

Sofie Skeie

# **I hvilken grad påvirker lokal arbeidsledighet andelen uføre?**

En empirisk analyse av effekten av arbeidsledighet på andelen uføre i norske fylker i tidsperioden 2000-2014

Masteroppgave i samfunnsøkonomi  
Veileder: Kåre Johansen  
Trondheim, juni 2018

Norges teknisk-naturvitenskapelige universitet  
Fakultet for økonomi  
Institutt for samfunnsøkonomi

**NTNU**

Norges teknisk-naturvitenskapelige universitet.

Masteroppgave i samfunnsøkonomi.

Fakultet for Økonomi  
Institutt for Samfunnsøkonomi.

Sofie Skeie

Printet av NTNU-trykk

---

## Forord

Denne masteroppgaven er en avslutning på en 5-årig integrert mastergrad i samfunnsøkonomi ved Institutt for samfunnsøkonomi på NTNU. Jeg ønsker å rette en stor takk til min veileder Kåre Johansen for veldig gode og kjappe tilbakemeldinger som har hjulpet meg mye i denne prosessen. Videre vil jeg takke familien min: pappa, mamma, Jens og Johanne for god støtte i løpet av dette semesteret og hele studietiden min. Til slutt vil jeg takke Lars Martin Haugland for hans uunnværlige støtte og for at han alltid tror på meg.

Trondheim, 1.juni 2018.

Sofie Skeie



---

## Sammendrag

Norge er i en spesiell situasjon ved at ledighetsraten vår er betydelig lavere enn OECD-gjennomsnittet, mens andelen som mottar uføretrygd er over gjennomsnittet, og nært toppen, blant OECD-landene. Formålet med denne oppgaven er å undersøke i hvilken grad lokal arbeidsledighet påvirker andelen uføre. Oppgaven tar utgangspunkt i et paneldatasett for de 19 norske fylkene i tidsperioden 2000-2014. Den avhengige variabelen er andelen av befolkningen som mottar uføretrygd, mens fokusvariabelen er registrert arbeidsledighet. Resultatene viser at utviklingen i fylkesspesifikke ledighetsrater har signifikant positiv og robust effekt på andelen uføre. Effekten er derimot relativt lav og det kan konkluderes med at ledighet i fylkene i relativt liten grad påvirker andelen uføre. Sammenlignet med de øvrige faktorene som inkluderes i analysen tilsier resultatene at utdannings sammensetningen i fylkene, og da spesielt andelen med lavt utdanningsnivå, i størst grad påvirker andelen uføre. Resultatene viser også at det er en høy grad av persistens i andelen uføre i fylkene.



---

## Abstract

Norway is in an unusual situation where the unemployment rate is lower than the OECD-average, while the share of the population receiving disability pensions is above average, and near the top, amongst the OECD countries. The purpose of this thesis is to examine to what extent local unemployment affects the share of the population receiving disability pensions. The thesis is based on a panel data set for the Norwegian counties in the time period 2000-2014. The dependent variable is the share of the population receiving disability pensions, while the variable of focus is registered unemployment. The results show that development in county-specific unemployment rates have significant positive and robust effects on the share of the population receiving disability pensions. However, the effect is relatively low and it can be concluded that unemployment in the counties to a relatively low extent affects the share of the population receiving disability pensions. Compared to the other factors included in the analysis the results suggest that the educational composition in the counties, and especially the share of the population with a low level of education, to the highest extent affect the share of the population receiving disability pensions. The results also show that there is a high degree of persistence in the share of the population receiving disability pensions in the counties





---

# Innhold

<b>Forord</b>	<b>i</b>
<b>Sammendrag</b>	<b>iii</b>
<b>Abstract</b>	<b>v</b>
<b>1 Innledning</b>	<b>1</b>
<b>2 Bakgrunn</b>	<b>3</b>
2.1 Vilkår for mottak av uføretrygd . . . . .	3
2.2 Mottakere av uføretrygd . . . . .	3
2.3 Utviklingen i arbeidsledighet . . . . .	5
2.4 De norske fylkene . . . . .	6
<b>3 Tidligere litteratur</b>	<b>9</b>
3.1 Arbeidsmarkedet . . . . .	9
3.1.1 Norge . . . . .	9
3.1.2 USA . . . . .	11
3.2 Risikogrupper og demografi . . . . .	11
3.3 Arv . . . . .	12
3.4 Betragtninger om denne oppgaven . . . . .	13
<b>4 Datamaterialet</b>	<b>15</b>
4.1 Avhengig variabel . . . . .	16
4.2 Fokusvariabel . . . . .	16
4.3 Kontrollvariable . . . . .	16
4.3.1 Kontrollvariable inkludert i hovedmodell . . . . .	17
4.3.2 Næringsstruktur . . . . .	18
4.3.3 Fylkets attraktivitet . . . . .	19
4.3.4 Høyere utdanning . . . . .	20
4.3.5 Innvandring . . . . .	20
4.3.6 Politisk tilhørighet . . . . .	21
4.4 Deskriptiv analyse . . . . .	22
4.4.1 Utvalgs karakteristikk og variasjon . . . . .	22
4.4.2 Korrelasjon og mulig multikollinearitet . . . . .	23

4.4.3	Utelatt-variabel-problem . . . . .	25
<b>5</b>	<b>Metode</b>	<b>27</b>
5.1	Modellspesifikasjon . . . . .	27
5.1.1	Korrigerende av estimerte standardavvik . . . . .	29
5.2	Statiske paneldatamodeller . . . . .	29
5.2.1	Pooled OLS . . . . .	30
5.2.2	Fixed effects . . . . .	31
5.2.3	Random effects/GLS . . . . .	32
5.2.4	Hausman-test . . . . .	33
5.3	Dynamiske paneldatamodeller . . . . .	34
5.3.1	Fixed effects . . . . .	34
5.3.2	Arellano og Bond . . . . .	35
<b>6</b>	<b>Resultater</b>	<b>39</b>
6.1	Hovedmodell . . . . .	39
6.1.1	Valg mellom Random effects og Fixed effects . . . . .	40
6.1.2	Effekten av ledighet . . . . .	40
6.1.3	Effekten av kontrollvariable . . . . .	41
6.2	Sensitivitetsanalyse . . . . .	42
6.3	Inkludering av lagget endogen variabel . . . . .	45
6.4	Kontroll for næringstruktur - delperioder . . . . .	47
6.4.1	Periode 1 . . . . .	49
6.4.2	Periode 2 . . . . .	49
<b>7</b>	<b>Avslutning</b>	<b>51</b>
7.1	Avsluttende diskusjon . . . . .	53
7.2	Konklusjon . . . . .	54
	<b>Referanser</b>	<b>57</b>
	<b>Appendiks</b>	<b>59</b>

# 1 Innledning

Inspirasjonen til denne oppgaven kommer fra Norges sydligste by og min andre hjemby, Mandal. I Mandal er det relativt mange uføre; i 2016 mottok 15.6 prosent av befolkningen mellom 18-66 år uføretrygd. Mandal er også en liten kommune med et lite arbeidsmarked og mange innbyggere må derfor pendle til byer som Kristiansand for å finne arbeid. Dette fikk meg til stille meg selv spørsmålet om forholdene på arbeidsmarkedet i Mandal kan være en forklarende årsak for den høye andelen uføre i kommunen.

Mandal har en høyere andel uføre enn på landsnivå, men Norge ligger fremdeles godt over gjennomsnittet blant OECD-landene i andel av befolkningen som mottar uføretrygd (disability benefits).<sup>1</sup> Per september 2017 mottok 9.6 prosent av Norges befolkning mellom 16 og 67 år uføretrygd, noe som tilsvarer rundt 323 000 personer.<sup>2</sup> Samtidig har vi en lav arbeidsledighetsrate og på samme tid var andelen av arbeidsstyrken som var helt arbeidsledige (registrert ledighet) på 2.5 prosent. Dette tilsvarer 68 879 personer og er godt under gjennomsnittet blant OECD-landene.<sup>3</sup> Denne forskjellen mellom andelen uføretrygdde og andelen arbeidsledige setter Norge i en spesiell situasjon sammenlignet med andre land og gir grunn til debatt og spørsmål rundt våre trygdeordninger. Den norske velferdsstatens formål er å ta vare på samfunnet og landets innbyggere, noe som blant annet innebærer at de som grunnet sykdom, skade eller lyte ikke kan delta på arbeidsmarkedet kan få hjelp til livets opphold gjennom å motta uføretrygd. Dermed skal forhold på arbeidsmarkedet egentlig ikke ha noen innvirkning på andelen som blir uføre. Likevel finner Bratsberg, Fevang og Røed (2010) at arbeidsledighet og uføretrygd er nære substitutter og at færre muligheter på arbeidsmarkedet øker sannsynligheten for å bli ufør. I denne oppgaven ønsker jeg derfor å undersøke dette nærmere. Jeg beholder det lokale perspektivet ved å se på de norske fylkene og oppgaven har følgende problemstilling:

*”I hvilken grad påvirker lokal arbeidsledighet andelen uføre?”*

Jeg forsøker å besvare problemstillingen ved en empirisk undersøkelse av effekten av arbeidsledighet på andelen uføre i norske fylker i tidsperioden 2000-2014. Jeg bruker et selvkonstruert paneldatasett for de 19 norske fylkene i analysen.<sup>4</sup>

Det er blitt gjort flere tidligere studier, både i Norge og utlandet, som ser på uføretrygd

---

<sup>1</sup>OECD (2010)

<sup>2</sup>NAV (2017b)

<sup>3</sup>OECD (2017)

<sup>4</sup>Fylkesstruktur i 2014

og forhold på arbeidsmarkedet. I likhet med Bratsberg et al. (2010) ser Rege, Telle og Votruba (2005) på effekten av nedleggelser av bedrifter og individers tilgang til uføretrygd. Bratsberg et al. (2010) finner at nedleggelser i firmaer de har analysert står for over 20 prosent av nye uførekrav, mens Rege et al. (2005) finner at personer som opprinnelig var ansatt i fabrikk som la ned i løpet av en periode på 5 år hadde en 28 prosent større sannsynlighet for å benytte seg av uføretrygd. Bragstad og Hauge (2008) undersøker geografisk variasjon i uførepensjonering i norske kommuner og effekter av forhold på arbeidsmarkedet. Forfatterne finner i dette tilfellet at økt lokal arbeidsledighet gir økt uføretilgang for kvinner, men ikke for menn. I USA finner Charles, Li og Stephens (2017) at verre arbeidsmarkedsforhold i en region øker uføreraten i den regionen og at individer i stor grad substituerer lønninger med uføretrygd når arbeidsmarkedsforhold forverres. En mer detaljert diskusjon av tidligere litteratur på området følger senere i oppgaven.

Oppgaven består av syv kapitler og har følgende oppbygging. I kapittel 1 motiverer jeg oppgaven og problemstillingen spesifiseres nærmere. I kapittel 2 diskuterer jeg bakgrunn for de viktigste temaene i oppgaven. I kapittel 3 presenterer jeg tidligere litteratur på området. I kapittel 4 presenterer jeg datagrunnlaget for analysen samt deskriptiv statistikk. I kapittel 5 diskuterer jeg de økonometriske metodene som benyttes i analysen. I kapittel 6 presenterer jeg resultater for analysen. I kapittel 7 avslutter jeg oppgaven med en oppsummering av funnene samt en diskusjon rundt det viktigste funnene og en helhetlig konklusjon.

## 2 Bakgrunn

*”Formålet med uføretrygd er å sikre inntekt for personer som har fått sin inntektsevne varig nedsatt på grunn av sykdom, skade eller lyte.”* (Folketrygdeloven - § 12- 1 Formål)

Uføretrygd inngår som del av folketrygden som ble opprettet i 1967. Som hovedregel er alle som er bosatt eller arbeider som arbeidstaker i Norge eller på faste eller flyttbare anlegg og innretninger på den norske kontinentalsokkelen pliktige medlemmer i folketrygden. I delkapittel 2.1 og 2.2 vil jeg se nærmere på villkår for mottak og mottakere av uføretrygd i Norge. Videre vil jeg undersøke utviklingen i arbeidsledighet i delkapittel 2.3. Til slutt presenterer jeg noen viktige karakteristikk ved de norske fylkene i delkapittel 2.4. Hensikten med dette kapitlet er å gi leseren nødvendig bakgrunnsinformasjon om uføretrygd, arbeidsledighet og situasjonen i de norske fylkene for å belyse den senere analysen.

### 2.1 Villkår for mottak av uføretrygd

Uføretrygd ytes til medlemmer i trygden som er mellom 18 og 67 år og som har vært medlemmer i minst tre år umiddelbart frem til uførheten inntrådte. Det er obligatorisk at mottaker har gjennomgått hensiktsmessig behandling og atferd og at varig nedsatt inntektsevne på grunn av en varig funksjonsnedsettelse er av slik grad at den fremstår som hovedårsak til tap av inntektsevne. Uføretrygd het frem til 1.1.2015 uførepensjon.<sup>5</sup>

### 2.2 Mottakere av uføretrygd

Som nevnt i innledingen av denne oppgaven er andelen som mottar uføretrygd i Norge høy i forhold til mange andre land. I denne oppgaven ser jeg spesifikt på andelen uføre i de norske fylkene. I figur 1(a) rapporteres andelen uføre i de 19 norske fylkene, samt på landsbasis, for årene 2000, 2008 og 2014. Vi kan se fra figuren at det er stor variasjon mellom fylkene i andelen av befolkningen som mottar uføretrygd. I 2014 skiller Østfold seg ut som fylket med høyest andel uføre, mens Oslo er fylket med lavest andel uføre. Vi ser samtidig at mens andelen uføre har økt i Østfold mellom 2000 og 2014 har andelen i Oslo i samme periode redusert seg betraktelig. I andre fylker som for eksempel Hordaland og Sogn og Fjordane ser vi svært små endringer i andelen uføre i løpet av tidsperioden. På landsbasis har andelen uføre redusert seg litt. Samlet sett kan det observeres at denne

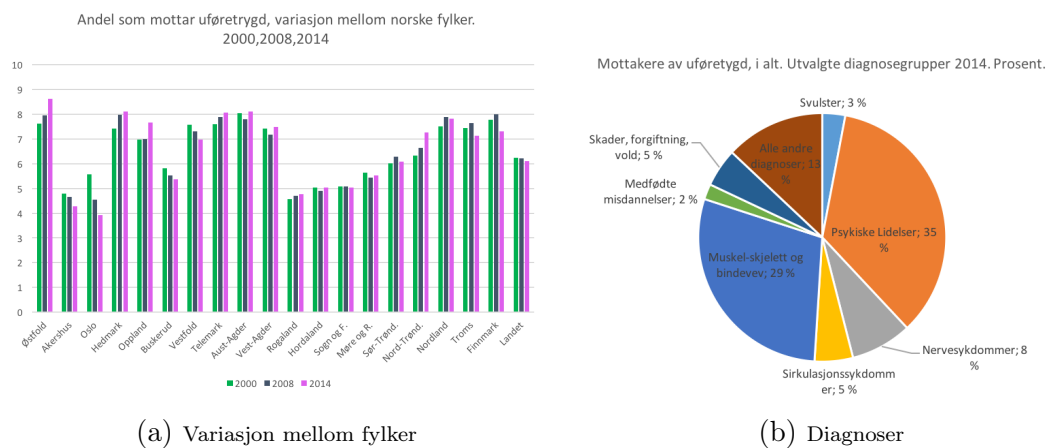
---

<sup>5</sup>Folketrygdeloven - § 12-4 og § 12-5, alder og hensiktsmessig behandling og arbeidsrettede tiltak

## 2. BAKGRUNN

variasjonen i andelen uføre mellom fylkene har holdt seg relativt stabil i tidperioden 2000 til 2014.

Figur 1



Kilde: NAV

Kilde: NAV

Ser vi videre på mottakere av uføretrygd på landsbasis er det en høyere andel kvinner enn menn som mottar uføretrygd, med 11.2 prosent kvinner mot 7.7 prosent menn per desember 2014. Denne forskjellen mellom kjønnene har økt de siste årene. Derimot er andelen blant menn noe høyere enn blant kvinner for de under 30 år som mottar uføretrygd.<sup>6</sup> Det er også varierende diagnoser blant mottakere av uføretrygd på landsbasis, og diagnosemønsteret varierer med hensyn til kjønn og alder. I figur 1(b) kan vi se at per desember 2014 var den største diagnosegruppen ”psykiske lidelser” (35 prosent), etterfulgt av ”muskel-skjelettsykdommer” (29 prosent).<sup>7</sup> Det er en høyere andel menn enn kvinner som faller under kategorien ”psykiske lidelser”, 39 prosent av menn mot 32 prosent for kvinner. Motsatt er det en høyere andel kvinner enn menn som faller i kategorien ”muskel-skjelettsykdommer”, 32 prosent av kvinner mot 21 prosent av menn. Som vi kan forvente er det en positiv sammenheng mellom alder og muskel-skjelettsykdommer, ettersom dette er sykdommer som kommer med alderen og svekket helsetilstand. Derimot observeres det en motsatt sammenheng i kategorien ”psykiske lidelser” der det er en høyest andel unge. Denne sammenhengen er mer uforklarlig utifra et biologisk ståsted. Mellom 2009 og 2014 er det observert en kraftig økning for diagnosegruppen psykiske lidelser på 4.2 prosentpoeng. Samtidig er det observert en reduksjon på 4.7 prosentpoeng for diagnosegruppen muskel-skjelettsykdommer. Dette kan være en del av forklaringen for at andelen unge uføre har økt, mens andelen eldre uføre har blitt redusert.

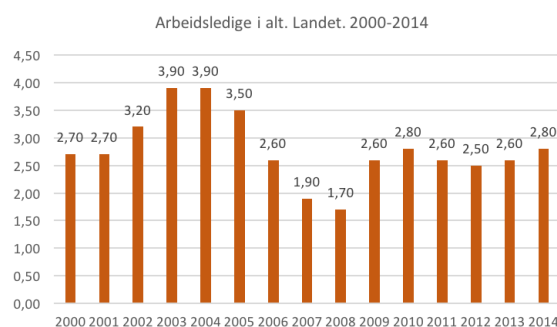
<sup>6</sup>NAV (2015)

<sup>7</sup>NAV (2017a)

Det har vært en stabil utvikling i andelen uføre de siste årene. Samtidig er det en lang-siktig trend at andelen uføre øker blant de unge og reduseres blant de over 55 år. Denne nedgangen for de over 55 år har bidratt sterkt til at utviklingen har holdt seg stabil på tross av økningen blant de unge. Nedgangen i andelen uføre blant de eldste startet på begynnelsen av 2000-tallet og kan trolig forklares av blant annet bedre helse, økt utdanningsnivå og pensjonsreformen i 2011 som åpnet for å ta ut alderspensjon fra 62 år.<sup>8</sup> Pensjonsreformen innførte fleksible uttaksregler for alderspensjon der pensjonen blir justert for uttaksalder slik at utsatt uttak gir en høyere årlig pensjon. I teorien kan det tenkes at denne ordningen kunne ført til økt bruk av blant annet uførepensjon ettersom flere kan tenkes å vente med å ta ut alderspensjon og heller søke om andre ytelser dersom de har et helseproblem. Denne bekymringen kan derimot ikke bekreftes da det i Jacobsen (2014) ikke finnes bevis for at den stemmer og antall mottakere av blant annet uførepensjon har utviklet seg stabilt og så å si likt over tid innenfor privat og offentlig sektor.

### 2.3 Utviklingen i arbeidsledighet

Fokus i denne oppgaven er effekten av lokal arbeidsledighet på andelen uføre i norske fylker. I figur 2 kan vi observere registrert arbeidsledighet på landsbasis i tidsperioden 2000-2014, som er tidsperioden jeg bruker i denne oppgaven. For variasjonen i andel arbeidsledige på fylkesnivå viser jeg til figur 5 i appendiks.



Figur 2: Variasjon i andelen arbeidsledige 2000-2014

Kilde: NAV

Vi kan se fra figur 2 at registrert arbeidsledighet i denne perioden har variert mye, og da spesielt mellom 2000 og 2009, med et toppunkt i 2003 og 2004. Videre stabiliserer andelen arbeidsledige seg mer mellom 2009 og 2014. Dette mønsteret viser seg også å holde i fylkene, som kan ses fra figur 5 i appendiks. I denne oppgaven observerer jeg og benytter

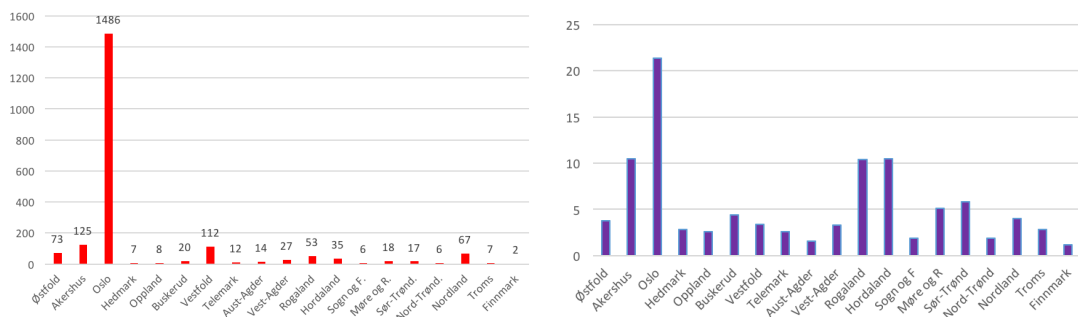
<sup>8</sup>NAV (2017c)

meg av arbeidsledigheten generelt, det vil si at jeg ikke undersøker effekten av spesielle hendelser som har økt eller redusert arbeidsledigheten. Dermed gir endringene i perioden meg muligheten til å observere effekten av naturlig variasjon i andelen arbeidsledige. Selv om det observeres variasjoner i arbeidsledighet har vi i Norge en generelt lav andel arbeidsledige i forhold til andre OECD-land, og ligger under gjennomsnittet for OECD-landene.<sup>9</sup>

## 2.4 De norske fylkene

I denne oppgaven ser jeg på arbeidsledighet og andelen uføre på fylkesnivå for de 19 norske fylkene med den fylkesstrukturen som gjaldt i 2015. Fylkene varierer i størrelse og befolkningsantall og strekker seg følgelig over hele landet. De har ulike karakteristikk og historie innenfor landbruk, fiske, industri osv. som til en viss grad har formet fylket slik det fremstår i dag. Befolkningstettheten i fylkene varierer, med Oslo som det soleklart tettest befolkede fylket. Dette kan observeres i figur 3(a). Her finner vi 1486 personer per  $km^2$  i Oslo, mot 2 personer per  $km^2$  i Finnmark, det minst tettbefolkede fylket. Dette følger naturlig fra at Finnmark har det største landarealet og det laveste befolkningstallet, mens Oslo har det minste arealet og det største befolkningstallet.

