

Forord

Jeg vil takke min veileder Per Tovmo som har vist stor tålmodighet under denne prosessen.

Jeg vil også takke mine foreldre Anne og Rolf-Egil, samt mine grandtanter Marit og Solveig.

Deres støtte har hjulpet meg over målstreken.

Alle feil er mine egne.

Sammendrag

Sykefravær er kostbart for samfunnet og det er derfor ønskelig å forstå variasjonene bedre. I den sammenheng er det foreslått at det er en sammenheng mellom konjunktorene og sykefraværet. I en lavkonjunktur disiplineres arbeiderne til å skjerpe sykefraværet sitt samtidig som marginale arbeidere med dårligere helse holdes utenfor arbeid. Dette fører etter teorien til en negativ sammenheng mellom arbeidsledigheten og sykefraværet. I denne oppgaven undersøker jeg dette forholdet på fylkesnivå med et paneldatasett for 4. kvartal 2000-2016. Analysen bygger på en statisk teorimodell for etterspørsel etter konsum og fritid. Jeg benytter også oljeprisfallet i 2014 som et naturlig eksperiment til å bygge opp under analysen. Mot forventning finner jeg et positivt forhold mellom arbeidsledigheten og sykefraværet. Jeg finner også at reformen for legemeldinger i 2004 bidro til å redusere fraværet, samtidig som den positive effekten av arbeidsledigheten på sykefraværet ble større. Jeg argumenterer for at dette stemmer overens med at både disiplinerings- og sammensettingshypotesen påvirker sykefraværet, men at de marginale arbeiderne bidrar til å trekke sykefraværet ned istedenfor opp. Jeg har derimot ikke muligheten til å teste denne konklusjonen direkte med mine data. Jeg foreslår derfor en utvidelse av teorimodellen som tar høyde for denne muligheten som et startpunkt for videre analyse.

Abstract

Sickness absence represents a large cost to society and it is therefore important to understand what contributes to it. A relationship between the business cycle and sickness absence has been suggested. When the economy is doing poorly workers are disciplined to reduce their absence, while marginal workers with poor health are less likely to find employment. In theory this leads to a negative relationship between unemployment and sickness absence. In this paper I examine this relationship on the county level in Norway using panel data for the 4th quarter 2000-2016. My econometric model is based on a theoretical model of the demand for consumption and leisure. I also include a natural experiment based on the disproportionate effect of the 2014 oil price shock in some counties. Contrary to expectations I find a positive relationship between unemployment and absence. I also find that the 2004 reform on doctor approved sickness absence practices had a negative effect on absence, while also increasing the positive effect of unemployment. I

argue that this can be explained by both discipline- and composition effects being important, but that the composition effect is in fact that absence decreases when marginal workers are employed. I am not able to test this directly using my data. I therefore suggest an extension to the theory model which accounts for this, which is intended to be a starting point for further investigation.

INNHold

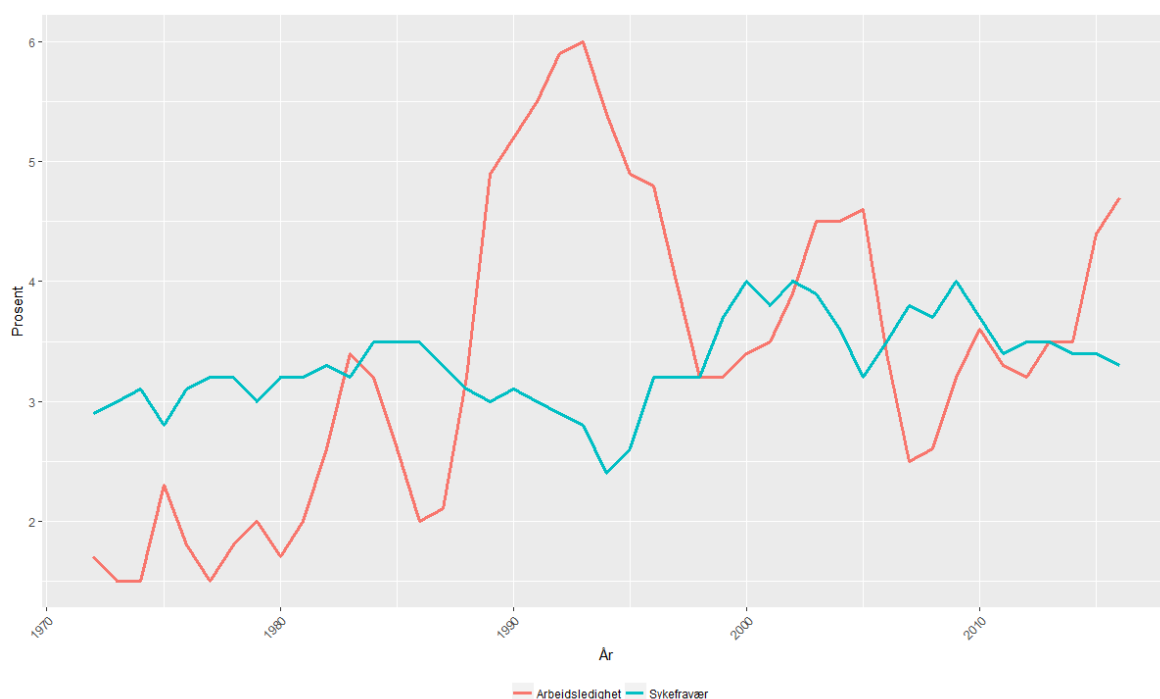
1	Innledning.....	1
2	Litteratur	3
2.1	Sykefravær og konjunkturer	3
2.2	Konjunkturer og sykefravær i Norge	5
2.3	Kjønnforskjeller i sykefraværet	6
2.4	Andre faktorer som påvirker sykefraværet	7
3	Teori.....	9
3.1	Sykefravær på fylkesnivå	12
4	Empirisk spesifikasjon og datamateriale.....	15
4.1	Valg av estimeringsmetode	15
4.1.1	Pooled Minste Kvadraters Metode	15
4.1.2	Fixed Effects	16
4.1.3	Random effects	17
4.1.4	Valget mellom fixed or random effects	18
4.2	Spesifikasjonstester	19
4.2.1	Durbin-Watson for paneldata	19
4.2.2	Wooldridge test for ikke-observerte effekter.....	19
4.3	Mer om variablene	20
4.3.1	Sykefravær.....	22
4.3.2	Arbeidsledighet	24
4.3.3	Kjønn.....	28
4.3.4	Alder	29
4.3.5	Næring.....	31

4.3.6	Utdanningsnivå.....	33
4.3.7	Dummyvariabler.....	34
4.4	Kvasieksperiment.....	34
5	Resultater	39
5.1	Kvasieksperiment.....	44
6	Oppsummering og konklusjon	47
7	Litteratur	49
8	Vedlegg.....	53
8.1	Fylkeseffekter og deskriptiv statistikk på fylkesnivå	53
8.2	Datakilder	56
8.3	Komparative statiske resultater	57

1 INNLEDNING

Diskusjoner rundt sykefraværet motiveres gjerne av økonomiske forhold. Et statlig sykeforsikringssystem som vi har i Norge er direkte kostbart for staten og dermed også alle skattebetalere. I 2016 var utgiftene til sykepenger 40,8 milliarder ifølge statsbudsjettet. I tillegg kan de økonomiske effektene av sykefravær være betydelige for bedrifter. Hem (2011) finner i sin rapport for NHO at en ukes sykefravær koster bedriftene 13000 kroner.

