

Petter Ihlebæk

# Sammenhengen mellom arbeidsproduktivitet og smitteverntiltak i Norge under koronapandemien

Masteroppgave i Samfunnsøkonomi

Veileder: Leif Anders Thorsrud

Juni 2023



Petter Ihlebæk

# **Sammenhengen mellom arbeidsproduktivitet og smitteverntiltak i Norge under koronapandemien**

Masteroppgave i Samfunnsøkonomi  
Veileder: Leif Anders Thorsrud  
Juni 2023

Norges teknisk-naturvitenskapelige universitet  
Fakultet for økonomi  
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden



## Sammendrag

Denne oppgaven undersøker sammenhengen mellom arbeidsproduktiviteten til en rekke sektorer i Norge med grad av smitteverntiltak i landet, eller restriksjoner som følge av koronapandemien. Oppgaven består av to deler. Den første delen predikerer arbeidsproduktiviteten i sektorene under pandemien, basert på data fra årene før pandemien. Den andre delen tar utgangspunkt i differansen mellom den faktiske arbeidsproduktiviteten til sektorene under pandemien og arbeidsproduktiviteten som blir predikert i den første delen. Det er rimelig å anta at denne differansen kan ha sammenheng med restriksjoner innført av myndighetene, og dette undersøkes med paneldatametoder. Forutsetningen er at prediksjonene er troverdige. Differansene skaper et paneldatasett.

På paneldataen benyttes LASSO for å selektere variabler, med hensikt om at en begrenset modell med sterke forklaringsvariabler skal velges fra et større sett. Den avhengige variabelen i paneldatasettet er differansen mellom den faktiske arbeidsproduktiviteten og den predikerte arbeidsproduktiviteten gjennom åtte kvartaler i 2020 og 2021. Hypotesen er at paneldatametodene vil finne et signifikant forhold mellom denne differansen og grad av smitteverntiltak/restriksjoner i samfunnet.

Fire modeller analyseres, to fixed effects-modeller og to dummy-regresjoner. Kun én av modellene gir en signifikant sammenheng, og det er svakt grunnlag for å bekrefte hypotesen. Den signifikante sammenhengen er dessuten av liten størrelse, og det diskuteres om den i det hele tatt er av økonomisk betydning. Allikevel er det mange faktorer som ikke tas hensyn til og det er trolig utelatte variabler. Videre er det spørsmål knyttet til problemstillingens vinkling og alternativkostnader.

## Abstract

This thesis investigates the relationship between labour productivity in a number of sectors in Norway and the corona policy stringency in the country. The thesis consists of two parts. In the first part, labour productivity in each sector is predicted through the pandemic years, based on data prior to the pandemic. The second part is based on the difference between the actual labour productivity in the sectors during the pandemic and the predicted labour productivity from the first part. It is reasonable to assume that this difference is associated with the government's policy stringency. A prerequisite is that the predictions from part one are credible. The differences creates a panel dataset.

When working with the panel data, LASSO is applied to select variables with the intention to select a parsimonious model with strong predictors from a larger set of variables. The dependent variable is the difference between actual labour productivity and predicted labour productivity for eight quarters through 2020 and 2021. The hypothesis is that the panel data methods will find a significant relationship between this difference and the corona policy stringency.

Four models are analyzed, two fixed effects models and two dummy-regressions. Only one of the models results in a significant relationship, and the support for confirming the hypothesis is weak. The significant relationship is rather small, and it is discussed whether or not it is economically significant. There are several factors which are not taken into account and probably omitted variables. Furthermore, there are questions regarding the thesis' approach and alternative costs.

## Forord

Denne oppgaven er skrevet som en del av en mastergrad i samfunnsøkonomi. Jeg vil takke min veileder Leif Anders Thorsrud for faglige innspill og diskusjoner. Det samme kan sies om studiekamerater. Deres kritikk og idéer har vært avgjørende for retningen oppgaven har tatt.

Petter Ihlebæk

Trondheim 31.05.2023

# Innhold

Sammendrag

Abstract

Forord

1. Innledning .....	1
1.1 Oppgavens oppbygning .....	2
1.2 Arbeidsproduktiviteten .....	3
1.3 Motivasjon og litteratur .....	4
2. Datasett og deskriptiv statistikk .....	6
2.1 ARIMA-data – Arbeidsproduktiviteten: 1995Q1 – 2019Q4 .....	6
2.2 Paneldata: 2020Q1 – 2021Q4 .....	7
2.2.1 Stringency index .....	9
2.2.2 Andre variabler på pandemirestriksjoner .....	10
2.2.3 De resterende kontrollvariablene .....	11
2.3 Sektorenes utvikling .....	12
2.3.1 Om valg av sektorer .....	15
3. ARIMA-modeller .....	16
3.1 Del 1: Autoregressive integrated moving average (ARIMA) - Teori .....	17
3.2 Estimering og prediksjon med ARIMA-modellene .....	20
4. Del 2: Paneldata – metode og analyse .....	23
4.1 Least absolute shrinkage operator (LASSO) .....	24
4.2 Variabelseleksjon med LASSO .....	25
4.4 Fixed effects: Teori og forutsetninger .....	26
4.4.1 Simultanitet .....	28
4.4.2 Multikollinearitet .....	30
5. Resultater .....	31
5.1 Fixed effects .....	31
5.2 Regresjon med dummy-variabler .....	34
5.3 Justering av modellen – flere observasjoner .....	36
5.4 Oppsummering av resultater .....	39
6. Diskusjon .....	41
7. Konklusjon .....	45
Referanseliste .....	47



## 1. Innledning

Under koronapandemien innførte norske myndigheter restriksjoner for å hindre smitte av viruset, som førte til at virksomheter måtte tilpasse seg. Dette endret måten arbeidstakerne utførte oppgavene sine på i mange tilfeller, og hvordan tilpasningen skjedde hang sammen med hva slags type virksomhet det var snakk om.

Hjemmekontor ble en nødvendighet under pandemien og var mulig for mange arbeidstakere i kontorarbeid. Derimot var det i andre sektorer nødvendig å ha personale til stede på arbeidsplassen, som en i en restaurant, butikk, eller byggeplass. Arbeidsproduktiviteten i de ulike sektorene kan derfor tenkes å være påvirket av restriksjonene. Således kan pandemien ha gitt kortsiktige og/eller langsiktige endringer i arbeidsproduktiviteten. Om den forventede arbeidsproduktiviteten under pandemien skilte seg fra den faktiske arbeidsproduktiviteten, er det nærliggende å tenke at pandemien var årsak til dette avviket. Denne oppgaven består av to deler relatert til dette:

1. Skape verdier for forventet arbeidsproduktivitet under pandemien.
2. Undersøke om avviket mellom forventede verdier og faktiske verdier kan forklares av myndighetenes restriksjoner/nedstengingstiltak.

I den første delen brukes data fra før pandemien til å predikere arbeidsproduktiviteten under pandemien. Dette gjøres for 12 ulike sektorer i Norge. De predikerte verdiene kan ses på som et anslag for normal utvikling i arbeidsproduktivitet. Verdiene sammenlignes så med den faktiske utviklingen under pandemien, og differansen mellom de predikerte verdiene og de faktiske verdiene vil representere avvik fra normalsituasjon. For å predikere verdier benyttes ARIMA (Auto regressive moving average). Hensikten med å gjøre disse prediksjonene er at det reduserer problemet med utelatte variabler i neste del.

I den andre delen forsøkes det å identifisere hvilke faktorer som påvirker avvikene i arbeidsproduktivitet og estimerer så effekten av disse. Avvikene i arbeidsproduktivitet er differansen mellom ARIMA-modellenes prediksjoner og de faktiske verdiene 2020Q1 – 2021Q4. Altså består denne delen av et panel med 12 sektorer og 8 perioder. Hensikten er å se hvilke faktorer som ga et unormalt nivå på arbeidsproduktivitet, med en idé om at

koronarestriksjoner er en del av forklaringen. Den interessante variabelen er en indeks som beskriver hvor strenge restriksjonene i Norge var gjennom pandemien. Hypotesen er at denne indeksen har en signifikant sammenheng med avvik i arbeidsproduktivitet. I denne delen benyttes LASSO (Least absolute shrinkage and selection operator) for variabelseleksjon, deretter brukes Fixed effects for estimering.

## 1.1 Oppgavens oppbygning

Siden oppgaven består av to deler vies litt plass til å forklare oppbygningen videre.

I begynnelsen defineres arbeidsproduktivitet, som er variabelen som analyseres gjennom hele oppgaven. Deretter følger motivasjonen til å skrive om dette og om tidligere publisert litteratur med lignende problemstillinger.

Etter dette kommer kapittel 2. Her presenteres all deskriptiv statistikk, selv om noe av den deskriptive statistikken i del 2 er beregnet basert på resultatene av analysen i del 1. I del 2 er avvik i arbeidsproduktivitet den avhengige variabelen og denne er beregnet ved hjelp av ARIMA-prediksjonene i del 1. Det er mest ryddig å samle den deskriptive statistikken før presentasjon av modeller og analyser. Sammen med beskrivelsen av datasettet følger generelle beskrivelser av sektorene som er inkludert og hvorfor, samt litt om deres utvikling. Formålet med kapittel 2 er å gi en god oversikt over hva slags data som er benyttet i både kapittel 3 og kapittel 4.

I kapittel 3 kommer en forklaring av hva ARIMA er og nødvendige forutsetninger, og deretter rapporteres resultatene fra modellene. ARIMA-modellene er kun ment som et ledd for å danne et paneldatasett, og er ikke en del av hovedresultatet, selv om de selvsagt vil være en del av diskusjonsdelen. Modellene er en forutsetning for den videre analysen. I dette kapittelet argumenteres det også for hvorfor denne delen kan styrke analysen i neste kapittel. Kapittel 3 er altså det som i innledningen ble omtalt som del 1.

Resultatene fra kapittel 3 benyttes til å danne et panel, som analyseres i kapittel 4. Her benyttes fixed effects og forutsetninger og egnethet av denne panelmetoden vurderes. Resultatene rapporteres i kapittel 5. I kapittel 6 er det viet god plass til diskusjon og

refleksjon over problemstillingen og tolkning av resultater. Til slutt kommer en konklusjon og deretter vedlegg og referanser.

## 1.2 Arbeidsproduktivitet

Arbeidsproduktivitet er den eneste variabelen som benyttes i den første delen, som strekker seg fra 1995 til 2019.

Arbeidsproduktivitet (AP) defineres som:

$$\text{Arbeidsproduktivitet}_i = \frac{\text{Bruttoprodukt}_i}{\text{Utførte timeverk}_i}, i = \text{sektor} \quad (1)$$

(Statistisk sentralbyrå 2017)

Statistisk sentralbyrå definerer bruttoprodukt som:

*Verdiskaping og opptjent bruttoinntekt fra innenlandsk produksjonsaktivitet i en næring eller sektor (eller totalt for alle næringer/sektorer), avledet og definert som produksjon minus produktinnsats. Bruttoprodukt publiseres i basisverdi, dvs. at produktsubsidier er inkludert, men ikke merverdiavgift eller andre produktskatter (se basisverdi). I offentlig forvaltning og annen ikke-markedsrettet virksomhet bestemmes bruttoprodukt som sum lønnskostnader, netto produksjonsskatter og kapitalslit.*

(Statistisk sentralbyrå 2023)

Utførte timeverk defineres som:

*Antall timeverk utført av alle sysselsatte personer (lønnstakere og selvstendige) i løpet av et år i innenlandsk produksjonsaktivitet. Timeverkene gjelder arbeid utført innenfor effektiv normalarbeidstid, med tillegg for utført overtid og fradrag for fravær pga. sykdom, permisjon, ferie og eventuelle arbeidskonflikter. Antall utførte timeverk er også påvirket av kalendermessige forhold (bevegelige helligdager og skuddår). Antall arbeidsdager vil kunne variere med inntil 3 dager fra ett år til det neste.*

(Statistisk sentralbyrå 2023)

Arbeidsproduktivitet vil ikke fange opp endringer i andre innsatsfaktorer enn arbeid, her timeverk, som for eksempel realkapital. Derfor vil AP heller ikke fange opp substitusjonseffekter i produksjonen. Fordelen ved å benytte seg av AP-begrepet i denne konteksten er at det er vanskelig å hensynta andre innsatsfaktorer når man jobber med sektor-data.

Et mer riktig bilde av produktivitetens utviklingen kan gis av total faktorproduktivitet (TFP). Forholdet mellom TFP og AP er beskrevet i en NOU fra 2015 om produktivitetsvekst (1 2015). Vekst i TFP er forskjellen mellom produksjonsveksten og veksten i alle innsatsfaktorer som bedriften betaler for. Altså er TFP lik AP pluss bidragene fra andre innsatsfaktorer. AP er dog mye enklere å regne ut, spesielt med tanke på produktivitet på sektornivå. Dessuten argumenterer overnevnte NOU med at AP gjerne kan brukes i analyser, siden AP og TFP har svært høy korrelasjon. Derfor er AP et begrep som kan bidra til økonomisk innsikt gjennom en analyse som denne oppgaven byr på. Generelt er produktivitet et tema det forskes mye på siden det er et viktig mål på vekstmuligheter i en økonomi.

### 1.3 Motivasjon og litteratur

Å undersøke hvordan arbeidsproduktiviteten kan ha vært påvirket av koronapandemien er en relevant problemstilling fordi det kan gi implikasjoner for hvordan lignende situasjoner kan håndteres i fremtiden. Myndighetene bør unngå å innføre restriksjoner som gir et negativt utslag i arbeidsproduktiviteten, spesielt om effekten er langvarig eller om det er kostbart å innhente seg fra. På en annen side kan det vise seg at arbeidsproduktiviteten henter seg raskt inn og er lite påvirket av restriksjoner. Hvor strenge og inngripende restriksjonene til enhver tid skulle være under pandemien var, og er, omdiskutert. Fordelene restriksjoner ga ved redusert smitte må veies opp mot konsekvensene det fikk for samfunnet. Denne oppgaven undersøker en av de mulige konsekvensene.

