Test-statistikkerne tilhørende omsætning, salg samt salgspris sammenholdes med den kritiske værdi, tilhørende et 5% signifikansniveau, for at undersøge, hvorvidt de udvalgte variable lider af unit

root. Det viser sig, at test-statistikkerne for omsætning og salg er lavere end den kritiske værdi tilhørende et 5% signifikansniveau på -3.45, hvilket betyder, at nulhypotesen, H_0, kan forkastes til fordel for alternativhypotesen, H_1. Der er således statistisk evidens for, at der ikke er unit root i disse tidsserier. Hvis salgsprisen derimod betragtes, så er test-statistikken større end den kritiske værdi, og der er statistisk evidens for, at denne tidsserie lider af unit root. Unit root elimineres ved at tage den første differens af tidsserien. Operationen foretages på alle tre variable, da dette er nødvendigt for at kunne opsætte en brugbar model. Derudover tages der logaritmen af variablene for at reducere variansen og af andre modeltekniske årsager såsom at betragte procentuelle ændringer i afhandlingen.

Unit Root Test (ADF)	*Test-statistik:*
Omsætning	-5.20
Salg	-4.61
Salgspris	-1.70

Tabel A.2: Test af unit root i tidsserier.

Kilde: Output fra R

ADF-testen udføres efter, at der er taget den første differens samt logaritmen af salgspris-tidsserien, og der fås en test-statistik på -3.10 jf. Tabel A.3. Test-statistikken sammenholdes med den kritiske værdi, tilhørende et 5% signifikansniveau på -1.95. Da test-statistikken er lavere end den kritiske værdi, kan nulhypotesen, H_0, forkastes til fordel for alternativhypotesen, hvilket implicerer, at salgspris-tidsserien ikke lider af unit root.

Unit Root Test (ADF)	*Test-statistik:*
Salgspris	-3.10

Tabel A.3: Test af unit root i salgspris-tidsserien.

Kilde: Output fra R

A.3 Opsætning af SVAR-model

I sektionen opsættes SVAR-modellen på baggrund af de ovennævnte variable, som er stationære efter at have taget den første differens og anvendt logaritmen på dem. Opsætningen af SVAR-modellen indebærer indledningsvist et valg af lags i modellen, som bestemmes ud fra *VARselect* algoritmen i R. Der tages med andre ord udgangspunkt i informationskriteriet, AIC, som er beskrevet i afhandlingens metode under Afsnit 3.3.4. Dernæst testes modellen for stationaritet og seriel korrelation.

I SVAR-modellen indgår samarbejdsvirksomhedens salg af plastprodukter samt salgspris som variable, da formålet, kort fortalt, er at undersøge implikationerne af energikrisen for virksomhedens omsætning igennem benyttelsen af impulse response funktioner til estimering af egenpriselasticiteter. Ud fra algoritmen, *VARselect*, minimeres AIC-informationskriteriet ved at tilføje ét lag af hver endogen variabel i systemet. I Tabel A.4 vises estimeringsresultaterne for samarbejdsvirksomhedens salg, *lnsales*, og heri kan det ses, at lagget af salget selv er højsignifikant. Konstanten inkluderes naturligt i modellen, da der skal være en god grund til at ekskludere denne.

	Afhængig variabel:		
	lnsales		
	Estimate	Pr(>|t|)	Signifikans
lnprice.l1	-0.191632	0.74068	
lnsalg.l1	-0.441747	0.00352	***
const	-0.001681	0.93276	
Frihedsgrader			43
Adjusted R^2			0.1458

Note: *p<0.1; **p<0.05; ***p<0.01

Tabel A.4: Estimationsresultater for ligningen, lnsales.

Kilde: Output fra R

Før SVAR-modellen kan anvendes til at undersøge afhandlingens problemstilling, er det nødvendigt at kontrollere for stationaritet og seriel korrelation. I R benyttes koden, *roots*, til at undersøge

modellens rødder, hvor det kræves, at disse ligger indenfor enhedscirklen for at opfylde stationaritetskriteriet for en VAR-model jf. Afsnit 3.12.