Figur 3



(a) Befolkningstetthet

Kilde: SSB

(b) Bruttoprodukt for Norske fylker 2014

Kilde: SSB

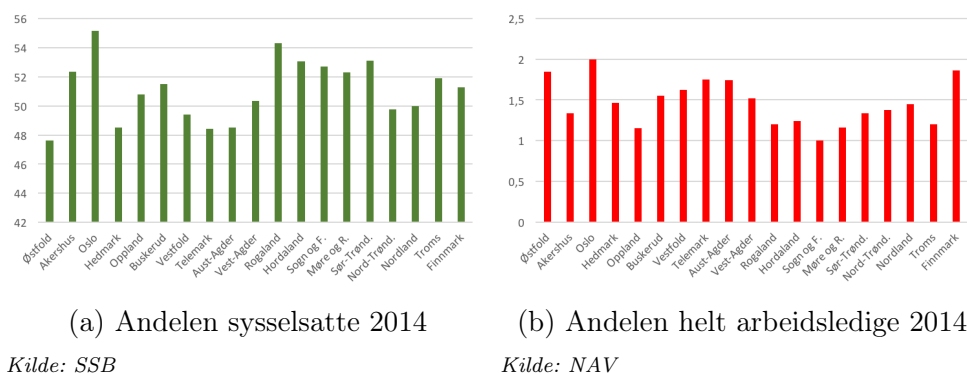
Ser vi på de øvrige fylkene varierer det mellom 6 personer per  $km^2$  (Sogn og Fjordane, Nord-Trøndelag) opp til 125 personer per  $km^2$  i Akershus. Dette er et betydelig mindre intervall der både størrelse og befolkningstall varierer. Det kan tenkes at tett befolkede fylker har et tettere arbeidsmarked, mens arbeidsmarkedene i fylkene med lavere befolkningstetthet har et mer spredt arbeidsmarked. Som et mål på verdiskaping i fylkene kan

<sup>9</sup>OECD (2017)



vi se på bruttoprodukt for fylkene. I figur 3(b) presenteres bruttoprodukt for de norske fylkene i 2014 som andel av landtotalen. Det er tydelig at Oslo har det største bruttoproduktet (21.4 prosent) og dermed kan betraktes som det mest verdiskapende fylket. På delt andreplass finner vi Oslos nærmeste nabo Akershus, samt vestlandsfylkene Rogaland og Hordaland der oljeindustrien står sterkt. Blant fylkene med lavest bruttoprodukt og følgelig de minst verdiskapende fylkene finner vi Finnmark helt nederst (1.2 prosent) med fylker som Nord-Trøndelag og Aust-Agder ikke langt over. Altså er det et ganske bredt intervall mellom fylkene fra 1,2 til 21,4 prosent og dermed store variasjoner i verdiskapning mellom fylkene. Andelen sysselsatte og arbeidsledige varierer også mellom fylkene og presenteres i henholdsvis Figur 4(a) og 4(b). Oslo har den høyeste andelen sysselsatte, og samtidig den høyeste andelen arbeidsledige. Dette kan muligens forklares av at det i Oslo bor veldig mange mennesker og at det er mange på jakt etter arbeid samtidig som det er mange i arbeid. Selv om arbeidsmarkedet er det største i landet kan det tenkes at arbeidsstyrken også er det.

Figur 4



(a) Andelen sysselsatte 2014

Kilde: SSB

(b) Andelen helt arbeidsledige 2014

Kilde: NAV

Fra figur 4(b) kan vi se at det også var betydelig variasjon i andelen arbeidsledige mellom fylkene i 2014, det siste året i tidsperioden som brukes i denne oppgaven. Dette er med på å indikere at det er forskjeller mellom de lokale arbeidsmarkedene i fylkene og forskjeller mellom arbeidsmuligheter.

I denne oppgaven ser jeg spesifikt på effekten av arbeidsledighet på fylkesnivå på andelen uføre. I tillegg til variasjon i andelen arbeidsledige mellom fylkene er det også variasjon i arbeidsledighet på fylkesnivå over tid.<sup>10</sup> Dette diskuteres nærmere i kapittel 4. Med bakgrunnsinformasjonen som er presentert i dette kapittelet i bakhodet vil jeg nå se nærmere på tidligere litteratur på dette området, noe som også er viktig for min analyse.

<sup>10</sup>Appendiks - Figur 5



### 3 Tidligere litteratur

Det er gjort mange studier rundt trygdeøkonomi i Norge og i utlandet, og på tema uføretrygd (disability benefits) spesifikt. Som spesifisert ligger fokuset i denne oppgaven på arbeidsmarkedet og da spesifikt effekten av arbeidsledighet på uføretrygd. I delkapittel 3.1 begynner jeg derfor med å presentere studiene jeg mener er mest relevante innenfor dette rammeverket med tema arbeidsmarked. Her presenterer jeg tidligere litteratur fra både Norge og USA. I delkapittel 3.2 presenterer jeg studier som belyser andre faktorer som påvirker andelen uføre, da spesifikt demografi og risikogrupper. Dette er hensiktsmessig ettersom jeg da i min analyse, så godt det lar seg gjøre, kan kontrollere for faktorer og karakteristikk som er med på å bestemme andelen uføre. I delkapittel 3.3 presenteres en studie som undersøker om uførhet er ”arvelig”. Dette er mer utenfor mitt rammeverk, men inkluderes for å belyse at det finnes faktorer som kan være med på å bestemme andelen uføre som jeg ikke har muligheten til å kontrollere for i min analyse. Tilslutt, i delkapittel 3.4 gjør jeg noen betraktninger rundt de tidligere studiene presentert i forhold til denne oppgaven og hvordan de er relevante for min studie.

#### 3.1 Arbeidsmarkedet

Jeg presenterer her studier som ser på arbeidsmarkedsforhold og effekten på uførhet. Jeg begynner med å presentere studier fra Norge i delkapittel 3.1.1, da dette er mest relevant og nær min oppgave som også ser på Norge. I delkapittel 3.1.2 presenterer jeg en studie fra USA som også har relevante funn.

##### 3.1.1 Norge

I Bratsberg et al. (2010), undersøker forfatterene om arbeidsledighet og uføretrygd er substitutter ved å se på forhold i arbeidsmarkedet og sannsynligheten for å bli ufør. De stiller spørsmålet om uførhet i befolkningen er et ledighetsproblem i forkledning ettersom økonomier med lav registrert arbeidsledighet ofte har høye uførerater. Som nevnt i innledningen er dette tilfellet i Norge. Forfatterene bruker data fra perioden 1992-2007 som består av persondata/individdata, beskrivelse av firmaer med hensyn til sammensetting av ansatte og økonomisk ytelse og informasjon om naturen av nedleggelser av firmaer. Datasettet dekomponeres inn i tre fireårs tidsperioder. Forfatterene analyserer, ved bruk av logit-modeller, sannsynligheten for å bli arbeidsledig og sannsynligheten for at den arbeidsledige finner nytt arbeid innen et år. De finner at mindre muligheter på arbeids-

### 3. TIDLIGERE LITTERATUR

---

markedet, grunnet nedbemanninger, øker sannsynligheten for å bli ufør og at en stor andel uførekrav i Norge direkte kan attribueres til sjokk i jobbmuligheter/ansettelsesmuligheter. Spesifikt finner forfatterene at nedleggelser står for 20 prosent av nye uførekrav. Videre finner de at for menn vil tap av arbeid mer enn doble risikoen for å bli permanent ufør. Forfatterene konkluderer med at arbeidsledighet og uføretrygd er nære substitutter. Resultatene er konsistente med Rege et al. (2005).

I Rege et al. (2005) undersøkes effekten av nedbemanninger i fabrikker på bruken av uførepensjon i Norge. Forfatterene finner at arbeidere som originalt var ansatt i fabrikker som ble nedlagt mellom 1993 og 1998 hadde en 27.9 prosent større sannsynlighet for å benytte seg av uførepensjon i 1999 i forhold til sammenlignbare arbeidere i fabrikker som ikke la ned. Videre finner forfatterene at effekten av nedbemanninger ikke er lineær da arbeidere som originalt var ansatt i fabrikker som nedbemannet 65-95 prosent av arbeidsstyrken hadde større sannsynlighet for å ende opp som uføre enn arbeidere i fabrikker som la helt ned. Effekten av nedbemanninger ble også undersøkt i Fevang og Røed (2006) der forfatterene så på personer registrert med et arbeidsforhold i 1992 og videre undersøkte om denne bedriften hadde nedbemannet i perioden frem til 2000. Resultatene viser at risiko for å bli ufør øker ettersom graden av nedbemanning øker. Et forsiktig anslag fra forfatterene tyder på at nedbemanninger foretatt i denne perioden i bedrifter som eksisterte i 1992 forårsaket omlag 5 prosent av tilstøtningen til uføretrygd i perioden 1993-2003. Øvrige deler av denne studien presenteres nærmere i delkapittel 3.2.

I Bragstad og Hauge (2008) undersøkes effekten av forhold på arbeidsmarkeder i kommuner i Norge på uføretrygding. Forfatterene bruker et paneldatasett i tidsperioden 1993-2004 for alle kommuner i Norge. De studerer uføretrygding av kvinner og menn separat, der det er mulig, og har inkludert informasjon om arbeidsledighet og omstillinger i industri og bergverk i inntil fem år tilbake i tid. Forfatterene finner at økt lokal arbeidsledighet slår ut i økt uføretilgang tre år senere for kvinner. For menn finner de derimot ikke en tydelig sammenheng. En mulig forklaring som pekes ut kan være at gifte kvinner er mindre mobile enn menn grunnet familieforpliktelser. Videre finner forfatterene at omstilling knyttet til nedgang i sysselsetting i industri og bergverk gir liten eller ingen effekt for kvinner, mens for menn finnes det en negativ sammenheng mellom uføretilgangen og sysselsettingsnedgangen tre år tidligere. Forfatterene finner også at for menn er det en klar effekt av endringer i andelen skilte menn tre år tidligere på uføretrygding. Dette kan ha sammenheng med forhold på arbeidsmarkedet da studier har funnet at menn som mister

jobben i større grad enn kvinner opplever skilsmisser.<sup>11</sup>

#### 3.1.2 USA

I Charles et al. (2017) undersøker forfatterene hvordan lokale arbeidsmarkedsforhold påvirker tilstrømming til uføretrygd i USA (US Social Security Disability Benefits - SSDI). Forfatterene utnytter variasjon i lokale lønninger generert av sjokk i olje og gass produksjon fra eksogene bevegelser i verdenspriser på olje og gass. Dette gjør de for å komme rundt det potensielle problemet med omvendt kausalitet fra at lokale områder med kronisk dårlig helse mest sannsynlig vil ha en høyere tilstøtning til uføretrygd og lavere lønninger. Forfatterene bruker data på county-nivå og ser på perioden 1970-2011. Analysen er en videreføring av tidligere arbeid av blant annet Black, Daniel og Sanders (2002). Hovedfunnet i studien er at verre arbeidsmarkedsforhold i et county øker uføreraten i gitte county. Videre finner de at individer i stor grad substituerer lønninger med uføretrygd når arbeidsmarkedsforhold forverres. Dette ble også funnet i Black et al. (2002), men som det påpekes var denne studien basert på et utvalg med lavt utdannede personer og dermed ble resultatene funnet kun gjeldende for denne gruppen. Derimot bruker Charles m.fl. et mer nasjonalt representativt utvalg av stater, av mer utdannede personer og for en nyere tidperiode. Dette resultatet viser dermed at tilstrømmingen til uføretrygd er lik over områder og personer.

## 3.2 Risikogrupper og demografi

I Fevang og Røed (2006) kartlegges det videre hvilke grupper i befolkningen som har høyest risiko for å bli uføretrygdet. Forfatterene undersøker om individer som blir uføre har karakteristikk som skiller seg fra den øvrige befolkningen. Forfatterene har tilgang på registerdata fra ulike etater for perioden 1993-2003, og undersøker variasjoner i uføretilbøyelighet med en logit-modell som beskriver sannsynligheten for å være ufør i 2003 som en funksjon av relevante forklaringsvariable. De finner at det er store variasjoner i uførerisiko mellom ulike grupper. Forfatterene finner at kvinner er mer utsatte enn menn, eldre er mer utsatte enn yngre, personer med lav utdanning er mer utsatte enn personer med høy utdanning og blant innvandrere er personer fra Nord-Afrika og Midtøsten mer utsatt enn andre. De finner også at endring i familiesituasjon, derunder skilsmisser/separasjon er forbundet med økt sannsynlighet for å bli ufør. Det er imidlertid mer vanskelig å si hvorvidt denne sammenhengen er kausal.

---

<sup>11</sup>Rege, Telle og Votruba (2007)

### 3. TIDLIGERE LITTERATUR

---

I tillegg til å undersøke effekten av forhold på lokale arbeidsmarked ser Bragstad og Hauge (2008) også på andre faktorer som kan forklare geografisk variasjon i andelen uføre. Som nevnt tidligere finnes det en effekt av endring i andelen skilsmisser for menn, men dette finnes ikke for kvinner. Videre finner forfatterene at verken endringer i fordelingen i utdanningsnivå eller endringer i inntektsnivået i kommunen ser ut til å påvirke uførepensjoneringen, men for kvinner ser det ut til at et høyere inntektsnivå blant kvinner i kommunen gir flere uførepensjonister. Nettoinnflytting i kommunen ser ut til å ha en negativ sammenheng med uførepensjonering og forfatterene finner at jo høyere nettoinnflytting gir desto lavere uføretilgang.

I Bratsberg og Røed (2011) ser forfatterene på befolkningens alders- og utdanningssammensetning og hvorvidt endringer i disse kan forklare veksten i bruken av uførepensjon og andre trygdeytelser som har funnet sted siden tidlig 1990-tallet. Forfatterene bruker data for personer født i Norge med to norskfødte foreldre som er bosatt i landet hele året og som er mellom 30 og 61 år. De avgrensner til personer over 29 år på grunnlag av argumentasjon om at personer under denne alderen ikke enda har ”avslørt” sin høyeste fullførte utdanning. For analysen bruker forfatterene en Blinder-Oaxaca-Dekomponering som har grunnlaget at utfallene av interesse kan analyseres som en lineær funksjon av forklaringsvariablene. Forfatterene fokuserer på utfallene ”mottak av uføretrygd” og ”mottak av uføretrygd eller en midlertidig helse relatert ytelse”. De kan isolert sett fastslå at endring i alderssammensetningen mellom 1994 og 2006 tilsier økte uførerater, åpenbart gjennom høyere alder sin naturlig forverrende effekt på helsetilstanden som forårsaker økt uførhet. Derimot tilsier økt utdanningsnivå lavere uføretilbøyelighet og økende utdanningsnivå har trukket ned økningen i uføreraten. Dermed finner forfatterene at disse faktorene samlet sett forklarer lite, om noe, av endringene som har funnet sted. Det avsluttes med å sette søkelyset på arbeidsmarkedet som et mulig område som kan forklare de økende uføreratene.

### 3.3 Arv

I Bratberg, Nilsen og Vaage (2012) ønsker forfatterene å finne svar på om barns eksponering for foreldre som mottar uføretrygd fører til at barnet i fremtiden har en høyere uføretilbøyelighet. Forfatterene bruker registerdata for hele befolkningen som inneholder forelder-barn identifikasjoner, så vel som longitudinell data som inkluderer informasjon om mottakere av uføretrygd. I analysen betraktes sannsynligheten for å bli mottaker av uføretrygd før fylte 40 år for individer født 1951-1962, med uførestatus målt i 1991-2003

og forfatterene bruker en enkel lineær sannsynlighetsmodell. Bratberg m.fl. finner en positiv sammenheng mellom sannsynligheten for å motta uføretrygd mellom foreldre og barn og at denne effekten er sterkest for barn som har opplevd dette over lengst tid. Altså er det en "spillover" effekt mellom foreldre og barn når det gjelder uføretrygd og dersom barn har foreldre som mottar uføretrygd er det en høyere sannsynlighet for at de selv vil motta uføretrygd i fremtiden. Dermed er det ikke bare en bekymring for generasjonen som mottar uføretrygd i dag, men også for deres barn, den fremtidlige generasjonen.

### 3.4 Betraktninger om denne oppgaven

De foregående studiene er med på å underbygge at forhold på arbeidsmarkedet har effekt på uføretilbøyelighet og at arbeidsledighet er en nær substitutt til uføretrygd. I motsetning til tidligere studier har jeg ikke det samme datagrunnlaget, da jeg verken har tilgang på individdata eller data spesifikt om nedbemanninger. Dermed vil denne oppgaven skille seg fra tidligere studier rundt arbeidsledighet og uføretrygd ved at datagrunnlaget er mer aggregert og ved at jeg ønsker å se effekten av arbeidsledighet generelt. Med andre ord ser jeg altså ikke på spesifikke hendelser som påvirker arbeidsledighet/fører til arbeidsledighet for en gruppe slik som i tidligere studier, men på faktisk arbeidsledighet. Videre kan jeg kontrollere for mange av variablene som har vist seg å være viktige, som blant annet utdanningsnivå og skilsmisser. Dette presenteres nærmere i kapittel 4. Teorien om at det eksisterer en "ufør-kultur" kan også til dels underbygges av studien til Bratsberg m.fl.(2012), og at det muligens eksisterer en slik kultur innad i fylker med med en høy andel uføre. Dette er det derimot ikke mulig for meg å kontrollere for, men det kan tenkes at dette er en del av grunnlaget for relativt stabile forskjeller i andelen uføre mellom fylkene.





## 4 Datamaterialet

I dette kapitlet presenterer jeg datagrunnlaget for analysen i oppgaven. I delkapittel 4.1 presenterer jeg den avhengige variabelen, i delkapittel 4.2 og 4.3 presenterer jeg henholdsvis fokusvariabelen og kontrollvariablene som inngår i analysen. Tilslutt presenterer jeg den deskriptive analysen av data i delkapittel 4.4.

Jeg har konstruert et balansert paneldatasett for de 19 norske fylkene for en tidsperiode på 15 år, fra og med 2000 til og med 2014. At jeg har et balansert paneldatasett vil si at jeg har observasjoner for alle fylkene for alle de 15 årene. Data er hentet fra Norsk samfunnsvitenskapelig datatjeneste (NSD) sin kommunedatabase, Statistikkbanken SSB og NAV.<sup>12</sup> I utgangspunktet bruker jeg datasettet for hele tidsperioden i analysen, men dersom jeg ønsker å kontrollere for næringsstruktur er ikke dette mulig. I 2008 endret SSB sine næringskoder fra SN2002 til SN2007.<sup>13</sup> Denne endringen medfører at data for næringsstruktur i fylkene ikke kan brukes samlet før og etter 2008. Jeg har derfor valgt å dele datasettet inn i to separate tidsperioder som kan benyttes når jeg inkluderer kontrollvariable for næringsstruktur i fylkene. For hovedanalysen benyttes datasettet for hele tidsperioden da det kun er variabler for næringsstruktur som skiller de to separate periodene. Jeg har da datasett for tre ulike perioder som kan brukes i analysen.

Tabell 1: Tidsperioder

Hele perioden	2000 - 2014
Delperiode 1	2000 - 2008
Delperiode 2	2008 - 2014

Jeg har beregnet de fleste variablene som andel av hele befolkningen i fylke  $i$  for hvert år  $t$ . For de variable dette ikke gjelder vil jeg bemerke det spesifikt. Det kan argumenteres for at det ville være mer hensiktsmessig å benytte data for befolkningen mellom 18-67 år da denne aldersgruppen er kvalifiserte for å motta uføretrygd, men jeg har ikke tilgang på data for mange av de øvrige variablene mine i denne aldersgruppen. Dermed er den avhengige variabelen, i tillegg til de øvrige variablene, gitt som andel av hele befolkningen i fylket for hvert år  $t$ .

<sup>12</sup>Data er tilgjengelig med tillatelse fra NSD. Verken NSD, SSB eller NAV er ansvarlige for analysen av data eller for de tolkingene som er gjort i denne oppgaven.

<sup>13</sup>SSB (2007)

### 4.1 Avhengig variabel

Den avhengige variabelen er det vi i analysen ønsker å se effekten *på* av de andre variablene i modellen. Den er dermed den endogene variabelen i modellen fordi den bestemmes av modellen. Den avhengige variabelen i analysen min er andelen uføre. Data er hentet fra NAV som antall av hele befolkningen i fylket og videre beregnet som andelen av hele befolkningen i fylket for hvert år. Dette kan jeg presentere mer formelt som:

$$\frac{\text{Antall\_uføre}_{it}}{\text{Hele\_befolkningen}_{it}} = \text{Andel\_Uføre}_{it}$$

Der underskriftene *i* og *t* gir henholdsvis fylke og år og vil gjelde for alle variable i data-settet. Som vi har sett tidligere er det betydelig variasjon i andelen uføre mellom fylkene, men ikke veldig store endringer i denne variasjonen over tid. Dette diskuteres nærmere i den deskriptive analysen i delkapittel 4.4.