Motivasjonen for oppgaven går også på det rent metodiske. Spesielt er det interessant å bruke LASSO for å selektere variabler. Dette er en statistisk teknikk som brukes i mange

fagfelt. I økonomien er det for eksempel brukt til å selektere variabler som forklarer oljepris (Zhang, Ma et al. 2019). Tidligere masteroppgaver har også benyttet LASSO som et verktøy for variabelseleksjon, blant annet til å forklare kapitalstruktur i bedrifter (Karlsen and Mathisen 2021) og til å analysere effekten av petroleumsinvesteringer på kommuner (Vågbø 2019). Det er derfor, for undertegnede, en ny måte å spesifisere en økonometrisk modell på. Et alternativ som er mye brukt under studiene er å selektere variabler med bakgrunn i økonomisk teori og intuisjon. Selvsagt vil datasettet det kan selekteres fra inneholde variabler med bakgrunn i intuisjon og teori, men det er allikevel spennende å se om LASSO kan selektere gode forklaringsvariabler. Det er absolutt relevant i økonomifaget, som ikke er noen vitenskap, men et fagfelt der man stadig bør utfordre teorier og etablerte sannheter.

Naturligvis vil litteraturen om lignende problemstillinger være begrenset siden det er kort tid siden pandemien. Norges bank har skrevet om produktiviteten i de ulike sektorene og viser til at produktivitetsveksten falt i begynnelsen av pandemien, for så å hente seg inn raskt, og at fallet var størst i sektorer med allerede lavt produktivetsnivå (Brubakk and Hagelund 2022). Internasjonalt finnes flere studier som undersøker sammenhengen mellom bruken av hjemmekontor og produktivitet under pandemien. En studie gjort på industrier i USA, Storbritannia og Frankrike finner at industrier med mye hjemmekontor ikke hadde ulik produktivetsutvikling enn industrier med lite hjemmekontor i 2020 (Vries, Erumban et al. 2021). Det er et mulig argument for at de eventuelle positive effektene av nedstenging på produktivitet, som man kanskje skulle tro hjemmekontor var, ikke er til stede. Tvert om kan smitteverntiltak som fører til hjemmekontor ha en negativ effekt på produktiviteten, fordi hjemmekontor kan være et hinder for samarbeid (Yang, Jaffe et al. 2020).

Det finnes en del litteratur internasjonalt som undersøker sammenhengen mellom økonomiske størrrelser og smitteverntiltak. For eksempel en rapport fra National bureau of economic research som sammenlignet hvordan nedstengninger påvirket utførte timeverk i en rekke stater tidlig i pandemien (Bartik, Bertrand et al. 2020). De skriver blant annet at tiltak som hindret mobilitet førte til mange permisjoner og derfor færre timeverk. I tillegg trekker de frem noe som også vil være viktig i diskusjonen av min oppgave, nemlig at effekten av å lette på et tiltak ikke nødvendigvis har like stor motsatt effekt som å innføre tiltaket. Argumentet er at ved å innføre tiltak vil noen bedrifter gå konkurs og disse vil ikke returnere når tiltakene lettes på. Med andre ord er det trolig en viss asymmetri mellom

positive og negative endringer i smitteverntiltak/restriksjoner. En annen måte å si det på er at innføringen av restriksjoner kan gi permanente endringer.

Noe av litteraturen over er også referert til av koronakommisjonen, som er det mest omfattende arbeidet på samfunnsøkonomiske konsekvenser av smitteverntiltakene i Norge (Holden, Brasch et al. 2022). Ekspertgruppen bak rapporten ser på blant annet helsetap og velferdstap som følge av restriksjoner, men de ser lite på produktivitet. Men, de skriver at hjemmekontor gir redusert transporttid og kan derfor gi høyere produktivitet. Om verdiskaping skriver de at den falt brått da den første smittebølgen traff våren 2020, før den hentet seg inn ganske kjapt. Omikron-varianten førte til en ny smittebølge og verdiskapingen falt på nytt, men ikke like mye som ved den første smittebølgen. Kanskje kan det forklares av at landet var bedre forberedt når den andre smittebølgen kom, samt at deler av befolkningen hadde oppnådd immunitet gjennom enten sykdom eller vaksinerings. Kommisjonen sier også noe om arbeidstimer, som er med i beregning av arbeidsproduktivitet. De observerte at antall arbeidstimer falt da restriksjoner ble innført, og at det tok tid før timene hentet seg opp igjen da restriksjonene ble lettet på.

Med dette som bakteppe vil nå det selvstendige arbeidet begynne. I kapittel 2 gis innsikt i dataene som benyttes.

## 2. Datasett og deskriptiv statistikk

I de følgende avsnittene gis en oversikt over datasettene som brukes i de ulike modellene. ARIMA-modellene er bygget på data fra og med 1995Q1 til og med 2019Q4. Dette er del 1. Påfølgende panel inneholder kun data fra og med 2020Q1 til og med 2021Q4. Dette er del 2.

### 2.1 ARIMA-data – Arbeidsproduktivitet: 1995Q1 – 2019Q4

Det er totalt 12 ARIMA-modeller, én for hver sektor. Det er kun én variabel i hver modell, nemlig arbeidsproduktiviteten (AP) i den gitte sektoren. AP er beregnet for hver sektor

basert på sektorens bruttoprodukt (Statistisk sentralbyrå 2023) og utførte timeverk (Statistisk sentralbyrå 2023), kvartalsvis fra 1995 til 2019. Det vil si at det er 100 observasjoner av AP for hver sektor. Hensikten med å stoppe i 2019 er å predikere AP under pandemien basert på data fra før pandemien. Under følger en deskriptiv tabell.

Sektor	Observasjoner	Gjennomsnitt	Standardavvik	min	max
Jord_og_skogbruk	100	189,26	58,57	105,91	335,38
Fiske_Akvakultur	100	968,11	433,99	365,23	1755,43
Utvinning_Petroleum	100	6390,57	2263,90	3233,94	10050,77
Industri	100	584,32	90,15	432,86	712,08
Elektrisitets_Vann_Gass	100	1763,93	338,02	909,10	2250,51
Bygg_Anlegg	100	607,24	28,21	567,46	690,88
Overnatting_Servering	100	378,41	17,64	332,38	428,98
Finans_Forsikring	100	1543,33	478,27	715,98	2356,78
Offentlig_adm_Forsvar	100	603,28	65,95	492,55	721,31
Undervisning	100	578,25	18,36	515,02	637,73
Helse_Omsorg	100	552,13	19,06	520,84	604,64
Kultur_Underholdning	100	560,72	54,22	478,13	658,79

Tabell 1: Tall i antall kroner bruttoprodukt per timeverk

Tabellen viser at det er valgt ut sektorer med store forskjeller i arbeidsproduktiviteten. Jord- og skogbruk viser seg å ha den laveste gjennomsnittlige arbeidsproduktiviteten og petroleumsutvinning har den klart høyeste. Lenger ned, i sektorenes utvikling, undersøkes noen sektorer for å vise at en potensiell koronaeffekt ikke er like åpenbar for alle.

Siden arbeidsproduktiviteten ikke sier noe om hvilke produksjonsfaktorer har bidratt, men kun hva produktet per timeverk er, vil man ikke kunne si hva det er i produksjonen som har gitt endring. Faktorer som kan bidra til bruttoproduktet er for eksempel innsats, humankapital, produksjonskapital, kapasitetsutnyttelse og teknologi. Altså tar ikke arbeidsproduktiviteten hensyn til substitusjonseffekter i faktorinnsatsen, som tidligere nevnt.

## 2.2 Paneldata: 2020Q1 – 2021Q4

Den avhengige variabelen i panelet vil hentes fra prediksjonene til ARIMA i kapittel 3, og defineres som differansen mellom ARIMA-prediksjonen og den faktiske arbeidsproduktiviteten. Dette gjøres for åtte perioder, 2020Q1 – 2021Q4. Da vil det for hver

sektor regnes ut 8 differanser og med 12 sektorer er det 96 observasjoner som utgangspunkt. La denne avhengige variabelen hete  $APav$ , for *Arbeidsproduktivet avvik*.

$$APav_{it} = AP_{it} - \widehat{AP}_{it} \quad (2)$$

*= Avvik i arbeidsproduktivet for sektor i i kvartal t*

$$AP_{it} = \text{Faktisk AP} \quad \widehat{AP}_{it} = \text{Predikert AP}$$

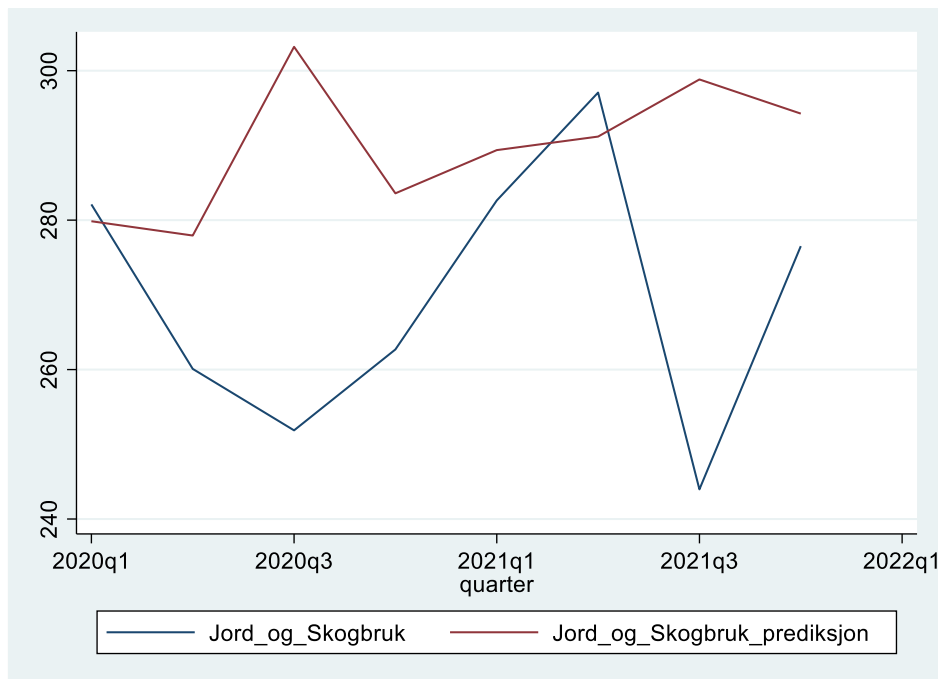
I analysen av panelet vil altså den avhengige variabelen være generert av ARIMA-modellen. Den kan oppsummeres slik:

Avvik i arbeidsproduktivet			
Gjennomsnitt	Standardavvik	Min	Max
21,99	96,14	-202,78	394,75

Tabell 2

Spredningen er ganske stor og det er synlig at ARIMA-prediksjonene fører til både positive og negative avvik. Dette er å forvente på tvers av sektorer, men også på tvers av tid, siden arbeidsproduktivet bør bli mer "normal" etter hvert som pandemien slipper taket. I grafen nedenfor ser man at AP i jord- og skogbruket var betydelig lavere i pandemiens startfase enn det ARIMA-modellen har predikert. Etter hvert henter AP'en seg inn igjen, før den på nytt faller tidlig i 2021, da restriksjonene på nytt ble strengere, se figur 2 lenger ned.





Figur 1

Datasettet inneholder videre 28 selvstendige mulige forklaringsvariabler som LASSO-operatoren selekterer fra. Blant disse er 14 relatert til nedstengingstiltak og de resterende er makroøkonomiske størrelser og variabler knyttet til arbeidsmarked. En fullstendig oversikt finnes i vedlegg 1. Variabelen av størst interesse er *Stringency index*, forkortet *s\_index\_start*.

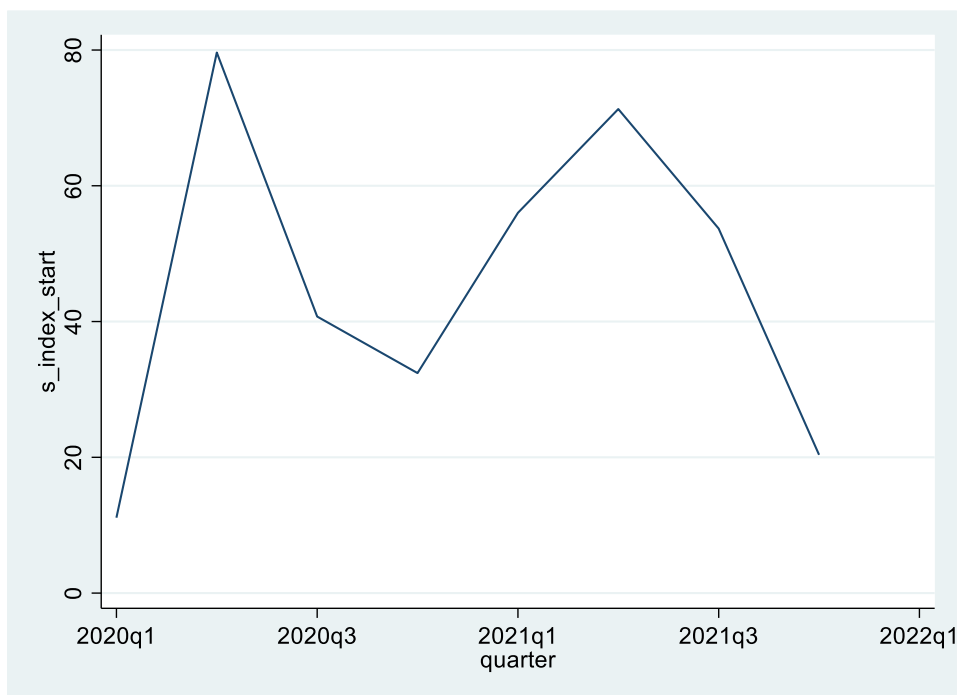
### 2.2.1 Stringency index

Hypotesen i oppgaven er at denne variabelen har en signifikant effekt på avvik i arbeidsproduktivitet, men hypotesen er åpen for både negativ og positiv effekt. Datainnsamlingen og utviklingen av indeksen er gjort av Blatvanik school of Government, en del av universitetet i Oxford (Hale, Angrist et al. 2021). Indeksen er regnet ut med grunnlag i de andre variablene som går på nedstengingstiltak. Derfor gir den en indikasjon på hvor strenge myndighetenes tiltak til enhver tid var, hver dag gjennom pandemien. Siden denne oppgaven bruker data som har observasjoner hvert kvartal, inkluderes indeksens verdi på kvartalets første dag i datasettet, og denne kalles *s\_index\_start*. Unntaket er 2020Q1, der det ikke er observert noen indeks før i februar. Hensikten med å inkludere den første verdien i kvartalet er at det mest sannsynlig tar tid før endringer i indeksen eventuelt gir signifikante endringer i avvik i arbeidsproduktivitet.