Ved at aflæse Tabel A.5 kan det konkluderes, at alle egenværdierne er mindre end 1 i absolut værdi, hvilket implicerer, at rødderne ligger indenfor enhedscirklen. SVAR-modellen opfylder således stationaritetskriteriet.

	Rodens værdi:
Rod 1	0.4270307
Rod 2	0.1416776

Tabel A.5: Undersøgelse af rødder for SVAR-modellen.
Kilde: Output fra R

Afslutningsvist undersøges det, hvorvidt der er seriel korrelation tilstede i modellen ved brug af Portmantaeu-testen. Ved at betragte p-værdien i Tabel A.6, som ligger langt over et 5% signifikansniveau, kan nulhypotesen, H_0, ikke forkastes til fordel for alternativhypotesen, H_1. Dette implicerer således, at der er statistisk evidens for, at der ikke er seriel korrelation i modellen. Modellen kan nu benyttes til at foretage en impulse response analyse.

	P-værdi:
Portmanteau Test	0.8885

Tabel A.6: Test af seriel korrelation for SVAR-modellen.
Kilde: Output fra R

I Tabel A.7 illustreres, B_0-matricen for SVAR-modellen, hvor de samtidige effekter mellem de endogene variable indgår. Herudfra kan det ses, at der er pålagt én restriktion i overensstemmelse med Afsnit 3.28, hvor salgsprisen, $lnprice$, optræder som den mest eksogene variabel i systemet og salget, $lnsales$, som den mest endogene variabel. Når disse samtidige effekter ganges med 100, skal

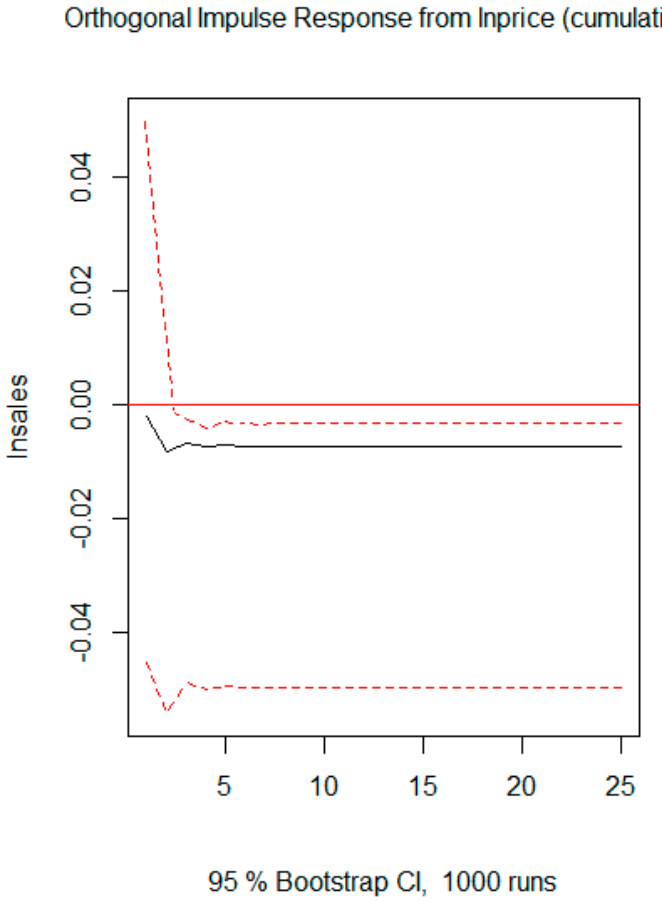
et éngangsstød til samarbejdsvirksomhedens salgspris på ét procentpoint i den første måned tolkes som et fald i virksomhedens salg på 0,21 procentpoint i den pågældende måned.

	lnprice	*lnsales*
lnprice	0.036910446	0
lnsalg	-0.002091669	0.133366

Tabel A.7: B_0-matricen for SVAR-modellen.
Kilde: Output fra R

SVAR-model: Kumulative Impulse Response Funktioner

I dette afsnit vedlægges de kumulative impulse response funktioner i form af figurer og tal. Disse funktioner er identiske med dem, der præsenteres i Afsnit 5.1.3 til beregning af en kort- og langsigtet egenpriselasticitet.