Figur 1: Arbeidsledighet og sykefravær (AKU)



I denne sammenhengen har forholdet mellom arbeidsledigheten og sykefraværet vært mye studert. Figur 1 viser arbeidsledigheten og sykefraværet som målt i Arbeidskraftundersøkelsen fra 1972-2016. Dette forholdet har ofte vært observert som negativt, og hovedhypotesen er at forholdene i arbeidsmarkedene virker disiplinerte på arbeidere. Arbeiderne blir straffet for sitt sykefravær gjennom for eksempel høyere sannsynlighet for å miste jobben. Denne effekten tenker man at er sterkere under en lavkonjunktur av naturlige årsaker. Videre vil sammensetningen av arbeidere variere mellom en lav- og høykonjunktur, spesielt med tanke på andelen ansatte som kan betraktes som marginale arbeidere. Her

antas det at disse arbeiderne av forskjellige årsaker har høyere sykefravær enn andre, noe som også kan bidra til en negativ sammenheng mellom arbeidsledigheten og sykefraværet. Tidligere norske studier har funnet noe støtte for disiplineringshypotesen, men har i mindre grad funnet støtte for sammensetningshypotesen. Intensjonen med denne oppgaven er å studere den totale sammenhengen på fylkesnivå.

For å oppnå dette har jeg satt sammen et panel med data for 4. kvartal fra 2000-2016. Dette er ikke mange perioder, og en tidsserieanalyse på landsbasis vil trolig være ganske svak. Dermed håper jeg at å utnytte variasjonen i både ledigheten og fraværet innad i fylkene kan gi bedre estimater. Ideelt sett får jeg på denne måten 19 ganger så mange observasjoner som jeg ville hatt på landsbasis, men siden mye av variasjonen er delt blir nok ikke resultatet fullt så bra. Jeg vil også gjennomføre et kvasiekperiment basert på at arbeidsmarkedene i fylkene har forskjellig eksponering til oljeprisen. Denne falt kraftig og ganske uventet i 2014 og representerer dermed et naturlig eksperiment.

Samlet finner jeg en positiv sammenheng mellom arbeidsledigheten og sykefraværet. Jeg skiller ikke mellom disiplinerings- og sammensetningshypotesen formelt, men argumenterer for at funnene mine kan stemme overens med at begge effektene eksisterer. Argumentet er at sammensetningseffekten er positiv i retning. Dette kan for eksempel skyldes at de marginale arbeiderne blir straffet hardere hvis de er fraværende, og derfor vil ha lavere fravær for et gitt nivå på arbeidsledigheten. Jeg foreslår derfor at det bør skilles mellom individer også i straffefunksjonen i teorimodellen. Jeg mener også det hadde vært en interessant ting å studere direkte som en oppfølging til det jeg har gjort her.

2 LITTERATUR

2.1 SYKEFRAVÆR OG KONJUNKTURER

To sentrale hypoteser om sammenhengen mellom konjunkturer og sykefravær kan spores tilbake til Leigh (1985): når arbeidsledigheten er høy er det sannsynlig at (1) de som mister jobben har høyere sykefravær enn de som beholder den, og (2) de som beholder jobben sin vil skjerpe inn sykefraværet sitt av frykt for å miste den. Dette blir referert til som henholdsvis sammensetnings- og disiplineringshypotesen.

Leigh (1985) finner støtte for hypotesen sin på både mikro- og makronivå – begge hentet fra universitetet i Michigan sin «Panel Study of Income Dynamics». For mikroanalysen sammenlignet han individers sykefraværsadferd i 1976 og 1977 basert på intervjuer i 1977 og 1978. For makroanalysen aggregerte han data for individer til et nasjonalt mål på sykefravær fra 1967 til og med 1978.

Markham (1985) brukte nasjonale månedlige data fra 1976 til 1982 fra «the Bureau of National Affairs» (BNA). Fraværet som ble sett på var alt ikke-planlagt fravær, det vil si eksklusive ferier, helligdager eller annet planlagt fravær, som varte i en hel dag eller lengre. Dette ble sammenlignet med arbeidsledighetstall fra «the U.S. Bureau of Labor Statistics» (BLS). Data for arbeidsledighet var sesongjustert, mens data for fravær ikke var det. Dette ble kontrollert for i modellen med sesong-dummyer.

På et regionalt nivå brukte Markham (1985) BNA sine data for de fire geografiske regionene nordøst, sør, sentral og vest¹. For arbeidsledigheten ble det konstruert et mål for de samme regionene basert på en vektning av individuelle stater i området sin arbeidsledighet og arbeidsstyrke.

¹ Jeg kan bare anta dette er en inndeling av fastlands-USA.

Markham (1985) fortsetter helt ned til et organisatorisk nivå. Én fabrikk i tekstilindustrien ble valgt ut. Denne ga tilgang på månedlig data som passet data fra BNA. Fabrikken hadde for det meste kvinner blant sine 400 ansatte, og tilbydde ingen sykelønnsordning.

Resultatene i Markham (1985) viste en klar negativ sammenheng mellom arbeidsledigheten og sykefraværet. Dette var tilfellet på alle tre nivå: nasjonalt, regionalt og for den utvalgte fabrikken. Det eneste unntaket var nordøst-USA. Forfatteren spekulerer i at dette kan ha noe å gjøre med at regionen har få stater med «right-to-work»-lover, og at regionen har den høyeste konsentrasjonen av fagforeninger. Dette er derimot ikke en del av analysen.

Yaniv (1991) utvikler en teoretisk modell som tar høyde for dynamiske faktorer. En av disse faktorene er at sykefraværsadferden i dag avhenger av arbeidsforhold i fremtidige perioder. Når tiden passerer er det nødvendigvis færre perioder igjen i arbeidslivet, f.eks. til pensjonsalderen, noe som betyr at kostnaden ved sykefravær går ned med tiden. Å være arbeidsledig i ett år er, alt annet likt, bedre enn å være arbeidsledig i to år når det kommer til inntjening. Derfor kan man si at sykefravær er mindre risikabelt i en senere periode. På den andre siden går sannsynligheten for å være i jobb i fremtidige perioder opp med tiden, gitt at man har god sykefraværshistorie. I dette rammeverket finner forfatteren at sykefraværet vil ha en negativ sammenheng med arbeidsledigheten, noe som støtter Leigh (1985) sine funn.

En god oppsummering av den tidligere litteraturen finnes i Brown og Sessions (1996). Forfatterne trekker frem viktigheten av å forstå arbeidsfravær, samtidig som de uttrykker en viss misnøye med hvor lang tid det har tatt før temaet har fått en seriøs behandling av samfunnsøkonomer. De presenterer en statisk neoklassisk modell for arbeidstilbud som typisk for økonomiske analyser så langt, og poengterer at i et slikt system vil det være vanskelig å identifisere den direkte effekten av forskjellige tiltak og variabler. Dette er fordi arbeidsgivere og arbeidstakere optimaliserer sine valg avhengig av hvordan den andre aktøren opererer.