Indeks første dag i kvartalet			
Gjennomsnitt	Standardavvik	Min	Max
45,66	22,52	11,11	79,63

Tabell 3 s\_index\_start

Grafen under viser indeksen utvikling gjennom pandemien, der observasjonene er første dag i kvartalet. Den gjør det synlig at minimumspunktet er den første observasjonen, i februar 2020. Gjennom pandemiens første fase stiger den kraftig og når maksimumspunktet ved inngangen til andre kvartal 2020 før den begynner å falle. Inn i 2021 ble det en ny runde med svært restriktive tiltak, men gjennom året ble tiltakene lettet på. På vei inn i 2022 begynte pandemien og nærme seg slutten i Norge, med få restriktive tiltak fra myndighetene.



Figur 2

## 2.2.2 Andre variabler på pandemirestriksjoner

Selv om indeksen anses for å være den mest interessante variabelen med tanke på problemstillingen, blir det inkludert en rekke spesifikke variabler som er hentet fra samme

datainnsamling. Oversikten finnes i tabellen nedenfor og de aller fleste er ordinale eller dummyer.

Variable	Obs	Mean	Std. Dev.	Min	Max
<i>r rate</i>	96	1.212	.66	.77	2.93
<i>new cases</i>	96	49282.25	62207.839	4238	204828
<i>new deaths</i>	96	163.375	129.565	24	445
<i>icu adm</i>	96	181.25	124.3	16	448
<i>hosp adm</i>	96	898.625	539.397	150	1931
<i>positive rate</i>	84	.037	.039	.005	.13
<i>school closures</i>	96	1.375	.861	0	3
<i>public events closures</i>	96	1	.871	0	2
<i>stay at home</i>	96	.5	.503	0	1
<i>face coverings</i>	96	1.125	.932	0	2
<i>public campaigns</i>	96	1.75	.665	0	2
<i>int dom travel</i>	96	.5	.503	0	1
<i>testing policy</i>	96	1.75	.834	0	3
<i>income support</i>	96	1.75	.665	0	2

Tabell 4

Variabelen *r-rate* er reproduksjonsraten, *icu adm* og *hosp adm* er henholdsvis innleggelser på intensivavdeling og innleggelser på sykehus. Disse variablene kan tenkes å ha en selvstendig signifikant effekt på AP, uavhengig av indeksens verdi. For eksempel kan det hende *stay at home* og *r-rate* vil ha en sterkere sammenheng med AP enn *testing policy*, men da vil ikke nødvendigvis indeksen være signifikant, siden den er en funksjon av både de restriksjonene som eventuelt er viktige og de som eventuelt ikke er viktige for å forklare AP.

Samtlige variabler presentert i dette delkapittelet er hentet fra datasettene konstruert av Blatvanik school of government og tilgjengelig på ourworldindata.org. Det er to datasett, ett for alle variablene fra og med *r rate* og ned til og med *positive rate* i tabell 4 (Mathieu, Ritchie et al. 2020) og ett for de resterende variablene i tabellen (Mathieu, Ritchie et al. 2020). Førstnevnte inneholder også Stringency index.

### 2.2.3 De resterende kontrollvariablene

Inkludert i datasettet er en rekke felles og sektorspesifikke kontrollvariabler som kan tenkes å påvirke arbeidsproduktiviteten fra kvartal til kvartal. Tabellen under viser den deskriptive statistikken til disse, og en fullstendig oversikt finnes i vedlegg 1. En generell utfordring ved

mange av variablene er at det er åpenbar simultanitet og kollinearitet med AP. Mer om dette i kapittel 4.

Variable	Obs	Mean	Std. Dev.	Min	Max
ledige stillinger	88	3111.364	2484.871	300	9000
norske varer u energi	96	110.377	.955	108.567	112.133
energivarer	96	144.55	44.997	97.267	226.4
importerte varer	96	116.275	15.579	106.7	157.1
varekonsumindeks	96	148.533	7.813	132.733	156.933
privat konsum	96	385001.25	20929.581	348384	419805
offentlig konsum	96	231897	7362.188	217372	241483
etterspørsel fast	96	806787.13	29966.122	752105	857979
eksport alt	96	287004.88	13541.499	268150	308343
import alt	96	289472	18671.561	259382	317089
bruttoprodukt fastland	96	678258	22079.548	634277	711823
midlertidige ansatte	56	19.214	20.786	1	74
antall virksomheter	80	29258.456	24563.536	1623	71195

Tabell 5

Tabellen avslører at det er et ubalansert panel. De sektorspesifikke variablene er *ledige stillinger*, *midlertidige ansatte* og *antall virksomheter*. Grunnen til at Min og Max er runde tall for *ledige stillinger* er at det er rundet av til nærmeste 100. Videre er *norske varer u energi*, *energivarer* og *importerte varer* prisindekser, og *midlertidige ansatte* er oppgitt i antall tusen. For *antall virksomheter* var kun årlige observasjoner tilgjengelig, men en observasjon for hvert kvartal ble funnet med lineær interpolasjon. Alle variablene lagges også én gang, og disse laggene er inkludert i settet LASSO selekterer fra.

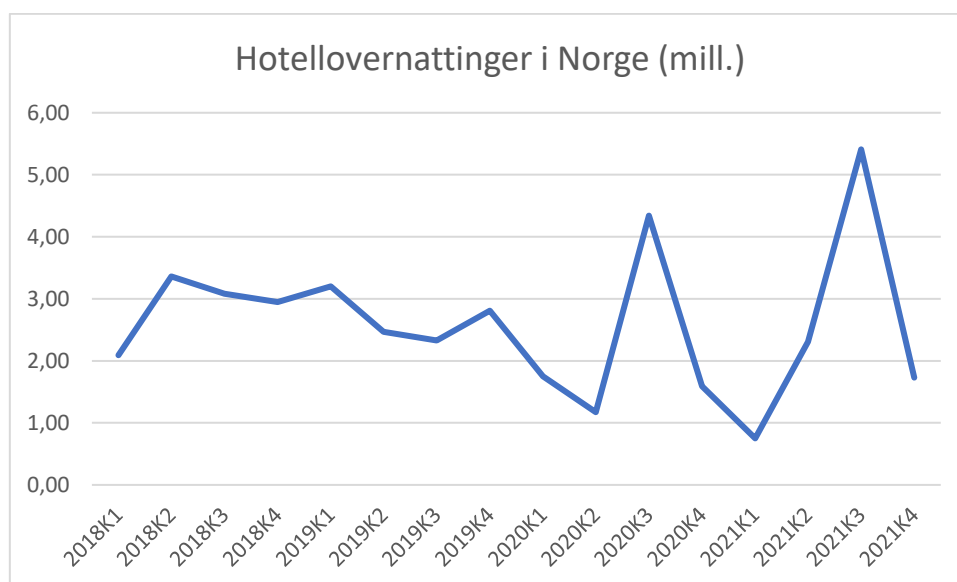
Helt til slutt har er det inkludert noen interaksjoner mellom indeksen og andre pandemivariabler. Interaksjonene er med *stay at home*, *new deaths* og *r-rate*. Alle disse er inkludert både i inneværende og i forrige periode (lag).

Samtlige variabler presentert i dette delkapittelet er hentet fra SSB's statistikkbank (Statistisk sentralbyrå 2023). I referanselisten ligger lenker til tabeller for hver variabel.

## 2.3 Sektorenes utvikling

Å visualisere sektorenes utvikling kan bidra til å synliggjøre den eventuelle “koronaeffekten” og dermed problemstillingens relevans. En komplett grafisk oversikt finnes i vedlegg 2. Der vises alle sektorenes AP fra 1995 og ut pandemien. Formålet med eksemplene som vises her er å illustrere at det er ulikheter mellom sektorene.

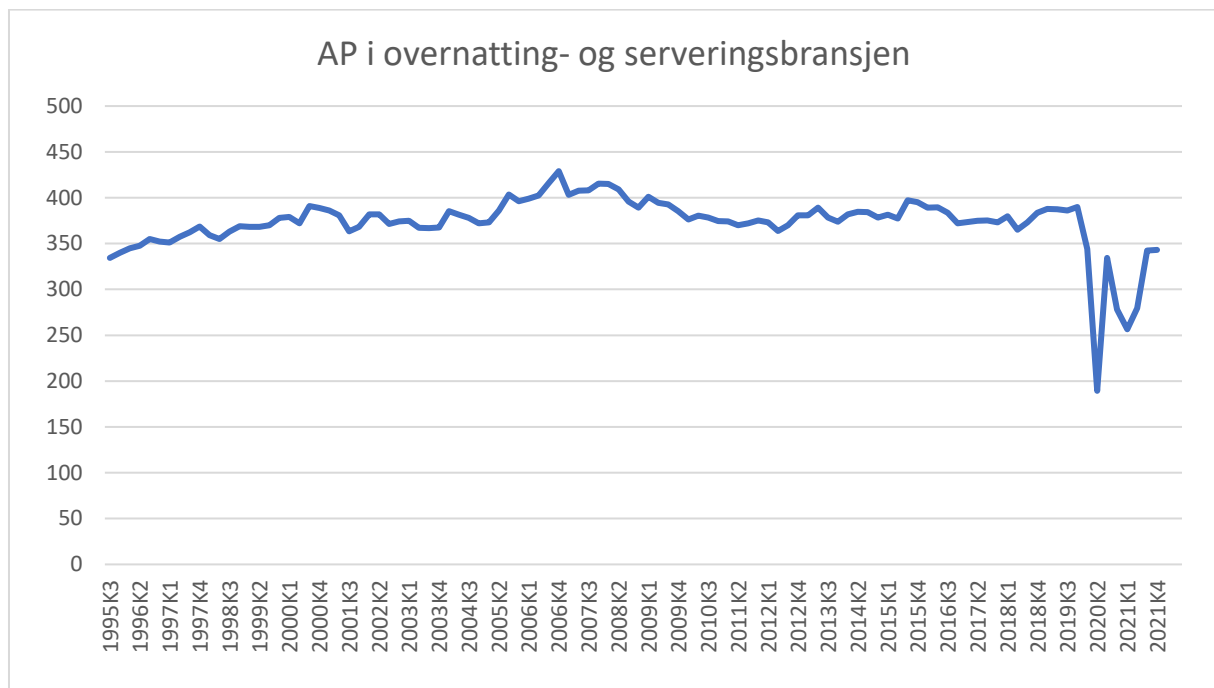
La oss først se på overnatting- og servering, en av de tolv sektorene. Fra tabell 1 ser man at denne er sektoren med nest lavest gjennomsnittlig AP fra 1995 – 2019. Kun jord- og skogbruk har lavere. Da pandemien kom, traff den overnatting og servering hardt. Det ble mindre reising og dermed redusert etterspørsel etter hotellovernattinger. Grafen under viser antall millioner hotellovernattinger i Norge fra 2018 og gjennom pandemien (Statistisk sentralbyrå 2023):



Figur 3

Det er synlig at nedgangen i hotellovernattinger begynner før pandemien. Den når et toppunkt i tredje kvartal 2020. Dette kan ha sammenheng med den lave indeksen dette kvartalet, se figur 2. En visuell sammenligning av figur 2 og figur 3 antyder en ganske sterk negativ korrelasjon mellom antall hotellovernattinger og indeks for restriksjoner. Det er ikke urimelig å anta en del av variasjonen i hotellovernattinger skyldes variasjonen i indeksen. Da vil også bruttoproduktet i overnattings- og serveringsbransjen være påvirket, og dermed AP.

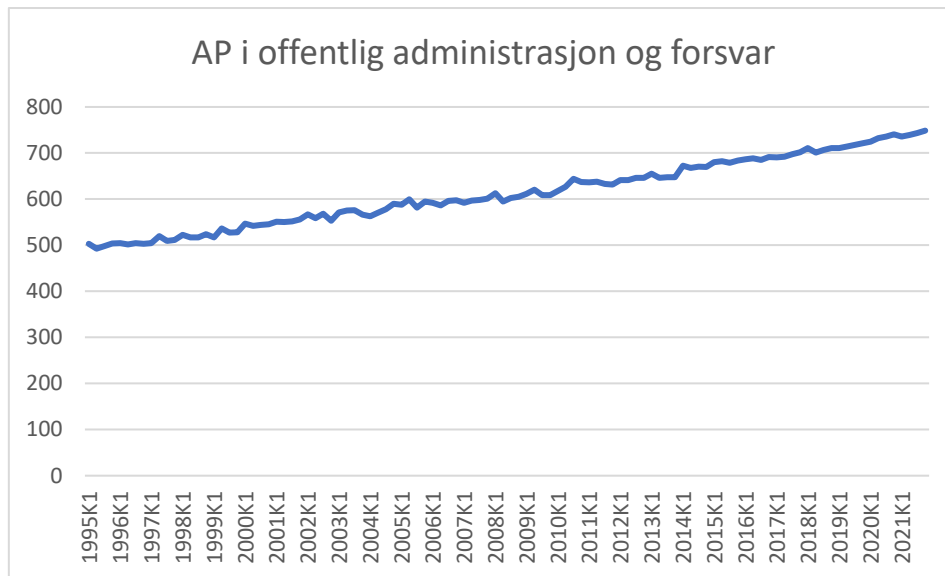
Under er fremstillingen av arbeidsproduktiviteten til sektoren, fra 1995 og gjennom pandemien.



Figur 4

Grafen viser at de store variasjonene ikke kommer før pandemiens inntog. Før 2020 er det lite variasjon, med en topp i 2006. Både antall hotellovernattinger og arbeidsproduktiviteten synes å korrelere ganske sterkt negativt med indeks for restriksjoner, men om det finnes grunnlag for å påstå at det foreligger kausalitet gjenstår å se.

Offentlig administrasjon og forsvar er derimot en sektor som ikke får noen større variasjon under pandemien. Arbeidsproduktiviteten i denne sektoren utvikler seg litt som forventet, om man kun ser på de historiske dataene. Dette kan være intuitivt ettersom offentlige organer står i en særstilling sammenlignet med konkurranseutsatte sektorer, og dessuten er mange av disse organene samfunnskritiske, som for eksempel politi og rettsvesen. Figur 5 nedenfor viser denne normale utviklingen. Offentlig administrasjon og forsvar har hatt en ganske jevn produktivitetsvekst siden midten av nittitallet.