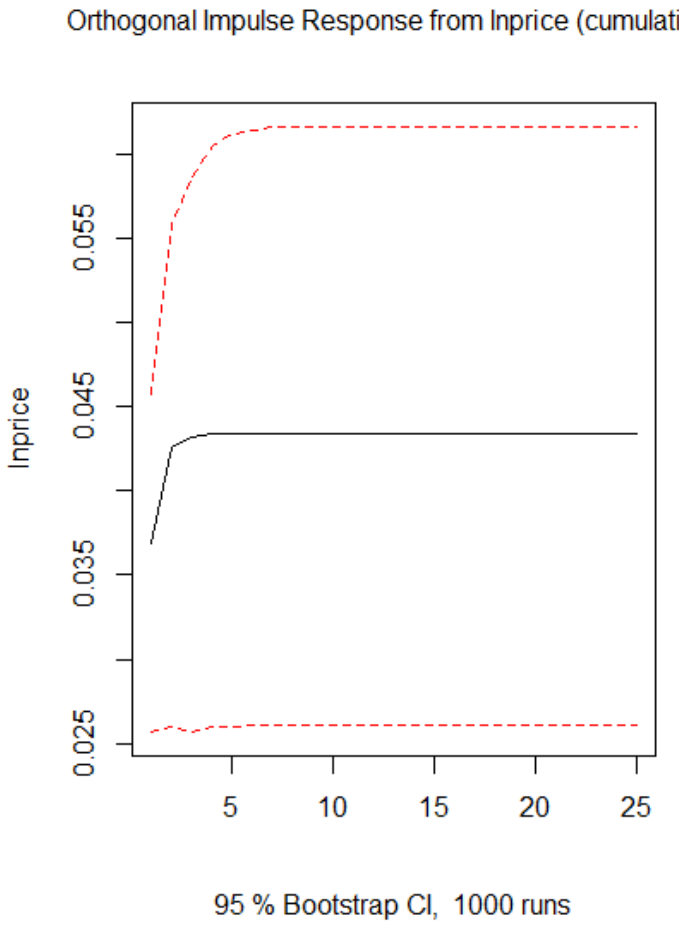
Figur A.4: Kumulative impulse responses af salgspris på salg.

Kilde: Plot fra R

	lnsales
1. måned	-0.002091669
2. måned	-0.008240911
3. måned	-0.006612753
4. måned	-0.007449386
5. måned	-0.007112146
6. måned	-0.007258995
7. måned	-0.007196688
8. måned	-0.007223352
9. måned	-0.007211974
10. måned	-0.007216834
11. måned	-0.007214759
12. måned	-0.007215645

Tabel A.8: Kumulative impulse response af lnprice på lnsales.

Kilde: Output fra R



Figur A.5: Kumulative impulse responses af salgspris på salgspris.

Kilde: Plot fra R

	lnprice
1. måned	0.03691045
2. måned	0.04258931
3. måned	0.04320193
4. måned	0.04337069
5. måned	0.04335960
6. måned	0.04337298
7. måned	0.04336849
8. måned	0.04337058
9. måned	0.04336971
10. måned	0.04337008
11. måned	0.04336993
12. måned	0.04336999

Tabel A.9: Kumulative impulse response af lnprice på lnprice.

Kilde: Output fra R

A.4 Opsætningen af ARIMA-model

I det følgende gennemgås opsætningen af ARIMA-modellen, hvor det bagvedliggende formål er at danne et in-sample forecast af samarbejdsvirksomhedens omsætningen ved brug af omsætningen selv, hvilket betegnes som et univariate setup. Indledningsvist defineres den bedste model ved brug af *auto.arima* algoritmen i R, hvor der tages udgangspunkt i informationskriteriet, AIC. Herudover kontrolleres ARIMA-modellen for stationaritet, seriel korrelation, ARCH og residualernes fordeling, hvilket er essentielt for modellens brugbarhed.