### 4.2 Fokusvariabel

Fokusvariabelen er den variabelen i analysen som vi ønsker å se effekten *av* på den avhengige variabelen. Denne inngår altså som en eksogen forklaringsvariabel som bestemmes utenfor modellen. Fokusvariabelen i analysen min vil være andelen helt arbeidsledige. Jeg bruker tall for registrerte ledige og ikke fra SSBs utvalgsbaserte arbeidskraftsundersøkelser (AKU), fordi statistikken over registrerte ledige kan brytes ned på fylkesnivå. Denne er beregnet som et årsgjennomsnitt for hvert år og er hentet fra NAV. Videre har jeg beregnet den som andelen av hele befolkningen i fylke *i*, år *t*.

$$\frac{\text{Ledig\_gjennomsnitt}_{it}}{\text{Hele\_befolkningen}_{it}} = \text{Ledig}_{it}$$

Jeg har valgt denne variabelen som fokusvariabel fordi dette er en variabel som gir et godt mål på forholdet på arbeidsmarkedet. Vi har også sett fra figur 5 i appendiks at arbeidsledigheten på fylkesnivå har vist mye variasjon i tidsperioden 2000-20014.

### 4.3 Kontrollvariable

I tillegg til den avhengige variabelen og fokusvariabelen inkluderer jeg et sett med kontrollvariable i analysen som jeg presenterer i de følgende delkapitlene. Kontrollvariablene er inkludert som eksogene forklaringsvariable i analysen, altså de bestemmes utenfor mo-

dellen. De inkluderes ikke nødvendigvis fordi vi ønsker å se effekten av variablene på den avhengige variabelen, men fordi de er med på å bestemme andelen uføre. Da de påvirker den avhengige variabelen er det fornuftig å kontrollere for disse i analysen. Videre er det spesielt viktig å inkludere kontrollvariable som i tillegg til å påvirke andelen uføre også er korrelert med fokusvariabelen for arbeidsledighet. Hvis slike variable blir utelatt, vil det være et problem med utelatt-variabel-skjevhet i analysen. Dette diskuteres nærmere i delkapittel 4.4. Samtidig kan det være interessant å se effektene av kontrollvariablene selv om det ikke er hovedfokuset i analysen. I delkapittel 4.3.1 presenterer jeg kontrollvariablene som er inkludert i hovedmodellen og videre presenterer jeg de øvrige kontrollvariabelen i de resterende delkapitlene.

#### 4.3.1 Kontrollvariable inkludert i hovedmodell

Kontrollvariablene jeg inkluderer i hovedmodellen er valgt med bakgrunn i tidligere litteratur og hva jeg mener er mest hensiktsmessig. Jeg har valgt de ”fylkeskarakteristikkene” som det er viktigst å kontrollere for da de er med på å bestemme andelen uføre. Alle variablene er som tidligere beregnet som andel av hele befolkningen, bortsett fra variablene for inntekt og bruttoprodukt. Jeg har valgt å beregne bruttoprodukt som andel av landstotalen fordi jeg da får et mål på hvor stor andel av landets verdiskapning som skapes i hvert fylke for hvert år. Mer formelt er variabel for bruttoprodukt beregnet:

$$\frac{\text{Bruttoprodukt\_fylke}_{it}}{\text{Bruttoprodukt\_landet}_t} = \text{Brutto}_{it}$$

Videre er inntekt gitt som gjennomsnittlig bruttoinntekt i fylket for hvert år. Jeg har valgt å ikke beregne denne som andel av befolkningen fordi jeg mener det ikke er hensiktsmessig. For de andre variablene, som for eksempel arbeidsledighet, gir ikke antallet arbeidsledige så mye informasjon hvis man ikke ser på det i sammenheng med befolkningsantall. Dette mener jeg derimot at ikke gjelder for inntekt. Samtlige variable som er inkludert i hovedmodellen, presenteres i tabell 2. Variabel for sykefravær i fylkene inkluderes, da personer som blir uføre blant annet har vært sykemeldt tidligere. Videre inkluderer jeg bruttoprodukt som et mål på verdiskapning i fylket som kan gi en indikator for arbeidsmarkedet. Jeg inkluderer videre inntekt i fylket, selv om Bragstad og Hauge (2008) kun fant en effekt av inntekt for kvinner. Fevang og Røed (2006) finner at både skilsmisse/separasjon og det å miste ektefellen øker sannsynligheten for uførepensjonering, og dermed vil det være hensiktsmessig å kontrollere for andel skilsmisser i fylket. Samme studie i tillegg til Bratsberg et al. (2010) finner også at personer med lavt utdanningsnivå er mer utsatte

#### 4. DATAMATERIALET

---

for å bli uføre enn personer med høy utdanning og at økt utdanningsnivå tilsier lavere uføretilbøyelighet, dermed er andelen med lav utdanningsnivå i fylkene med på å bestemme andelen uføre.

Tabell 2: Hovedvariable

Variabel	Beskrivelse
Sykefrav	Sykefravær i fylket (Kilde: SSB).
Bruttoprod	Bruttoprodukt for fylket som andel av landstotal. <b>Mål på verdiskapning i fylket.</b> (Kilde: SSB)
Innt.Br	Gjennomsnittlig bruttoinntekt i fylket. (Kilde: SSB)
Skilsm	Skilsmisser i fylket (Kilde: NSD)
UtdL	Andel med høyeste oppnådde utdanning på grunnskole- eller videregående nivå (Kilde: NSD).

#### 4.3.2 Næringsstruktur

Som tidligere nevnt har jeg muligheten til å kontrollere for næringsstruktur i fylkene. Jeg har tilgang på data fordelt på aggregerte hovedgrupper for næringer, samt på et mer detaljert nivå. Etter egen vurdering finner jeg at det er mest hensiktsmessig å benytte data for de aggregerte hovedgruppene. Dette er fordi jeg mener data for de mer detaljerte næringsgruppene blir overflødig på fylkesnivå, da mange av gruppene har svært få eller tilnærmet ingen observasjoner for mange av fylkene. Videre er det mer hensiktsmessig for effektiviteten og enkelheten av analysen å ikke ha 100 variable for sysselsetting innad i fylkene. Som nevnt tidligere har jeg ikke muligheten til å bruke datasettet samlet for hele perioden når jeg kontrollerer for næringsstruktur. Dermed er settet av kontrollvariable for næringsstruktur delt inn i to perioder.

I data for periode 1 som er under SN2002-koden er det 11 aggregerte hovedgrupper, mens for periode 2 som er under SN2007-koden er det hele 17 aggregerte hovedgrupper. Oversikt over hovedgruppene for både SN2002 og SN2007 finnes i tabell 15 i appendiks. Selv om jeg ikke har mulighet til å sette sammen de to periodene til et datasett, ønsker jeg å behandle dem mest mulig likt. Jeg har dermed, så godt det har latt seg gjøre, aggregert opp variablene slik at jeg har 10 hovedgrupper for begge periodene. Jeg gjorde selvfølgelig et forsøk på å sette sammen de to periodene etter å ha utført denne aggregeringen, men dette lot seg ikke gjøre da det ble for stor forskjell på verdiene. Jeg valgte likevel å beholde de aggregerte variablene i de to periodene for å få mest mulig likhet mellom datasettene for de to periodene. Videre gjør dette data mer håndterlig og jeg ser på det som mest hensiktsmessig for analysen. Jeg regner med at disse variablene ikke endrer

seg mye over tid, ettersom næringsstrukturen i et fylke er ganske ”satt” og ikke endrer seg særlig på kort sikt. Derimot observeres det variasjon mellom fylkene, ettersom det er ulike næringsstrukturer mellom fylkene. I tabell 3 presenteres de ti ulike variablene for henholdsvis periode 1 og periode 2 som karakteriserer næringsstrukturen i fylkene.

Tabell 3: Variable etter næringskode - Periode 1 og 2

Variabelnavn	Variabel Periode 1 / Periode 2	SN2002	SN2007
JSF1	Jordbruk, skogbruk og fiske	01-05	01-03
IUB2	Industri, utvinning og bergverksdrift	10-11,12-37	05-09,10-33
KV3/EVR3	Kraft- og vannforsyning / Elektrisitet, vann og renovasjon	40-41	35-39
BA4	Bygge- og anleggsvirksomhet	45	41-43
VHR5	Varehandel, hotell og restaurantvirksomhet	50-55	45-47,55-56
TK6/TLIK6	Transport og kommunikasjon / Transport og lagring, Informasjon og kommunikasjon	60-64	49-53,58-63
FT7/FF7	Finansiell tjenesteyting / Finansiering og forsikring	65-67	64-66
FTE8/FTTE8	Forretningsmessig tjenesteyting, eiendomsdrift / Forretningsmessig- og teknisk tjenesteyting, eiendomsdrift	70-74	68-75,77-82
OFT9	Offentlig forvaltning og annen tjenesteyting	75-99	84-88,90-99
U10	Uoppgitt	00	00

Kilde for alle variable er statistikkbanken til SSB.

### 4.3.3 Fylkets attraktivitet

Bragstad og Hauge (2008) m.fl finner at en høyere nettoinnflytting til en kommune gir en lavere uføretilgang. Jeg velger derfor å inkludere nettoinnflytting i fylkene. Videre mener jeg det er hensiktsmessig å inkludere nettoinnpendling til fylkene forbindelse med arbeid som kan gi et mål på attraktiviteten til arbeidsmarkedet i fylket. Nettoinnflytting kan gi et mål på attraktiviteten til fylket som helhet, som åpenbart også inkluderer arbeidsmarkedet. Attraktiviteten av arbeidsmarkedet er et hensyn som må tas før man velger å bosette seg et sted. Variablene presenteres i tabell 4.

Tabell 4: Attraktivitet

Variabel	Beskrivelse
Nettoinp	Nettoinnpendling til fylket i forbindelse med arbeid. <b>Mål på attraktivitet av arbeidsmarkedet i fylket.</b> (Kilde: NSD)
Nettoinnfly	Nettoinnflytting til fylket. <b>Mål på attraktivitet av fylket.</b> (Kilde: SSB)

#### 4.3.4 Høyere utdanning

Fra tidligere litteratur og som tidligere nevnt vet jeg at utdanningsnivå påvirker uføretilbøyligheten til individer.<sup>14</sup> Jeg mener derfor det er hensiktsmessig å kontrollere for utdanningsnivå i fylkene, og da ”utdanningsfordelingen” i fylkene. I tillegg til andel med lavt utdanningsnivå som er inkludert i hovedmodellen vil det også være hensiktsmessig å senere i analysen inkludere variable for andel med høyere utdanning. Dette fordi jeg da kan få et mål på utdanningssammensettingen i fylket. I tabell 5 presenteres variablene. Det er to mulige nivå, og data er gitt slik at høyeste oppnådde utdanning en person har bestemmer nivå.

Tabell 5: Høyere utdanning

Variabel	Beskrivelse
UtdHkort	Andel med høyeste oppnådde utdanning bachelorgrad. (Kilde NSD)
UtdHlang	Andel med høyeste oppnådde utdanning master- eller Phd-grad. (Kilde: NSD)

Som tidligere variable er disse variabelene har jeg hentet ut i antall hele befolkningen og beregnet som andel av befolkningen i fylket i år t.

#### 4.3.5 Innvandring

Fra tidligere litteratur vet vi også at innvandringsstatus har betydning for uføretilbøyligheten til en person. Som jeg har presentert har det blitt funnet at blant innvandrere skiller personer fra Nord-Afrika og Midtøsten seg ut som mer utsatte enn andre, mens personer med vestlig og øst-asiatisk opphav er mindre utsatt for å bli uføre enn andre.<sup>15</sup>

Tabell 6: Innvandring

Variabel	Beskrivelse
Eur. u/T	Andel innvandrere i fylket fra Europa, uten Tyrkia. (Kilde:SSB)
Af.	Andel innvandrere i fylket fra Afrika (Kilde: SSB)
Asia m/T	Andel innvandrere i fylket fra Asia, med Tyrkia. (Kilde: SSB)
N-Am.	Andel innvandrere i fylket fra Nord-Amerika (Kilde: SSB)
S/M Am.	Andel innvandrere i fylket fra Mellom- og Sør-Amerika (Kilde: SSB)
Oz.	Andel innvandrere fra Oceania (Australia og NZ) (Kilde: SSB)

Jeg har ikke tilgang til datamaterialet på et mer detaljert nivå enn verdensdelene og jeg kan dermed ikke spesifikt kontrollere for Nord-Afrika og Midtøsten. Jeg mener likevel

---

<sup>14</sup>Fevang og Røed (2006), Bratsberg et al. (2010)

<sup>15</sup>Fevang og Røed (2006)

det kan være hensiktsmessig å kontrollere for andelen innvandrere. Som tidligere er disse variablene hentet ut i antall og videre beregnet som andel av hele befolkningen i fylket i år t. Variablene presenteres i tabell 6.

#### 4.3.6 Politisk tilhørighet

Jeg har valgt å inkludere variable for å kartlegge partitilhørighet i fylkene ved fylkesvalg. Jeg har kun valgt fire politiske parti fordi jeg mener disse representerer hver sin side av politikken, høyre- og venstresiden, på en god måte. Høyre og Arbeiderpartiet (Ap) mener jeg det er en selvfølge å inkludere da de er de største politiske partiene. Disse partiene kan begge ansees som sentrumsnære, men skiller fremdeles på hver sin side med Høyre på høyresiden og Ap på venstresiden. Videre har jeg inkludert Sosialistisk Venstreparti (SV) og Fremskrittspartiet (Frp) som mer ”ekstreme” mål da de representerer ytterpunktene av høyre og venstresiden på en god måte. selv om det kan argumenteres at partiet Rødt også kan være et godt mål for et slikt ytterpunkt og kan betraktes som mer ytterliggående enn SV på venstresiden, mener jeg partiet er for lite til å ha betydning i denne analysen. Jeg mener derfor SV er et bedre mål på den ytterligliggende venstresiden i politikken. I tabell 7 følger en oversikt over valg-variable:

Tabell 7: Politisk tilhørighet

Variabel	Beskrivelse
StemSVF.	Andel godkjente stemmer i fylket til SV i fylkesvalg. (Sosialistisk Venstreparti) (Kilde: NSD)
StemAPF.	Andel godkjente stemmer i fylket til AP i fylkesvalg. (Det Norske Arbeiderparti) (Kilde: NSD)
StemHF.	Andel godkjente stemmer i fylket til H i fylkesvalg. (Høyre) (Kilde: NSD)
StemFrpF.	Andel godkjente stemmer i fylket til Frp i fylkesvalg. (Fremskrittspartiet) (Kilde: NSD)

Til en viss grad mener jeg at det å inkludere variable for stemmefordeling kan fungere som et mål på den ”politiske kulturen” i et fylke. Dersom et fylke veier tyngre på høyre- eller venstresiden, kan dette gjenspeile innbyggerenes prioriteringer og meninger. Tradisjonelt sett har partier på venstresiden av politikken jobbet for en sterk stat og brede velferdsordninger. Det kan dermed tenkes at dersom fylker veier tyngre på venstresiden av politikken vil det være flere som mottar uføretrygd. Videre har tidligere litteratur vist at det er en ”spillover”-effekt mellom foreldre og barn når det gjelder uføretrygd og det kan derfor tenkes at det kan eksistere såkalte ”ufør-kulturer”.<sup>16</sup> Dette har jeg som sagt

<sup>16</sup>Bratberg et al. (2012)

ikke muligheten til å kontrollere for med mitt datagrunnlag (og det er generelt vanskelig å kontrollere for), men jeg mener at det beste forsøket jeg kan gjøre er å kartlegge den ”politiske kulturen” i fylkene.

### 4.4 Deskriptiv analyse

I dette delkapittelet vil jeg presentere deskriptiv analyse for variablene i hovedmodellen. I delkapittel 4.4.1 presenterer jeg og diskuterer variasjon og andre utvalgskarakteristikker av variablene. I delkapittel 4.4.2 ser jeg på korrelasjon og undersøker etter mulig multi-kollinearitet. Tilsvarende analyse for øvrige kontrollvariable finnes i appendiks.

#### 4.4.1 Utvalgskarakteristikker og variasjon

Jeg begynner den deskriptive analysen med å undersøke variasjonen i variablene, i tabell 8. Siden jeg bruker et paneldatasett er dette viktig. I paneldata har variablene variasjon mellom tverrsnittsenhetene (between groups) og over tid (within groups). For mitt datasett har vi altså variasjon mellom fylkene (between) og over årene i tidsperioden (within). Vi kan også observere total variasjon (overall). Videre rapporteres standardavvik for variablene, som forteller oss om spredningen i utvalget fra gjennomsnittet. Tilslutt rapporteres gjennomsnitt for koeffisientene, samt minimums- og maksimumsverdier. Dette gir oss en god oversikt over de ulike variablene. Ettersom datasettet består av mange variable kan tabell med verdier for alle variable finnes i appendiks.<sup>17</sup>

Fra tabell 8 ser jeg at den gjennomsnittlige verdien til variabelen for andelen uføre er på rundt 6.7 prosent. Videre er between-variasjon (mellom fylker) målt ved standardavvik på rundt 1.3 prosentpoeng, within-variasjonen (over tid) målt ved standardavvik på rundt 0.3 prosentpoeng og total-variasjon på 1.3 prosentpoeng. Within-variasjon utgjør rundt 4 prosent av variasjonen og dermed utgjør between-variasjon 96 prosent. Altså er det stor variasjon mellom fylkene, men denne variasjonen er svært lav over tid. Videre finner jeg at den gjennomsnittlige verdien til andelen arbeidsledige er 1.5 prosent. Her finner jeg at within-variasjon utgjør 60 prosent av variasjonen. Det vil si at andelen arbeidsledige endres relativt mye over tid, i tillegg til å variere mellom fylkene med en between-variasjon på følgelig 40 prosent. Videre for de øvrige kontrollvariablene har sykefravær et gjennomsnitt på 2.7 prosent der within-variasjonen utgjør 37 prosent av variasjonen. Bruttoprodukt har en gjennomsnittsverdi på 5.3 prosent, altså 5.3 prosent av den totale

---

<sup>17</sup>Tabell 18-22



verdiskapningen i landet. Her ser jeg at det er ganske stor forskjell mellom minimums- og maksimums-verdiene. Dette vil tilsa at mens noen fylker utgjør store deler av landets verdiskapning, utgjør andre fylker små deler. Her utgjør within-variasjonen kun 0.4 prosent av den totale variasjonen, altså er denne tilnærmet konstant over tid. Videre finner jeg at gjennomsnittlig verdi for inntekt er 311329 kroner, dvs gjennomsnittsinntekten i fylkene er på 311 329 kroner. Her finner jeg en within-variasjon på hele 78 prosent av total variasjon, altså varierer denne mye over tid. Gjennomsnittsverdien på andelen skilsmisser er lav med 0.2 prosent, og varierer en del over tid med en within-variasjon på 32 prosent av total variasjon. Tilslutt er det gjennomsnittlig 60 prosent av befolkningen i fylket som har høyeste oppnådde utdanning på grunnskole- eller videregående skolenivå. Within-variasjon utgjør 13 prosent av total variasjon og denne endrer seg reelt lite over tid.

Tabell 8: Utvalgskarakteristikker og variasjon

Variabel	Variasjon	Gjennomsnitt	Standardavvik	Min	Max
<b>And.U</b>	Overall	0.0665325	0.0130254	0.0380261	0.0863471
	Between		0.0130654	0.0467735	0.080999
	Within		0.0027153	0.0574269	0.076001
<b>Ledgj</b>	Overall	0.0148126	0.0043751	0.0056965	0.0306064
	Between		0.0028395	0.0103159	0.0224255
	Within		0.0033877	0.0074895	0.0233439
<b>Sykefrav</b>	Overall	0.0265059	0.0028063	0.0173127	0.0325461
	Between		0.002286	0.0210024	0.0294162
	Within		0.001705	0.0204934	0.0298659
<b>Brutto</b>	Overall	0.0526281	0.0469385	0.0113083	0.2299532
	Between		0.0480381	0.0119893	0.2152826
	Within		0.0030532	0.0381813	0.0672987
<b>Inntbr</b>	Overall	311329.1	63066.77	196600	481000
	Between		30089.21	30089.21	378420
	Within		55827.3	201795.8	436795.8
<b>Skilsm</b>	Overall	0.0021429	0.0003651	0.0012298	0.0029706
	Between		0.0003093	0.0013879	0.0026329
	Within		0.0002058	0.0014498	0.0026401
<b>UtdL</b>	Overall	0.5960377	0.0476108	0.4051984	0.6735435
	Between		0.0455077	0.4563549	0.6470302
	Within		0.0172606	0.5448812	0.6481928

De variable som viser liten tidsvariasjon er faktorer som ikke endrer seg mye på kort sikt. For eksempel vil ikke bruttoproduktet til et fylke endres på kort sikt da det er høyst usannsynlig at et fylke plutselig får sterkt økt verdiskapning.

#### 4.4.2 Korrelasjon og mulig multikollinearitet

Jeg har et stort datasett med mange variable jeg har muligheten til å inkludere. Jeg velger her å undersøke korrelasjonen mellom variablene inkludert i hovedmodellen og herfra un-

#### 4. DATAMATERIALET

---

dersøke muligheten for multikollinearitet. Vi kan se på korrelasjonen i korrelasjonsmatrisen gitt av tabell 9. Jeg begynner med å se på korrelasjonen mellom den avhengige variabelen og de øvrige inkluderte variablene. Som forventet er ledighet og sykefravær positivt korrelert med andelen uføre, i likhet med skilsmisser og andel med lavt utdanningsnivå. Derimot er inntekt og bruttoprodukt negativt korrelert med andelen uføre. Dette er som forventet. Disse verdiene gir meg ingen resultater vi kan ta med oss videre, men er en interessant indikasjon.