Figur 5

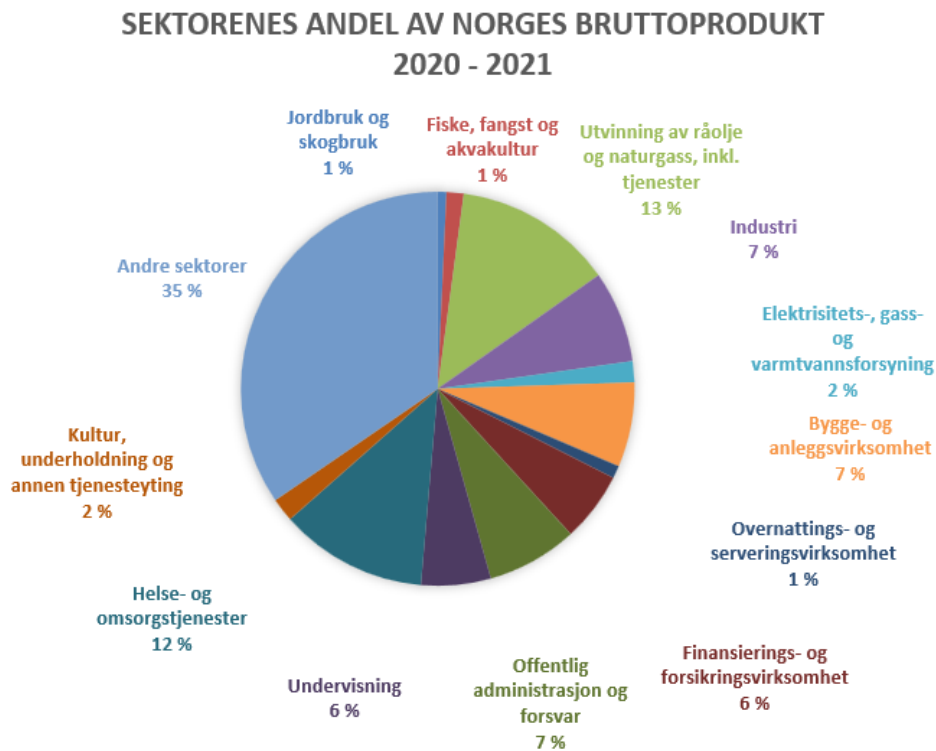
De overnevnte eksemplene på sektorenes utvikling viser at panelet består av sektorer som trolig i varierende grad ble påvirket av pandemien. Dette gir utfordringer når panelanalysen gjennomføres, siden det kanskje er naivt å tro man skal finne en signifikant sammenheng mellom avvik i arbeidsproduktivitet og indeks for restriksjoner når de forskjellige sektorene er påvirket ulik grad og retning. Dette utdypes i neste kapittel.

### 2.3.1 Om valg av sektorer

De tolv sektorene som er valgt ut er et godt utgangspunkt for analysen fordi de utgjør en betydelig del av norsk økonomi og de følger en inndeling SSB også følger når de fører statistikk på en rekke relevante variabler. Dessuten er det variasjon i nivå på arbeidsproduktivitet og i hvilken grad sektorene ble rammet av smitteverntiltak. Derfor kan en samlet analyse gi et godt bilde av norsk økonomi.

Sektorene er hentet fra SSB's standard for næringsgruppering (Statistisk sentralbyrå 2009). De tolv sektorene utgjorde omtrent 65% av Norges samlede bruttoprodukt i 2020 og 2021 (Statistisk sentralbyrå 2023). Nedenfor vises en figur som illustrerer de ulike sektorenes

relative størrelse i forhold til norsk økonomi totalt sett, målt i bruttoprodukt, gjennom 2020 og 2021.



Figur 6

Diagrammet viser at de største sektorene er helse- og omsorgstjenester og utvinning av råolje og naturgass. Utenom disse er de resterende sektorene relativt små, spesielt overnattings- og serveringsvirksomhet, jordbruk og fiske.

### 3. ARIMA-modeller

I de følgende avsnittene presenteres teorien bak modellene som brukes i oppgaven, samt estimere modellene og presentere resultatene fra ARIMA. Dette kapittelet skaper den avhengige variabelen i kapittel 4.



### 3.1 Del 1: Autoregressive integrated moving average (ARIMA) - Teori

ARIMA er en type tidsseriemodell som brukes til å predikere en variabel basert på tidligere observasjoner og feilledd. I makroøkonomien er produktivitet noe som gjerne forklares av humankapital og teknologi, men ARIMA tar kun hensyn til empirisk relevant informasjon om selve variabelen. Årsaken til at dette gjøres er at prediksjonene vil gi verdier for arbeidsproduktivitet i en normalsituasjon. Altså bør prediksjonene ARIMA gjør kunne forklares av de variablene som normalt forklarer AP. Når den faktiske AP'en avviker fra prediksjonen bør dette kunne forklares med pandemirelaterte variabler, siden prediksjonene tar hensyn til andre variabler.

Modellen kan forklares som summen av tre deler:

#### **Autoregressiv prosess (AR)**

Den autoregressive delen av modellen beskriver forholdet mellom en variabel og observasjoner av variabelen i tidligere perioder, pluss et feilledd. I mitt tilfelle vil  $y$  være arbeidsproduktiviteten i en sektor (AP):

$$AP_t = \alpha + \beta_1 AP_{t-1} + \beta_2 AP_{t-2} + \dots + \beta_p AP_{t-p} + u_t \quad (3)$$

$\Rightarrow$

$$AP_t = \alpha + \sum_{i=1}^p \beta_i AP_{t-p} + u_t \quad (4)$$

Egenskapene til denne prosessen bestemmer om tidsserien er stasjonær eller ikke, noe den bør være for å få forventningsrette estimater i ARIMA. Denne utfordringen er beskrevet i større detalj lenger ned.

#### **Glidende gjennomsnitt (MA)**

En MA-prosess er en prosess der en variabel forklares av feilleddene i tidligere perioder.

Anta først at feilleddene har forventningsverdi lik 0 og har konstant varians:

$$E(u_t) = 0 \quad \text{var}(u_t) = \sigma^2 \quad (5)$$

$$AP_t = \alpha + u_t + \gamma_1 u_{t-1} + \gamma_2 u_{t-2} + \dots + \gamma_q u_{t-q} \quad (6)$$

⇒

$$AP_t = \alpha + \sum_{i=1}^q \gamma_i u_{t-i} + u_t \quad (7)$$

Variabelen AP vil ha konstant gjennomsnitt, siden feilleddenes forventningsverdi er 0:

$$E(AP_t) = \alpha \quad (8)$$

Variansen vil også være konstant, siden vært av feilleddene har forventningsverdi lik 0, noe AP også har, om man ser bort ifra konstantleddet:

$$var(AP_t) = \sigma^2(1 + \gamma_1^2 + \gamma_2^2 + \dots + \gamma_p^2) \quad (9)$$

Siden MA-leddet inneholder uobserverte data, nemlig tidligere residualer, estimeres ARIMA som inneholder minst ett MA-led med Maximum Likelihood, ikke OLS.

### **Integrated**

I'en i ARIMA henviser til differensiering av variabelen AP, for å sikre stasjonæritet. En stasjonær variabel er en variabel som over tid har konstant gjennomsnitt, konstant varians og konstant autokovarians med alle tidligere observasjoner. Under forrige avsnitt ble det vist at MA-prosessen iallfall tilfredsstiller de to første kravene, gitt forutsetninger om feilleddet. Her beskrives hvordan AR-prosessen kan testes for å avdekke variabelens egenskaper.

Stasjonæritet innebærer at tidsserien er lite persistent, slik at sjokk ikke er vedvarende. Om man har en ikke-stasjonær prosess vil man i en ARIMA-modell få skjeve estimater fordi variabler som egentlig ikke har en sammenheng vil synes å ha en sammenheng over tid. I tillegg vil ikke den ikke-stasjonære variabelen følge en t-distribusjon, slik at t-verdien vil være feil og inferens ugyldig. Få variabler er stasjonære i utgangspunktet, men mange blir det av å differensiere av første orden:

$$\Delta AP_t = AP_t - AP_{t-1} \quad (10)$$

Hvorvidt tidsserien er stasjonær ved ett lag kan testes med en Dickey Fuller test. Om man tar utgangspunkt i ligning (1), kan denne differensieres og gi ligningen som testes:

Ligning (1) med ett lag, uten konstant:

$$AP_t = \beta AP_{t-1} + u_t$$

$$\Rightarrow \Delta AP_t = \theta AP_{t-1} + u_t \quad \beta - 1 = \theta \quad (11)$$

$$H_0: \theta = 0$$

$$H_1: \theta < 0$$

$$t = \frac{\hat{\theta}}{SE(\hat{\theta})} \sim \text{Dickey Fuller distribution}$$

Om absoluttverdien til teststatistikken er større enn absoluttverdien til den kritiske verdien forkastes nullhypotesen. Da er tidsserien stasjonær. I tabell 6 lenger ned er Dickey Fuller-testen gjort på alle sektorene.

(Brooks 2019)

ARIMA-modellen består altså av AR og MA, og må differensieres (I), om AP ikke er stasjonær. Med AR(1) og MA(1):

$$AP_t = \alpha + \beta AP_{t-1} + \gamma u_{t-1} + u_t \quad (12)$$

Om  $AP_t$  ikke er stasjonær blir den differensiert:

$$\Delta AP_t = \alpha + \beta AP_{t-1} + \gamma u_{t-1} + u_t \quad (13)$$

### Seleksjon av beste ARIMA – AIC:

Da er det tre elementer som må bestemmes:

p : Antall perioder tilbake variabelen lagges.

d : Antall ganger variabelen differensieres.

q : Antall perioder tilbake feilledet lagges.

For å finne disse beregnes Akaike's Information Criteria (AIC) for aktuelle modeller.

$$AIC = \ln(\hat{\sigma}) + \frac{2k}{T} \quad (14)$$

$$\hat{\sigma} = \text{Residualvarians} \quad k = p + q + 1 \quad T = \text{Antall observasjoner}$$

(Brooks 2019)

Programmeringsspråket Python brukes for å sammenligne AIC-verdien til en rekke kombinasjoner av p, d og q. Modellen som velges er den med lavest AIC.

Hver sektor tilsvarer en tidsserie, og det vil derfor estimeres en modell for hver sektor i Python/STATA.

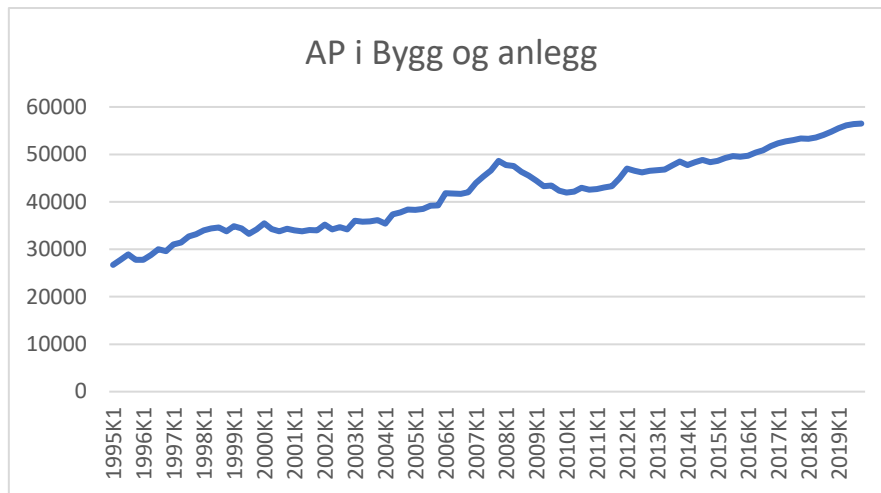
## 3.2 Estimering og prediksjon med ARIMA-modellene

Her presenteres ARIMA-modellene. For hver variabel undersøkes det først om den førstedifferensierte er stasjonær med en Dickey Fuller-test. Resultatene er presentert her.

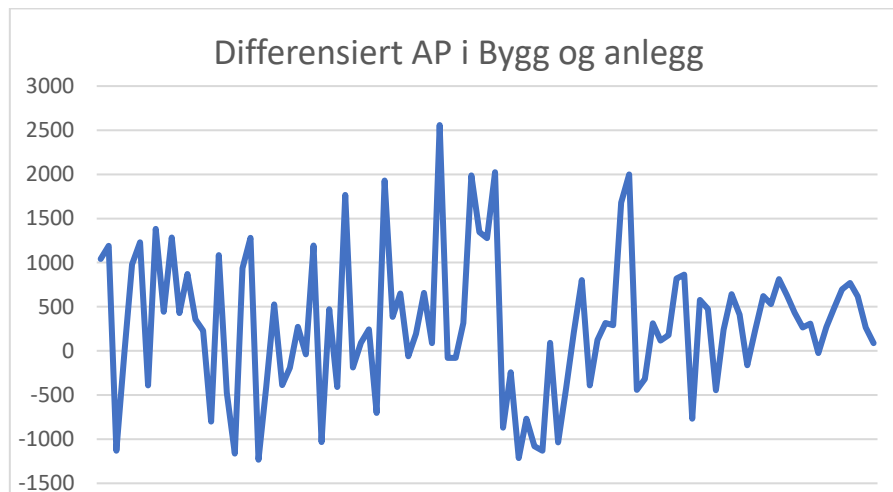
Differensiert variabel	p-verdi DF
Jord_og_skogbruk	9.044830072997563e-24
Fiske_Akvakultur	8.98709599858561e-06
Utvinning_Petroleum	1.0519892030534613e-17
Industri	1.4999692355671555e-23
Elektrisitets_Vann_Gass	3.107105630212051e-07
Bygg_Anlegg	1.0389351944910287e-09
Overnatting_Servering	0.01865723245085774
Finans_Forsikring	1.5318802580121856e-19
Offentlig_adm_Forsvar	2.0436145127183606e-13
Undervisning	5.3394645036308296e-17
Helse_Omsorg	7.156973981483055e-15
Kultur_Underholdning	1.944500634214243e-21

Tabell 6

Siden alle de differensierte variablene har en p-verdi mindre enn 0.05, er konklusjonen at alle er stasjonære. Det betyr at det er en restriksjon på det videre modellvalget, nemlig at variablene skal differensieres én gang. Altså har vi at  $d = 1$  i  $ARIMA(p,d,q)$  for samtlige sektorer. Nedenfor er en grafisk fremstilling av arbeidsproduktiviteten i bygg og anlegg, før og etter differensiering, for å gi et visuelt inntrykk av stasjonærhet.