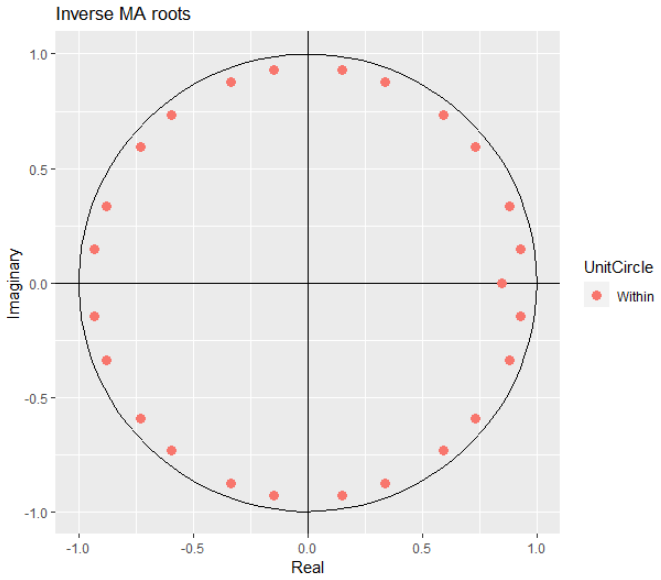
I Tabel A.10 rapporteres den bedste model, hvilket er baseret på modellen med den laveste AIC-værdi. Modellen kan karakteriseres som værende en ren MA-proces af første integrationsordre med to sæsonkomponenter til MA-leddet.

	auto.arima:
Bedste model:	ARIMA(0,1,1)(0,0,2)[12] with drift
AIC:	4025.15

Tabel A.10: auto.arima

Kilde: Output fra R

Det næste trin er at undersøge modellen for stationaritet, hvilket udføres ved at betragte et *autoplot* herunder. I Figur A.6 kan det ses, at alle de inverse MA-rødder ligger indenfor enhedscirklen, og stationaritetskriteriet er således opfyldt.

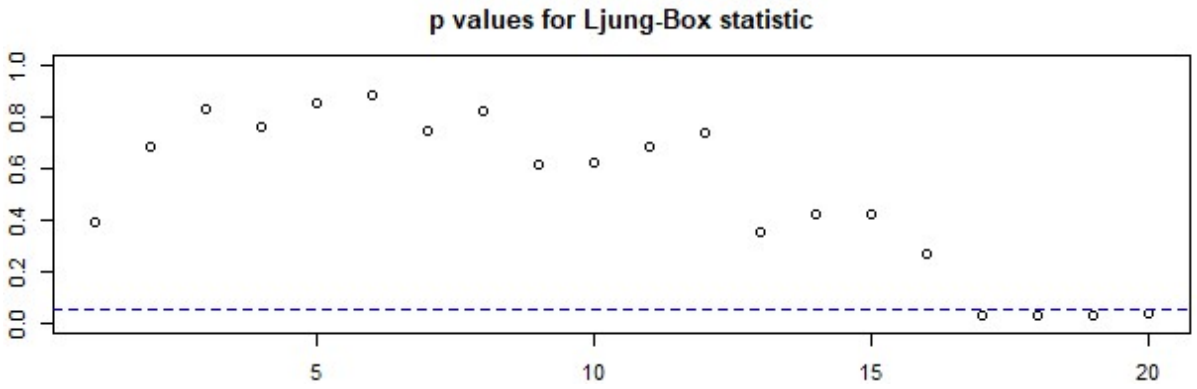


Figur A.6: Undersøgelse af stationaritet (1).

Kilde: Plot fra R

Herefter kontrolleres modellen for seriel korrelation, hvilket indebærer en undersøgelse af, om fejlleddet i denne periode er korreleret med dets tidligere perioder. Der anvendes en Ljung-Box test,

og der undersøges op til 20 lags målt i måneder. Nulhypotesen, H_0, udtrykker, at der er ingen seriel korrelation tilstede i modellen, hvilket opstilles imod alternativhypotesen, H_1, som påpeger, at seriel korrelation derimod er tilstede. I Figur A.7 kan det ses, at nulhypotesen, H_0, kan forkastes til fordel for alternativhypotesen ved det 17, 18, 19 og 20. lag, da p-værdierne ligger under et 5% signifikansniveau. Der eksisterer således en statistisk signifikant korrelation mellem fejlleddet i denne periode og disse sidste perioder. Løsningen på dette er en manuel tilføjelse af flere lags og led til modellen, indtil at denne serielle korrelation er statistisk insignifikant.