Tabell 9: Korrelasjonsmatrise

	Andel_U	Ledig	Sykefr	Brutto	Inntbr	Skilsm	UtdL
Andel_U	1.000						
Ledig	0.2766	1.000					
Sykefr	0.4856	-0.1789	1.000				
Brutto	-0.6661	0.0671	-0.3736	1.000			
Inntbr	-0.3697	-0.2504	-0.0826	0.3949	1.000		
Skilsm	0.2286	0.3420	0.0469	0.2149	-0.1423	1.000	
UtdL	0.6327	-0.0409	0.3308	-0.8207	-0.6804	-0.0833	1.0000

Jeg fokuserer så på korrelasjon mellom forklarings-/kontrollvariablene. Grunnen til at jeg undersøker korrelasjon mellom forklaringsvariable er at høye korrelasjoner kan være en indikator for mulig multikollinearitet, som det er viktig å avdekke. Multikollinearitet er graden av lineær sammenheng mellom flere forklaringsvariabler i en multippel regresjonsmodell, og dersom vi har *perfekt* multikollinearitet vil dette bryte med Gauss-Markov teoremet. Derimot vil dette ikke gjelde ved sterk multikollinearitet, men det vil kunne by på utfordringer i analysene og kan resultere i feil estimert signifikans av en variabel grunnet en ”blåst opp” varians. For å undersøke dette bruker jeg variansinflasjonsfaktoren, som vi på kort kan kalle VIF. Denne kvantifiserer hvor mye variansene av de estimerte koeffisientene er ”blåst opp”. Dermed forteller VIF-verdiene oss til hvilken grad variansen er ”blåst opp”. Formelt finner vi VIF-verdien som:

$$VIF_i = \frac{1}{1 - R_i^2}$$

Den generelle tommelfingerregelen her er at dersom vi har en VIF verdi på 10 eller over er dette et tegn på alvorlig multikollinearitet. Det er viktig å huske på at dette bare er en tommelfingerregel, men jeg velger å forholde meg til den her. I tabell 10 presenteres VIF-verdiene for variablene inkludert i hovedmodellen. Jeg kan se her at ingen av variablene har en VIF verdi som overstiger 10, men andelen med lav utdanning har en verdi på 6.16 som er relativt høyt. Etter å ha undersøkt VIF-verdier for hovedvariabelen konkluderer

jeg med at jeg kan inkludere alle hovedvariablene i analysen. Derimot har jeg avdekket høye VIF-verdier for andre kontrollvariable (VIF) som indikerer at jeg har et multikollinearitetsproblem i datasettet mitt. Dette kan ses i tabell 16 i appendiks. Det viktigste er at jeg er klar over dette og kan ta hensyn til det når jeg tolker resultater fra analysen. Tegn og konsekvenser kan være at estimerte koeffisienter skifter fortegn eller har store endringer i verdi ved inkludering av variable.

Tabell 10: VIF-verdier

Variabel	VIF
UtdL	6.16
Brutto	3.94
Inntbr	2.63
Ledig	1.34
Skillsm	1.28
Sykefr	1.27
Gjennomsnittlig VIF	2.77

#### 4.4.3 Utelatt-variabel-problem

Til slutt retter jeg oppmerksomheten tilbake til korrelasjonsmatrisen gitt i tabell 9 og undersøker korrelasjonen mellom fokusvariabelen for arbeidsledighet og de øvrige inkluderte kontrollvariablene. Som nevnt tidligere er det viktig å inkludere variable som i tillegg til å påvirke andelen uføre også er korrelert med fokusvariabelen for arbeidsledighet. Dersom slike variable utelates vil jeg potensielt ha et utelatt variabel-problem i analysen som vil gi utelatt variabel-skjevhet. Denne skjevheten kommer av at modellen tillegger effekten av de utelatte variablene til de estimerte effektene av de inkluderte variablene. Dermed vil de estimerte koeffisientene rapportere høyere eller lavere effekter enn den sanne effekten fordi estimatoren er korrelert med restleddet. Av definisjon fanger restleddet opp de utelatte variablene. Jeg kan se fra tabell 9 at samtlige variable er korrelert med arbeidsledighet, og da spesielt inntekt, skilsmisser og sykefravær. Denne korrelasjonen er en indikator for at disse variablene skal inkluderes i modellen. Ettersom hovedmodellen min er en 'naiv' modell som skal tillegges ytterligere kontrollvariable, kan det tenkes at hovedmodellen vil ha et problem med utelatt variabel skjevhet. Dette kan jeg ta hensyn til og redusere gjennom sensitivitetsanalysen, der jeg tillegger flere kontrollvariable på modellen, da det kan tenkes at hovedmodellen er underspesifisert. I tabell 18 i appendiks rapporteres korrelasjonene mellom fokusvariabelen og de øvrige kontrollvariablene. I forhold til tidligere studier er mitt datamateriale mer aggregert og begrenset. Det er derfor faktorer som det ikke er mulig å inkludere, som for eksempel tilstedeværelsen av en "ufør-kultur" i fylkene. Dermed vil jeg potensielt ha et utelatt variabel-problem, selv

#### 4. DATAMATERIALET

---

med de øvrige kontrollvariablene.

Denne gjennomgangen av datamaterialet, samt den deskriptive analysen, er viktig i lys av neste kapittel der jeg presenterer de ulike modellspesifikasjonene samt de økonometriske metodene som benyttes i analysen.

## 5 Metode

I denne oppgaven bruker jeg et balansert paneldatasett av de 19 norske fylkene for tidsperioden 2000-2014. Jeg begynner med å presentere modellspesifikasjonen min for denne oppgaven i delkapittel 5.1. selv om paneldata har mange spennende egenskaper kan estimering av paneldatamodeller by på utfordringer. I dette kapitlet vil jeg videre gå igjennom problemer ved estimering av paneldatamodeller og hvordan disse kan håndteres med ulike metoder. I delkapittel 5.2 presenterer jeg estimeringsmetoder for statiske paneldatamodeller og i delkapittel 5.3 presenterer jeg estimeringsmetoder for dynamiske paneldatamodeller. For denne delen av kapitlet tar jeg utgangspunkt i Verbeek (2012)<sup>18</sup> og Wooldridge (2013)<sup>19</sup>.

### 5.1 Modellspesifikasjon

Jeg vil nå presentere de ulike modellspesifikasjonen jeg skal benytte meg av i analysen. Alle modellene er på på formen log-log som illustreres ved at  $\ln$  er inkludert i variabelnavnene. Dette vil si at variablene er log-transformerte og koeffisientene  $\beta_j$ ,  $j = 1, \dots, 6$  tolkes som elastisiteter som sier hvor mange prosent andelen uføre endres med når hver forklaringsvariabel øker med én prosent. Jeg begynner analysen med hovedmodellen min som er spesifisert som:

$$\begin{aligned} \ln \text{Andel\_Uit} = & \beta_0 + \beta_1 \ln \text{Ledig}_{it} + \beta_2 \ln \text{Sykefr}_{it} + \beta_3 \ln \text{Brutto}_{it} \\ & + \beta_4 \ln \text{Inntbr}_{it} + \beta_5 \ln \text{Skillsm}_{it} + \beta_6 \ln \text{UtdL}_{it} + \alpha_i + \eta_{it} \end{aligned} \quad (1)$$

Den avhengige variabelen er, som tidligere spesifisert, andelen uføre. Videre er fokusvariabelen for arbeidsledighet inkludert som forklaringsvariabel. Tilslutt er de øvrige kontrollvariablene sykefravær, bruttoprodukt, inntekt, skilsmisser og andelen med lav utdanning inkludert ettersom de er med på å bestemme andelen uføre. Alle variablene er gitt i samme tidsperiode og dermed vil en endring i en forklaringsvariabel i dag med én prosent gi endringen i andelen uføre i dag. Med andre ord ser vi hele tiden på modellen i likevektstilstand.

Videre utvider jeg modellen ved å inkludere laggede verdier for én og to perioder for

<sup>18</sup>Kapittel 10

<sup>19</sup>Kapittel 13 og 14

arbeidsledighet og sykefravær som forklaringsvariable i modellen. Det vil si at ledighet lagget med én periode gir ledigheten i fjor, mens andelen uføre fremdeles er gitt i dag. Dette velger jeg å gjøre fordi det kan tenkes at tilpasningen mellom ledighet og sykefravær til uførhet ikke nødvendigvis skjer i samme periode, altså vil det være en tregere tilpasning mellom variablene. Modellen i dette tilfellet er gitt av (2):

$$\begin{aligned} \ln Andel\_Uit = & \beta_0 + \beta_1 \ln Ledig_{it} + \beta_2 \ln Ledig_{it-1} + \beta_3 \ln Ledig_{it-2} + \beta_4 \ln Sykefr_{it} \\ & + \beta_5 \ln Sykefr_{it-1} + \beta_6 \ln Sykefr_{it-2} + \beta_7 \ln Brutto_{it} + \beta_8 \ln Inntbr_{it} \\ & + \beta_9 \ln Skillsm_{it} + \beta_{10} \ln UtdL_{it} + \alpha_i + \eta_{it} \end{aligned} \quad (2)$$

De øvrige variablene fra hovedmodellen er inkludert som tidligere, men det nye elementet er de laggede verdiene av ledighet og sykefravær. Parametrene tolkes fremdeles som elastisiteter, men tolkningen av de laggede variablene er litt annerledes siden modellen ikke lengre ses på i en likevektstilstand. Ledighet lagget med to én periode tolkes slik at én prosents økning i ledighet i fjor gir en endring i andelen uføre i dag. Tilpasningen skjer altså ikke momentant.

Videre er det hensiktsmessig å inkludere flere kontrollvariable i modellen i lys av den tidligere diskusjonen rundt utelatt variabel-problem og påfølgende skjevhet. Jeg kan også undersøke robustheten til resultatene fra analysen ved å inkludere øvrige kontrollvariable i modellen. Som presentert i kapittel 4 har jeg tilgang på mange kontrollvariable i datasettet og disse inkluderes gruppevis i analysen. Modellen i dette tilfellet er gitt av (3) og fremstår slik den gjør i analysen. Det vil si at jeg har ekskludert de variablene som samlet sett ikke har noen signifikant effekt i modellen. Jeg inkluderer kontrollvariable som gir mål på attraktiviteten til fylket, utdanningssammensetting, innvandring og stemmefordeling som her betegnes som "videre kontrollvariable".

$$\begin{aligned} \ln Andel\_Uit = & \beta_0 + \beta_1 \ln Ledig_{it-1} + \beta_2 \ln Sykefr_{it-1} + \beta_3 \ln Sykefr_{it-2} + \beta_4 \ln Brutto_{it} \\ & + \beta_5 \ln Inntbr_{it} + \beta_6 \ln Skillsm_{it} + \beta_7 \ln UtdL_{it} + \text{Videre kontrollvariable} + \alpha_i + \eta_{it} \end{aligned} \quad (3)$$

For å videre utnytte de dynamiske egenskapene til paneldatasettet kan jeg undersøke treghet i tilpasning av den avhengige variabelen ved å inkludere lagget endogen variabel som høyresidevariabel i modellen. Grunnen for å gjøre dette er at det kan tenkes at andel

uføre tidligere år har en påvirkning på andelen uføre i dag, altså at det er persistens i andelen uføre. Modellen i dette tilfellet er gitt av (4):

$$\begin{aligned} \ln Andel\_U_{it} = & \beta_0 + \beta_1 \ln Andel\_U_{it-1} + \beta_2 \ln Ledig_{it-1} + \beta_3 \ln Sykefr_{it-1} \\ & + \beta_4 \ln Sykefr_{it-2} + \beta_5 \ln Brutto_{it} + \beta_6 \ln Inntbr_{it} + \beta_7 \ln Skillsm_{it} \\ & + \beta_8 \ln UtdL_{it} + \text{Videre kontrollvariable} + \alpha_i + \eta_{it} \end{aligned} \quad (4)$$

### 5.1.1 Korrigerer av estimerte standardavvik

Estimeringsmetodene jeg benytter meg av i analysen forutsetter at det ikke forekommer seriekorrelasjon eller heteroskedastisitet da dette vil føre til at de estimerte standardavvikene ikke blir korrekte. Ved fravær av seriekorrelasjon og heteroskedastisitet menes det at vi har en konstant varians over tid og individ. Det kan tenkes at standardavvikene innen fylker er korrelert på en ukjent måte. For å korrigere for dette velger jeg å benytte meg av cluster-korrigerede standardavvik i analysen og alle rapporterte standardavvik er korrigert for clustering.

## 5.2 Statiske paneldatamodeller

Jeg begynner med å ta utgangspunkt i den følgende enkle paneldatamodellen:

$$y_{it} = \beta_0 + \beta_1 x_{it} + u_{it} \quad (5)$$

Paneldata kjennetegnes ved at variablene varierer både over tid og mellom individer, som i mitt tilfelle vil si over 15 år og mellom 19 fylker. I den enkle modellen er  $y_{it}$  gitt som den avhengige variabelen i analysen som i mitt tilfelle er andelen uføre. Videre antar vi at  $x_{it}$  er en  $1 \times k$  linjevektor som består av aktuelle forklaringsvariable med den tilhørende koeffisientvektoren  $\beta_1$ . Restleddet  $u_{it}$  er sammensatt av to komponenter:

$$u_{it} = \alpha_i + \eta_{it} \quad (6)$$

der  $\eta_{it}$  er den idiosynkratiske restleddskomponenten som varierer både mellom individer og over tid (between og within). Komponentene vi er interessert i er  $\alpha_i$  som fanger opp uobserverbar heterogenitet som kan være med å påvirke den avhengige variabelen i modellen. Denne varierer kun mellom individer og er dermed konstant over tid. I vårt tilfelle

er den fylkesspesifikk. Kan formulere følgende forutsetninger om restleddet i modellen:

$$E(\eta_{it}|x_{i1}, x_{i2}, \dots, x_{iT}) = 0 \quad (7a)$$

$$E(\eta_{it}\eta_{js}|x_{it}) = \sigma_u^2 \quad \text{hvis } i = j \quad \text{og } t = s, \quad 0 \quad \text{ellers} \quad (7b)$$

$$E(\alpha_i\alpha_j|x_{i1}, x_{i2}, \dots, x_{iT}) = \sigma_\alpha^2 \quad \text{hvis } i = j, \quad 0 \quad \text{ellers} \quad (7c)$$

$$E(\eta_{it}\alpha_j) = 0 \quad \text{for alle } i, j, t \quad (7d)$$

$$E(\alpha_i|x_{i1}, x_{i2}, \dots, x_{iT}) = 0 \quad (7e)$$

$$E(\alpha_i|x_{i1}, x_{i2}, \dots, x_{iT}) \neq 0 \quad (7f)$$

Forutsetning (7a) sier at ingen av forklaringsvariablene systematisk avhenger av den idiosynkratiske restleddskomponenten over tid eller på tvers av individ. Fra forutsetning (7b) og (7c) ser vi at begge restleddskomponentene krever en konstant varians, altså at det er fravær av heteroskedastisitet (konstant varians over individ) og seriekorrelasjon (konstant varians over tid). Forutsetning (7d) sier at restleddskomponentene er ukorrelerte over tid og individ. Restleddsforutsetning (7e) sier at den enhetsspesifikke restleddskomponenten er ukorrelert med forklaringsvariabelen, og motsatt sier (7f) at dette ikke holder. Det er (7e) og (7f), altså om den fylkesspesifikke komponenten er korrelert med forklaringsvariabelen eller ikke, som vil være gjeldene for valg av estimeringsmetode.

### 5.2.1 Pooled OLS

OLS, som kalles Pooled OLS når vi bruker paneldata, byr ofte på problemer ved estimering av paneldatamodeller. Dersom vi ønsker å estimere (5) med OLS er det nødvendig at forutsetning (7e) holder, altså at en fylkesspesifikke restleddskomponenten er ukorrelert med de inkluderte forklaringsvariablene. Dersom denne ikke holder og vi istedet har at forutsetning (7f) gjelder, vil OLS på (5) gi forventningsskjeve og inkonsistente estimatorene ved at estimatoren  $\widehat{\beta}_1$  ikke konvergerer i sannsynlighet mot den sanne parameterverdien  $\beta_1$  når utvalget går mot uendelig.

$$plim\widehat{\beta}_1 \neq \beta_1 \quad (8)$$

Som oftest vil det være lite realistisk å anta at forutsetning (7e) holder og det vil mest sannsynlig være utelatte fylkesspesifikke variable som inngår i den fylkesspesifikke komponenten  $\alpha_i$ . At variablene er utelatte vil si at de egentlig skulle vært inkludert i (5), men



ikke er det grunnet at de ikke er observerbare for oss eller at vi av andre årsaker ikke har muligheten til å inkludere dem, som diskutert tidligere. Videre, ettersom vi i paneldata-modellen ser på de samme fylkene over flere år er det lite sannsynlig at restleddene er ukorrelerte over de 15 årene, altså vil vi ha et problem med seriekorrelasjon.

### 5.2.2 Fixed effects

Fixed effects-metoden (FE-metoden) lar meg håndtere problemet med eventuell uobserverbar heterogenitet ved at effekten av variable som ikke varierer i et fylke, såkalte fylkesfaste effekter, bli transformert bort. Det er viktig å merke at parametrene i modellen er identifisert kun basert på within-dimensjonen av datamaterialet og dermed vil ikke denne metoden utnytte all informasjon som er tilgjengelig i paneldatamaterialet. For å nærmere forklare metoden tar jeg utgangspunkt i den samme enkle paneldatamodellen gitt av (5):

$$y_{it} = \beta_0 + \beta_1 x_{it} + u_{it} \quad (5)$$

Som tidligere er restleddet  $u_{it}$  sammensatt av de to komponentene, den idiosynkratiske og den fylkesspesifikke:

$$u_{it} = \alpha_i + \eta_{it} \quad (6)$$

Vi antar her at forutsetning (7e) ikke holder, og dermed at (7f) holder. Dersom dette er tilfellet, vet vi at  $\alpha_i$  er korrelert med en eller flere av forklaringsvariablene og vi har et endogenitetsproblemet. I dette tilfellet bør Fixed effects-metoden (FE-metoden) anvendes.<sup>20</sup> Idéen med FE-metoden er å transformere bort den fylkesspesifikke restleddskomponenten som gir oss problemer. For å oppnå FE-estimatoren begynner vi med å beregne individuelle gjennomsnitt for samtlige variable. Det vil si at vi tar hvert fylke sitt gjennomsnitt over tid som vi illustrerer her med den avhengige variabelen:

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it} \quad (9)$$

Ved bruk av de individuelle gjennomsnittene finner vi den rene tverrsnittsfunksjonen der variablene med strek over er de individuelle gjennomsnittene:

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + \alpha_i + \bar{\eta}_{it} \quad (10)$$

---

<sup>20</sup>Verbeek (2012)

Vi trekker så tverrsnittsfunksjonen (11) fra den opprinnelige modellen vår gitt av (5), som gir oss følgende:

$$y_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + \eta_{it} - \bar{\eta}_i \quad (11)$$

Ved å anvende OLS direkte på (12) får vi altså FE-estimatoren som under våre forutsetninger vil være en konsistent estimator for  $\beta_1$ . Som vi kan se er den problematiske fylkesspesifikke restleddskomponenten blitt transformert bort fra modellen.

### 5.2.3 Random effects/GLS

I motsetning til FE-metoden utnytter Random effects-metoden (RE-metoden også kalt feasible GLS), all variasjon i data både mellom fylker og over tid. Metoden er altså mindre ekstrem enn FE-metoden. Vi antar her at forutsetning (7e) holder, altså at fylkesspesifikke restleddskomponenten er ukorrelert med forklaringsvariabelen. Som nevnt tidligere kan vi bruke OLS dersom (7e) holder, men vi vil ha et problem med seriekorrelasjon. Dette korrigerer GLS/RE-metoden for. Tar utgangspunkt i samme modell som tidligere:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \underset{=\alpha_i + \eta_{it}}{u_{it}} \quad (5)$$

Variansen til restleddet kan da skrives som:

$$E(u_{it}^2) = \sigma_\alpha^2 + \sigma_\eta^2 \quad (12)$$

og kovariansen:

$$\begin{aligned} E(u_{it}u_{is}) &= E(\alpha_i + \eta_{it}, \alpha_i + \eta_{is}) \\ \text{der } E(\alpha_i\alpha_i) &= \sigma_\alpha^2 \\ \Rightarrow E(u_{it}u_{is}) &= \sigma_\alpha^2 > 0 \end{aligned} \quad (13)$$

Fra (13) ser vi at kovariansen mellom restleddene er korrelert og vi har altså at restleddene  $u_{it}, u_{is}$  vil være seriekorrelerte med hverandre. Det vil si at vi får en varians-kovarians matrise som ikke er på formen *konstant* x  $I_N$ , altså at varians-kovarians matrisen multiplisert med identitetsmatrisen ikke er lik kovariansmatrisen. Dette er illustrert av (34) og (35) i appendiks. Dette kan korrigeres for ved å bruke GLS/RE-estimatoren. Ved bruk av GLS-metoden transformeres variablene slik vi får en varians-kovariansmatrise på formen *konstant* x  $I_N$  som vil si at den transformerte modellen oppfyller Gauss-Markov betingel-

sene. Transformasjon gir:<sup>21</sup>

$$y_{it} - \theta \bar{y}_i = \beta_0(1 - \theta) + \beta_1(x_{it} - \theta \bar{x}_i) + u_{it} - \theta \bar{u}_i \quad (14)$$

Der  $\theta$  er gitt som:

$$\theta = 1 - \left( \frac{\sigma_\eta^2}{\sigma_\eta^2 + T\sigma_\alpha^2} \right)^{\frac{1}{2}} \quad (15)$$

Dersom  $\theta = 0$  går vi tilbake til pooled OLS systemet, da vil altså RE-estimatoren være ekvivalent med OLS-estimatoren. Denne kollapser til 0 dersom  $\sigma_\alpha^2 = 0$ , altså er  $\alpha_i$  uviktig og vi kan benytte oss av pooled OLS uten bekymring for seriekorrelasjon. Videre kan vi se at dersom  $\theta = 1$  vil RE-estimatoren være ekvivalent med FE-estimatoren. Dette skjer dersom  $\sigma_\alpha^2 \rightarrow \infty$ . Dette betyr at effekten av  $\alpha_i$  er stor og vi må benytte oss av FE som estimeringsmetode. Vanligvis har vi at  $0 \leq \theta \leq 1$ .