Goddard og Audas (2001) undersøker konjunktoreffekter på sykefraværet med en modell ikke ulik effektivitetslønnslitteraturen². De anvender månedlig tidsseriedata fra USA for perioden 1979-1993 på makronivå for å teste sine antakelser. I estimeringene sine inkluderer de en konjunkturindikator både på industri og nasjonalt nivå. De argumenterer for at produkt- og arbeidsetterspørsel påvirker fraværet på et makronivå fordi skyggekostnadene for bedriftene assosiert med fravær varierer med konjunktorene. De finner en negativ sammenheng mellom den nasjonale arbeidsledigheten og fraværet, noe som de bemerker er i samsvar med litteraturen så langt.

2.2 KONJUNKTURER OG SYKEFRAVÆR I NORGE

Dyrstad og Lysø (1998) finner at det er stor forskjell mellom effekten av arbeidsledigheten på kort- og langtidsfraværet. De skiller også mellom kortsiktige og langsiktige effekter i begge tilfeller. For korttidsfraværet finner de signifikante kortsiktige elastisiteter på $-0,17$ for menn og $-0,14$ for kvinner, og signifikante langsiktige elastisiteter på $-0,17$ og $-0,18$. De finner altså en signifikant negativ sammenheng mellom arbeidsledighet og korttidssykefraværet som ikke er signifikant forskjellig mellom kvinner og menn. For langtidsfraværet på lang sikt finner de kun en statistisk signifikant sammenheng for kvinner, og denne er betydelig mindre med en elastisitet på $-0,05$. På kort sikt finner de at det er hvor raskt ledigheten endres, det vil si $\Delta u_t - \Delta u_{t-1}$, som er relevant for sykefraværet. Denne elastisiteten er $-0,09$ for kvinner og $-0,04$ for menn.

Askildsen, Bratberg og Nilsen (2000) bruker data fra «Klientstrømmer Inn i, Rundt i og Ut av Trygdesystemet» (KIRUT) til å undersøke både disiplinering- og sammensetningshypotesen. Datamaterialet er begrenset til sykefravær som varer i mer enn to uker. Det vil si sykefravær som er betalt av trygdesystemet. De finner ingen støtte for sammensetningshypotesen i sin analyse, og bemerker at marginale arbeidere generelt har lavest sannsynlighet for å være sykemeldte. Derimot finner de støtte for at atferdsendringer, spesielt blant stabile arbeidere, forklarer økningen i sykefraværet under en høykonjunktur.

² Se Shapiro og Stiglitz (1984) for en teoretisk introduksjon til effektivitetslønnsmodeller.

Dyrstad og Ose (2002) studerer samme spørsmål med et litt annet datasett. De har brukt aggregerte kvartalsdata publisert av Næringslivets Hovedorganisasjon (NHO) fra første kvartal 1971 til andre kvartal 1998. Datasettet gir forfatterne mulighet til å skille mellom kort- og langtidsfravær. For langtidsfravær finner de at kun disiplineringseffekter er til stede. For korttidsfravær skiller de mellom menn og kvinner. For kvinner finner de at ingen av effektene er til stede, mens det er mulig at begge effektene er tilstede for menn.

Dale-Olsen og Markussen (2010) studerer utviklingen i sykefraværet fra 1972-2008. De bruker data fra Arbeidskraftundersøkelsen (AKU) og registerdata fra sykepengeregisteret. Med data fra AKU finner de at endret alderssammensetning, yrkestilhørighet og utdanningsnivå ikke forklarer økningen i kvinner sitt fravær. Med data fra sykepengeregisteret studerer de utviklingen i sykefraværet for konkrete diagnoser. De har valgt ut brudd, forstuinger og dislokasjoner. De finner at mens forekomsten av disse diagnosene stort sett er uendret siden 1995, har varigheten på sykemeldingene økt med hele 20 prosent. De argumenterer for at dette demonstrerer holdningsendringer.

2.3 KJØNNSFORSKJELLER I SYKEFRAVÆRET

Som studiene over indikerer er kjønnsforskjeller et gjengående tema i studier av sykefravær uansett hvilke forhold man er interessert i å studere. Allen (1981) finner en positiv sammenheng mellom sykefravær og familiestørrelse for gifte kvinner, samt en positiv sammenheng for kvinner i arbeiderklassen. Ellers finner han at forskjellen i sykefraværet mellom men og kvinner kan forklares av forskjeller i marginalinntekt og fleksibilitet i arbeidstid.

Basert på data fra en spørreundersøkelse studerer Vistnes (1997) forskjellen i fravær fra arbeid mellom menn og kvinner. Forfatteren finner at kvinner med unge barn har en høyere sannsynlighet for å være fraværende, mens det har ingen effekt for menn. Dette samsvarer med tidligere studier. I motsetning til andre studier fant forskeren derimot at for menn med unge barn i barnehage økte antall fraværsdager når mannen først var borte.

Knutsson og Goine (1998) studerte langsiktig sykefravær i to len i Sverige. De valgte ut 12 yrker som representerte typiske kvinne- og mannsyrker. Etter kontroll for alder og

arbeidsledighet finner de fortsatt store forskjeller i sykefravær mellom yrkene. De finner også at sykefraværet øker med alder. Når det kommer til sammenhengen mellom arbeidsledighet og sykefraværet var det kun statistisk signifikant for menn, hvor de fant en negativ sammenheng.

Brage, Nygård og Tellnes (1998) studerer langtidssykefravær (her sykefravær lengre enn 14 dager) med data fra sykepengeregisteret. De kontrollerer bl.a. for alder, inntekt, ekteskapsstatus og diagnose (for diagnoser relatert til muskler og skjellet). De finner som forventet at kjønn er en sterk forklaringsvariabel for langtidssykefravær grunnet disse lidelsene, hvor kvinner har høyere kumulativ forekomst og lengre fravær. Fraværsfrekvensen var derimot lik. De trekker frem inntekt og inntektsrelaterte faktorer som ofte er oversett, og som bidrar til at kjønnsforskjellene blir overdrevet i noe av forskningen. Den kumulative forekomsten var lav for begge kjønn i den høyeste inntektsgruppen (inntekt på over 200000 kr i året). I gruppen 150-200000 kr hadde menn betydelig høyere forekomst enn kvinner. I den laveste inntektsgruppen (under 100000 kr i året) hadde kvinner den laveste forekomsten

Mastekaasa (2012) undersøker kjønnsforskjeller i sykefraværet i perioden 1974 til 2010 motivert av økningen i sykefraværet blant kvinner. Datamaterialet er hentet fra SSB sin Arbeidskraftundersøkelse (AKU). Forfatteren finner at store deler av økningen kan forklares av endringer i hvilke typer jobber kvinner har. På 1970-tallet var de fleste kvinner i mindre fraværsfremmende yrker enn menn. Forfatteren argumenterer selv for at «i dag har kvinners og menns yrkesmessige plassering ingen betydning for kjønnsforskjeller i fravær». Dette er beroligende fordi det vil bety at økningen fra 1970-tallet, hvor kvinner nå har et betydelig høyere sykefravær enn menn, trolig ikke vil fortsette.