Figur A.7: Undersøgelse af seriel korrelation (1).

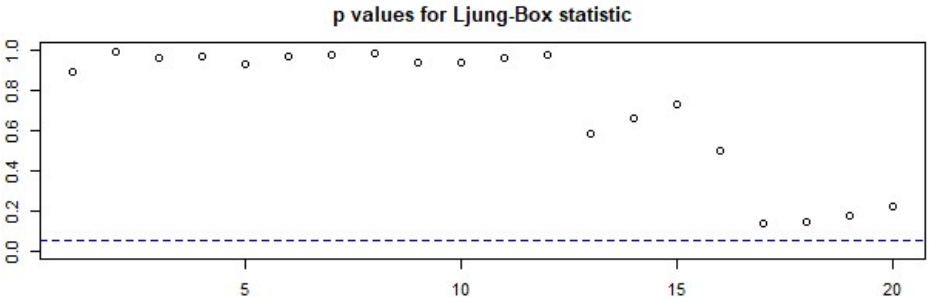
Kilde: Plot fra R

Modelopsætningen, som formår at eliminere den ovenfor viste seriel korrelation, er en SARIMA af første integrationsordre, hvor der inkluderes ét AR-led og to MA-led, som vist i Tabel A.11 I denne forbindelse tilføjes der ét sæsonkomponent til AR-leddet samt de to komponenter til MA-leddene fra tidligere. I Figur A.8 og A.9 vises det, at seriel korrelationen er statistisk insignifikant, og at den nye model opfylder stationaritetskriteriet.

	Manuel tilføjelse af ekstra led:
SARIMA:	ARIMA(1,1,2)(1,0,2)[12] with drift
AIC:	4015.98

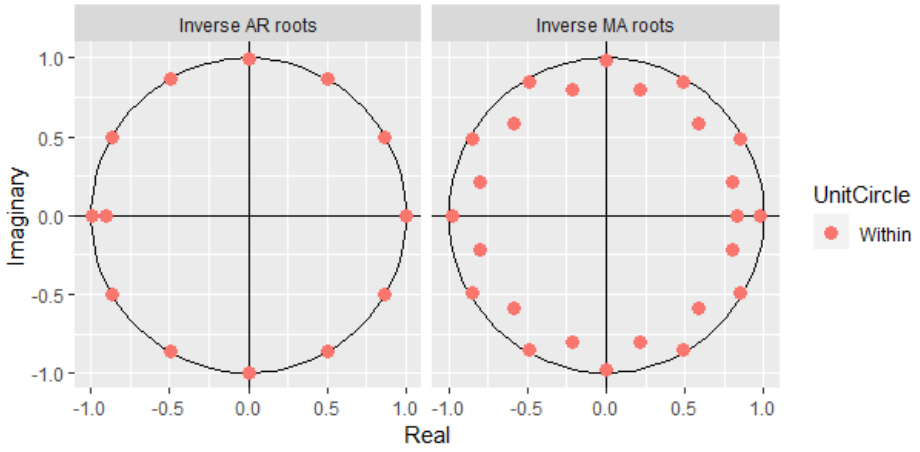
Tabel A.11: Modelopsætning: Løsning på seriel korrelation.

Kilde: Output fra R



Figur A.8: Undersøgelse af seriel korrelation (2).