#### 5.2.4 Hausman-test

Valget av estimeringsmetode avhenger av den fylkesspesifikke restleddskomponenten  $\alpha_i$  og om denne er korrelert med forklæringsvariablene eller ikke, altså om vi har et problem med uobserverbar heterogenitet. Dersom denne er ukorrelert med inkluderte forklaringsvariable i modellen kan jeg benytte meg av RE-metoden, dersom dette derimot ikke holder bruker jeg FE-metoden. For å teste dette kan vi benytte oss av den såkalte Hausman-testen. Hausman (1978) formulerer følgende hypoteser:

$$H_0 : E(\alpha_i | x_{it}) = 0$$

$$H_A : E(\alpha_i | x_{it}) \neq 0$$

Under nullhypotesen er RE-estimatoren konsistent og effisient, mens FE-estimatoren er konsistent. Under den alternative hypotesen er kun FE-estimatoren konsistent. Den generelle idéen er at de to estimatorene sammenlignes. En signifikant forskjell mellom de to estimatorene indikerer at det er lite sannsynlig at nullhypotesen holder.<sup>22</sup> Testobservatoren for Hausman-testen er  $\chi^2$  - fordelt og er gitt:

$$\chi^2 = (\hat{\beta}_{FE} - \hat{\beta}_{RE}) [\hat{V}(\hat{\beta}_{FE} - \hat{V}(\beta_{RE}))]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \quad (16)$$

<sup>21</sup>Wooldridge (2013)

<sup>22</sup>Wooldridge (2013)

Vi kan forkaste nullhypotesen dersom testobservatoren er større en kritisk verdi. Dersom vi forkaster  $H_0$  vil dette tilsi at FE-metoden er å foretrekke.

### 5.3 Dynamiske paneldatamodeller

En av fordelene med paneldata er at jeg har muligheten til å inkludere dynamikk i modellen ettersom paneldata også varierer over tid. Dette vil si at jeg kan undersøke treghet i tilpasning av den avhengige variabelen ved å inkludere lagget verdi av den avhengige variabelen som høyresidevariabel i modellen. Grunnen for å gjøre dette er at det kan tenkes at andel uføre tidligere år har en påvirkning på andelen uføre i dag, altså det er persistens i andelen uføre. For enkelthets skyld tar jeg utgangspunkt i følgende modell der kun den laggede endogene variabelen inngår som høyresidevariabel i den videre diskusjonen:

$$y_{it} = \beta_0 + \beta_1 y_{i,t-1} + \alpha_i + u_{it} \quad (17)$$

der restleddet fremdeles består av en idiosynkratisk komponent, samt den fylkesspesifikke komponenten. (17) impliserer at:

$$y_{it-1} = \beta_0 + \beta_1 y_{it-2} + u_{it-1}$$

der restleddskomponenten er gitt:

$$u_{it-1} = \alpha_i + \eta_{it-1} \quad (18)$$

Det vil si at det er en positiv korrelasjon mellom andelen uføre forrige periode og den fylkesspesifikke restleddskomponenten som ikke varierer over tid. Dersom vi velger å bruke OLS på (18) vil vi få en inkonsistent estimator og  $\beta_1$  vil være positivt forventningsskjev, altså vil den overvurdere effekten av den laggede endogene variabelen.

#### 5.3.1 Fixed effects

Det kan også være problematisk å estimere dynamiske paneldatamodellen med FE-metoden. For å illustrere dette tar vi utgangspunkt i (18) og transformerer den slik at den er gitt som avvik fra fylkesspesifikke gjennomsnitt (FE-transformert):

$$y_{it} - \bar{y}_i = \beta_1 (y_{i,t-1} - \bar{y}_{i-1}) + u_{it} - \bar{u}_i \quad (19)$$

der:

$$\begin{aligned}\bar{y}_i &= \frac{1}{T}(y_{i1} + y_{i2} + \dots + y_{iT}) \\ \bar{y}_{i-1} &= \frac{1}{T}(y_{i0} + y_{i1} + \dots + y_{iT-1}) \\ \bar{u}_i &= \frac{1}{T}(u_{i1} + u_{i2} + \dots + u_{iT})\end{aligned}\tag{20}$$

Det vil si at høyresidevariabelen og restleddet henholdsvis er gitt:

$$\begin{aligned}y_{i,t-1} - \frac{1}{T}(y_{i0} + y_{i1} + \dots + y_{it} + \dots + y_{iT-1}) \\ u_{it} - \frac{1}{T}(u_{it} + u_{i2} + \dots + u_{it-1} + \dots + u_{iT})\end{aligned}\tag{21}$$

Vi kan se at  $-\frac{1}{T}y_{it}$  i høyresidevariabelen er negativt korrelert med  $u_{it}$  og  $-\frac{1}{T}u_{it-1}$  også er negativt korrelert med  $y_{it-1}$  i høyresidevariabelen. Videre kan vi se at  $-\frac{1}{T}u_{it-1}$  og  $-\frac{1}{T}y_{it-1}$  er positivt korrelert. Problemet er at de negative korrelasjonene dominerer de positive, slik at forklaringsvariabelen er korrelert med restleddet. FE-estimatoren vil dermed undervurdere effekten av den laggede endogene variabelen,  $y_{it-1}$ . Dette kalles for Nickell-skjevhet og Nickell (1981) viste at FE-metoden vil gi inkonsistente estimatører så lenge tidsperioden (T) er endelig. Derimot vil skjevheten reduseres jo større T, altså jo lengre tidsperiode vi har.

### 5.3.2 Arellano og Bond

En mulig metode for å estimere dynamiske paneldatamodeller er Arellano og Bond (1991) sin GMM-metode. Metoden er bygget på Anderson og Hsiao (1981) sin metode og idéen er å utvide settet av instrumenter. For å illustrere førstedifferensierer vi (17):

$$\begin{aligned}y_{it} - y_{it-1} &= \beta_1(y_{it-1} - y_{it-2}) + u_{it} - u_{it-1} \\ \Delta y_{it} &= \beta_1 \Delta y_{it-1} + \Delta u_{it}\end{aligned}\tag{22}$$

Vi antar fravær av andregrads seriekorrelasjon i de differensierte restleddene, og dermed gjelder de følgende momentrestriksjonene i hver periode:

$$\begin{aligned}
 E[\Delta u_{i2}, y_{i0}] &= 0 \\
 E[\Delta u_{i3}, y_{i1}] &= E[\Delta u_{i3}, y_{i0}] = 0 \\
 E[\Delta u_{i4}, y_{i2}] &= E[\Delta u_{i4}, y_{i1}] = E[\Delta u_{i4}, y_{i0}] = 0 \\
 &\vdots \\
 E[\Delta u_{iT}, y_{ij}] &= 0 \quad \text{for } j = 1, 2, \dots, T-2
 \end{aligned} \tag{23}$$

Dette gir et sett av momentrestriksjoner som kan brukes når vi utleder en estimator for (23). Momentrestriksjonene impliserer et sett  $Z_i$  med instrumenter. Momentrestriksjonene, der vi setter inn for det differensierte restleddet, kan skrives:

$$\begin{aligned}
 E(Z_i' \Delta u_i) &= 0 \\
 E[Z_i' (\Delta y_i - \beta_1 \Delta y_{i-1})] &= 0
 \end{aligned} \tag{24}$$

Vi har en situasjon der antallet momentrestriksjoner (instrumenter) er større enn nødvendig og modellen er overidentifisert. Vi kan konstruere en estimator for  $\beta_1$ , som er konsistent og forventningsrett, ved å erstatte de teoretiske med de empiriske momentene i restriksjonen og minimere en vektet sum av disse. Vi minimerer følgende kvadratiske sum der  $W_i$  er en positiv og symmetrisk vektingsmatrise:

$$\min_{\beta_1} Q = \left[ \frac{1}{N} \left( \sum_{i=1}^N Z_i' (\Delta y_i - \beta_1 \Delta y_{i-1}) \right)' W_i^{-1} \left[ \frac{1}{N} \sum_{i=1}^N Z_i' (\Delta y_i - \beta_1 \Delta y_{i-1}) \right] \right] \tag{25}$$

Ved å differensiere med hensyn på  $\beta_1$  og løse ut for denne får vi:

$$\hat{\beta}_1^{GMM} = \left[ \left( \sum_{i=1}^N \Delta y_{i-1}' Z_i \right) W_i \left( \sum_{i=1}^N Z_i' \Delta y_{i-1} \right) \right]^{-1} \left[ \left( \sum_{i=1}^N \Delta y_{i-1}' Z_i \right) W_i \left( \sum_{i=1}^N Z_i' \Delta y_i \right) \right] \tag{26}$$

Som gir oss FD-GMM-estimatoren. Ulike vektingsmatriser ( $W_i$ ) gir ulike estimatører for  $\beta_1$ , men alle er konsistente. Den optimale vektingsmatrisen er asymptotisk proporsjonal til den inverse varians-kovarians matrisen av momentrestriksjonene og bør tilfredstille:

$$\text{plim} W_i = E(z_i' \Delta u_i \Delta u_i' z_i)^{-1} \tag{27}$$

Metoden kan brukes i 2-steps eller 1-steps prosedyrer. Dersom vi kun pålegger restriksjon om fravær av seriekorrelasjon bruker vi 2-steg. Vi oppnår en konsistent estimator for

$\beta_1$ , for eksempel basert på  $W_i = I$ . Vi beregner residualene og estimerer det empiriske motstykket til (28) gitt som:

$$\widehat{W}_i^{OPT} = \left( \frac{1}{N} \sum_{i=1}^N Z_i' \Delta \widehat{u}_i \Delta \widehat{u}_i' Z_i \right)^{-1} \quad (28)$$

I andre trinn bruker vi  $\widehat{W}_i^{OPT}$  til å beregne GMM-estimatoren.

Dersom vi ønsker å pålegge restriksjon om fravær av heteroskedastisitet i tillegg til fravær av autokorrelasjon kan vi bruke en-steps metoden.<sup>23</sup> I dette tilfellet kan vi skrive variansen og kovariansen til restleddet som:

$$\begin{aligned} Var(\Delta u_{it}) &= E[(u_{it} - u_{it-1})(u_{it} - u_{it-1})] \\ Var(\Delta u_{it}) &= E[u_{it}^2 - 2u_{it}u_{it-1} + u_{it-1}^2] \\ Var(\Delta u_{it}) &= \sigma^2 - 0 + \sigma^2 = 2\sigma^2 \end{aligned} \quad (29)$$

$$\begin{aligned} Cov(\Delta u_{it}, \Delta u_{it-1}) &= E[(u_{it} - u_{it-1})(u_{it-1} - u_{it-2})] \\ Cov(\Delta u_{it}, \Delta u_{it-1}) &= E[u_{it-1}u_{it-1} - u_{it-1}^2 - u_{it}u_{it-2} + u_{it-1}u_{it-2}] \\ Cov(\Delta u_{it}, \Delta u_{it-1}) &= -\sigma^2 \end{aligned} \quad (30)$$

Og vi kan da skrive varians-kovariansmatrisen som:

$$E(\Delta u_i \Delta u_i') = \sigma_u^2 G = \sigma_u^2 \begin{bmatrix} 2 & -1 & 0 & \dots \\ -1 & 2 & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ \vdots & 0 & -1 & 2 \end{bmatrix} \quad (31)$$

Den optimale vektingsmatrisen for GMM-estimatoren i dette tilfellet blir da:

$$\widehat{W}_i^{OPT} = \left[ \frac{1}{N} \sum_{i=1}^N Z_i' G Z_i \right]^{-1} \quad (32)$$

Vektingsmatrisen inneholder her ingen ukjente parametere kan brukes direkte når vi beregner GMM-estimatoren, under restriksjonene om fravær av autokorrelasjon og heteroskedastisitet. De resterende forklaringsvariablene som inkluderes i modellen vil operere

<sup>23</sup>Verbeek (2012)

## 5. METODE

---

som instrumenter for seg selv.<sup>24</sup>

Jeg begynner med å estimere hovedmodellen med pooled OLS, RE- og FE- metodene for å teste meg frem til den foretrukne estimeringsmetoden som jeg fortsetter med i den videre analysen. Videre benytter jeg meg av både FE- og Arellono og Bonds GMM-metode når jeg estimerer modellen med inkludert lagget endogen variabel. Samtlige resultater presenteres og kommenteres i neste kapittel.

---

<sup>24</sup>Verbeek (2012)



## 6 Resultater

I dette kapitlet presenterer jeg resultatene fra den økonometriske analysen utført med metodene jeg illustrerte i forrige kapittel. I delkapittel 6.1 presenterer jeg resultater fra estimering av hovedmodellen, valg mellom Fixed effects og Random effects/GLS samt inkludering av laggede eksogene variable. I delkapittel 6.2 presenterer jeg resultatene for sensitivetsanalyse av modellen der jeg inkluderer ytterligere kontrollvariable. I delkapittel 6.3 presenterer jeg resultatene fra modellen der lagget endogen variabel er inkludert som høyresidevariabel i modellen. Til slutt presenterer jeg resultatene for kontroll av næringsstruktur i delkapittel 6.4.

### 6.1 Hovedmodell

I tabell 11 rapporteres resultatene for estimering av hovedmodellen min.

Tabell 11: Hovedmodell

	RE (1)	FE (2)	FE (3)	FE (4)
	lnAndeluføre	lnAndeluføre	lnAndeluføre	lnAndeluføre
lnLedig <sub>it</sub>	0.0310** (0.00874)	0.0276** (0.00949)	0.0135 (0.0111)	
lnLedig <sub>it-1</sub>			0.0280** (0.00911)	0.0544*** (0.00945)
lnLedig <sub>it-2</sub>			0.0399* (0.0174)	
lnSykefr <sub>it</sub>	-0.0356 (0.0309)	-0.0635 (0.0310)	-0.00673 (0.0275)	
lnSykefr <sub>it-1</sub>			0.121** (0.0377)	0.0787** (0.0262)
lnSykefr <sub>it-2</sub>			0.0916* (0.0366)	0.0593* (0.0261)
lnBrutto <sub>it</sub>	-0.0705 (0.0581)	-0.0277 (0.0897)	-0.0296 (0.101)	-0.0586 (0.101)
lnInntbr <sub>it</sub>	0.286*** (0.0635)	0.306*** (0.0669)	0.243*** (0.0574)	0.237** (0.0604)
lnSkillsm <sub>it</sub>	0.140*** (0.0366)	0.123** (0.0369)	0.0659 (0.0345)	0.0714 (0.0347)
lnUtdL <sub>it</sub>	1.873*** (0.0355)	2.015*** (0.397)	1.911*** (0.335)	1.899*** (0.353)
_cons	-4.725*** (0.577)	-4.997*** (0.705)	-3.400*** (0.839)	-3.741*** (0.649)
N	285	285	247	247

Standardavvik korrigerert for clustering i parentes

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Modellene estimert i (1) og (2) samsvarer med modellspesifikasjon (1), og er estimert med henholdsvis GLS/RE-metoden og FE-metoden. I modell (3), som samsvarer med modellspesifikasjon (2), er modellen estimert med FE-metoden og jeg har inkludert laggede

verdier for én og to perioder av ledighet og sykefravær. I modell (4) har jeg ekskludert de variable som samlet ikke har noen signifikant effekt i modellen.

I delkapittel 5.1.1 ble bruken av OLS på paneldatamodeller diskutert og presentert som en uegnet estimeringsmetode for modellen. Jeg har estimert hovedmodellen med OLS som presenteres i tabell 23 i appendiks. For å undersøke om bruken av pooled OLS er uegnet i mitt tilfelle, har jeg benyttet meg av den såkalte Breusch-Pagan LM-testen for heteroskedastisitet. Leser henvises til appendiks under tabell 23 for formelle resultater av testen. I testen kan jeg klart forkaste  $H_0$ , som tilsier at RE-metoden benyttes videre. Dermed er resultatene for RE presentert her, da jeg som neste test vil bruke Husman-testen for å velge mellom RE og FE.

### 6.1.1 Valg mellom Random effects og Fixed effects

I modell (1) og (2) er hovedmodellen estimert med henholdsvis RE/GLS-metoden og FE-metoden. Jeg begynner med å utføre en Hausman-test for å velge mellom disse to estimeringsmetodene. Under nullhypotesen er RE-estimatoren konsistent og effisient, mens FE-estimatoren er konsistent. Under den alternative hypotesen er kun FE-estimatoren konsistent. Det vil si at dersom jeg kan forkaste nullhypotesen velger jeg å fortsette med Fixed effects som estimeringsmetode. Testobservatoren er  $\chi^2$ -fordelt med 8 frihetsgrader. Kritisk verdi er  $\chi_{0.05}^2(8) = 15.51$  og testobservatoren er  $\chi_{0.05}^2(8) = 39.92$ . Jeg kan altså klart forkaste  $H_0$  og jeg velger derfor å fortsette analysen med Fixed effects som estimeringsmetode. Dermed er modellene (3) og (4) estimert med Fixed effects.

### 6.1.2 Effekten av ledighet

Jeg begynner med å kommentere effekten av andelen arbeidsledige på andelen uføre i den enkle modellen estimert med Fixed effects som er gitt av (2). Jeg finner her en svakt positiv, men signifikant, effekt av ledighet på andelen uføre med 0.0276 prosent. Den korrekte måten å tolke dette resultatet på er at én prosents økning i andelen arbeidsledige øker andelen uføre med  $\approx 0.03$  prosent. Videre inkluderer jeg variable for arbeidsledighet lagget med én periode og lagget med to perioder i modell (3). I modell (3) finner jeg ingen signifikant effekt av ledighet uten lag. Derimot gir ledighet lagget med én periode en signifikant positiv effekt med  $\approx 0.03$  prosent på andelen uføre, tilsvarende med ledighet uten lag fra modell (2). Den korrekte måten å tolke dette resultatet på er at én prosents økning i andelen arbeidsledige i fjor, gir 0.03 prosents økning i andelen uføre i dag. Videre finner jeg også en signifikant positiv effekt av ledighet lagget med to perioder på andelen uføre

med  $\approx 0.04$  prosent. I modell (4) er variable for ledighet, og ledighet lagget to perioder (i tillegg til sykefravær uten lag) ekskludert da disse samlet ikke har noen signifikant effekt i modellen. Dermed står jeg i modell (4) kun igjen med ledighet lagget med én periode. Jeg finner at denne har en sterkt signifikant (0.001 nivå) positiv effekt på andelen uføre med 0.0544 prosent,  $\approx 0.05$  prosent.

### 6.1.3 Effekten av kontrollvariable

Jeg begynner med å kommentere effekten av sykefravær. I modell (2) der den enkle modellen er estimert med Fixed effects gir denne en negativ effekt, men effekten er ikke signifikant. I modell (3) inkluderes laggede verdier for sykefravær med én og to perioder. I denne modellen finner jeg at effekten for sykefravær uten lag fremdeles er negativ og ikke er signifikant. Derimot finner jeg signifikante positive effekter for sykefravær lagget med én og to perioder på  $\approx 0.1$  prosent. I modell (4) ekskluderes sykefravær uten lag (sammen med ledighet og ledighet lagget to perioder). Jeg finner her en positiv og signifikant effekt av ledighet lagget med én periode på  $\approx 0.08$  prosent og for ledighet lagget med to perioder på  $\approx 0.06$  prosent, litt svakere effekter enn i modell (2).

Jeg ser videre på effektene av de øvrige kontrollvariablene. Bruttoprodukt gir ingen signifikant effekt i noen av modellene, men kommer ut med et negativ fortegn som forventet. Videre gir inntekt en signifikant og positiv effekt i alle tre modellene med en effekt på  $\approx 0.2$  prosent i modell (3). Dette vil tilsi at én prosents økning i gjennomsnittlig inntekt vil øke andelen uføre med 0.2 prosent, og effekten er dermed motsatt enn forventet. Skillisser gir en positiv signifikant effekt i modell (1) med  $\approx 0.1$  prosent, men denne holder seg ikke signifikant i modell (2) og (3). Derimot får jeg en positiv og signifikant effekt av andelen med lav utdanning i alle tre modellen med en effekt på hele  $\approx 1.9$  prosent i modell (3). Det vil tilsi at dersom andelen med grunnskole- eller videregående nivå som høyeste oppnådde utdanning øker med én prosent i fylket, vil andelen uføre øke med hele 1.9 prosent.

I denne delen av analysen har jeg funnet en signifikant og positiv effekt av ledighet lagget med én periode på  $\approx 0.05$  prosent. Videre har jeg funnet positive og signifikante effekter av sykefravær lagget med én og to perioder på henholdsvis 0.08 og 0.06 prosent. Til slutt har jeg funnet positive og signifikante effekter av inntekt på 0.2 prosent og andelen med lavt utdanningsnivå på hele 1.9. Den endelige modellen fra denne delen av analysen er gitt av (4) og jeg benytter meg av denne i den følgende sensitivitetsanalysen.

## 6.2 Sensitivitetsanalyse

For å undersøke sensitiviteten eller robustheten til resultatene mine inkluderer jeg videre kontrollvariable for å se hvordan effekten av fokusvariabelen min og de inkluderte kontrollvariabelene ”reagerer”.