2.4 ANDRE FAKTORER SOM PÅVIRKER SYKEFRAVÆRET

Mange andre faktorer kan tenkes å påvirke sykefraværet, og litteraturen reflekterer dette. Leigh (1995) finner at sigarettøyking har en positiv effekt på sykefraværet for både menn og kvinner. Dette er en av mange studier som ser på forholdet mellom dårlige helsevaner, som røyking og rusavhengighet, og sykefravær på individnivå. Andre studier fokuserer f.eks. på

alder, som i Thomson, Griffiths og Davidson (2000) som finner en positiv effekt av alder på godkjent fravær, og en negativ effekt av alder på ikke-godkjent fravær. Det første funnet samsvarer med Paringer (1983) som finner en positiv sammenheng mellom alder og lengden på sykefravær. O'Reilly og Stevens (2002) oppsummerer litteraturen knyttet til influensaviruset og sykefravær. Generelt finner tidligere studier at 2-4 arbeidsdager går tapt per influensatilfelle.

Ose og Dyrstad (2003) studerer sammenhengen mellom overtidsarbeid og sykefravær med paneldata for 263 bedrifter kvartalsvis fra 1990 til 1996. De definerer korttidsfravær som fravær som varer opp til tre dager, og langtidsfravær som fravær som varer mer enn tre dager. Også i denne studien spiller kjønn en viktig rolle. I datasettet de benytter har kvinner betydelig høyere fravær enn menn, med 7% høyere korttidsfravær og 50% høyere langtidsfravær. Menn hadde over dobbelt så mange overtidstimer per ansatt enn kvinner.

De empiriske resultatene indikerer at kjønnsforskjellen er reell. Kvinner har høyere fravær i alle spesifikasjoner både for kort- og langtidsfravær. Resultatene indikerer også at forskjellen er større for langtidsfraværet. Forskjeller mellom fraværslengdene finnes også for lønningsvariablene. For høyere normal lønn er den negative effekten på fraværet større for langtidsfraværet. For overtidspremien snur effekten fra signifikant negativ for korttidsfravær til ikke signifikant og positiv for langtidsfraværet. De finner at mer overtidsarbeid øker både kort- og langtidsfraværet, med en større effekt på langtidsfraværet. Økt fravær fører igjen til mer overtidsarbeid. Trolig skyldes dette at arbeidet som ikke har blitt gjort i fraværperioden må tas igjen.

Ose, Jensberg, Reinertsen, Sandsund og Dyrstad (2006) sammenligner sykefraværet i Norge med det i andre land, og konkluderer med at det var noe høyt i Norge den gangen. De trekker frem at sykefraværet var historisk høyt i årsskiftet 2003-2004, og at det så falt kraftig i 2004. Forskerne poengterer at dette kan skyldes endringer i regelverket for sykemeldinger, men understreker at vi ikke vet det sikkert. Ose (2010) følger opp rapporten fra 2006 med fokus på litteratur fra Norge. En av konklusjonene er at vi med stor sikkerhet kan si at mye av nedgangen i 2004 skyldes regelendringene.

3 TEORI

Det er vanlig å ta utgangspunkt i en standard modell for individuelt arbeidstilbud når man studerer fraværsbeslutninger.³ Denne modellen forutsetter at beslutningen om å være fraværende fra arbeidet er et spørsmål om tidsallokering mellom arbeid og fravær, og at denne beslutningen er økonomisk motivert. Jeg legger frem en enkel versjon av denne arbeidstilbudsmodellen og aggregere den fra individ- til fylkesnivå. I denne modellen har individene nyttefunksjon:

$$V = V(C, L; S) \tag{3.1}$$

Hvor C er konsum, L er fritid og S er en variabel som representerer helsetilstanden til individet. Jeg antar at V er konkav i C og L . Videre antar jeg at $\partial^2 V / \partial C \partial L$ og $\partial^2 V / \partial L \partial C$ ikke er negative. Dette er så fremt jeg kan finne ikke eksplisitt gjort i f.eks. Dyrstad og Lysø (1998), men jeg ser ikke hvordan de har kommet frem til entydige komparative statistiske resultater uten disse antakelsene. Det synes uansett rimelig å anta at nytten av høyere konsum ikke er fallende i fritid og motsatt.

Inklusjonen av helsevariabelen S fortjener nærmere kommentar. Denne variabelen fanger opp forskjeller i individer, men kan også tenkes å være forskjellig mellom grupper som f.eks. mellom kjønnene. Jeg har valg å la denne variabelen være helt eksogen, slik som den er presentert i Dyrstad og Lysø (1998), men dette er av praktiske hensyn. Det kan for eksempel tenkes at økt konsum av fritid fører til bedre helse fordi man går flere turer i fjellet. På samme måte kan det tenkes at høyere konsum fører til dårligere helse på grunn av overvekt. Disse effektene har jeg ikke tatt høyde for. Jeg antar at helsetilstanden kan ta verdier mellom 0 og 1, hvor 1 indikerer at individet er helt arbeidsufør. Videre antas det at grensenytten av fritid er økende i S mens konsum ikke er det. Det vil si at syke individer har høyere nytte av fritid enn friske.

³ Se bl.a. Brown og Sessions (1996) og Dyrstad og Lysø (1998). Jeg følger sistnevnte nærme.

Individet har også en budsjettrestriksjon gitt ved (3.2) hvor W er realtimelønn, H er planlagt arbeidstid, k er kompensasjonsgrad ved sykefravær, A er sykefraværet og U er arbeidsledigheten:

$$C = W[H - (1 - k)A] - P(A, U) \quad (3.2)$$

Budsjettrestriksjonen sier at konsumet er begrenset av tre komponenter. Den første er WH som beskriver hvor mye individet vil tjene hvis det jobber alle de planlagte timene. Den andre er $(1 - k)WA$ som beskriver hvor mye personen taper direkte på fraværet sitt A . Merk at hvis $k = 1$, det vil si at individet får samme lønn ved sykefravær som ved å arbeide, reduserer disse uttrykkene til kun WA . I Norge er kompensasjonsgraden i mange praktiske tilfeller lik én. Det siste leddet $P(A, U)$ er en straffefunksjon. Denne straffen kan tenkes å representere tapte muligheter til forfremmelse, økt sannsynlighet for oppsigelse, eller andre ikke-direkte faktorer som vil påvirke individets konsum. Jeg antar at straffen er konveks og økende i A , det vil si at $\partial P/\partial A$ og $\partial^2 P/\partial A^2$ begge er positive. Videre antar jeg at $\partial P/\partial U$ og $\partial^2 P/\partial A \partial U$ også er positive siden tilstanden i arbeidsmarkedet vil være avgjørende for straffefunksjonen. Jeg antar også at straffefunksjonen er null hvis sykefraværet er null, det vil si $P(0, U) = 0$.

I tillegg til budsjettrestriksjonen står individet ovenfor en tidsrestriksjon

$$L = T - H + A, \quad (3.3)$$

som sier at fritid er bestemt av total tid T , avtalte arbeidstimer H og arbeidsfravær A . Jeg antar at avtalte arbeidstimer H er eksogen slik at fritiden i sin helhet er bestemt av arbeidsfraværet.