Kilde: Plot fra R

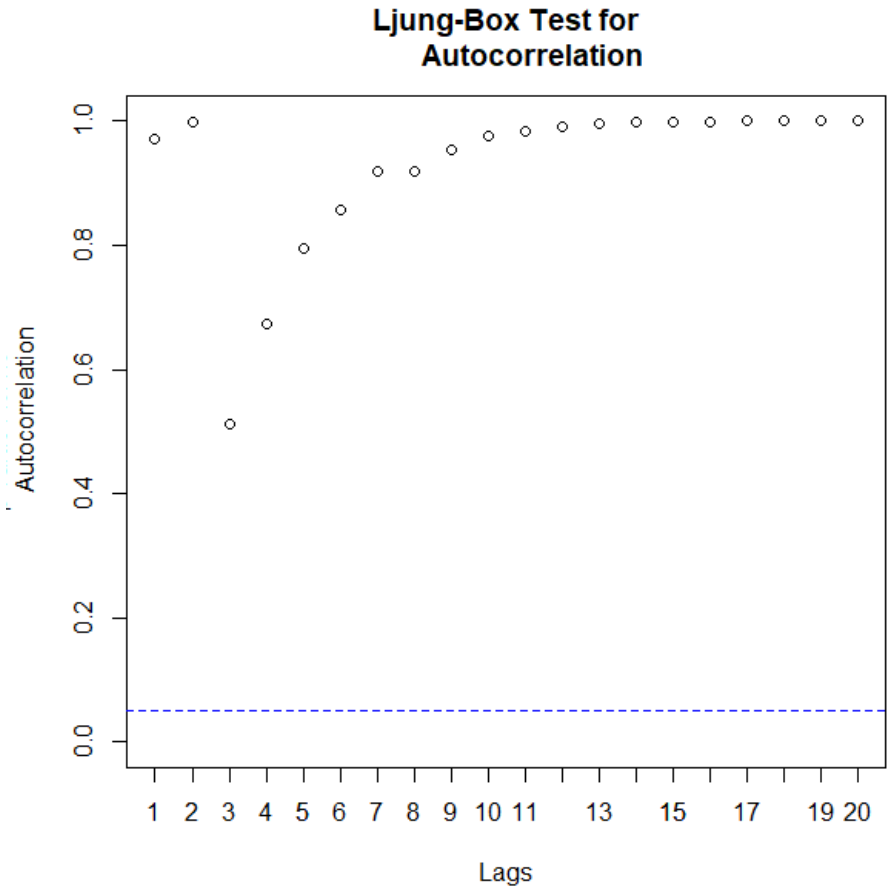


Figur A.9: Undersøgelse af stationaritet (2).

Kilde: Plot fra R

I det følgende undersgøes SARIMA-modellen for ARCH (Auto-regressive conditional heteroskedasticity), hvilket implicerer en undersøgelse af, hvorledes residualernes varians er invariant over tid dvs. korreleret eller ej over tid. Tilstedeværelsen af ARCH medfører en øget varians i modellen og derigennem inefficiente standard errors. Undersøgelsen af ARCH tager udgangspunkt i en Ljung-Box test, hvor nulhypotesen, H_0, som udtrykker, at der er ARCH, opstilles imod alternativhypotesen, H_1, som påpeger, at der ikke er ARCH tilstede i modellen. Ved at betragte Figur A.10 vises det, at nulhypotesen ikke kan forkastes til fordel for alternativhypotesen, da p-værdien ligger over et 5%

signifikansniveau for alle tidsperioder. Der eksisterer altså ikke en statistisk signifikant korrelation mellem fejlleddet i denne periode og dets tidligere perioder set i forhold til deres varians.



Figur A.10: Undersøgelse af ARCH.