Figur 7



Figur 8

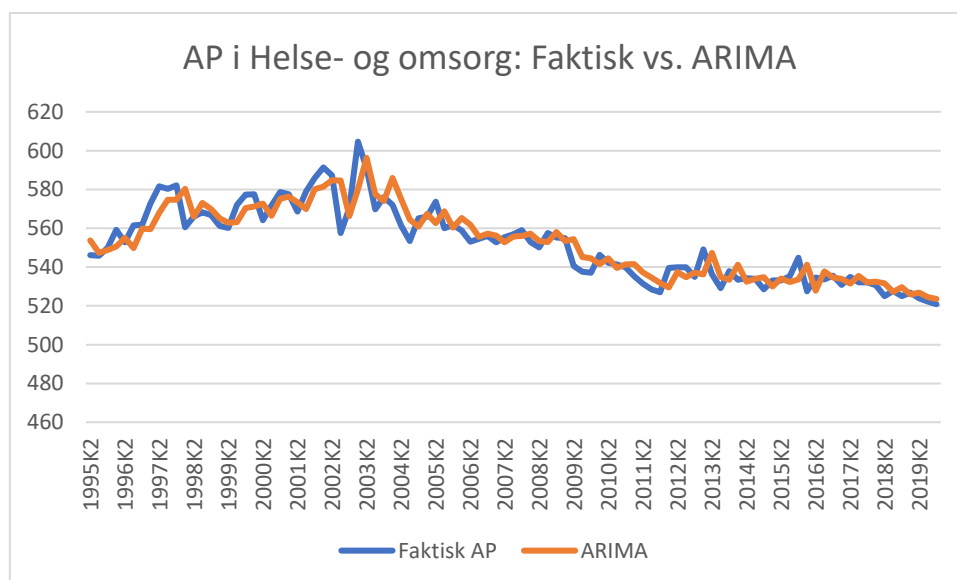
Tabellen nedenfor viser den valgte modellen for hver variabel med hensyn på å minimere AIC, og med restriksjon om at variablene differensieres én gang. Konstanten viser en deterministisk trend i tidsserien, og en konstant gir lavere AIC i noen av modellene. Som et mål på modellenes presisjon brukes Mean Absolute Percentage Error (MAPE). Den gir absoluttverdiene til residualene i prosent. Det betyr at negative og positive residualer ikke nuller hverandre ut, samt at det er mulig å sammenligne modeller der nivået på variablene er ulikt, som er tilfellet her.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{AP_t - \widehat{AP}_t}{AP_t} \right| \quad AP_t = \text{Observerte verdier} \quad \widehat{AP}_t = \text{predikerte verdier} \quad (15)$$

Differensiert variabel	p	d	q	Konstant	MAPE
Jord_og_skogbruk	3	1	1	Ja	6,83 %
Fiske_Akvakultur	1	1	1	Ja	8,42 %
Utvinning_Petroleum	1	1	1	Nei	3,77 %
Industri	2	1	2	Nei	1,46 %
Elektrisitets_Vann_Gass	2	1	2	Nei	5,48 %
Bygg_Anlegg	2	1	1	Ja	1,35 %
Overnatting_Servering	2	1	3	Nei	1,43 %
Finans_Forsikring	1	1	1	Ja	3,01 %
Offentlig_adm_Forsvar	1	1	1	Ja	0,86 %
Undervisning	1	1	2	Nei	1,95 %
Helse_Omsorg	3	1	0	Nei	1,04 %
Kultur_Underholdning	2	1	2	Nei	1,28 %

Tabell 7

MAPE er lav for alle modellene, noe som tilsier at in-sample-prediksjonene passer de faktiske verdiene godt. Det er også tydelig grafisk at ARIMA har høy presisjon når vi ser bakover i tid. Grafen under viser ARIMA-modellen til Helse- og omsorgssektoren mot de faktiske verdiene for arbeidsproduktiviteten.



Figur 9

Neste steg er å bruke de overnevnte modellene til å gjøre prediksjoner for 8 kvartal, fra og med 2020Q1. Prediksjonene starter ved å ta i bruk den faktiske dataen og ikke modellenes

estimerte AP-verdier. Lenger ut i prediksjonen vil de predikerte AP-verdiene brukes til å gjøre prediksjonene. Deretter tas differansen mellom modellenes prediksjoner og den observerte arbeidsproduktiviteten i disse kvartalene, som er avhengig variabel i neste del,  $AP_{av}$  (se ligning 2). Denne differansen er også rapportert gjennom MAPE under. Tabell med komplett fremstilling av in-sample-prediksjoner, pandemi-prediksjoner (out of sample) og faktiske verdier er tilgjengelig i vedlegg 3. Modellene i sin helhet, med koeffisienter og signifikansnivåer er rapportert i vedlegg 3.1.

Nedenfor ses MAPE for out of sample-prediksjonene gjennom pandemien.

Variabel	MAPE
Jord_og_skogbruk	8,66 %
Fiske_Akvakultur	10,42 %
Utvinning_Petroleum	4,64 %
Industri	3,18 %
Elektrisitets_Vann_Gass	4,66 %
Bygg_Anlegg	1,10 %
Overnatting_Servering	37,09 %
Finans_Forsikring	1,80 %
Offentlig_adm_Forsvar	1,10 %
Undervisning	1,69 %
Helse_Omsorg	4,35 %
Kultur_Underholdning	13,64 %

Tabell 8

Legg merke til at MAPE er høyere gjennom pandemien enn før pandemien, i tabell 7. Dette er å forvente, siden modellen frem til 2019 vil kunne predikere basert på de faktiske observasjonene. Etter 2019 vil modellen kunne basere seg på faktiske observasjoner til å begynne med, men vil etter hvert bruke predikert arbeidsproduktivitet som input for neste prediksjon. Da vil naturligvis presisjonen svekkes.

## 4. Del 2: Paneldata – metode og analyse

## 4.1 Least absolute shrinkage operator (LASSO)

LASSO er en metode som er hensiktsmessig når man ønsker å selektare variabler fra et sett variabler. Det bidrar til å finne en modell som ikke har for mange variabler. Metoden ble introdusert av Robert Tibshirani i 1996 (Tibshirani 1996). Følgende problem blir minimert:

$$\arg \min \sum_{i=1}^N \left( y_i - a - \sum_{j=1}^K \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^K |\beta_j| \quad (16)$$

, der N er antall observasjoner og K er antall parametere.

Det første leddet er det samme som minimeres ved OLS-regresjon, altså summen av de kvadrerte residualene. Det andre leddet er summen av absoluttverdiene til koeffisientene, multiplisert med lambda. Når denne ligningen minimeres vil koeffisientene nærme seg null, og noen vil bli null. Derfor kalles lambda straffeleddet: det straffer koeffisienter mot null, og små koeffisienter vil bli null. Hvor streng straffen er avhenger av størrelsen på lambda. Desto større lambda, desto strengere straff. Det vil si at større lambda gir skjevare estimer og mulig flere som bli null. Lambdaen bestemmes ganske likt som ved ARIMA, nemlig ved AIC-kriteriet.

$$AIC = N * \log\left(\frac{RSS}{N}\right) + 2 * df \quad (17)$$

$N = \text{Observasjoner}$      $RSS = \text{Sum av kvadrerte residualer}$      $df = \text{frihetsgrader}$

(Ahrens, Hansen et al. 2022)

Som forklart vil LASSO trekke i retning av en "enkel" modell, men dette går på bekostning av estimatenes egenskaper. De estimerte koeffisientene vil ikke være forventningsrette, altså  $\hat{\beta} \neq E(\beta)$ . Det skyldes, som sett, at straffeleddet skyver koeffisientene mot null.

Alternativt kan lambda velges med kryssvalidering, noe som er ganske vanlig. Kryssvalidering innebærer å dele datasettet opp i trening- og testdata, noe som i dette tilfellet anses som lite hensiktsmessig siden datasettet består av såpass få observasjoner. Da blir både trening- og testdataen svært begrenset og lambdaen blir bestemt på bakgrunn av svært lite data. Med AIC blir hele datasettet utnyttet.



## 4.2 Variabelseleksjon med LASSO

I STATA er prosessen over gjennomført på datasettet fra 2020 og 2021, presentert i kapittel 2.2.1 – 2.2.3. Allikevel legges en restriksjon, nemlig at variabelen  $s\_index\_start$ , som måler grad av restriksjoner i Norge, skal være med. Denne er også med i lagget form. Tabellene under viser alle variablene LASSO selv velger. For å gjøre det oversiktlig er det én tabell for koronarestriksjoner og én tabell for andre variabler.

Pandemivariabler	
Variabel	Beskrivelse
Int dom travel	Dummy for restriksjoner på reising
L. stay at home	Dummy for restriksjoner på å forlate hjemmet
L. hosp adm	Antall pasienter innlagt på sykehus

Tabell 9: Benevnelsen "L." refererer til den laggede variabelen

Andre kontrollvariabler	
Variabel	Beskrivelse
ledige stillinger	Ledige stillinger per sektor rundet av til nærmeste 100
midlertidige ansatte	Antall midlertidige ansatte per sektor
antall virksomheter	Antall virksomheter per sektor
L. avvik i arbeidsproduktivitet	
L. ledige stillinger	
L. norske varer u energi	Konsumprisindeks
L. importerte varer	Konsumprisindeks
L. midlertidige ansatte	
L. antall virksomheter	

Tabell 10: Benevnelsen "L." refererer til den laggede variabelen

Koeffisientene LASSO estimerer er skjeve og derfor ikke nødvendig å tolke i dette tilfellet. Av den grunn rapporteres de heller ikke her. Hensikten med LASSO i denne oppgaven er variabelseleksjon, ikke estimering og inferens, som det også kan brukes til.

I de neste avsnittene vil det vises at LASSO-metoden og variablene den selekterer ikke kan stoles blindt på. Siden Fixed effects benyttes må det tas hensyn til en rekke forutsetninger

om modellens spesifikasjon, blant annet simultanitet, heteroskedastisitet og multikollinearitet. Dette er forutsetninger som ligger til grunn for konsistente og forventningsrette estimater.

#### 4.4 Fixed effects: Teori og forutsetninger

Med variablene over og  $s\_index\_start$  inkludert, er følgende modell utgangspunktet foran fixed effects:

$$APav_{it} = \alpha + \delta_1 s\_index\_start_{it} + \delta_2 L.s\_index\_start_{it} + \beta X_{it} + u_{it} \quad (18)$$

, der  $X$  er en kolonnevektor som inneholder alle variabler i tabell 9 og tabell 10 og  $\beta$  er tilhørende radvektor med koeffisienter.

Fixed effects (FE) er en velkjent økonometrisk metode når man ønsker å fjerne påvirkningen av uobservert heterogenitet i variabler som ikke varierer over tid, men kun mellom sektorer. FE-estimerer oppnås gjennom OLS-regresjon på en ligning som er endret slik:

Anta den opprinnelige regresjonslikningen (med én  $\beta$ ):

$$AP_{it} = \alpha_i + \beta_1 x_{it} + u_{it} \quad t = 1, 2, \dots, 8 \quad i = 1, 2, \dots, 12$$

$$AP_{it} - \overline{AP}_i = \beta_1 (x_{it} - \bar{x}_i) + u_{it} - \bar{u}_i \quad (19)$$

$$\overline{AP}_i = T^{-1} \sum_{t=1}^{12} AP_{it}$$

(Wooldrige 2015)

Min modell transformeres slik som over og ligning (19) estimeres så med OLS. Eventuelle variabler som ikke varierer over tid,  $\alpha_i$ , er nullet ut, siden deres gjennomsnitt over tid vil være likt verdien på et hvilket som helst tidspunkt. Om en sektor har uobserverte variabler som ikke varierer over tid, så vil ikke dette gi skjevhet i estimatene ved bruk av FE, selv om de korrelerer med de andre uavhengige variablene. Det er godt mulig at slike variabler

eksisterer for de ulike sektorene, som påvirker avvik i arbeidsproduktiviteten. Det kan for eksempel være egenskaper ved arbeidskontrakter og forskjeller mellom fysisk arbeid og kontorarbeid, eller tilgang på ressurser. Da restriksjoner ble innført var det slik at enkelte sektorer kunne fortsette arbeidet nesten som normalt, men andre sektorer kunne det ikke. Noen av disse forholdene, som det å sitte på serveringssteder med avstand, eller det å ha hjemmekontor, er potensielle uobserverte variabler som varierte lite gjennom pandemien. Derfor kan FE være en god metode for å svare på problemstillingen oppgaven reiser.

Følgende forutsetninger gjelder for at FE skal gi forventningsrette og konsistente estimater:

1. Den avhengige variabelen kan beskrives som en lineær sammenheng av parametre.

$$AP_{it} = \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_k x_{itk} + u_{it}$$

2. Observasjonsutvalget fra i'ene er tilfeldig. De n observasjonene er et tilfeldig utvalg av populasjonen.
3. Ingen perfekt multikollinearitet. Ingen uavhengig variabel har en perfekt lineær sammenheng med andre uavhengige variabler. Hver variabel har noe variasjon over tid.
4. Feilledet, ved et hvert tidspunkt t, har forventningsverdi lik null, gitt de uavhengige variablene i alle tidsperioder.

$$E(u_{it}|X_i, \alpha_i) = 0$$

Dette impliserer at de uavhengige variablene er eksogene.

5. Feilledet har konstant varians over tid, gitt de uavhengige variablene. Homoskedastisitet.

$$var(u_{it}|X_i, \alpha_i) = var(u_{it}) = \sigma_u^2 \quad \text{for alle } t = 1, 2, \dots, 8$$

6. For  $t \neq s$  er feilledene ikke korrelert, gitt de uavhengige variablene.

$$Cov(u_{it}, u_{is}|X_i, \alpha_i) = 0$$

7. Gitt de uavhengige variablene er feilledet normalfordelt med 0 i gjennomsnitt og konstant varians. Det betyr at t-statistikkene til estimatene og F-statistikken faktisk følger t- og F-distribusjon.