Jeg løser optimeringsproblemet via Lagrange-funksjonen:

$$\mathcal{L} = V(C, L; S) - \lambda[W(H - (1 - k)A) - P(A, U) - C] \quad (3.4)$$

Individet velger C og A . Førsteordensbetingelsene er da:

$$\frac{\delta \mathcal{L}}{\delta C} = \frac{\delta V}{\delta C} + \lambda = 0, \quad \lambda = -\frac{\delta V}{\delta C} \quad (3.5)$$

$$\frac{\delta \mathcal{L}}{\delta A} = \frac{\delta V}{\delta L} + \lambda \left[W(1 - k) + \frac{\delta P}{\delta A} \right] = 0 \quad (3.6)$$

Ved å sette (3.5) inn i (3.6) gir dette en tilpasning hvor den marginale substitusjonsraten mellom fritid og konsum er lik prisen av fritid:

$$\frac{\delta V / \delta L}{\delta V / \delta C} = W(1 - k) + \frac{\delta P}{\delta A} \quad (3.7)$$

Da har jeg to ligninger som beskriver det optimale valget av to variabler C og A , én for budsjettlinjen fra $\delta \mathcal{L} / \delta \lambda$ og én fra (3.7), som i optimum begge er lik null. Jeg definerer to funksjoner fra disse som henholdsvis f_1 og f_2 i henhold til implisittfunksjonsteoremet, og setter opp et ligningssystem $Ax = b$ med de partiellderiverte med tanke på U :

$$\begin{bmatrix} \frac{\delta f_1}{\delta C} & \frac{\delta f_1}{\delta A} \\ \frac{\delta f_2}{\delta C} & \frac{\delta f_2}{\delta A} \end{bmatrix} \begin{bmatrix} \frac{\delta C}{\delta U} \\ \frac{\delta A}{\delta U} \end{bmatrix} = \begin{bmatrix} -\frac{\delta f_1}{\delta U} \\ -\frac{\delta f_2}{\delta U} \end{bmatrix}$$

Determinanten til den første matrisen må sjekkes for at den ikke er null, og jeg trenger å vite om den er positiv eller negativ for å bruke Cramers metode. $f_{1,C} = 1$ og $f_{1,A} = W(1 - k) + P_A > 0$. For de to andre finner jeg:

$$\frac{\delta f_2}{\delta C} = \frac{V_{LC}V_C - V_{CC}V_L}{V_C^2} > 0$$

$$\frac{\delta f_2}{\delta A} = \frac{V_{LL}V_C - V_{CL}V_L}{V_C^2} < 0$$

Jeg finner altså en positiv størrelse for $f_{2,C}$ og en negativ størrelse for $f_{2,A}$, og finner da at $\det(A) = f_{2,A} - [W(1 - K) + P_A]f_{2,C} < 0$. For å komme videre trenger jeg bare å finne

$\det(A_U)$ hvor A_U er matrisen A over med den andre kolonnen byttet ut med b . Nytt her er $-f_{1,U} = -P_U$ og $-f_{2,U} = P_{AU}$. Da finner jeg $\det(A_U) = P_{AU} + P_U f_{2,C} > 0$. Etter Cramers regel har jeg da, siden telleren er positiv og nevneren er negativ:

$$\frac{\delta A}{\delta U} = \frac{\det(A_U)}{\det(A)} < 0$$

Den samme prosessen repeteres for de andre variablene av interesse⁴. Fra dette kan individets fraværsbeslutning beskrives på redusert form som:

$$A = (W, k, H, U, S) \tag{3.8}$$

Hvor W, k, U, H og S alle er eksogene variabler. Jeg konkluderer på bakgrunn av utregningen over at forholdet mellom arbeidsledighet og sykefraværet er negativt, det vil si at $\partial A / \partial U < 0$. Dette beskriver disiplineringshypotesen. Med forutsetningene jeg har gjort om S er $\partial A / \partial S > 0$. Dermed kan sammensettingseffekten tenkes å oppstå når individer med forskjellig helsetilstand S er inkludert i eller holdt utenfor gruppen av sysselsatte. Forholdet mellom fravær og kompensasjonsgraden er også entydig bestemt med $\partial A / \partial k > 0$. På bakgrunn av antakelsene jeg har gjort om de blandede partiellderiverte av nyttefunksjonen V er $\partial A / \partial H > 0$. For W trekker inntekts- og substitusjonseffektene i motsatt retning, men i tilfellet hvor $k = 1$ er også W entydig bestemt med $\partial A / \partial W > 0$.

3.1 SYKEFRAVÆR PÅ FYLKESNIVÅ

Jeg ønsker å estimere effekten av arbeidsledighet på sykefravær på fylkesnivå. Jeg lar H og A representere dagsverk istedenfor timer. Kompensasjonsgraden bør i prinsippet være lik for alle fylker, og er uansett i prinsippet lik én i Norge, og jeg setter derfor $k = 1$. Merk at dette gir entydige prediksjoner for effekten av lønn. De $i = 1, 2, \dots, N$ individspesifikke fraværsfunksjonene aggregeres i hver periode $t = 1, 2, \dots, T$ til fylkesnivå. Da får jeg at:

⁴ De resterende utregningene presenteres som vedlegg.

$$A_{jt} = \sum_{i=1}^N A_{it}(W_{it}, H_{it}, S_{it}, U_{jt}) \quad (3.9)$$

Som sier at tapte dagsverk i fylke $j = 1, 2, \dots, 19$ i tidsperiode t er summen av tapte dagsverk for individene. Hvis $H_{jt} = \sum_{i=1}^N H_{it}$ er summen av planlagte dagsverk i fylke j i periode t er sykefraværsprosenten:

$$syk_{jt} = \frac{A_{jt}}{H_{jt}} \quad (3.10)$$

Merk at fordi de i individene i hvert fylke, som i praksis vil si de sysselsatte, ikke nødvendigvis er de samme i hver periode, holder ikke resonnementene jeg har gjort tidligere rundt effekten av de eksogene variablene. Under antakelse om at sammensetningshypotesen ikke er relevant, noe bl.a. Askildsen et al. (2000) og Dyrstad og Ose (2002) peker mot, holder resultatene for alle variablene med unntak av H . Da kan jeg tolke de empiriske resultatene mine opp mot disiplineringshypotesen. Hvis komposisjon kommer inn i bildet gir ikke teorimodellen klare prediksjoner. Sammensetningshypotesen mer generelt predikerer samme effekt på fraværet som disiplineringshypotesen. Dette er fordi man antar at det er marginale arbeidere som faller bort når arbeidsledigheten øker, altså arbeidere som trolig har høyere verdi av helsetilstandsvariabelen S , og at disse har høyere sannsynlighet for å ta sykefravær. Dermed blir det naturlig å predikere en lik retning på forholdet på fylkesnivå selv om begge hypotesene er relevante. Jeg har i mitt empiriske oppsett ingen strategi for å skille disse effektene, og hvilke som er relevante basert på mine resultater blir derfor ren spekulasjon.

4 EMPIRISK SPESIFIKASJON OG DATAMATERIALE

4.1 VALG AV ESTIMERINGSMETODE

Generelt ønsker jeg å estimere en modell for sykefraværet syk_{jt} som i ligning (3.10). Dette operasjonaliserer jeg som en funksjon av U_{jt} og et sett med n kontrollvariabler $\mathbf{x}_{jt} = (1, x_{1jt}, x_{2jt}, \dots, x_{njt})'$:

$$syk_{jt} = \alpha_j + U_{jt}\beta_1 + \mathbf{x}'_{jt}\boldsymbol{\beta} + \varepsilon_{jt} \quad (4.1)$$

Denne modellen tar høyde for at det kan være uobserverte forskjeller mellom fylkene α_j som påvirker sykefraværet. Det kan f.eks. være at helsetilstandsvariabelen i ett fylke er generelt høyere (det vil si befolkningen er generelt sykere) enn i et annet. Dette kan f.eks. skyldes at folk oftere bor i byer i det ene fylket, hvor det kan tenkes at forurensning og stress fører til dårlige helsetilstand.⁵ Et sentralt spørsmål er derfor hvordan jeg velger å behandle disse uobserverte faste effektene α_j . Jeg vil diskutere forskjellige tilnærminger til dette basert på antakelser om restleddet.⁶

4.1.1 Pooled Minste Kvadraters Metode

Den ideelle situasjonen er hvis modellen vår er så godt spesifisert at vi kan estimere den med pooled minste kvadraters metode (MKM). I dette tilfellet har vi 323 datapunkter med variasjon i arbeidsledighet og sykefravær som effektivt estimeres av vanlig MKM. Det er derimot lite sannsynlig at dette er tilfellet for mine data.