Kilde: Plot fra R

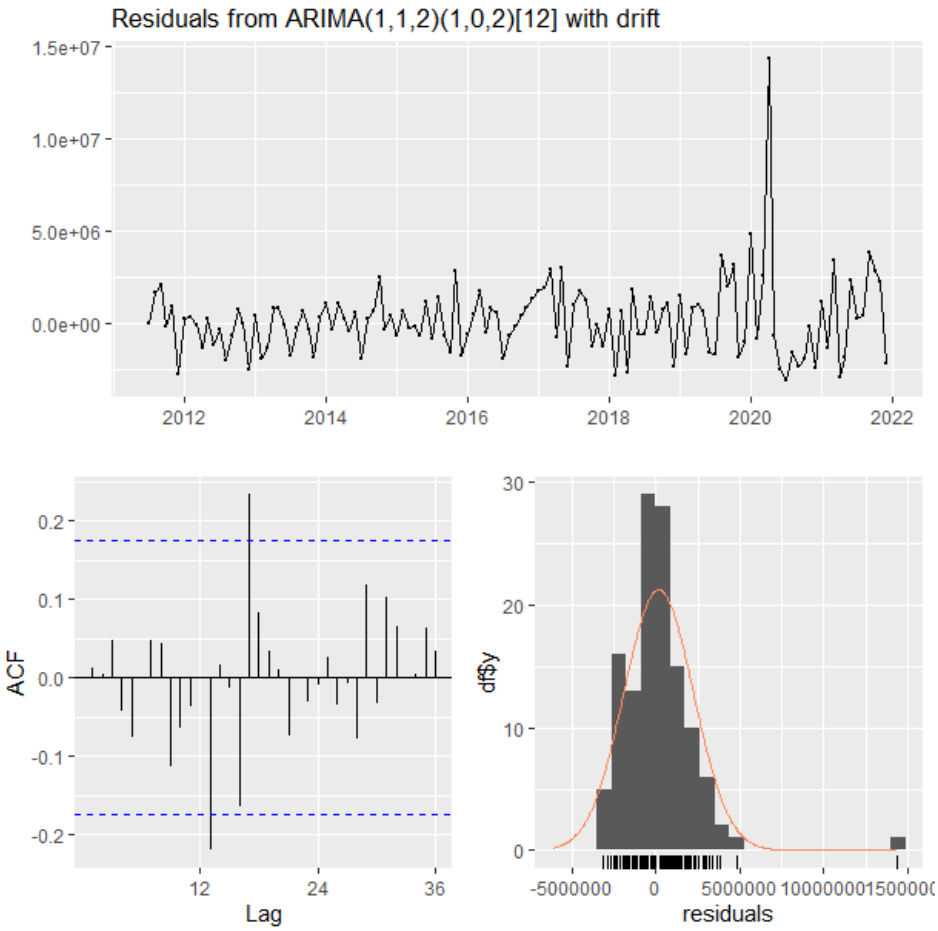
Afslutningsvist undersøges det, hvorvidt residualerne følger en normalfordeling, da dette er et krav for at t- og f-statistikker er troværdige. Undersøgelsen tager udgangspunkt i en Ljung-Box test, hvor nulhypotesen, H_0, som udtrykker, at residualerne følger en normalfordeling, opstilles imod alternativhypotesen, H_1, som påpeger, at dette ikke er tilfældet. Hvis Tabel A.12 betragtes, ses det, at p-værdien ligger over et 5% signifikansniveau, hvilket implicerer, at nulhypotesen, H_0, ikke

kan forkastes til fordel for alternativhypotesen, H_1, og der er således statistisk evidens for, at residualerne følger en normalfordeling.

	P-værdi:
Ljung-Box test	0.08276

Tabel A.12: Undersøgelse af normalitet i fejlleddet.
Kilde: Output fra R

I Figur A.11 betragtes SARMA-modellens residualer nærmere. Til trods for, at der er statistisk evidens for, at residualerne følger en normalfordeling, forekommer der en outlier ude til højre fra gennemsnittet. Dette skyldes et strukturelt brud i år 2020, herunder coronakrisen. I det følgende kontrolleres der for bruddet ved at tilføje to dummies som kun gør sig gældende fra år 2020-03 til 2020-05 og 2020-05 til 2020-07. I denne forbindelse er der forsøgt med forskellige dummies for at dække dette brud bedst muligt.



Figur A.11: SARMA-modellens residualer.

Kilde: Plot fra R

I Tabel A.13 defineres den endelige SARMA-model, hvor der nu er tilføjet to dummies for at kontrollere for coronakrisen i år 2020. Der opnås således den laveste AIC-værdi på 3977.35.

SARIMA:	ARIMA(1,1,2)(1,0,2)[12] with drift
AIC:	3977.35

Tabel A.13: SARMA-model med to dummies.

Kilde: Output fra R

A.4.1 Opsætning af VAR-model

I det følgende opsættes VAR-modellen, som anvender samarbejdsvirksomhedens salg, salgspris og omsætning til forecasting af omsætningen selv, hvilket betegnes som et multivariate setup. Indledningsvist vælges antallet af lags i modellen ud fra *VARselect* algoritmen i R, som anvender AIC. Afslutningsvist testes modellen for stationaritet og seriel korrelation.