Tabell 12: Sensitivitetsanalyse

	(1)	(2)	(3)	(4)	(5)
	lnAndel_U	lnAndel_U	lnAndel_U	lnAndel_U	lnAndel_U
lnLedig <sub>it-1</sub>	0.0489*** (0.00940)	0.0650*** (0.0117)	0.0519*** (0.0124)	0.0660*** (0.0115)	0.0666*** (0.0116)
lnSykefr <sub>it-1</sub>	0.0691** (0.0222)	0.124*** (0.0260)	0.0700** (0.0239)	0.118*** (0.0246)	0.120*** (0.0252)
lnSykefr <sub>it-2</sub>	0.0603* (0.0271)	0.0662* (0.0297)	0.0519 (0.0274)	0.0571 (0.0284)	0.0577 (0.0287)
lnBrutto <sub>it</sub>	-0.0763 (0.108)	0.0679 (0.107)	0.0648 (0.110)	0.0762 (0.103)	0.0767 (0.102)
lnInntbr <sub>it</sub>	0.205** (0.0645)	0.323** (0.104)	0.234 (0.111)	0.340** (0.106)	0.341** (0.106)
lSkilsm <sub>it</sub>	0.0713 (0.0345)	0.0503 (0.0312)	0.0717* (0.0336)	0.0602 (0.0312)	0.0604 (0.0303)
lUtdL <sub>it</sub>	1.716*** (0.324)	1.775*** (0.238)	1.819*** (0.179)	1.724*** (0.237)	1.720*** (0.236)
Nettoinnfl <sub>it</sub>	-3.130* (1.219)	-3.743* (1.470)	-3.875* (1.416)	-3.675* (1.398)	-3.659* (1.390)
Nettoinnpen <sub>it</sub>	0.939 (0.971)				
lUtdHkort <sub>it</sub>		0.460* (0.200)	0.282 (0.251)	0.490* (0.196)	0.489* (0.198)
lnUtdHlang <sub>it</sub>		-0.334*** (0.116)	-0.423** (0.126)	-0.354*** (0.115)	-0.354*** (0.115)
lnEurUT <sub>it</sub>			0.0701 (0.0287)		
lnAf <sub>it</sub>			0.0457 (0.0289)		
lnAsMT <sub>it</sub>			0.0289 (0.0440)		
lnNAM <sub>it</sub>			0.00196 (0.0296)		
lnSMAM <sub>it</sub>			0.0161 (0.0363)		
lnOz <sub>it</sub>			-0.0111 (0.0142)		
StemSVF <sub>it</sub>				0.768*** (0.187)	0.736*** (0.168)
StemAPF <sub>it</sub>				-0.200** (0.0556)	-0.195*** (0.0422)
StemHF <sub>it</sub>				0.0260 (0.0496)	
StemFrpF <sub>it</sub>				-0.0378 (0.123)	
_cons	-3.535*** (0.719)	-4.616** (1.568)	-3.672 (1.767)	-4.821** (1.614)	-4.820** (1.610)
N	247	247	247	247	247

Standardavvik i parentes, korrigeret for clustering.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Videre ønsker jeg å ta hensyn til potensiell utelatt variabel-skjevhet ved å inkludere variable som også er korrelert med fokusvariabelen. Dersom jeg tillegger en variabel og den estimerte koeffisienten til ledighet endrer seg mye, indikerer dette at den inkluderte variabelen er relatert til både ledighet og andelen uføre og denne variabelen skal dermed inkluderes i modellen for å unngå utelatt variabel-skjevhet. Denne delen av analysen samsvarer med modellspesifikasjon (3). Jeg tar utgangspunkt i modell (4) fra tabell 11 (hovedmodell) og tillegger kontrollvariable på denne. I modell (1) inkluderer jeg variable for nettoinnpendling og nettoinnflytting, som mål på attraktivitet av fylket. I modell (2) inkluderer jeg variable for høyere utdanning, slik at jeg kontrollerer for utdanningsfordeling i fylket. I modell (3) inkluderer jeg variable for innvandring og til slutt, i modell (4), inkluderer jeg variable for andelen godkjente stemmer ved fylkesvalg til partiene SV, Ap, Høyre og Frp. I modell (5) rapporteres resultatene for den endelige modellen. Resultatene presenteres i tabell 12.

Jeg begynner med å se at effekten av ledighet lagget med én periode på andelen uføre holder seg sterkt signifikant i alle modellene, med effekter på rundt 0.05 - 0.06 prosent. I modell (5) finner jeg en positiv effekt på 0.0666 som tilsvarer  $\approx 0.07$  prosent. Det vil tilsi at dersom andelen arbeidsledige økte med én prosent i fjor, vil andelen uføre øke med 0.07 prosent i dag. Siden effekten holder seg sterkt signifikant og ikke endres drastisk i verdi, vil det tilsi at effekten av ledighet på andelen uføre er robust. Ettersom effekten ikke endrer seg spesielt i verdi, tilsier dette at hovedmodellen ikke er underspesifisert. Hadde modellen vært underspesifisert ville verdien på koeffisienten endret seg i større grad, som følge av redusert utelatt variabel-skjevhet.

Videre gir sykefravær lagget med én periode i modell (5) en signifikant effekt på 0.12 prosent på andelen uføre. Den effekten holder seg signifikant i alle modellene og endrer seg noe i verdi. Derimot finner jeg at effekten av sykefravær lagget med to perioder kun er svakt signifikant i modell (1) og (2) og ikke signifikant i resten av modellene. Effekten endrer seg ikke mye i verdi og ligger på rundt 0.06 prosent. Inntekt gir en positiv og signifikant effekt på  $\approx 0.3$  prosent i modell (5), og verdien endrer seg ikke mye mellom modellene. Effekten av inntekt er derimot ikke signifikant i modell (3). For bruttoprodukt er det ingen signifikante effekter og denne endrer også fortegn og rapporterer et positivt fortegn i modell (5). For skillsmisser finner jeg kun en svakt signifikant effekt i modell (3) på 0.07 prosent. Tilslutt finner jeg at effekten av andelen med lav utdanning holder seg sterkt signifikant i alle modellene. I modell (5) finner jeg en positiv effekt av andelen med lav utdanning på andelen uføre med 1.72 prosent. Dette er en reativt stor effekt

og underbygger tidligere tidligere funn som finner at personer med lav utdanning har en høyere uføretilbøyelighet.<sup>25</sup>

I modell (1) inkluderer jeg kontrollvariable for nettoinnflytting og nettoinnpendling som mål på attraktiviteten til fylket. Jeg finner ingen signifikant effekt av nettoinnpendling og finner i tillegg at denne kan ekskluderes videre. Nettoinnflytting gir en svakt signifikant negativ effekt i alle modellene og i modell (5) finner jeg at én enhets økning i nettoinnflytting gir en reduksjon i andelen uføre på  $\approx -3.7$  prosent. Denne effekten er svært høy og samsvarer med Bragstad (2008). I modell (2) inkluderer jeg kontrollvariable for høyere utdanning og finner at begge nivå har signifikante effekter. I modell (5) finner jeg en positiv og signifikant effekt av andelen med kort høyere utdanning (bachelor), øker andelen uføre med  $\approx 0.5$  prosent og effekten av andelen med lang høyere utdanning (master, Phd), reduserer andelen uføre med  $\approx -0.4$  prosent. Sammenlignet med andelen med lav utdanning tilsier dette at høyere utdanning fører til færre uføre, og da spesielt lang høyere utdanning. Altså jo høyere utdanningsnivå i fylket, jo lavere andel uføre. I modell (3) inkluderer jeg kontrollvariable for andelen innvandrere etter opphav fra verdensdelene. Jeg finner ingen individuelle signifikante effekter og variablene samlet sett har ingen signifikant effekt i modellen. I modell (4) inkluderer jeg til slutt kontrollvariable for andelen godkjente stemmer til SV, Ap, Høyre og Frp ved fylkesvalg i perioden. Jeg finner ingen signifikante effekter av godkjente stemmer til Høyre og Frp. Derimot finner jeg signifikante effekter for SV og Ap. I modell (5) er effektene av godkjente stemmer til SV og Ap sterkt signifikante med effekter på henholdsvis  $\approx 0.7$  og  $\approx -0.2$  prosent. Det vil tilsi at én prosent økning i andelen godkjente stemmer til SV øker andelen uføre med 0.7 prosent, mens tilsvarende for Ap reduserer andelen uføre. Ettersom begge disse partiene ligger på samme side i politikken, dog i hver sin ende, er dette et interessant funn. Det kan tenkes at denne forskjellen i effekter er fordi Ap er såpass sentrumsnært i forhold til SV som befinner seg langt ute på venstresiden av politikken.

I denne delen av analysen har jeg funnet at effekten av ledighet lagget med én periode er robust med en sterkt signifikant effekt på 0.0666 prosent i den endelige modellen. Videre har effekten av andelen med lavt utdanningsnivå holdt seg signifikant i alle modellene med en effekt på 1.72 prosent i den endelige modellen. Av de nye kontrollvariabelene som jeg inkluderte i modellen har jeg funnet en svakt signifikant effekt av nettoinnflytting på -3.7 prosent for én enhets økning i nettoinnflytting i fylket. Videre fant jeg signifikante effekter av høyere utdanning med 0.5 prosent for bachelor-nivå og -0.4 for master/Phd-nivå. Til

---

<sup>25</sup>Fevang og Røed (2006)

slutt fant jeg at andelen godkjente stemmer til SV og Ap i fylkesvalg ga signifikante effekter på henholdsvis 0.7 og -0.2 prosent i den endelige modellen. Den endelige modellen fra denne delen av analysen er gitt av modell (5) og jeg benytter meg av denne i den neste delen av analysen.

### 6.3 Inkludering av lagget endogen variabel

Som tidligere diskutert i kapittel 5, kan det tenkes at andelen uføre tidligere år har en effekt på andelen uføre i dag, altså at det er persistens i andelen uføre i fylkene. Dette kan jeg undersøke ved å inkludere en lagget verdi av den avhengige variabelen som høyre-sidevariabel i modellen. Ved å estimere modellspesifikasjon (4) kan jeg fange opp både kortids- og langtidseffekter til forklaringsvariablene. Langtidseffekten (LE) finner jeg på følgende måte for alle variable untatt sykefravær:

$$LE = \frac{\beta_j}{1 - \beta_1} \quad j = 1, \dots, 7$$

For sykefravær har jeg to variable inkludert, sykefravær lagget med én periode og lagget med to perioder. Dermed vil langtiseffekten for sykefravær være gitt som:

$$LE_{Sykefr} = \frac{\beta_j + \beta_i}{1 - \beta_1} \quad j = 1, \dots, 7$$

Jeg antar videre at vi har en stabil dynamisk prosess som vil innebære at  $\beta_1$  vil ligge i intervallet mellom 0 og 1  $\Rightarrow 0 < \beta_1 < 1$ . Jo nærmere  $\beta_1$  ligger 1, jo tregere er tilpasningen i andelen uføre fra et år til det neste, altså jo tregere er tilpasningen fra en likevekt til en annen. Jeg presenterer resultater for både Fixed effects-estimering og Arellano og Bond. Ettersom jeg har en lang tidsperiode (T=15) vil den såkalte Nickell-biasen være liten. Resultater presenteres i tabell 13. Resultater for FE presenteres i modell (1) og resultater for AogB presenteres i modell (2). Langtidseffekter (LE) presenteres vedsiden av begge modellene.

Ved første øyekast ser jeg at begge modellene gir en sterkt signifikant effekt av den laggede avhengige variabelen, som vil tilsi en høy grad av persistens i andelen uføre. I modell (1) der FE-estimering er benyttet finner jeg en effekt av lagget andel uføre på 0.692 som vil tilsi en relativt treg tilpasning. Med andre ord vil tidligere andel uføre ha en effekt på  $\approx 0.7$  på andelen uføre i år. Videre er variabelen for andelen arbeidsledige lagget med én

## 6. RESULTATER

periode sterkt signifikant og har en effekt på  $\approx 0.03$  prosent på kort sikt og 0.09 på lang sikt. Det vil si at én prosent økning i andelen arbeidsledige i fjor vil da øke andelen uføre med 0.03 prosent på kort sikt og 0.09 prosent på lang sikt. Jeg finner at effektene for sykefravær lagget med én og to perioder begge er sterkt signifikante med korttidseffekter på  $\approx 0.07$  prosent for begge to, og en samlet langtidseffekt på 0.45. Det kan bemerkes at langtidseffekten av sykefravær er relativt høyere enn på kort sikt.

Tabell 13: Inkludering av lagget endogen variabel

	FE		AB	
	(1)	LE	(2)	LE
	lnAndel_U <sub>it</sub>		lnAndel_U <sub>it</sub>	
lnAndel_U <sub>it-1</sub>	0.692*** (0.0570)		0.494*** (0.0775)	
lnLedig <sub>it-1</sub>	0.0264*** (0.00564)	0.09	0.0318*** (0.00634)	0.06
lnSykefr <sub>it-1</sub>	0.0658*** (0.0142)	0.45	0.0556** (0.0178)	0.24
lnSykefr <sub>it-2</sub>	0.0723*** (0.0126)		0.0647*** (0.0158)	
lnBrutto <sub>it</sub>	-0.0120 (0.0383)	-0.04	-0.0428 (0.0536)	-0.09
lnInntbr <sub>it</sub>	0.0554 (0.0526)	0.18	0.0723 (0.0555)	0.14
lnSkilsm <sub>it</sub>	0.0101 (0.0194)	0.03	0.0272 (0.0183)	0.05
lnUtdL <sub>it</sub>	0.582** (0.173)	1.89	0.662** (0.201)	1.31
Nettoinnfl	-1.529 (0.846)	-4.96	-0.587 (0.685)	-1.16
lnUtdHkort <sub>it</sub>	0.0675 (0.0917)	0.22	0.0914 (0.113)	0.18
lnUtdHlang <sub>it</sub>	-0.0318 (0.0734)	-0.10	-0.0545 (0.0877)	-0.11
StemSVF	0.474** (0.155)	1.54	0.510*** (0.136)	1.00
StemAPF	-0.0872 (0.0433)	-0.28	-0.103* (0.0406)	-0.20
_cons	-0.577 (0.815)		-1.352 (0.932)	
N	247		228	

Standardavvik i parentes, korrigert for clustering

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Av de øvrige kontrollvariablene finner jeg kun signifikante effekter for andelen med lav utdanning og godkjente stemmer til SV ved fylkesvalg. Andelen med lavt utdanningsnivå viser en signifikant effekt på andelen uføre med 0.582 prosent på kort sikt, og en langtidseffekt på 1.89 prosent som samsvarer mer med tidligere resultater. Andelen godkjente stemmer til SV ved fylkesvalg har en signifikant positiv korttidseffekt på  $\approx 0.5$  prosent og en langtidseffekt på hele 1.54 prosent. Modellen estimert med pooled OLS finnes i tabell 24 i appendiks. Som teorien tilsier kan det virke som pooled OLS overvurderer effekten



av lagget uavhengig variabel mens FE undervurderer den.<sup>26</sup>

I modell (2) er modellen estimert med Arellano og Bonds GMM-metode, som jeg diskuterte i delkapittel 5.2. Her finner jeg en litt svakere effekt av den laggede avhengige variabelen med 0.494. Det vil si at tilpasningen fra en likevekt til en annen verken er spesielt kjapp eller treg. Resultatet er sterkt signifikant og det kan konkluderes at andelen uføre er relativt avhengig av tidligere andel uføre. Videre finner jeg også her en sterkt signifikant effekt av ledighet lagget med én periode. Denne gir en korttidseffekt på 0.03 prosent og en langtidseffekt på 0.06 prosent som samsvarer med resultatene fra de endelige modellene i tabell 11 og 12. Effektene av sykefravær lagget med én og to perioder er begge signifikante med korttidseffekter på  $\approx 0.06$  prosent for begge to og en samlet langtidseffekt på 0.24. Av de øvrige kontrollvariablene finner jeg også her en signifikant effekt av andelen med lav utdanning med 0.662 prosent på kort sikt og 1.3 prosent på lang sikt. Videre finner jeg en sterkt signifikant effekt av andelen godkjente stemmer til SV ved fylkesvalg på 0.51 prosent på kort sikt og 1 prosent på lang sikt. I tillegg finner jeg her en svakt signifikant effekt av andelen stemmer til Ap på -0.1 prosent på kort sikt og -0.2 prosent på lang sikt. Som tidligere vil det tilsi at økte stemmer til SV øker andelen uføre, mens stemmer til Ap reduserer andelen uføre. I denne modellen finner jeg at det er mindre avvik mellom kortids- og langtidseffektene enn i FE-estimeringen.

Jeg har i denne delen av analysen funnet at det er høy grad av persistens i andelen uføre. Effekten av fokusvariabelen var sterkt signifikant i begge estimeringsmetodene og langtidseffekten i modell (2) samsvarte med tidligere effekter av ledighet lagget med én periode. Modellen fra denne delen av analysen benyttes i neste del av analysen da jeg ønsker å kontrollere for næringsstruktur i fylkene og dermed må dele datasettet opp i to separate tidsperioder.

## 6.4 Kontroll for næringsstruktur - delperioder

Jeg har tidligere presentert variable i datasettet som gir andelen sysselsatte i fylket etter næring. Ved bruk av disse variablene ønsker jeg å kontrollere for næringsstrukturen i fylkene. Det kan tenkes at disse holder seg relativt stabile over tid ettersom næringsstrukturen i et fylke ikke endrer seg på kort sikt. Jeg har derimot ikke mulighet til å bruke datasettet samlet i hele tidsperioden (2000-2014) grunnet SSB sine endringen i næringskoder fra SN2002 til SN2007. Dermed bruker jeg nå datasettet delt inn i to perioder.

---

<sup>26</sup>Verbeek (2012)

## 6. RESULTATER

Resultatene presenteres i tabell 14. Det henvises til tabell 25 appendiks for resultater fra estimering av hovedmodellen og modell inkludert laggede verdier for ledighet og sykefravær for de to periodene. Jeg bruker her modell (3) og (6) fra tabell 25 i appendiks. I modell (1) og (3) inkluderer jeg alle de ti aggregerte variablene for næringsstruktur og i modell (2) og (4) har jeg ekskludert de variablene som ikke har noen samlet signifikant effekt i modellen. Det henvises også til tabell 26 og 27 appendiks for videre sensitivitetsanalyser av modellene for delperiodene.

Tabell 14: Kontroll for næringsstruktur

	Periode 1 (1)	Periode 1 (2)	Periode 2 (3)	Periode 2 (4)
	lnAndel_U	Andel_U	Andel_U	Andel_U
lnLedig <sub>it-1</sub>	0.0134 (0.0232)	0.0523*** (0.0110)		
lnLedig <sub>it-2</sub>			0.0475** (0.0125)	0.0511*** (0.0108)
lnSykefr <sub>it-1</sub>	0.0371 (0.0225)	0.0708* (0.0296)	-0.0696 (0.0825)	-0.0834 (0.0691)
lnSykefr <sub>it-2</sub>	0.0853 (0.0425)	0.0631*** (0.0148)		
lnBrutto <sub>it</sub>	-0.00613 (0.0894)	-0.0337 (0.0916)	-0.240 (0.128)	-0.185 (0.136)
lnInntbr <sub>it</sub>	0.0575 (0.0774)	0.216** (0.0557)	0.172 (0.127)	0.0435 (0.0923)
lnSkilsm <sub>it</sub>	0.0313 (0.0295)	0.0123 (0.0287)	0.0497 (0.0246)	0.0566* (0.0269)
lnUtdL <sub>it</sub>	1.980*** (0.354)	2.093*** (0.449)	0.370 (0.255)	0.279 (0.235)
lnJSF1 <sub>it</sub>	-0.0962 (0.0598)		0.152 (0.0913)	0.123* (0.0500)
lnIUB2 <sub>it</sub>	-0.134 (0.0985)		0.00693 (0.0751)	
lnKV3/EVR3 <sub>it</sub>	0.00568 (0.0218)		0.101 (0.105)	
lnBA4 <sub>it</sub>	0.0155 (0.108)		-0.155 (0.204)	
lnVHR5 <sub>it</sub>	-0.233 (0.153)	-0.334** (0.106)	-0.0909 (0.131)	
lnTK6/TLIK6 <sub>it</sub>	0.0236 (0.0995)		0.289 (0.161)	
lnFT7/FF7 <sub>it</sub>	-0.172 (0.0824)		0.0675 (0.124)	
lnFTE8/FTTE8 <sub>it</sub>	0.0513 (0.0905)		0.167 (0.121)	
lnOFT9 <sub>it</sub>	0.385 (0.201)	0.388 (0.236)	1.086* (0.390)	1.194*** (0.303)
lnUoppgitt <sub>it</sub>	0.0259 (0.0287)		0.0768 (0.0391)	0.0816*** (0.0271)
_cons	-2.795* (1.183)	-3.847*** (0.893)	-0.787 (1.850)	-0.511 (0.867)
N	133	133	95	95

Standardavvik i parentes,

korrigert for clustering

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

### 6.4.1 Periode 1

Jeg begynner med å kommentere resultatene i periode 1, der ledighet lagget med én periode er inkludert. Dette er tilsvarende med modellen brukt for hele perioden. I modell (1) der alle variable for næringsstruktur inkluderes finner jeg en svært svak positiv effekt av ledighet lagget med én periode på andelen uføre med 0.0134 prosent. Denne effekten er derimot ikke signifikant. I modell (2), der variable uten samlet effekt er ekskludert, finner jeg derimot en positiv og sterkt signifikant effekt av ledighet lagget med én periode på andelen uføre med  $\approx 0.05$  prosent. Dette er svært nært effekten jeg har funnet i analysen for hele perioden.

Videre undersøker jeg effektene av variablene for næringsstruktur. I modell (1) finner jeg ingen signifikante effekter av næringsstruktur på andelen uføre. Samlet sett har alle variablene en signifikant effekt i modellen og jeg kan derfor ikke ekskludere alle. Jeg finner at jeg må beholde variable for VHR (varehandel, hotell og restaurantvirksomhet) og OFT (Offentlig forvaltning og annen tjenesteyting) som da er inkludert i modell (2). I modell (2) finner jeg en signifikant og negativ effekt av VHR med -0.334 prosent. Det vil tilsi at én prosents økning i andelen sysselsatt innenfor næringen varehandel, hotell og restaurantvirksomhet i fylket vil redusere andelen uføre med  $\approx -0.3$  prosent. OFT gir en positiv effekt på 0.388 prosent, men denne effekten er ikke signifikant.

Til slutt ser jeg på effekter av de øvrige kontrollvariablene. Jeg finner en sterkt signifikant og positiv effekt av andel med lav utdanning på andelen uføre  $\approx 2$  prosent. Dette tilsvarer med analysen for hele perioden. I modell (1) er ingen andre kontrollvariable signifikante. I modell (2) finner jeg en sterkt signifikant positiv effekt av sykefravær lagget med to perioder på  $\approx 0.06$  prosent, mens jeg finner en svakere signifikant effekt av sykefravær lagget med én periode på  $\approx 0.07$  prosent. Til slutt gir inntekt en positiv og signifikant effekt i modell (2) på  $\approx 0.2$  prosent, tilsvarende med hovedanalysen.