Antakelsen som trolig bryter ned ved estimering av (4.1) med MKM er antakelsen om at restleddet ikke er korrelert med noen av forklaringsvariablene. Dette er tilfellet hvis α_j er korrelert med U_{jt} eller \mathbf{x}_{jt} . Hvis α_j ikke er korrelert med U_{jt} og \mathbf{x}_{jt} får jeg at $E[(e_{jt} + \alpha_j) | U_{jt}, \mathbf{x}_{jt}] = 0$. Dermed gir MKM konsistente estimater av $\boldsymbol{\beta}$. Restleddet er derimot

⁵ Dette er ikke strengt tatt en tidsfast effekt da det kan tenkes at andelen som bor i by i et gitt fylke endres over tid. I prinsippet er det vanskelig å tenke seg mange helt tidsfaste effekter, da selv geografi endrer seg over en lang nok tidshorisont. For denne oppgaven vil jeg snakke om tidsfaste effekter for den begrensede perioden jeg observerer på 17 år.

⁶ Jeg vil ikke utlede estimatorene i sin helhet. For detaljer se f.eks. Verbeek (2004) som jeg har basert meg på.

seriekorrelert (så lenge $\alpha_j \neq 0$) og må estimeres med seriekorrelasjonsrobuste standardavvik. Dermed er MKM benyttet på (4.1) ikke den mest effektive estimatoren selv om $E[(e_{jt} + \alpha_j)|U_{jt}, \mathbf{x}_{jt}] = 0$ holder.

4.1.2 Fixed Effects

En logisk alternativ estimeringsmetode, og kanskje den vanligste estimeringsmetoden for paneldata, er fixed effects. Fixed effects estimatoren er også kjent som «within»-estimatoren. En svakhet med datagrunnlaget jeg har brukt er at få variabler er inkludert. Dette hadde ideelt sett vært variabler som oppfyller noen av kriteriene: (1) at de har teoretisk relevans, (2) vi er interessert i å estimere de, det vil si vi er interessert i mulige effekter de har på sykefraværet, eller (3) de kontrollerer for deler av variansen i sykefraværet som hjelper oss å bedre estimere effekten av arbeidsledighet. I teorien skulle vi ønske at listen over variabler var utfyllende med tanke på punkt tre, slik at ingen relevante variabler er utelatt, men dette er vanskelig i praksis. Med en fixed effects modell korrigerer vi for utelatte tidsfaste variabler som kun varierer mellom enhetene i panelet, men man mister også mye av den totale variasjonen i panelet. I dette tilfellet er enhetene fylker, og det gir intuitivt mening at det finnes grunnleggende variasjoner mellom fylkene som kan ha en effekt på sykefraværet. Det kan være geografiske forskjeller som temperatur og eksponering til sollys, forskjeller i sykefraværskultur, eller lignende.

I prinsippet er fixed effects estimering så enkelt som å estimere (4.1) med $j - 1$ konstantledd α_j , og kan estimeres ved å inkludere $j - 1$ dummyvariabler (hvor j representerer fylker i mitt tilfelle) og estimere denne ligningen med MKM. Dette er ekvivalent med å estimere en ligning som uttrykker (4.1) som avvik fra gjennomsnittsverdier:

$$syk_{jt} - \overline{syk}_j = (U_{jt} - \bar{U}_j)\beta_1 + (\mathbf{x}_{jt} - \bar{\mathbf{x}}_j)' \boldsymbol{\beta} + (\varepsilon_{jt} - \bar{\varepsilon}_j) \quad (4.2)$$

For å estimere (4.2) konsistent kreves det at $E[\varepsilon_{jt}|(U_{jt} - \bar{U}_j), (\mathbf{x}_{jt} - \bar{\mathbf{x}}_j)] = 0$. Tilstrekkelige betingelser for dette er at U_{jt} og \mathbf{x}_{jt} ikke er korrelert med ε_{jt} (som alltid) og at \bar{U}_j og $\bar{\mathbf{x}}_j$ ikke er korrelert med restleddet. Dette er tilfellet hvis U_{jt} og \mathbf{x}_{jt} er strengt eksogene. Det vil si at

forklaringsvariablene ikke er avhengige av restleddet i noen periode. Denne antakelsen bryter sammen hvis det f.eks. inkluderes en lagget avhengig variabel i modellen.

Fordelen med fixed effects estimatoren er at den eliminerer de faste effektene α_j . Dermed tillater estimering med fixed effects at α_j er korrelert med de andre forklaringsvariablene. I tillegg er parameterne α_j estimert forventningsrett som $\hat{\alpha}_j = \bar{y}_j - \bar{U}_j \beta_1 - \bar{x}_j' \boldsymbol{\beta}_{FE}$, og kan være interessante i seg selv.

4.1.3 Random effects

Random effects estimatoren er best illustrert ved å tenke på α_j og ε_{jt} som et sammensatt restledd. Det vil si $\eta_{jt} = \alpha_j + \varepsilon_{jt}$. Som nevnt for pooled MKM betyr det at restleddene er seriekorrelerte siden de faste effektene α_j er inkudert i restleddet i alle tidsperioder. Det kan vises at $Corr(\eta_{jt}, \eta_{js}) = \sigma^2_\alpha / (\sigma^2_\alpha + \sigma^2_\varepsilon)$ for $t \neq s$, hvor $\sigma^2_\alpha = Var(\alpha_j)$ og $\sigma_\varepsilon = Var(\varepsilon_{jt})$. Jeg utleder ikke GLS transformasjonen her,⁷ men resultatet er en variabel theta:

$$\theta = 1 - \left[\frac{\sigma^2_\varepsilon}{\sigma^2_\varepsilon + \sigma^2_\alpha} \right]^{1/2} \quad (4.3)$$

Denne variabelen tar verdiene [0,1] og med denne dannes den transformerte ligningen:

$$syk_{jt} - \theta \overline{syk}_j = (1 - \theta)\beta_0 + (U_{jt} - \theta \bar{U}_j)\beta_1 + (\mathbf{x}_{jt} - \theta \bar{\mathbf{x}}_j)' \boldsymbol{\beta} + (\varepsilon_{jt} - \theta \bar{\varepsilon}_j) \quad (4.4)$$

Denne ligningen er illustrativ. Hvis θ er null estimeres ligning (4.1). Hvis θ derimot er én estimeres fixed effects ligningen (4.2). Dermed er random effects en vekting av denne transformasjonen. Hvis theta er (0,1) er estimatorene forskjellige. Fordelen med random effects, som benytter et GLS-rammeverk, er at estimatoren er effektiv. Ulempen, kontra fixed effects, er at \mathbf{x}_{jt} ikke kan være korrelert med α_j .