Under forutsetning 1 – 6 er FE-estimatene *best linear unbiased estimator (BLUE)*. At estimatorene er forventningsrette (unbiased) betyr at forventningsverdien til estimatet er lik den faktiske verdien som passer hele populasjonen.

$$E(\hat{\beta}_i) = \beta_i \text{ for alle } i = 1, 2, \dots, k$$

(Wooldridge 2015)

I de neste avsnittene vurderes de mest sentrale forutsetningene og på bakgrunn av disse vurderingene må den LASSO-selekterte modellen justeres.

#### 4.4.1 Simultanitet

Simultanitet fører til endogenitet og derfor skjeve estimater, slik at forutsetning 4 brytes. De selekterte variablene i tabell 9 og tabell 10 vil naturlig vekke mistanke om dette. For eksempel kan det tenkes at antall virksomheter i en sektor påvirkes av hvor godt bedriftene i sektoren klarer seg, i form av arbeidsproduktivitet. Et stort fall i arbeidsproduktivitet kan føre til dårligere konkurransekraft og konkurs, derfor færre virksomheter.

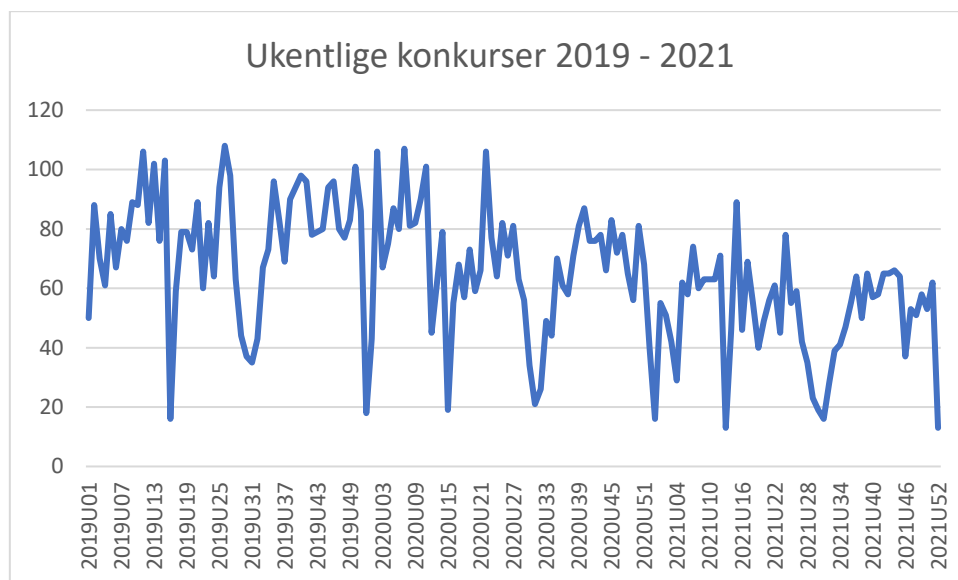
Et positivt avvik, slik det er definert i ligning 2, innebærer at sektoren har høyere arbeidsproduktivitet enn det prediksjonen tilsier. Et avvik kan øke i positiv retning fra én periode til en annen av tre grunner:

1. Prediksjonen er konstant, men faktisk arbeidsproduktivitet øker.
2. Den faktiske arbeidsproduktiviteten er konstant, men prediksjonen minker.
3. Den faktiske arbeidsproduktiviteten øker, og prediksjonen minker.

Sett i lys av dette er det verdt å merke seg at den faktiske arbeidsproduktiviteten aldri er konstant over to perioder eller mer, men det hender prediksjonene er det. Tilfelle 2 er derfor ikke reelt i mitt datasett. Allikevel er det ikke i prediksjonene det er eventuell simultanitet, men i den faktiske arbeidsproduktiviteten. Tilfelle 1 og 3 er derfor mulige, og innebærer endringer i arbeidsproduktivitet som kan ha en effekt på samtlige variabler i tabell 9 og 10.

I tillegg til antall virksomheter er det naturlig å mistenke simultanitet i antall ledige stillinger og antall midlertidige ansatte, siden det kan være en sammenheng mellom disse variablene. Av denne grunn utelates disse variablene, men deres laggede variabel er inkludert, for å redusere simultanitet. Allikevel er det slik at lagging ikke eliminerer simultanitet (Reed 2014).

Men, det kan det også argumenteres for at simultaniteten er liten ved å inkludere disse variablene. Det er mulig antall virksomheter kan påvirkes gjennom AP ved at det fører til konkurser, men dataene for antall åpne konkurser i Norge fra 2019 gjennom 2021 viser ikke noe stort skifte da pandemien kom, se graf under (Statistisk sentralbyrå 2023):



Figur 10

Noen koronaeffekt på antall konkurser er derfor ikke åpenbar, og kanskje vil ikke arbeidsproduktiviteten, eller avviket i arbeidsproduktiviteten i en gitt sektor, påvirke antall virksomheter i den sektoren.

Antall midlertidige ansatte kan også ha liten simultanitet med avvik i arbeidsproduktiviteten. En rekke faktorer bestemmer antall midlertidige ansatte, og avviket i arbeidsproduktiviteten er kanskje en faktor, men trolig en liten en. Det samme kan sies for *ledige stillinger*. Veksten i en sektor vil trolig korrelere med arbeidsproduktiviteten, men andre faktorer vil stå for mye av variasjonen i *ledige stillinger*. Men, det kan ikke utelukkes at det med disse variablene

foreligger et simultanitetsproblem som fører til en endogen modell. Å kun inkludere de laggede variablene har flere fordeler: Det blir færre variabler med potensiell simultanitet. For det andre tas det hensyn til at variablene bruker tid på å påvirke  $AP_{av}$ . Til slutt så reduserer det multikollinearitet fordi variablene i periode  $t$  korrelerer sterkt med variablene i periode  $t-1$  for de nevnte variablene.

Pandemivariablene er det derimot lite sannsynlig at bestemmes simultant med avvik i AP. Det er lite trolig at sektorenes arbeidsproduktivitet gir sterke implikasjoner for myndighetenes restriksjoner på såpass kort sikt, men muligens kan det påvirke gjennom koronastøtte til virksomheter, noe det gjerne skulle vært kontrollert for, men dessverre ikke er. Det skrives mer om i diskusjonsdelen.

#### 4.4.2 Multikollinearitet

Forutsetning 3 om multikollinearitet kan undersøkes ved å se på korrelasjoner, og det er høyst relevant med tanke på variablene i datasettet. Siden det er rimelig å forvente at pandemirestriksjonene endret seg samtidig og i samme retning vil det potensielt være vanskelig å se den isolerte effekten av hver pandemirelatert variabel.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(1) s_index_start	1.000													
(2) L.s_index_start	-0.430	1.000												
(3) int_dom_travel	<b>0.864</b>	-0.350	1.000											
(4) ledige_stillingerr	0.127	0.076	0.219	1.000										
(5) midl_ansatte	-0.007	0.052	0.025	0.583	1.000									
(6) antall_virksomheterr	-0.003	0.003	0.000	0.227	0.213	1.000								
(7) L.AP_av	0.021	-0.188	0.060	0.083	-0.012	-0.056	1.000							
(8) L.hosp_adm	0.499	0.244	<b>0.743</b>	0.345	0.072	0.004	-0.033	1.000						
(9) L.stay_at_home	-0.015	0.218	0.417	0.318	0.078	0.006	0.029	<b>0.792</b>	1.000					
(10) L.ledige_stillinger	-0.040	0.173	0.064	<b>0.808</b>	0.607	0.211	0.078	0.169	0.204	1.000				
(11) L.norske_varer_u_energi	-0.201	-0.174	-0.273	-0.088	-0.021	-0.000	0.089	-0.324	-0.245	-0.271	1.000			
(12) L.importerte_varer	0.371	0.176	0.324	0.290	0.041	0.002	-0.047	0.645	0.405	-0.014	0.221	1.000		
(13) L.midl_ansatte	-0.039	0.040	-0.004	0.586	<b>0.984</b>	0.219	0.018	0.041	0.072	0.620	-0.005	0.033	1.000	
(14) L. antall_virksomheter	-0.003	0.003	0.001	0.224	0.210	<b>1.000</b>	-0.056	0.005	0.007	0.208	-0.001	0.003	0.216	1.000

Tabell 11

Pandemivariablene i tabell 9 vil man forvente at har noe korrelasjon. Under pandemien var det gjerne slik når det ble strengere restriksjoner på et område, da ble det strengere restriksjoner på andre områder og. Variabelen *s\_index\_start* er dessuten en funksjon av flere av de enkeltstående restriksjonene. Over vises korrelasjonene for alle variablene i tabell 9 og 10, samt indeksen selv.

Her er det flere korrelasjoner som bør hensyntas. Indeksen har en korrelasjon på 0.864 med restriksjoner på reise i inn- og utland. Sistnevnte har også en korrelasjon på 0.743 med sykehusinnleggelser foregående kvartal. Variablene *ledige\_stillinger*, *midl\_ansatte* og *antall\_virksomheter* har alle høy korrelasjon med sin laggede verdi. I tillegg har *stay\_at\_home* og *hosp\_adm* sine laggede verdier høy korrelasjon. Med denne kunnskapen droppes derfor *int\_dom\_travel* og *L.hosp\_adm*, samt *ledige\_stillinger*, *midl\_ansatte* og *antall\_virksomheter*. Flere av disse ble bestemt droppet under forrige delkapittel, men som det også ble nevnt er multikollinearitet nok et argument for å droppe disse, og det kommer til syne her.

## 5. Resultater

### 5.1 Fixed effects

I tabellen nedenfor presenteres resultatene av fixed effects estimering med variablene fra tabell 9 og tabell 10 etter at en rekke av de er droppet som følge av problemene diskutert i avsnittene over. Den første modellen blir:

#### Modell 1

$$\begin{aligned}
 APav_{it} = & \alpha + \delta_1 s\_index\_start_{it} + \delta_2 L.s\_index\_start_{it} + \beta_1 L.APav_{it} \\
 & + \beta_2 L.stay\_at\_home_{it} + \beta_3 L.ledige\_stillinger_{it} \\
 & + \beta_4 L.norske\_varer\_u\_energi_{it} + \beta_5 L.importerte\_varer_{it} \\
 & + \beta_6 L.midlertidige\_ansatte_{it} + \beta_7 L.antall\_virksomheter_{it} + u_{it} \quad (19)
 \end{aligned}$$

Standardfeilene er robuste mot heteroskedastisitet, men ikke under like streng forutsetning som forutsetning 6. STATA tillater seriekorrelasjon i sine robuste estimater, slik at resultatene kun er robuste mot heteroskedastisitet mellom sektorer.

VARIABLES	(1) AP_av
s_index_start	-0.325* (0.138)
L.s_index_start	0.115 (0.221)
L.AP_av	-0.0183 (0.276)
L.stayathome	-9.660 (17.55)
L.ledige_stillinger	0.00558 (0.00451)
L.norske_varer_u_energi	9.716 (6.362)
L.importerte_varer	-0.0572 (0.238)
L.midl_ansatte	1.182 (1.026)
L.antall_virksomheter	0.00502 (0.00861)
Constant	-1,284 (705.3)
Observations	49
Number of id	7
R-squared	0.207

Figur 11

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

STATA utelater flere sektorer som følge av at panelet er ubalansert. Modellens forklaringskraft er på 20,7%, noe som ikke er særlig høyt. Kun én variabel er signifikant på 10% signifikansnivå, nemlig *s\_index\_start*. Ingen andre variabler viser seg å være signifikante.

Det sentrale resultatet fra denne modellen er at indeksen for restriksjoner har en statistisk signifikant effekt på avviket i arbeidsproduktivitet. Koeffisienten har en verdi på -0,325 og



den økonomiske tolkningen blir: Når indeksen øker én enhet i verdi, så synker avvik i arbeidsproduktivitet med 0,325 kroner bruttoprodukt per timeverk. For å gi et innblikk i den økonomiske signifikansen, eller betydningen, kan man regne ut den marginale effekten av å gå fra indeksens minimum (11,11), til indeksens maksimum (79,63). Det er en differanse på 68,52 og gir derfor en endring på  $-0,325 * 68,52 = -22,27$  kroner. Avviket blir sterkere negativt, det vil si differansen mellom faktisk arbeidsproduktivitet og prediksjoner. Tolkningen er derfor ikke rett frem. Prediksjonen kan også endre seg fra et kvartal til et annet og derfor endre størrelsen på avviket. Men, variasjonen i arbeidsproduktivitet gjennom pandemien er betydelig større enn variasjonen i prediksjonene. Derfor kan det meste av den marginale effekten tilskrives endring i arbeidsproduktivitet, uten at den kan kvantifiseres nøyaktig med en slik tolkning. Tolkningen vurderes nærmere i diskusjonsdelen, men for nå er resultatet at en signifikant endring i  $AP_{av}$  vil tilsvare en endring i  $AP$ .

Tatt i betraktning at analysen er gjort på et panel bestående av 7 sektorer med ulikt produktivetsnivå vil det være vanskelig å si mye om økonomisk signifikans. Det vil avhenge av hvilken sektor man ser på. Ta for eksempel overnatting og servering, som har et lavt produktivetsnivå, og finans og forsikring, som har et høyt produktivetsnivå. Overnatting- og serveringsbransjen har et gjennomsnittlig avvik på -94,45 under pandemien. Gjennomsnittet for finans- og forsikringsbransjen er 17,25. I snitt går det altså dårligere enn forventet med overnattings- og serveringsbransjen og bedre enn forventet med finans- og forsikringsbransjen. Fra indeksens laveste verdi til indeksens høyeste verdi kan man forvente en reduksjon i avvik på 22,27. Overnatting- og serveringsbransjen går da fra vondt til verre, og finans- og forsikringsbransjen ligger rundt forventet arbeidsproduktivitet, dersom endringene skjer fra gjennomsnittlig verdi. Tabell 1 viser dessuten at de to sektorene har vidt forskjellig nivå på arbeidsproduktiviteten og reduksjon i avvik er relativt sett større for sektorer med lavt produktivetsnivå sammenlignet med sektorer med høyt produktivetsnivå.

Eksempelet ovenfor illustrerer en svakhet ved analysen. Er det kanskje naivt å tro at man i en paneldatanalyse skal finne noen økonomisk signifikant sammenheng mellom indeks og avvik i arbeidsproduktivitet når panelet består av sektorer med så ulikt produktivetsnivå, og som dessuten ble rammet svært ulikt av pandemiens restriksjoner? Med andre ord vil det

være interessant å se sektorenes individuelle effekt på avvik i arbeidsproduktiviteten. Derfor utføres i neste del en regresjon med en dummy-variabel for hver sektor.