Ved brug af *VARselect* i R minimeres AIC-informationskriteriet, når der tilføjes fire lags af hver endogen variabel i systemet. Før VAR-modellen kan bruges til at danne et forecast, er det nødvendigt at kontrollere for stationaritet og seriel korrelation. I R anvendes koden, *roots*, til at undersøge modellens rødder, hvor det kræves, at disse ligger indenfor enhedscirklen.

Ud fra Tabel A.14 ses det, at stationaritetkriteriet for VAR-modellen opfyldes, da alle egenværdierne er mindre end 1 i absolut værdi, hvilket implicerer, at rødderne ligger indenfor enhedscirklen.

	Rodens værdi:
Rod 1	0.87526187
Rod 2	0.87526187
Rod 3	0.82058890
Rod 4	0.82058890
Rod 5	0.75042104
Rod 6	0.75042104
Rod 7	0.72065960
Rod 8	0.72065960
Rod 9	0.53074157
Rod 10	0.53074157
Rod 11	0.42693630
Rod 12	0.08603535

Tabel A.14: Undersøgelse af rødder for VAR-modellen.

Kilde: Output fra R

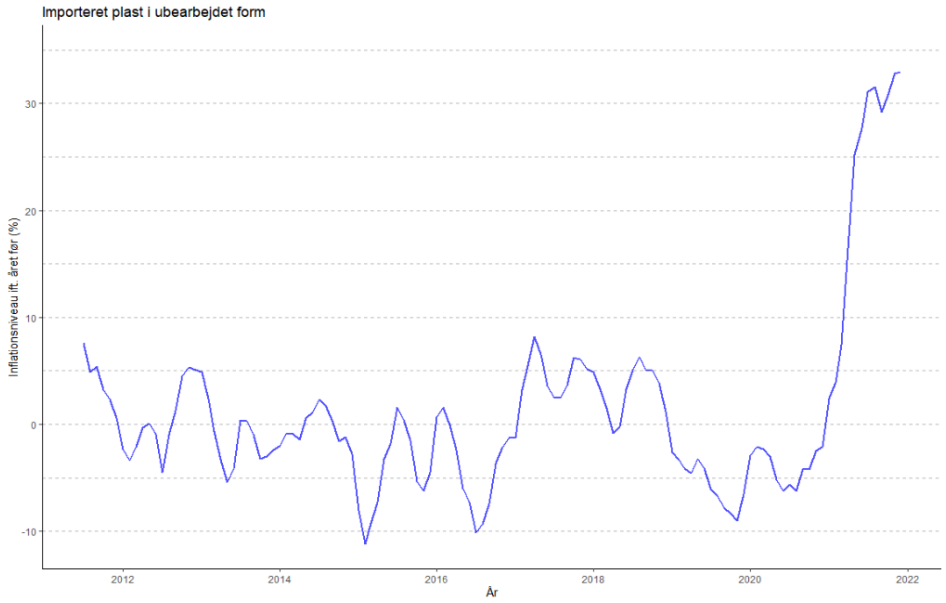
Afslutningsvist kontrolleres der for seriel korrelation i VAR-modellen ved brug af Portmantaeu-testen. I Tabel A.15 ses det, at p-værdien ligger langt over et 5% signifikansniveau, hvilket betyder, at nulhypotesen, H_0, ikke kan forkastes til fordel for alternativhypotesen, H_1. Der er således statistisk evidens for, at der ikke er seriel korrelation tilstede i VAR-modellen. På nuværende tidspunkt kan VAR-modellen benyttes til forecasting af samarbejdsvirksomhedens omsætning.

	P-værdi:
Portmanteau Test	0.9379

Tabel A.15: Test af seriel korrelation for VAR-modellen.

Kilde: Output fra R

A.5 Inflation i Importeret Plast



Figur A.12: Inflation i 'Importeret plast i ubearbejdet form'.

Kilde: Danmarks Statistik, 2015