⁷ Resonnementet er hentet fra Wooldridge (2013). Utledningen finnes i Verbeek (2004).

4.1.4 Valget mellom fixed or random effects

I mitt tilfelle kan det ligge en viss interesse i de faste effektene α_j . Dermed er det naturlig å foretrekke fixed effects estimering. Allikevel ligger hovedfokuset i oppgaven på forholdet mellom arbeidsledighet og sykefravær i seg selv. Derfor vil jeg vurdere muligheten for å estimere modellen med random effects.

For å undersøke dette estimerer jeg modellene mine for både fixed og random effects og benytter en test etter Hausman (1978) for å velge mellom disse. Testen er ofte kalt Durbin-Wu-Hausman testen og går ut på å sammenligne random og fixed effects modeller med antakelsen om at fixed effects modellen er konsistent. Dette hviler på antakelsen om at $E[\varepsilon_{it} | \mathbf{x}_{is}] = 0$ for alle s og t .⁸ Hvis \mathbf{x}_{it} heller ikke er korrelert med α_j er random effects estimatoren konsistent. Under nullhypotesen om at begge modellene er konsistente er det ønskelig å benytte random effects modellen fordi denne har mer effektive estimater. Under nullhypotesen om at random effects estimatoren er effisient kan det vises at kovariansmatrisen til forskjellen $\hat{\beta}_{FE} - \hat{\beta}_{RE}$ ganske enkelt er:⁹

$$V[\hat{\beta}_{FE} - \hat{\beta}_{RE}] = V[\hat{\beta}_{FE}] - V[\hat{\beta}_{RE}] \quad (4.5)$$

Dermed kan Hausman testobservatoren beregnes som:

$$\xi_H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' [\hat{V}[\hat{\beta}_{FE}] - \hat{V}[\hat{\beta}_{RE}]]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \quad (4.6)$$

Hvor \hat{V} er de estimerte kovariansmatrisene. Under nullhypotesen om at $\hat{\beta}_{FE} - \hat{\beta}_{RE} = 0$ er testobservatoren kjikvadratfordelt med K frihetsgrader, hvor K er antallet ukjente parametere i β når modellen ikke er feilspesifisert. Hvis estimatoren er større enn den kritiske verdien avvises nullhypotesen om at begge estimatorene er konsistente. Hvis dette er tilfellet foretrekkes fixed effects modellen.

⁸ Jeg vil her og heretter inkludere arbeidsledigheten i \mathbf{x} .

⁹ Disse er hentet fra Verbeek (2004).

4.2 SPESIFIKASJONSTESTER

I tillegg til Hausman-testen vil jeg benytte to andre spesifikasjonstester. Den første er Durbin-Watson testen for seriekorrelerte restledd og den andre er en test etter Wooldridge (2002) for uobserverte effekter.

4.2.1 Durbin-Watson for paneldata

Durbin-Watson er en vanlig test for seriekorrelerte restledd:

$$DW = \frac{\sum_{t=2}^N (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^N (\hat{\varepsilon}_t^2)} \quad (4.7)$$

Denne teststatistikken kan generaliseres til paneldata som:

$$DW_p = \frac{\sum_{j=1}^N \sum_{t=2}^N (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{j=1}^N \sum_{t=1}^N (\hat{\varepsilon}_t^2)} \quad (4.8)$$

Nullhypotesen er at $\rho = 0$ i $\varepsilon_{jt} = \rho \varepsilon_{j,t-1} + \varepsilon_{jt}$. Kritiske verdier av Durbin-Watson testobservatoren kan være uhåndterlige. Funksjonen *pdwtest* i R-pakken *plm*, som jeg benytter, rapporterer p-verdien direkte, og jeg anvender derfor denne.

4.2.2 Wooldridge test for ikke-observerte effekter

Wooldridge (2002) presenterer en testobservator for ikke-observerte effekter i paneldata:

$$W = \frac{\sum_{j=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{\eta}_{jt} \hat{\eta}_{js}}{\left[\sum_{j=1}^N \left(\sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{\eta}_{jt} \hat{\eta}_{js} \right)^2 \right]^{1/2}} \quad (4.9)$$

Under nullhypotesen om at restleddene ikke er seriekorrelerte er testobservatoren asymptotisk normalfordelt. Det er verdt å merke at nullhypotesen avvises for mange former for seriekorrelasjon. Sånn sett er det ikke en perfekt test for ikke-observerte effekter. En annen ulempe er at testen er avhengig av asymptotiske egenskaper for $N \rightarrow \infty$. Fordelen er at testen fungerer under heteroskedastisitet. Testen er inkludert som funksjonen *pwttest* i R-pakken *plm*, og jeg har brukt p-verdiene direkte.

4.3 MER OM VARIABLENE

Etter spørsmålet om hvordan jeg skal håndtere de fylkesfaste effektene er den andre problemstillingen hvordan jeg skal ta høyde for forskjellen i komposisjon av befolkningen i hvert fylke, og hvordan dette spiller inn på sykefraværet. Fra teorimodellen vil jeg korrigere for de tre variablene lønn, helsetilstand og planlagte dagsverk. Problemet er at jeg ikke har tall for de to første på fylkesnivå. Alternativet blir da å kontrollere for faktorer som kan tenkes å forklare noe av de samme variasjonene. Samtidig vil jeg motivere valg av kontrollvariabler med andre empiriske funn, selv om de ikke nødvendigvis inngår i (3.8) og (3.9). Innrømmelsen her er at disse ligningene er nyttige forenklinger, men at det er lite trolig at denne modellen dekker alle relevante faktorer. Som et resultat av disse problemstillingene vil modellene jeg faktisk estimerer være:

$$\begin{aligned} syk_{jt} = & \alpha_j + arb_{jt}\beta_1 + kvinner_{jt}\beta_2 + unge_{jt}\beta_3 + a74_{jt}\beta_4 \\ & + utd_{jt}\beta_5 + jord_{jt}\beta_6 + ind_{jt}\beta_7 + fintek_{jt}\beta_8 + hel_{jt}\beta_9 \\ & + r04_t\delta_1 + arb_{jt} * r04_t\beta_{10} + \text{årsdummier}_t + \varepsilon_{jt} \end{aligned} \quad (4.10)$$

I begge ligningene er $j = 1, 2, \dots, 19$ fylkene og $t = 1, 2, \dots, 17$ fjerde kvartal i årene jeg har data for (2000-2016). Variabelen for den registerbaserte ledigheten er ikke rapportert kvartalsvis men etter en referanseuke i november og representerer det beste målet jeg har på denne ledigheten i 4. kvartal. Avhengig variabel er syk_{jt} som er sykefraværsprosenten basert på det legemeldte sykefraværet. Sykefraværet kan ta verdiene $[0, 100]$, og alle andre prosentandeler er derfor skalert på samme måte (det vil si at jeg har 10 prosent som 10,0 og ikke 0,1). Jeg vil rapportere resultatene for to mål på arbeidsledigheten arb_{jt} . Det første målet er arbeidsledigheten som målt ved AKU, markert arb_{jt}^{AKU} , og det andre er arbeidsledigheten som målt ved registerdata fra NAV, markert arb_{jt}^{NAV} .