## 5.2 Regresjon med dummy-variabler

Overnevnte kritikk er et argument for å undersøke hvorvidt sektorene i seg selv er signifikante i å forklare avvik i arbeidsproduktiviteten. Ved å inkludere dummy-variabler for hver sektor og ellers estimere den samme modellen som over vil man få de samme koeffisientene og standardfeilene som ved fixed effects. Modellen blir nå:

### Modell 2

$$\begin{aligned} APav_{it} = & \alpha + \delta_1 s \text{ index } start_{it} + \delta_2 L.s \text{ index } start_{it} + \beta_1 L.APav_{it} \\ & + \beta_2 L.stay \text{ at } home_{it} + \beta_3 L.ledige \text{ stillinger}_{it} \\ & + \beta_4 L.norske \text{ varer u energi}_{it} + \beta_5 L.importerte \text{ varer}_{it} \\ & + \beta_6 L.midlertidige \text{ ansatte}_{it} + \beta_7 L.antall \text{ virksomheter}_{it} + \theta_i S_{it} \\ & + u_{it} \quad (20) \end{aligned}$$

Der S er en kolonnevektor med dummyer for hver sektor og  $\theta$  er tilhørende koeffisienter.

VARIABLES	(2) AP_av
s_index_start	-0.325 (0.297)
L.s_index_start	0.115 (0.280)
L.AP_av	-0.0183 (0.240)
L.stayathome	-9.660 (13.58)
L.ledige_stillinger	0.00558 (0.00395)
L.norske_varer_u_energi	9.716 (11.07)
L.importerte_varer	-0.0572 (0.456)
L.midl_ansatte	1.182 (1.006)
L.antall_virksomheter	0.00502 (0.0109)
Bygg_anlegg	-42.23 (57.21)
Overnatting_servering	146.4 (523.4)
Finans_forsikring	353.0 (662.1)
Offentlig_adm	305.3 (663.5)
Undervisning	198.2 (510.5)
Helse_omsorg	-50.55 (140.9)
Constant	-1,414 (1,387)
Observations	49
R-squared	0.675

Figur 12

Robust standard errors in parentheses  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Resultatene viser dog at standardfeilene ikke er de samme som i fixed effects. Dette er fordi STATA regner ut de robuste standardfeilene annerledes for en dummy-regresjon kontra en fixed effects-regresjon. Mer spesifikt vil STATA i fixed effects-regresjonen tillate seriekorrelasjon i feileddene over tid, men i dummy-regresjonen vil dette ikke tillates og

standardfeilene blir derfor større. Altså tar STATA hensyn til forutsetning 6 i dummy-regresjonen, noe som gir større standardfeil.

Siden standardfeilene nå har økt er ikke lenger  $s\_index\_start$  signifikant. For sektor-dummyene er jord- og skogbruk referansen, men ingen av dummyene er signifikante. Dette resultatet indikerer at innenfor denne modellen er ikke sektorene i seg selv signifikante forklaringsvariabler. Dette er et uventet resultat, men kan dels forklares av de store standardfeilene. Modellen med dummyer vil som forklart ovenfor være strengere og kun anta at feilleddene er uavhengige mellom sektorer, men ikke gjennom tid.

Samlet sett kan de to modellene oppsummeres som at det er et svakt signifikant forhold mellom avvik i arbeidsproduktivitet og indeks for restriksjoner under forutsetning om at feilleddene kan korrelere over tid, men at sektorene i seg selv ikke synes å kunne forklare de nevnte avvikene.

### 5.3 Justering av modellen – flere observasjoner

En svakhet ved de to modellene presentert ovenfor er at de kun har 49 observasjoner, syv sektorer. Det skyldes det ubalanserte panelet, og variabelen *midlertidige ansatte* er den som fører til at fem sektorer droppes. Derfor vil det her vises den samme modellen som under 4.1, men utelate *midlertidige ansatte*. Da vil ti sektorer inkluderes, men fortsatt ekskluderes to sektorer på grunn av manglende observasjoner av *antall virksomheter*.

#### Modell 3

$$\begin{aligned} APav_{it} = & \alpha + \delta_1 s\_index\_start_{it} + \delta_2 L.s\_index\_start_{it} + \beta_1 L.APav_{it} \\ & + \beta_2 L.stay\_at\_home_{it} + \beta_3 L.ledige\_stillinger_{it} \\ & + \beta_4 L.norske\_varer\_u\_energi_{it} + \beta_5 L.importerte\_varer_{it} \\ & + \beta_6 L.antall\_virksomheter_{it} + u_{it} \quad (21) \end{aligned}$$

En fixed effects-regresjon med robuste standardfeil gir følgende resultater:

VARIABLES	(3) APav
s_index_start	-0.0631 (0.361)
L.s_index_start	0.515 (0.350)
L.AP_av	-0.221 (0.160)
L.stay at home	-9.076 (13.64)
L.ledige_stillinger	0.00102 (0.00563)
L.norske_varer_u_energi	3.118 (8.648)
L.importerte_varer	-0.552 (0.524)
L.antall_virksomheter	0.0274* (0.0130)
Constant	-1,097 (797.9)
Observations	70
Number of id	10
R-squared	0.199

Figur 13

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Modellens forklaringskraft er knapt redusert på tross av at en variabel er fjernet og er på 19,9%, kontra 20,7% i forrige fixed effects-modell. I denne modellen er ikke lenger *s\_index\_start* signifikant. Kun *L.antall virksomheter* er signifikant på et 10% signifikansnivå. Ved å inkludere flere observasjoner og droppe en variabel blir altså resultatet noe annet enn under 5.1. Denne modellen tyder på at det ikke finnes noen signifikant sammenheng mellom avvik i arbeidsproduktivitet og indeks for restriksjoner. Under er det kjørt dummy-regresjonen fra 5.2, men også denne uten *midlertidige ansatte*. Resultatene viser noe annet enn i 5.2. Modell 4:

$$\begin{aligned}
 APav_{it} = & \alpha + \delta_1 s \text{ index start}_{it} + \delta_2 L.s \text{ index start}_{it} + \beta_1 L.APav_{it} \\
 & + \beta_2 L.stay \text{ at home}_{it} + \beta_3 L.ledige \text{ stillinger}_{it} \\
 & + \beta_4 L.norske \text{ varer u energi}_{it} + \beta_5 L.importerte \text{ varer}_{it} \\
 & + \beta_6 L.antall \text{ virksomheter}_{it} + \theta_i S_{it} + u_{it} \quad (22)
 \end{aligned}$$

VARIABLES	(4) APav
s_index_start	-0.0631 (0.390)
L.s_index_start	0.515 (0.352)
L.AP_av	-0.221 (0.175)
L.stayathome	-9.076 (17.28)
L.ledige_stillinger	0.00102 (0.00430)
L.norske_varer_u_energi	3.118 (11.21)
L.importerte_varer	-0.552 (0.600)
L.antall_virksomheter	0.0274** (0.0125)
Utvinning	2,049** (780.6)
Industri	1,194** (541.2)
Bygg	-118.8* (67.84)
Overnatting_Servering	1,221** (602.8)
Finans	1,715** (755.9)
Offentlig_adm_Forsvar	1,694** (752.5)
Undervisning	1,278** (565.9)
Helse	256.1** (120.1)
Kultur_Underholdning	940.1** (445.6)
Constant	-2,119 (1,380)
Observations	70
R-squared	0.815

Figur 14

Robust standard errors in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Jord- og skogbruk er referansen for sektor-dummyene. Denne modellen avslører at sektorene i seg selv viser seg å være signifikante i å forklare avvik i arbeidsproduktivitet. For eksempel har Bygg- og anleggsbransjen en marginaleffekt på -118,8 kroner bruttoprodukt per timeverk sammenlignet med jord- og skogbruk. Det kan tolkes som at strengere smitteverntiltak gikk hard utover denne sektoren. Både helsesektoren og kultursektoren har sterke positive marginaleffekter sammenlignet med jord- og skogbruk. Forklaringskraften er nå betydelig større enn ved fixed effects, 81,5% mot 19,9%. At dummyene er signifikante og at dummy-regresjonen har mye høyere forklaringskraft tyder på at innenfor denne modellen står sektorene for det meste av variasjonen i avvik i arbeidsproduktivitet. Dette kan videre tolkes som at det trolig er utelatt relevante variabler, fordi man skulle tro avviket kunne forklares av mer enn bare hvilken sektor avviket er observert i. Allikevel, med et datasett bestående av såpass få tidsperioder og relativt sett mange sektorer er det ikke overraskende at dummy-regresjonene gir høy forklaringskraft.

## 5.4 Oppsummering av resultater

Tabellen på neste side viser alle de fire modellene i én tabell.

VARIABLES	(1) APav	(2) APav	(3) APav	(4) APav
s_index_start	-0.325* (0.138)	-0.325 (0.297)	-0.0631 (0.361)	-0.0631 (0.390)
L.s_index_start	0.115 (0.221)	0.115 (0.280)	0.515 (0.350)	0.515 (0.352)
L.AP_av	-0.0183 (0.276)	-0.0183 (0.240)	-0.221 (0.160)	-0.221 (0.175)
L.stayathome	-9.660 (17.55)	-9.660 (13.58)	-9.076 (13.64)	-9.076 (17.28)
L.ledige_stillinger	0.00558 (0.00451)	0.00558 (0.00395)	0.00102 (0.00563)	0.00102 (0.00430)
L.norske_varer_u_energi	9.716 (6.362)	9.716 (11.07)	3.118 (8.648)	3.118 (11.21)
L.importerte_varer	-0.0572 (0.238)	-0.0572 (0.456)	-0.552 (0.524)	-0.552 (0.600)
L.antall_virksomheter	0.00502 (0.00861)	0.00502 (0.0109)	0.0274* (0.0130)	0.0274** (0.0125)
Utvinning				2,049** (780.6)
Industri				1,194** (541.2)
Bygg_Anlegg		-42.23 (57.21)		-118.8* (67.84)
Overnatting_Servering		146.4 (523.4)		1,221** (602.8)
Finans_Forsikring		353.0 (662.1)		1,715** (755.9)
Offentlig_adm_Forsvar		305.3 (663.5)		1,694** (752.5)
Undervisning		198.2 (510.5)		1,278** (565.9)
Helse_Omsorg		-50.55 (140.9)		256.1** (120.1)
Kultur				940.1** (445.6)
L.midlertidige_ansatte	1.182 (1.026)	1.182 (1.006)		
Constant	-1,284 (705.3)	-1,414 (1,387)	-1,097 (797.9)	-2,119 (1,380)
Observations	49	49	70	70
R-squared	0.207	0.675	0.199	0.815
Number of id	7		10	

Figur 15

Robust standard errors in parentheses  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1



Resultatene fra de fire modellene gir et svakt grunnlag for å si at det finnes en signifikant sammenheng mellom avvik i arbeidsproduktivitet og indeks for restriksjoner, og kun én modell finner denne sammenhengen med 10% signifikansnivå. Ved å utelate en variabel inkluderes flere observasjoner, og dummy-regresjonen i modell 4 viser at sektorene i seg selv er sterke forklaringsvariabler. Trolig er det flere utelatte variabler, noe som bidrar til den ganske lave forklaringskraften i modell 1 og modell 3. Datasettet LASSO selekterte fra bestod av flere titalls variabler, men gitt resultatene har ikke LASSO selektert gode forklaringsvariabler. Det kan være fordi LASSO ikke har egnet seg i denne sammenhengen, og/eller at datasettet LASSO har selektert fra er mangelfullt. I tillegg vil flere av pandemivariablene, som er ordinale eller dummyer, variere lite over tid, noe som gir fixed effects lite variasjon og dermed lite signifikante variabler.

## 6. Diskusjon

Ulike virkninger av pandemien på økonomien er relevante problemstillinger, men viene å håndtere med presisjon. Det er flere momenter som bør drøftes i vurderingen av om problemstillingen lar seg besvare med metodene som er benyttet i denne oppgaven. I de påfølgende avsnittene drøftes de mange forholdene som trekker resultatene i tvil, og hva som i lignende studier kan gi bedre økonomisk innsikt i fremtiden.

Først og fremst bygger hypotesen om at det er et signifikant forhold mellom avvik i arbeidsproduktivitet og indeks for pandemirestriksjoner på et ambisiøst premiss, nemlig at prediksjonene gjort med ARIMA-modellene er gode og "korrekte". Hensikten med å benytte denne metoden er at den skal anslå den riktige arbeidsproduktiviteten de neste 8 kvartalene når man står i begynnelsen av 2020. Dersom det finnes argumenter for at prediksjonene er dårlige, da kan det og argumenteres for at resultatene fra LASSO og Fixed effects er lite pålitelige. ARIMA vil bli mindre presis etter hvert som man predikerer flere fremtidige perioder. For eksempel vil en ARIMA(1,1,1), som er tilfellet for fiske- og akvakultur, kun benytte faktisk observert AP i én periode frem, altså 2020Q1. Verdiene fra 2020Q2 og utover vil være predikert med bakgrunn i forrige periodes prediksjon. Da vil det naturligvis bli

mindre presist i å predikere verdier etter hvert som tidshorizonten øker. Det fører til at avvik i arbeidsproduktivitet blir beregnet med mindre gode forutsetninger. Allikevel er dette et mulig alternativ til å kun analysere et panel fra og med 2020 med faktisk arbeidsproduktivitet som avhengig variabel. Årsaken er som tidligere nevnt at ARIMA-prediksjonene reduserer problemet ved utelatte variable, siden prediksjonene bør være resultatet av de relevante variablene. Da er hypotesen at avvikene som oppstår har sammenheng med smitteverntiltakene.

En annen grunnleggende del av oppgaven som gir utfordringer er det faktum at den forsøker å besvare kortsiktige endringer i produktivitet på sektornivå. Om man tar i betraktning alle pandemivariablene og spesielt kontrollvariablene, vil man kunne argumentere for at det er lite sannsynlig å se endringer i en hel sektor gitt kortsiktige endringer i for eksempel antall virksomheter. Det bør og nevnes at eventuell koronastøtte til bedrifter vil påvirke dette, og for de fleste bedrifter var det ikke slik at de måtte stenge/permittere med én gang det ble strengere restriksjoner. Koronastøtte til bedrifter kan ha hjulpet de minst produktive bedriftene til å holde hjulene i gang og på den måten gitt lavere AP i sektoren som helhet. Da vil antall virksomheter være høyere og arbeidsproduktivitet være lavere, noe som gjør det vanskeligere å isolere forholdet mellom arbeidsproduktiviteten og indeksvariabelen. I stedet for skulle det vært inkludert variabler som lot meg kontrollere for mottatt koronastøtte, men det viste seg vanskelig på sektornivå.