### 6.4.2 Periode 2

I periode 2 er variabel for ledighet lagget med to perioder inkludert, da ledighet uten lag og ledighet lagget med én periode (sammen med sykefravær uten lag og sykefravær lagget to perioder) ikke hadde noen samlet signifikant effekt i modellen. I modell (3) der alle variable for næringsstruktur er inkludert finner jeg en positiv og signifikant effekt av ledighet lagget med to perioder på  $\approx 0.05$  prosent. I modell (4) der variable for næringsstruktur uten samlet effekt er ekskludert finner jeg tilsvarende positiv effekt på  $\approx 0.05$  prosent av andelen

ledige på andelen uføre. Denne er sterkt signifikant i motsetning til modell (3). Dette vil tilsi at én prosents økning i andelen arbeidsledige for *to* år siden, vil gi en økning i andelen uføre med 0.05 prosent i dag. Resultatene samsvarer med resultater funnet for ledighet lagget med én periode i hovedanalysen for hele perioden og effekten i periode 1 gitt i modell (2).

Blant variablene for næringsstruktur finner jeg en svakt signifikant og positiv effekt av OFT (offentlig forvaltning og annen tjenesteyting) på  $\approx 1$  prosent i modell (3) der alle variable for næringsstruktur er inkludert. Jeg finner at jeg må beholde variablene for JSF (jordbruk, skogbruk og fiske), OFT og Uoppgitt næring. I modell (4), der kun disse tre av variablene for næringsstruktur er inkludert, finner jeg at effekten av OFT nå er sterkt signifikant med en positiv effekt på hele  $1.194 \approx 1.2$  prosent. Det vil tilsi at én prosents økning i andelen sysselsatte innen næringen OFT vil øke andelen uføre med 1.2 prosent. Videre finner jeg også en positiv, men svakere signifikant, effekt av JSF på  $\approx 0.1$  prosent. Tilslutt er også effekten av Uoppgitt signifikant med en positiv effekt på  $\approx 0.08$  prosent. Av de øvrige kontrollvaribale finner jeg ingen signifikante effekter i modell (3). I modell (4) finner jeg kun en svakt signifikant positiv effekt av skilsmisser på andelen uføre med 0.0566 prosent.

I denne delen av analysen fant jeg at effektene av ledighet lagget med én periode (i periode 1) og ledighet lagget med to perioder (i periode 2) samsvarte med effektene funnet i analysen for hele perioden. Videre fant jeg i periode 1 en signifikant effekt av andelen sysselsatte innenfor næringen varehandel, hotell og restaurantvirksomhet på -0.3 prosent. Her var også effekten av andelen med lavt utdanningsnivå sterkt signifikant med en effekt på hele 2 prosent. I periode 2 fant jeg en signifikant og sterk effekt av andelen sysselsatt innenfor næringen offentlig forvaltning og annen tjenesteyting på 1.2 prosent. Samlet sett var det liten effekt av kontroll for næringsstruktur.

Jeg har nå presentert og kommentert funnene i min analyse ved de ulike modellspesifikasjonene. I det påfølgende og avsluttende kapittelet vil jeg begynne med å oppsummere funnene slik at leseren får en mer helhetlig oversikt. Videre vil jeg diskutere de viktigste funnene før jeg konkluderer oppgaven.

## 7 Avslutning

Fomålet med denne oppgaven var å undersøke effekten av lokal arbeidsledighet på andelen uføre ved hjelp av et selvkonstruert paneldatasett for de norske fylkene i tidsperioden 2000-2014. Jeg begynte med å estimere hovedmodellen min der fokusvariabelen for andelen arbeidsledige var inkludert, i tillegg til kontrollvariable for sykefravær, bruttoprodukt, inntekt, skilsmisser og andelen med lavt utdanningsnivå. Jeg estimerte modellen med RE- og FE-metoden og valgte ved hjelp av en Hausman-test å fortsette med FE som estimeringsmetode. Hovedmodellen estimert med Fixed effects ga en positiv effekt av arbeidsledighet på andelen uføre på  $\approx 0.03$  prosent som var signifikant på 0.01 nivå. Resultatet tilsier en svak effekt av arbeidsledighet på andelen uføre ved at én prosent økning i arbeidsledighet i et fylke kun vil øke andelen uføre med 0.03 prosent. Jeg fant også positive og signifikante effekter av inntekt og skilsmisse på henholdsvis 0.3 prosent og 0.1 prosent. Effekten av skilsmisse var som forventet og samsvarer med funnene til Fevang og Røed (2006). Derimot var effekten av inntekt mer uventet, men kan ha sammenheng med Bragstad og Hauge (2008) sitt funn av at økt inntektsnivå blant kvinner gir flere mottakere av uføretrygd. Tilslutt fant jeg en sterkt signifikant effekt av andelen i fylket med lav utdanning som høyeste utdanningsnivå på andelen uføre med 2 prosent. Resultatet tilsier at utdanningsnivå i fylket har svært mye å si for andelen uføre. Effekten av sykefravær var ikke signifikant og rapporterte et negativt fortegn, noe som var uventet.

Jeg utvidet så hovedmodellen ved å inkludere laggede verdier for én og to perioder av ledighet og sykefravær. Etter å ha ekskludert de variablene som samlet sett ikke hadde noen signifikant effekt i modellen, besto den endelige modellen av fokusvariabelen for ledighet lagget med én periode og variable for sykefravær lagget med én og to perioder. Jeg fant her en sterkt signifikant (0.001 nivå) og positiv effekt av arbeidsledighet lagget med én periode på andelen uføre med 0.05 prosent. Effekten var noe høyere enn tidligere, men fremdeles lav, noe som tilsier at effekten av lokal ledighet påvirker andelen uføre i relativt liten grad. I denne modellen fant jeg signifikante og positive effekter av sykefravær på rundt 0.07 prosent, som tilsier at det var riktig å inkludere laggede verdier av sykefravær. De øvrige variablene endret seg ikke mye, men effekten av skilsmisse var ikke lengre signifikant. For å undersøke robustheten til resultatene, samt om modellen var underspesifisert, utførte jeg en sensitivitetsanalyse av modellen ved å inkludere øvrige kontrollvariable. Jeg inkluderte kontrollvariable for attraktiviteten til fylker, høyere utdanning for å kontrollere for utdannings sammensetting, innvandring og partitilhørighet i fylkene. Jeg fant her at effekten av arbeidsledighet var robust med hensyn til inkludering av ytteligere variable,

og effekten holdt seg sterkt signifikant på 0.001 nivå i alle modellene. Effekten varierte fra 0.0489 prosent til 0.0666 prosent i den endelige modellen, men som tidligere fant jeg at effekten var svært lav og tilsier at lokal arbeidsledighet i fylket i liten grad påvirker andelen uføre. Blant de mest interessante funnene fra kontrollvariablene var utdannings sammensetningen i fylket. Her fant jeg at én prosents økning i andelen med lavt utdanningsnivå hadde en effekt på 1.72 prosent på andelen uføre. Videre fant jeg at variable for høyere utdanning også ga signifikante effekter på andelen uføre med 0.5 prosent for utdanning på bachelornivå og -0.4 prosent for utdanning på master- og Phd-nivå. Effektene tilsier at jo høyere andel høyt utdannede som bor i fylket, jo lavere vil andelen uføre være. Effekten av de med lav utdanningsnivå er klart sterkest, men effekten viser å holde seg positiv med et lavere nivå også for de med utdanning på bachelornivå. Det er først ved én prosents økning i andelen med utdanning på master- eller Phd- nivå at effekten reduserer andelen uføre. Videre fant jeg ved kontroll for partifordeling i fylkene at én prosents økning i andelen stemmer til SV i fylkesvalg økte andelen uføre med 0.7 prosent, mens tilsvarende for Ap ga en reduksjon i andelen uføre med -0.2 prosent. Selv om begge partiene kan karakteriseres som partier på venstresiden i politikken er Ap sentrumsnært mens SV ligger lengre ute på venstresiden.

Videre tok jeg utgangspunkt i den endelige modellen fra sensitivitetsanalysen og inkluderte lagget endogen variabel som høyresidevariabel i modellen. Dette gjorde jeg for å undersøke persistensen i andelen uføre. Jeg estimerte modellen med FE- og Arellano og Bond's GMM-metode, for å ta hensyn til såkalt Nickell-skjevhet. Jeg fant at begge modellene tilsa en høy grad av persistens i andelen uføre. Med denne modellspesifikasjonen kunne jeg også beregne langtidseffekter for de inkluderte forklaringsvariablene i modellen. Jeg fant en korttidseffekt av arbeidsledighet på 0.03 prosent ved begge estimeringsmetodene og langtidseffekter på henholdsvis 0.09 og 0.06 for FE- og AB- metoden. Langtidseffekten fra AB-estimeringen samsvarte godt med de tidligere effektene jeg fant.

Til slutt kontrollerte jeg for næringsstrukturen i fylkene. I denne delen av analysen brukte jeg datasettet separat i de to periodene 2000-2008 og 2008-2014, på grunn av SSB sine endringer i næringskoder fra SN2002 til SN2007. Jeg estimerte hovedmodellene til de to periodene og inkluderte 10 variable for næringsstruktur. I periode 1 fant jeg kun en signifikant effekt av andelen sysselsatt innenfor "varehandel, hotell- og restaurantvirksomhet" på -0.334, som vil tilsa at én prosent økning i andelen sysselsatte innenfor denne næringen reduserer andelen uføre med -0.3 prosent. Effekten av ledighet lagget med én periode fant jeg at var 0.0523 prosent som samsvarte med den tidligere analysen. I periode 2 fant jeg

en svakt signifikant effekt av andelen sysselsatte innenfor næringen ”jordbruk, skogbruk og fiske” med 0.123 prosent. Videre fant jeg en sterkt signifikant effekt (0.001 nivå) av andelen sysselsatte innenfor næringen ”offentlig forvaltning og annen tjenesteyting” med hele 1.194 prosent. I tillegg var effekten av uoppgitt næring signifikant med en effekt på 0.08 prosent. Effekten av ledighet lagget med to perioder var her på 0.0511 prosent for én prosents økning i andelen ledige, som samsvarer med resultatene fra periode 1 og den tidligere analysen. Samlet sett ga kontrollen for næringsstruktur lite signifikante resultater i analysen.

## 7.1 Avsluttende diskusjon

I denne oppgaven har jeg sett på effekten av faktisk registrert arbeidsledighet på andelen uføre i fylkene. Det vil si at jeg i motsetning til andre studier ikke har sett på spesifikke hendelser som nedbemanninger i og nedleggelse av for eksempel bedrifter eller fabrikker, som igjen fører til arbeidsledighet for de det gjelder. selv om jeg finner en positiv signifikant og robust effekt av ledighet i fylkene på andelen uføre, er denne effekten relativt svak som tilsier at faktisk ledighet i liten grad påvirker andelen uføre. Dette indikerer at sammenhengen mellom uføretrygd og ledighet er mer individspesifikk og at effekten generelt ikke er spesielt stor. Videre kan det tyde på at det er andre faktorer som bestemmer andelen uføre i fylkene. Av de sterkeste effektene, er andelen med lav utdanning i fylkene. Jeg finner en effekt på  $\approx 1.3$  prosent på andelen uføre for én prosents økning i andelen med lavt utdanningsnivå i fylket. Utdanningsnivå gir en indikator på hvilke muligheter et individ har i arbeidsmarkedet, som følgelig er begrensede dersom man har et lavt utdanningsnivå. I Bratsberg et al. (2010) finner forfatterene spesifikt at mindre muligheter på arbeidsmarkedet grunnet nedbemanninger øker sannsynligheten for å bli ufør. Denne sammenhengen kan dermed tenkes å strekke seg lengre enn kun nedbemanninger ved at personer med lavt utdanningsnivå generelt har mindre muligheter på arbeidsmarkedet og dermed øker andelen uføre. Bratsberg et al. (2010) finner at ledighet og uføretrygd er nære substitutter og jeg kan til en viss grad bekrefte dette med mine resultater. At andelen med lavt utdanningsnivå øker andelen uføre kan ses på som at ledighet og uføretrygd er substitutter ved at personer med lav utdanning heller substituerer seg til uføretrygd grunnet mindre muligheter på arbeidsmarkedet. Det kan dermed ses på som et ’valg’ mellom arbeidsledighet og uføretrygd blant de lavt utdannede. Charles et al. (2017) finner at verre arbeidsmarkedsforhold fører til at individer i stor grad substituerer lønninger med uføretrygd. Det kan tenkes at dette også gjelder her ved at lønninger for jobber som krever

lavt utdanningsnivå er lave og dermed kan individer substituere de lave lønningene med uføretrygd. Videre finner jeg signifikante effekter av andel godkjente stemmer til SV i fylkesvalg med en langtidseffekt på 1 prosent. Dette vil tilsi at dersom oppslutningen til SV øker i et fylke vil andelen uføre øke og at stemmegivningen til SV er med på å bestemme andelen uføre i fylkene. Dette er det nærmereste jeg kommer til å finne en effekt av at en "kultur" i fylket kan påvirke andelen uføre i fylket, ved at den "politiske kulturen" i fylket har en effekt på andelen uføre. Derimot kan denne effekten være omvendt, og det kan tenkes at andelen uføre kan være med på å bestemme oppslutningen til SV i fylkene. Siden SV ligger relativt langt ute på venstresiden av politikken og tradisjonelt sett har arbeidet for brede velferdsordninger kan dette tyde på at uføretrygdede har preferanser for slik politikk. Dette indikerer at jeg har et endogentitetsproblem i analysen ved såkalt to-veis kausalitet. Denne effekten er dermed problematisk å tolke rett frem. Sykefravær i fylkene har også en signifikant effekt på andelen uføre som tilsier at dersom sykefraværet i et fylke øker vil andelen uføre øke. Dette er som forventet fordi personer som blir uføre vil gjerne ha vært mye sykemeldt i forkant. Dette ville derimot ha vært interessant å undersøke på et mer detaljert nivå for å undersøke om det er spesielle næringer som skiller seg ut med høyt sykefravær og påfølgende høy andel som blir uføre. Fra min kontroll av næringsstrukturen i fylkene fant jeg få signifikante effekter. I periode 2 var effekten av andelen sysselsatte innenfor "offentlig forvaltning og annen tjenesteyting" signifikant med en effekt på hele 1.2 prosent. Det vil tilsi at fylker med mange sysselsatte innenfor denne næringen vil ha en økt andel uføre. En mulig forklaring for dette kan være at sysselsatte innenfor det offentlige har veldig trygge rammer og høyt stillingsvern, samtidig som lønningene gjennomsnittlig er lavere. Det kan tenkes at det dermed er mindre "kostbart" for individer å ta ut mye sykefravær som etterhvert kan føre til at man substituerer seg til uføretrygd, heller enn å returnere til arbeid. Dette er derimot kun spekulasjon fra min side.

### 7.2 Konklusjon

Fra min analyse kan jeg konkludere med at utviklingen i fylkesspesifikke ledighetsrater har signifikant og positiv effekt på andelen uføre. Jeg har funnet at effekten av arbeidsledighet har holdt seg robust gjennom de ulike modellspesifikasjonene med en stabil effekt på rundt 0.05-0.06 prosent på andelen uføre. Fra mine resultater kan jeg dermed konkludere med at arbeidsledighet i fylkene i realtvt liten grad påvirker andelen uføre i fylkene, og effekten av en økning i arbeidsledighet er liten. Sammenlignet med de øvrige faktorene



jeg har sett på tilsier mine resultater at utdanningssammensetning, og da spesielt andelen med lavt utdanningsnivå i et fylke, i størst grad påvirker andelen uføre. Jeg kan også konkludere med at det er en høy grad av persistens i andelen uføre i fylkene, som vil si at andelen uføre i stor grad bestemmes av tidligere andel uføre. I tidligere studier der spesielle hendelser som fører til ledighet er analysert er det funnet sterke effekter av arbeidsledighet og mindre muligheter på arbeidsmarkedet på uføretilbøyighet. Jeg mener derfor det kan være interessant å se videre på effekten av faktisk arbeidsledighet og mottakere av uføretrygd. I fremtidige analyser kan det være interessant å se på samme problemstilling isolert på nye mottakere av uføretrygd (tilgang til uføretrygd) og mer spesifikt innenfor næringsstrukturen i fylkene. Videre kan tematikken ses på i sammenheng med andre forklarende forhold på arbeidsmarkedene enn arbeidsledighet.



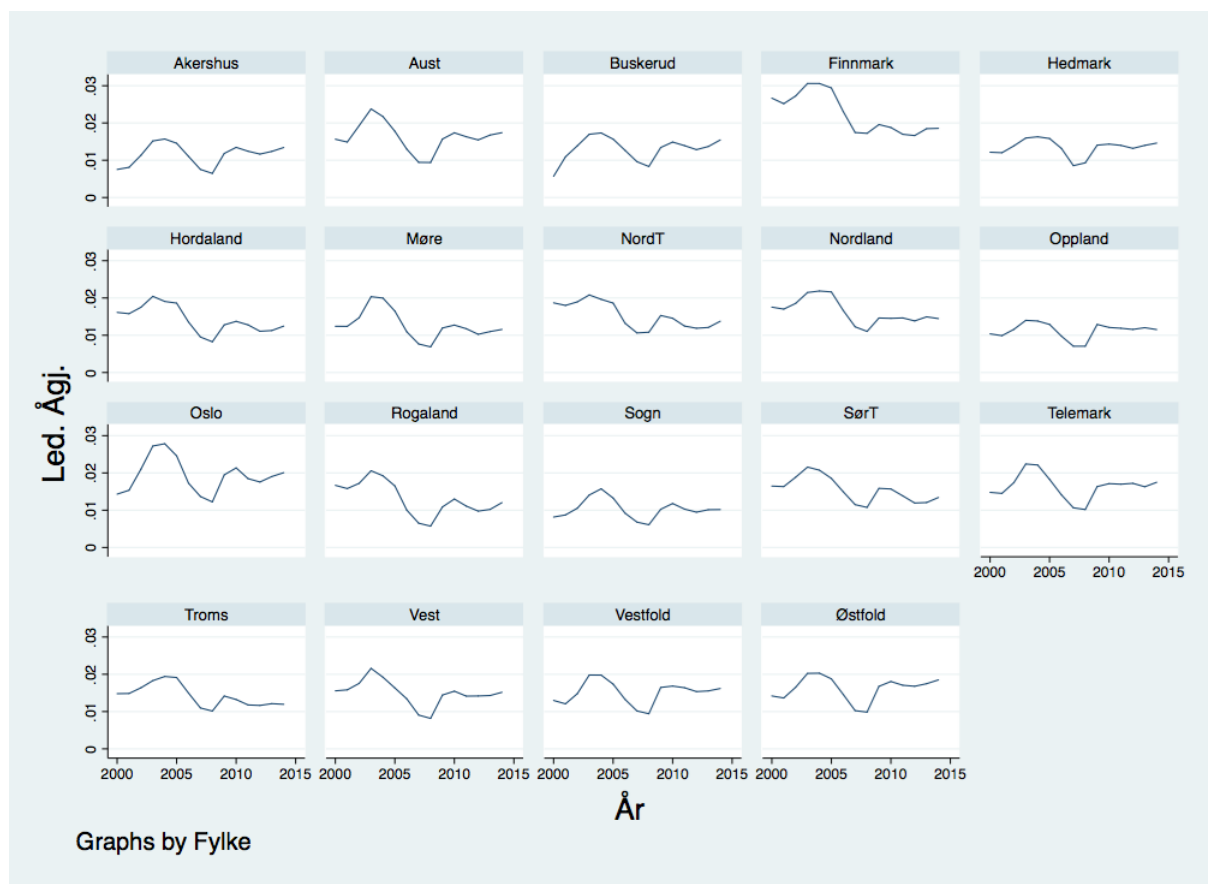
## Referanser

- Anderson, T. & Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American statistical Association*, 76(375), 598-606.
- Arellano, M. & Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Journal of the American statistical Association*, 58(2), 277-297.
- Black, D., Daniel, K. & Sanders, S. (2002, March). The impact of economic conditions on participation in disability programs: Evidence from the coal boom and bust. *American Economic Review*, 92(1), 27-50.
- Bragstad, T. & Hauge, L. (2008). Geografisk variasjon i uførepensjonering 1997-2004. *NAV rapport*(4).
- Bratberg, E., Nilsen, Ø.A. & Vaage, K. (2012, April). Is reciprocity of disability pension hereditary. *CESifo Working Paper*(3796).
- Bratsberg, B., Fevang, E. & Røed, K. (2010, April). Disability in the welfare state: An unemployment problem in disguise? *IZA DP*(4897).
- Bratsberg, B. & Røed. (2011). Kan demografi forklare veksten i uførhet? *Søkelys på Arbeidslivet*(1-2), 3-21.
- Charles, K.K., Li, Y. & Stephens, M.J. (2017). Disability benefit take-up and local labor market conditions.
- Fevang, E. & Røed, K. (2006). Veien til uføretrygd i norge. , 10. Hentet fra <https://www.ssb.no/a/publikasjoner/pdf/nosd383/nosd383.pdf>
- Jacobsen, O. (2014). Pensjonsreformen: Hvilken innvirkning har den hatt på bruken av helserelaterede ytelser? *Arbeid og velferd*(3).
- NAV. (2015). Utviklingen i uførepensjon per 31.desember 2014. *Statistikknotat*.
- NAV. (2017a, Mars). Utviklingen i uførediagnoser per 31. desember 2014. *Statistikknotat*.
- NAV. (2017b, September). Utviklingen i uføretrygd per 30. september 2017. *Statistikknotat*.
- NAV. (2017c, Mai). Utviklingen i uføretrygd per 31. mars 2017. *Statistikknotat*.
- OECD. (2010, November). Sickness, disability and work. *Sickness, Disability and Work: Breaking the Barriers*.
- OECD. (2017, Juni). Oecd employment outlook 2017.
- Rege, M., Telle, K. & Votruba, M. (2005). The effect of plant downsizing on disability pension utilization. *Discussion Papers, Statistics Norway, Research Department*(435).
- Rege, M., Telle, K. & Votruba, M. (2007, September). Plant closure and marital dissolu-

- tion. *Discussion papers*(514).
- SSB. (2007, November). Standard for næringsgruppering.
- Verbeek, M. (2012). *A guide to modern econometrics* (4. utg.). Wiley.
- Wooldridge, J.M. (2013). *Introductory econometrics a modern approach* (5. utg.). South-Western CENGAGE Learning.