Å se på avvik i arbeidsproduktivitet gjør også resultatene litt vanskeligere å tolke. Som nevnt under overskriften "Simultanitet", så er det i realiteten to måter avviket kan endres på. Enten ved at arbeidsproduktiviteten og prediksjonen endres, eller at kun arbeidsproduktiviteten endres. Med andre ord, de marginale effektene kan ikke direkte tolkes som en marginal effekt på arbeidsproduktiviteten, siden prediksjonen kan ha skapt endringen i avviket. Det kan allikevel påpekes at korrelasjonen mellom den faktiske arbeidsproduktiviteten i en sektor og avviket er svært høy. Det skyldes at prediksjonene varierer ganske lite, slik at en gitt endring i arbeidsproduktivitet vil korrelere med omtrent samme endring i avvik. Nedenfor vises korrelasjonene for vær sektor gjennom pandemien.

Sektor	Korr (AP , APav)
Jord_og_Skogbruk	0,9271
Fiske_Akvakultur	0,9874
Utvinning_Petroleum	1,0000
Industri	0,9048
Elektrisitet_Vann_Gass	0,9138
Bygg_Anlegg	0,9725
Overnatting_Servering	0,9997
Finans_Forsikring	0,8315
Offentlig_adm_Forsvar	0,7215
Undervisning	0,9993
Helse_Omsorg	0,9999
Kultur_Underholdning	0,9989

Figur 16

Derfor kan man nesten se på avvik i arbeidsproduktivitet som en slags proxy-variabel for arbeidsproduktivitet.

Arbeidet med oppgaven har etter hvert gått i en retning som gjør at problemstillingen ville latt seg bedre svare på om man hadde fokusert på bedrifter fremfor sektorer. Da ville det kanskje ha vært hatt tilgang til flere variabler med kortsiktig betydning og i tillegg vært mulig å ta hensyn til andre innsatsfaktorer i produksjonen. Samtidig kunne det, som det er gjort i oppgaven, også vært inkludert større makroøkonomiske variabler, men ikke minst kontroller for koronastøtte fra myndighetene og kanskje til og med diverse brudd i verdikjeder. Ved å observere bedrifter kunne det eksistert sterkere argumenter for å droppe ARIMA-delen og kun hatt et panel bestående av arbeidsproduktivitet som nevnt over, nettopp fordi det kunne vært lettere å inkludere sterke forklaringsvariabler. Ulempen ved denne tilnærmingen er at det kanskje ville gitt et mindre helhetlig bilde av norsk økonomi, noe som var hensikten med å analysere sektorer som utgjør store deler av økonomien til sammen.

I vurderingen av gevinsten og kostnaden av politiske og økonomiske tiltak bør man alltid ta hensyn til alternativkostnaden, og det er helt sentralt i drøftingen av denne oppgavens relevans og tolkning. For, hvordan hadde avviket i arbeidsproduktiviteten eventuelt endret seg om myndighetene ikke innførte restriksjoner? Spørsmålet er umulig å svare på med to streker under svaret, fordi de langsiktige konsekvensene ikke er kjent fullt ut. Kanskje hadde det vært kortsiktige gevinster, men la meg presisere at med gevinster menes økt arbeidsproduktivitet. Serveringsbransjen ville trolig gått bedre om de kunne tatt imot flere gjester og skjenket alkohol slik de ønsket. På en annen side ville fortsatt mange bedrifter

vært helt avhengige av farten i globale verdikjeder, og den ville ikke endret seg som følge av snill restriksjonspolitik i Norge. En slik politikk ville selvsagt ført til flere smittede og dermed flere alvorlig syke som ville gitt flere sykemeldte og kanskje flere utenfor arbeidsstyrken over tid. Kanskje er det også slik at snill kontra streng restriksjonspolitik ikke ville endret så mye i Norge uansett. Under pandemien var det en viss grad av selvregulering, for eksempel ved at folk holdt avstand selv når det ikke er krav om det, eller satt på hjemmekontor selv om ingen restriksjoner hindrer deg i å dra på kontoret.

Vurderingen av alternativkostnaden vil derfor trekke resultatene i tvil. Selv om det finnes en svak negativ sammenheng mellom avvik i arbeidsproduktivitet og indeks for restriksjoner i én av mine modeller, så kunne konsekvensene av pandemien på avvik i arbeidsproduktivitet vært enda større om restriksjonene ikke hadde blitt innført. Med dette i bakhodet er det fornuftig å være lite bastant i konklusjonene fra mine resultater.

En artikkel fra Deakin Business school i Australia hevder at restriksjoner og andre tiltak bidro til å redusere nedgangen i industriell produktivitet og sørget for at det var lettere for et land å hente seg tilbake til banen de lå på før pandemien på grunn av restriksjonene (lyke, Sharma et al. 2021). Resultatene deres er basert på et panel bestående av femti land. Blant annet argumenterer de for at usikkerheten som følge av pandemien kan påvirke produktiviteten negativt, men at myndighetenes handlekraft i form av restriksjoner og økonomisk støtte reduserte usikkerheten og dermed hindret negative endringer i produktivitet. Det er et argument i retning av at alternativkostnaden ved å ikke innføre restriksjoner er stor, fordi det ville ført til mer usikkerhet og kanskje også mindre koronastøtte, noe som ville hatt negativ effekt på produktiviteten.

Hensikten med dette kapittelet har vært å sette lys på de momentene som er mest sentrale å diskutere relatert til problemstillingen. Problemstillingen er vanskelig å besvare fordi den bygger på ARIMA-prediksjoner det er knyttet tvil til. Videre er det trolig flere utelatte variabler, som for eksempel selvregulering og koronastøtte. Dessuten er alternativkostnaden vanskelig å kvantifisere både på kort og lang sikt og den svake signifikante sammenhengen som finnes i én av mine fire modeller lar seg ikke tolke rett frem. Både fordi den avhengige variabelen er avvik i arbeidsproduktivitet og ikke arbeidsproduktivitet, og fordi effekten av å lette på tiltak ikke nødvendigvis har motsatt effekt av å innføre tiltak.

## 7. Konklusjon

Gitt resultatene som er presentert og diskutert er det ikke mulig å konkludere særlig bastant eller definitivt. Oppgaven kan allikevel oppsummeres ved å se på de to hoveddelene, hva hensikten med disse er og i hvilken grad de anvendte metodene besvarte oppgavens problemstilling.

Den første delen bygger ARIMA-modeller for 12 sektorer. Deretter brukes disse modellene til å predikere åtte kvartaler frem i tid, nærmere bestemt koronaårene. Troverdighetene til disse prediksjonene er avgjørende for oppgavens videre resultater. Modellene viser god in-sample fit, men det er grunn til å være skeptisk til ARIMA-modellenes prediksjoner etter hvert som de predikerer verdier lenger og lenger frem i tid. Det er som kjent ikke lett å forutse fremtiden.

I den neste delen undersøkes avvikene mellom prediksjonene og den faktiske arbeidsproduktiviteten under pandemien. Det kommer frem at noen sektorer gjør det bedre enn forventet og noen sektorer gjør det dårligere enn forventet, se vedlegg 3. Avvikene danner et panel som forsøker å modellere sammenhengen mellom avvik og indeks for restriksjoner. Panelet inkluderer mange forklarings- og kontrollvariabler, og LASSO blir benyttet med det håp om å velge gode forklaringsvariabler.

De fire modellene som brukes for å undersøke sammenhengen er fixed effects og en tilsvarende dummy-regresjon for å se på sektorspesifikke effekter. Forklaringsvariablene som er med i disse modellene er de som ble valgt av LASSO, men noen variabler er droppet fordi de fører til endogenitet og/eller sterk multikollinearitet. Den første fixed effects-modellen viser en svak signifikant sammenheng mellom avvik i arbeidsproduktivitet og indeks for restriksjoner. Når indeksen øker med én enhet, så reduseres avviket med 0,325 kroner bruttoprodukt per timeverk. Det betyr en svakere arbeidsproduktivitet, om man legger til grunn den sterke korrelasjonen mellom avvik og arbeidsproduktivitet, samt det faktum at for mange av sektorene er det lite variasjon i prediksjonene, men mye variasjon i arbeidsproduktiviteten. Modellen har lav forklaringskraft og indeksen er kun signifikant på et

10% signifikant nivå. Siden de tre andre modellene ikke finner noe signifikant sammenheng kan det ikke konkluderes med at sammenhengen er av noen økonomisk betydning.

## Referanseliste

1, N. (2015). Produktivitet – grunnlag for vekst og velferd — Produktivitetskomisjonens første rapport, Finansdepartementet.

Ahrens, A., et al. (2022, 12.12.2022). "Stata ML Page." Retrieved 21.03.2023, from [https://statalasso.github.io/docs/lassopack/help/lasso2\\_help/](https://statalasso.github.io/docs/lassopack/help/lasso2_help/).

Bartik, A. W., et al. (2020). MEASURING THE LABOR MARKET AT THE ONSET OF THE COVID-19 CRISIS.

Brooks, C. (2019). Introductory econometrics for finance, Cambridge university press: 246-292, 343-344.

Brubakk, L. and K. Hagelund (2022, 28.04.2022). "Produktivitetstutviklingen gjennom pandemien." from <https://www.norges-bank.no/bankplassen/arkiv/2022/produktivitetstutviklingen-gjennom-pandemien/>.

Hale, T., et al. (2021). "A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker)." Nature Human Behaviour **5**: 529-538.

Holden, S., et al. (2022). Samfunnsøkonomiske vurderinger av smitteverntiltak.

Iyke, B. N., et al. (2021). "COVID-19, POLICY RESPONSES, AND INDUSTRIAL PRODUCTIVITY AROUND THE GLOBE." Bulletin of Monetary Economics and Banking **24**(3): 365-382.

Karlsen, K. and N. Mathisen (2021). Capital Structure and Machine Learning Techniques in Scandinavia, Norwegian university of science and technology.

Mathieu, E., et al. (2020). "COVID-19: Stringency Index." from <https://ourworldindata.org/covid-stringency-index>.

Mathieu, E., et al. (2020). "Policy Responses to the Coronavirus Pandemic." from <https://ourworldindata.org/policy-responses-covid>.

Reed, R. (2014). On the Practice of Lagging Variables To Avoid Simultaneity.

Statistisk sentralbyrå (2009). "Standard for næringsgruppering (SN)." from <https://www.ssb.no/klasse/klassifikasjoner/6>.

Statistisk sentralbyrå (2017, 24.08.2017). "Produktivitetsberegninger for næringer." Retrieved 31.03.2023, from <https://www.ssb.no/nasjonalregnskap-og-konjunkturer/nasjonalregnskap/artikler/produktivitetsberegninger-for-naeringer>.

Statistisk sentralbyrå (2023). "05333: Varekonsumindeks (2005=100) 2000M01 - 2023M04." from <https://www.ssb.no/statbank/table/05333>.

Statistisk sentralbyrå (2023). "07205: Midlertidig ansatte, etter næring (SN2007) 2008K1 - 2023K1." from <https://www.ssb.no/statbank/table/07205>.

Statistisk sentralbyrå (2023). "08771: Ledige stillingar, etter næring (SN2007) 2010K1 - 2023K1." from <https://www.ssb.no/statbank/table/08771>.

Statistisk sentralbyrå (2023, 15.02.2023). "09171: Produksjon og inntekt, etter næring. Ujustert og sesongjustert 1978K1 - 2022K4." Retrieved 29.03.2023, from <https://www.ssb.no/statbank/table/09171>.

Statistisk sentralbyrå (2023, 15.02.2023). "09175: Lønn og sysselsetting, etter næring. Ujustert og sesongjustert 1995K1 - 2022K4." Retrieved 29.03.2023, from <https://www.ssb.no/statbank/table/09175>.

Statistisk sentralbyrå (2023). "09190: Makroøkonomiske hovedstørrelser. Ujustert og sesongjustert 1978K1 - 2023K1." from <https://www.ssb.no/statbank/table/09190>.

Statistisk sentralbyrå (2023). "11117: Konsumprisindeks for varer og tjenester, etter leveringssektor (2015=100) 2015M01 - 2023M04." from <https://www.ssb.no/statbank/table/11117>.

Statistisk sentralbyrå (2023). "Bruttoprodukt." Retrieved 24.03.2023, from <https://www.ssb.no/a/metadata/conceptvariable/vardok/1744/nb>.

Statistisk sentralbyrå (2023). "Opna Konkursar." from <https://www.ssb.no/statbank/table/12972/tableViewLayout1/>.

Statistisk sentralbyrå (2023, 24.02.2023). "Reiseundersøkelsen." from <https://www.ssb.no/statbank/table/11270/>.

Statistisk sentralbyrå (2023). "Utførte timeverk." Retrieved 24.03.2023, from <https://www.ssb.no/a/metadata/conceptvariable/vardok/1794/nb>.

Statistisk sentralbyrå (2023). "Virksomheter, etter næringshovedområde (SN2007) og antall ansatte (K) 2009 - 2023." from <https://www.ssb.no/statbank/table/10309>.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso." Royal Statistical Society **58**(1): 267-288.



Vries, K. d., et al. (2021). "Productivity and the pandemic: short-term disruptions and long-term implications." International economics and economic policy(18): 541-570.

Vågbø, P. C. B. (2019). The Economic Impact of Large Petroleum Investments on Local Regions in Norway: A Synthetic Control Analysis Using Machine Learning Techniques, Norwegian university of science and technology.

Wooldridge, J. M. (2015). Introductory econometrics: A modern approach, Cengage Learning: 457-458.

Wooldridge, J. M. (2015). Introductory econometrics: A modern approach, Cengage Learning: 435.

Yang, L., et al. (2020). How Work From Home Affects Collaboration: A Large-Scale Study of Information Workers in a Natural Experiment During COVID-19.

Zhang, Y., et al. (2019). "Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors?" Journal of Empirical Finance **54**: 97-117.