# Appendiks

Figur 5 - Variasjon i arbeidsledighet på fylkesnivå



Figur 5: Variasjon i ledighet 2000-2014, fylkesnivå

## Delkapittel 4.3.2 Aggregerte hovedgrupper SN2002 og SN2007

Tabell 15: Aggregerte hovedgrupper SN2002 og SN2007

SN2002	Beskrivelse	SN2007	Beskrivelse
01-05	Jordbruk, skogbruk og fiske	01-03	Jordbruk, skogbruk og fiske
11	Utvinning av råolje og naturgass	05-09	Bergverksdrift og utvinning
10,12-37	Industri og bergverksdrift	10-33	Industri
40-41	Kraft- og vannforsyning	35-39	Elektrisitet, vann og renovasjon
45	Bygge- og anleggsvirksomhet	41-43	Bygge- og anleggsvirksomhet
50-55	Varehandel, hotell- og restaurantvirksomhet	45-47	Varehandel, reparasjon av motorvogner
60-64	Transport og kommunikasjon	49-53	Transport og lagring
65-67	Finansiell tjenesteyting	55-56	Overnattings- og serveringsvirksomhet
70-74	Forretningsmessig tjenesteyting, eiendomsdrift	58-63	Informasjon og kommunikasjon
75-99	Offentlig forvaltning og annen tjenesteyting	64-66	Finansiering og forsikring
00	Uoppgitt	68-75	Teknisk tjenesteyting, eiendomsdrift
		77-82	Forretningsmessig tjenesteyting
		84	Off.adm., forsvar, sosialforsikring
		85	Undervisning
		86-88	Helse- og sosialtjenester
		90-99	Personlig tjenesteyting
		00	Uoppgitt

## Delkapittel 4.4.2) VIF-verdier for kontrollvariable

For å finne frem til VIF-verdier for kontrollvariable har jeg sett på alle variablene i en korrelasjonsmatrise og plukket ut de variablene som viste høye korrelasjoner mellom hverandre. VIF-verdier for disse variablene presenteres dermed i tabell 16.

Tabell 16: VIF-verdier for kontrollvariable

Variabel	VIF
UtdHkort	11.50
UtdHlang	7.97
StemAPF	7.44
StemFrpF	7.34
AsiaMT	6.90
Af	6.17
StemHF	5.82
StemSVF	5.20
EurUT	5.01
Oz	4.96
NAm	4.02
SMAm	4.02
Nettoinnpen	2.07

---

### Delkapittel 4.4.3) Korrelasjon for øvrige kontrollvariable

Tabell 17: Korrelasjon for øvrige kontrollvariable

	Andel_U	Ledig
Nettoinnfl	-0.1308	-0.1715
Nettoinnpen	-0.3344	0.2581
UtdHkort	-0.4699	-0.0297
UtdHlang	-0.5621	0.0271
EurUT	-0.3149	0.0071
Af	-0.3242	0.1346
AsiaMT	-0.4480	0.1293
NAm	-0.3903	0.0009
SMAm	-0.5315	-0.0007
Oz	-0.6049	-0.0198
StemSVF	0.0394	0.1577
StemAPF	0.0381	-0.0179
StemHF	-0.0835	-0.0180
StemFrpF	0.0125	-0.0180

### Delkapittel 5.1.3) Varians-kovarians materisen til (5):

$$\Omega_i = \begin{bmatrix} \sigma_\alpha^2 + \sigma_\eta^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 + \sigma_\eta^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 + \sigma_\eta^2 \end{bmatrix} \quad (33)$$

Som gir den komplette varians-kovariansmatrisen som er blokkdelt:

$$\Omega_i = \begin{bmatrix} \Omega_i & 0 & \cdots & 0 \\ 0 & \Omega_i & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \Omega_i \end{bmatrix} = \Omega_i \otimes I_N \quad (34)$$

## Delkapittel 4.4.1) Utvalgskarakteristikker og variasjon

Tabell 18: Øvrige kontrollvariable, hele perioden.

Variabel	Variasjon	Gjennomsnitt	St.avvik	Min	Max
Nettoinnp	Overall	-0.0182826	0.0576989	-0.1050996	0.2741262
	Between		0.0585144	-0.07582	0.2065824
	Within		0.0086029	-0.05619	0.0492612
Nettoinnf	Overall	-0.0010839	0.0041569	-0.0147315	0.0096528
	Between		0.004008	-0.0100367	0.0051106
	Within		0.0014169	-0.0057786	0.0037327
UtdHkort	Overall	0.1494121	0.0249707	0.1083519	0.2451932
	Between		0.020647	0.1315345	0.2199095
	Within		0.0147734	0.1231025	0.179228
UtdHlang	Overall	0.0412201	0.0222243	0.0186233	0.1509261
	Between		0.021109	0.0257359	0.111382
	Within		0.0083842	0.0135184	0.0807641
EurUT	Overall	0.0370346	0.0173231	0.0125689	0.1037665
	Between		0.0117889	0.0191846	0.0698703
	Within		0.0129601	0.0184124	0.0741634
Af	Overall	0.0071661	0.0060699	0.0010149	0.0359028
	Between		0.0060699	0.0039862	0.0293403
	Within		0.0028785	-0.0017827	0.0155
AsiaMT	Overall	0.0205343	0.0156512	0.0049879	0.086533
	Between		0.0152027	0.0093648	0.0776994
	Within		0.0050233	0.0061296	0.033823
NAm	Overall	0.0015254	0.0009297	0.000216	0.0041364
	Between		0.0009396	0.0003208	0.0034385
	Within		0.0001587	0.0009556	0.0022361
SMAm	Overall	0.0023445	0.0016725	0.0002518	0.0086483
	Between		0.0016375	0.0006133	0.0069582
	Within		0.0004981	0.0006596	0.004082
Oz	Overall	0.0002199	0.000156	0.0000405	0.0008919
	Between		0.0001399	0.0000943	0.0006605
	Within		0.0000757	-0.0000543	0.0004849
StemAPSt	Overall	0.04772	0.0822993	0	0.2679959
	Between		0.0094929	0.0330513	0.06669
	Within		0.0817772	-0.01897	0.2490259
StemHSt	Overall	0.026682	0.048429	0	0.1921967
	Between		0.0074571	0.0152249	0.0416716
	Within		0.04788	-0.0149896	0.1772071



Tabell 19: Periode 1 - Hovedmodell

Variabel	Variasjon	Gjennomsnitt	St.avvik	Min	Max
Andel uføre	Overall	0.0669726	0.0123073	0.0454571	0.0839382
	Between		0.0124452	0.0483113	0.081146
	Within		0.0019681	0.0604988	0.0716419
Ledig	Overall	0.015172	0.0052049	0.0056965	0.0306064
	Between		0.0034135	0.0102913	0.025266
	Within		0.0039984	0.0066485	0.0237262
Sykefr	Overall	0.0262747	0.0030548	0.0173127	0.0325461
	Between		0.0024247	0.0202374	0.0291312
	Within		0.0019311	0.0208862	0.0305624
Brutto	Overall	0.0526316	0.0469996	0.0113083	0.2299532
	Between		0.0480623	0.0117665	0.2184189
	Within		0.0027724	0.037175	0.0641659
Inntbr	Overall	273423.4	44328.32	196600	399200
	Between		29820.79	241177.8	344066.7
	Within		33430.05	215190.1	352390.1
Skillsm	Overall	0.0022379	0.000376	0.0012641	0.0029706
	Between		0.0003504	0.0013667	0.0028071
	Within		0.0001562	0.001844	0.0026772
UtdL	Overall	0.6072148	0.0433661	0.4458236	0.6735435
	Between		0.0428902	0.4781241	0.6566078
	Within		0.0112967	0.5710922	0.6376006

Tabell 20: Periode 2 - Hovedmodell

Variabel	Variasjon	Gjennomsnitt	St.avvik	Min	Max
Andel uføre	Overall	0.0658193	0.0139088	0.0380261	0.0863471
	Between		0.0141216	0.0407222	0.0833995
	Within		0.0017601	0.060353	0.070583
Ledig	Overall	0.0135674	0.0031372	0.0057307	0.0213441
	Between		0.0024612	0.0097449	0.0183131
	Within		0.002015	0.0070606	0.0165985
Sykefr	Overall	0.027115	0.0024675	0.0218007	0.0325461
	Between		0.0022953	0.0222849	0.0305604
	Within		0.0010293	0.0244476	0.0300137
Brutto	Overall	0.0526462	0.0468031	0.0113083	0.2141324
	Between		0.0478779	0.0121784	0.2094901
	Within		0.0015605	0.0439919	0.0572885
Inntbr	Overall	362797.7	40998.26	297800	481000
	Between		31874.43	327428.6	429657.1
	Within		26665.69	320554.9	418354.9
Skillsm	Overall	0.0020186	0.0003053	0.0012298	0.0026969
	Between		0.0002678	0.0014118	0.0023926
	Within		0.0001574	0.0016297	0.0024336
UtdL	Overall	0.5809692	0.0486217	0.4051984	0.6453148
	Between		0.0491408	0.4268613	0.6344711
	Within		0.0076826	0.5593062	0.5999314

Tabell 21: Periode 1 - Næringsstruktur

Variabel	Variasjon	Gjennomsnitt	Standardavvik	Min	Max
JSF1	Overall	0.0218281	0.0126372	0.0012287	0.0506026
	Between		0.0128503	0.0013038	0.0465572
	Within		0.0015297	0.0167258	0.0280096
IUB2	Overall	0.0689832	0.0214436	0.0334291	0.1178981
	Between		0.0215911	0.036106	0.1121729
	Within		0.0039485	0.0613955	0.0799156
KV3	Overall	0.0038569	0.0013486	0.0014109	0.0072033
	Between		0.0013068	0.0017347	0.0067084
	Within		0.0004374	0.0026988	0.0053943
BA4	Overall	0.0366245	0.0053087	0.0195438	0.0482528
	Between		0.004764	0.0212658	0.0440995
	Within		0.0025602	0.0324477	0.0438755
VHR5	Overall	0.0898588	0.0104532	0.0733943	0.1195812
	Between		0.0103307	0.0770601	0.1166647
	Within		0.0027508	0.0844672	0.0988393
TK6	Overall	0.0339093	0.0050478	0.024572	0.0490952
	Between		0.0049692	0.0260037	0.0453835
	Within		0.0013963	0.0310512	0.0381209
FT7	Overall	0.0086086	0.003711	0.0041437	0.0213869
	Between		0.0037781	0.0045617	0.0193949
	Within		0.0004117	0.0072868	0.0106006
FTE8	Overall	0.0459606	0.0189609	0.0230222	0.116874
	Between		0.0188375	0.027534	0.1033358
	Within		0.0046218	0.0366641	0.0607482
OFT9	Overall	0.190159	0.0185605	0.1570159	0.2457483
	Between		0.0177921	0.1663805	0.2457483
	Within		0.0065445	0.1779365	0.2030005
U10	Overall	0.0027973	0.0007167	0.0016607	0.0063599
	Between		0.0005731	0.0020619	0.0047549
	Within		0.0004479	0.0013681	0.0044023

Tabell 22: Periode 2 - Næringsstruktur

Variabel	Variasjon	Gjennomsnitt	Standardavvik	Min	Max
JSF'1	Overall	0.0179669	0.0108254	0.0007885	0.043271
	Between		0.0109269	0.0009465	0.0379827
	Within		0.001794	0.012978	0.0232553
IBU2	Overall	0.0604575	0.0220326	0.0216832	0.1163949
	Between		0.0223768	0.0223744	0.1127042
	Within		0.0027342	0.0538328	0.0703546
EVR3	Overall	0.0062741	0.0017428	0.0032008	0.0107833
	Between		0.0017651	0.0036381	0.010146
	Within		0.0002513	0.0055391	0.0069114
BA4	Overall	0.0420017	0.0051372	0.0232986	0.0516132
	Between		0.0051157	0.0250799	0.0493047
	Within		0.0011873	0.039541	0.0446007
VHR5	Overall	0.0884056	0.0085878	0.0758326	0.1176029
	Between		0.0083876	0.0778019	0.1097713
	Within		0.0025682	0.083407	0.0962372
TLIK6	Overall	0.0420678	0.0108344	0.0303728	0.0815991
	Between		0.0110191	0.0312891	0.0771016
	Within		0.0012177	0.0389142	0.0475186
FF7	Overall	0.0081098	0.0039802	0.0029784	0.0213762
	Between		0.0040454	0.0035086	0.0199022
	Within		0.0004699	0.0062843	0.0095837
FTTE8	Overall	0.0506689	0.014193	0.0335293	0.0947824
	Between		0.014437	0.034703	0.0913063
	Within		0.0015777	0.0469772	0.0548612
OFT9	Overall	0.1958738	0.0171407	0.1676701	0.2446272
	Between		0.0173644	0.1713033	0.241899
	Within		0.0024472	0.1858495	0.2009561
U10	Overall	0.0025782	0.0005002	0.0016805	0.0039941
	Between		0.0004561	0.0019765	0.003621
	Within		0.0002272	0.0020341	0.0031194

---

## Delkapittel 6.1) Hovedmodell - Pooled OLS vs. Random effects

---

Tabell 23: Hovedmodell hele perioden. Pooled OLS vs Random effects

	Pooled OLS (1)	RE/GLS (2)
	lnAndel_U	lnAndel_U
lnLedig <sub>it</sub>	0.180*** (0.0391)	0.0310*** (0.00874)
lnSykefr <sub>it</sub>	0.510** (0.135)	-0.0356 (0.0309)
lnBrutto <sub>it</sub>	-0.124** (0.0370)	-0.0705 (0.0581)
lnInntbr <sub>it</sub>	0.160 (0.0845)	0.286*** (0.0635)
lnSkilsm <sub>it</sub>	0.301** (0.101)	0.140*** (0.0366)
lnUtdL <sub>it</sub>	1.008* (0.365)	1.873*** (0.355)
_cons	-0.147 (1.034)	-4.725*** (0.577)
N	285	285

Standardavvik i parentes, korrigert for clustering

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

### Breusch-Pagan LM-test for Random effects

$$H_0 : \text{Var}(\mathbf{u}) = 0$$

$$\text{Testovervator } \chi_{0.05}^2 = 933.64$$

$$\text{P-verdi} = 0.0000$$

$H_0$  kan klart forkastes.

---

### Delkapittel 6.3) Inkludering av lagget endogenvariabel

---

Tabell 24: Inkludering av lagget endogen variabel - Modell estimert med pooled OLS

	(1)
	lnAndel_U
lnAndel_U <sub>it-1</sub>	0.958*** (0.0221)
lnLedig <sub>it-1</sub>	0.0151* (0.00615)
lnSykefr <sub>it-1</sub>	0.0207 (0.0143)
lnSykefr <sub>it-2</sub>	0.0320* (0.0117)
lnBrutto <sub>it</sub>	-0.00479 (0.00339)
lnInntbr <sub>it</sub>	0.0234 (0.0191)
lnSkilsm <sub>it</sub>	-0.00667 (0.0136)
lnUtdL <sub>it</sub>	0.101 (0.0610)
Nettoinnfl <sub>it</sub>	0.395 (0.401)
lnUtdHkort <sub>it</sub>	-0.0643 (0.0367)
lnUtdHlang <sub>it</sub>	0.0129 (0.00977)
StemSVF <sub>it</sub>	0.535** (0.143)
StemAPF <sub>it</sub>	-0.0794 (0.0384)
_cons	-0.241 (0.307)
N	247

Standardavvik i parentes, korrigert for clustering

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

## Delkapittel 6.4) Hovedmodell og sensitivetsanalyse for periode 1 og periode 2

Tabell 25: Hovedmodell, inkludering av laggede eksogene variable og endelig modell - Periode 1 og 2

	Periode 1 (1)	Periode 1 (2)	Periode 1 (3)	Periode 2 (4)	Periode 2 (5)	Periode 2 (6)
	lnAndel_U	lnAndel_U	lnAndel_U	lnAndel_U	lnAndel_U	lnAndel_U
LnLedig <sub>it</sub>	0.0536*** (0.0118)	0.00304 (0.0211)		-0.0274* (0.0126)	-0.0342 (0.0564)	
lnLedig <sub>it-1</sub>		0.0774*** (0.0192)	0.0672*** (0.0111)		-0.0839 (0.0755)	
lnLedig <sub>it-2</sub>		-0.0318 (0.0175)			0.0441 (0.0300)	0.0363** (0.0122)
lnSykefr <sub>it</sub>	-0.0528* (0.0187)	-0.0170 (0.0311)		-0.142 (0.0792)	-0.0691 (0.106)	
lnSykefr <sub>it-1</sub>		0.0616* (0.0244)	0.0886*** (0.0133)		-0.188 (0.0942)	-0.211* (0.0808)
lnSykefr <sub>it-2</sub>		0.0350 (0.0235)	0.0631** (0.0171)		0.0946 (0.101)	
lnBrutto <sub>it</sub>	0.0180 (0.0715)	-0.0607 (0.0916)	-0.0315 (0.0923)	-0.359** (0.118)	-0.426** (0.142)	-0.316* (0.129)
lnInntbr <sub>it</sub>	0.331*** (0.0637)	0.245*** (0.0517)	0.225*** (0.0570)	0.107 (0.0796)	-0.155 (0.109)	-0.127 (0.0881)
lnSkilsm <sub>it</sub>	0.0784* (0.0346)	0.0322 (0.0302)	0.0281 (0.0292)	0.101* (0.0354)	0.0403 (0.0351)	0.0492 (0.0285)
lnUtdL <sub>it</sub>	1.870*** (0.471)	1.815*** (0.426)	1.885*** (0.390)	0.607 (0.339)	0.739 (0.396)	0.489 (0.270)
_cons	-5.356*** (0.709)	-4.376*** (0.813)	-3.676*** (0.677)	-4.944*** (0.880)	-2.376 (1.363)	-2.160* (0.820)
N	171	133	133	133	95	95

Standardavvik i parents,

korrigert for clustering

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Tabell 26: Sensitivitetsanalyse periode 1 - inkludering av øvrige kontrollvariable.

	(1)	(2)	(3)	(4)
	lnAndel_U	lnAndel_U	lnAndel_U	lnAndel_U
lnLedig <sub>it-1</sub>	0.0523*** (0.0110)	0.0436** (0.0112)	0.0267* (0.0104)	0.0284** (0.00943)
lnSykefr <sub>it-1</sub>	0.0708* (0.0296)	0.0495 (0.0300)	0.0738* (0.0286)	0.0660* (0.0296)
lnSykefr <sub>it-2</sub>	0.0631*** (0.0148)	0.0703*** (0.0150)	0.107*** (0.0219)	0.110*** (0.0198)
lnBrutto <sub>it</sub>	-0.0337 (0.0916)	-0.0102 (0.0838)	0.0194 (0.0746)	0.00734 (0.0725)
lnInntbr <sub>it</sub>	0.216** (0.0557)	0.165** (0.0488)	0.0308 (0.0689)	-0.00209 (0.0554)
lnSkilsm <sub>it</sub>	0.0123 (0.0287)	0.0222 (0.0288)	0.0122 (0.0258)	0.00876 (0.0274)
lnUtdL <sub>it</sub>	2.093*** (0.449)	2.048*** (0.451)	1.923*** (0.460)	2.111*** (0.406)
lnVHR5 <sub>it</sub>	-0.334** (0.106)	-0.296* (0.106)	-0.306* (0.114)	-0.349** (0.114)
lnOFT9 <sub>it</sub>	0.388 (0.236)	0.471 (0.268)	0.173 (0.245)	0.146 (0.274)
Nettoinnfl <sub>it</sub>		-4.270** (1.417)	-4.074** (1.144)	-3.791** (1.247)
Nettoinnpen <sub>it</sub>		0.399 (0.856)		
lnUtdHkort <sub>it</sub>			0.603* (0.234)	0.484** (0.140)
lnUtdHlang <sub>it</sub>			-0.170 (0.217)	
_cons	-3.847*** (0.893)	-2.959** (0.806)	-1.068 (1.275)	-0.437 (0.829)
<i>N</i>	133	133	133	133

Standardavvik i parentes, korrigert for clustering.

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Tabell 27: Sensitivitetsanalyse periode 2 - inkludering av øvrige kontrollvariable

	(1)	(2)	(3)
	lnAndel_U	lnAndel_U	lnAndel_U
lnLedig <sub>it-2</sub>	0.0511*** (0.0108)	0.0514*** (0.0103)	0.0419** (0.0134)
lnSykefr <sub>it-1</sub>	-0.0834 (0.0691)	-0.0813 (0.0712)	-0.119 (0.0614)
lnBrutto <sub>it</sub>	-0.185 (0.136)	-0.188 (0.136)	-0.214 (0.127)
lnInntbr <sub>it</sub>	0.0435 (0.0923)	0.0598 (0.0831)	0.397 (0.279)
lnSkilsm <sub>it</sub>	0.0566* (0.0269)	0.0580* (0.0247)	0.0578 (0.0299)
lnUtdL <sub>it</sub>	0.279 (0.235)	0.247 (0.223)	0.806 (0.412)
lnJSF1 <sub>it</sub>	0.123* (0.0500)	0.138** (0.0443)	0.0997 (0.0576)
lnOFT9 <sub>it</sub>	1.194*** (0.303)	1.244*** (0.313)	1.153*** (0.279)
lnU10 <sub>it</sub>	0.0816** (0.0271)	0.0848** (0.0261)	0.0812* (0.0297)
Nettoinnfl <sub>it</sub>		2.051 (2.056)	
Nettoinnpen <sub>it</sub>		-0.966 (1.403)	
lnUtdHkort <sub>it</sub>			-0.391 (0.637)
lnUtdHlang <sub>it</sub>			-0.0624 (0.158)
_cons	-0.511 (0.867)	-0.577 (0.798)	-6.070 (3.874)
<i>N</i>	95	95	95

Standardavvik i parentes, korrigert for clustering

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$