Variabelen $kvinner_{jt}$ er andelen av sysselsatte som er kvinner i et gitt fylke j i tidsperiode t . Aldersvariablene er satt sammen fra en finere inndeling og jeg forsøker med det å ta høyde for alders effekter med så få variabler som mulig. *Unge* er definert som de opp til og med 24 år mens *a74* er de over 66 år. *Utd* representerer andelen med videregående eller høyere utdanning. Industrivariablene er diskutert i detalj senere. Dummyvariabelen *r04*

representerer en regelendring i praksis for legemeldt sykefravær og er kodet én for periodene 2004-2016 og null ellers.

Panelet jeg har brukt i denne oppgaven er satt sammen av flere tabeller fra Statistisk Sentralbyrå (SSB) sin statistikkbank, og er tilgjengelig for offentligheten. På grunn av diverse brudd i tidsseriene har jeg måttet sette sammen nye og avsluttede serier, noe jeg diskuterer nærmere i hvert tilfelle. I den påfølgende seksjonen vil jeg gå igjennom de forskjellige variablene i panelet og gjøre rede for hvilke tall som er brukt, hvorfor jeg har valgt disse tallene, og hvilke valg jeg har tatt når jeg har satt sammen panelet mitt. De registerbaserte variablene er generelt tilgjengelig per 4. kvartal for årene 2000-2016. For variabler som bryter denne trenden, f.eks. sykefraværet, har jeg derfor droppet observasjonene for 1. til 3. kvartal. Videre er den registerbaserte arbeidsledigheten kun rapportert etter en referanseuke i november.¹⁰ Figur 2 inneholder deskriptiv statistikk for selve panelet. For deskriptiv statistikk på fylkesnivå se vedlegg.

Figur 2: Deskriptiv statistikk

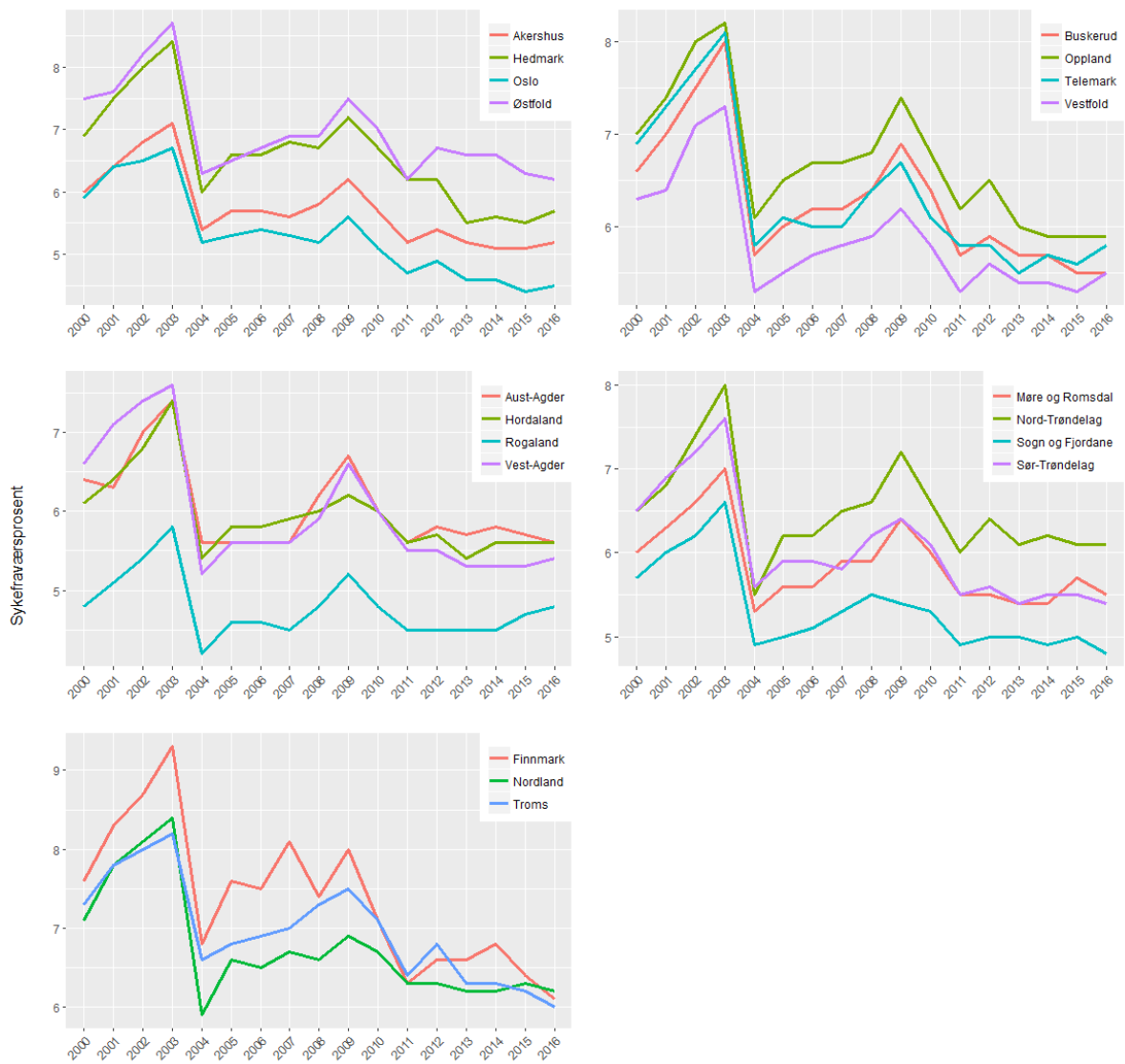
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	323	6.177	0.904	4.200	9.300
arbnav	323	2.666	0.795	0.900	5.900
arbaku	323	3.319	1.224	0.000	7.895
kvinner	323	47.005	0.586	45.219	48.450
a74	323	2.195	0.707	1.085	4.038
jord	323	3.898	2.390	0.131	10.696
ind	323	21.579	4.825	9.014	30.481
hel	323	20.073	2.143	14.432	25.788
r04	323	0.765	0.425	0	1
lsyk	323	1.810	0.144	1.435	2.230
fintek	323	11.034	3.543	5.779	23.884
utd	323	76.514	3.379	67.225	85.805
unge	323	13.508	1.183	10.338	16.794

¹⁰ Frem til 2014 var tallene rapportert månedlig.

4.3.1 Sykefravær

Den avhengige variabelen i den økonometriske modellen er sykefraværet. For det registerbaserte legemeldte sykefraværet er populasjonen alle arbeidstakere 16-69 år, og tallene er basert på registerdata hos NAV. Sykefraværstallene korresponderer til ligning (3.10). Dessverre publiserer ikke SSB tall for sykefraværet basert på egenmelding og legemelding hver for seg ned på fylkesnivå, og det finnes heller ikke et samlet tall som inneholder begge med denne fylkesinndelingen. Dermed er jeg begrenset til å se på effekten av arbeidsledigheten på det legemeldte sykefraværet.

Figur 3: Sykefravær



Ar

Det er tre ting som er verdt å merke seg ved sykefraværstallene. Det første er at undersøkelsen var frivillig til og med 1. kvartal 2003. Etter dette er den obligatorisk. I følge Statistisk Sentralbyrå (2017b) førte dette til en økning i svarprosenten fra 65 til 90 prosent. Det andre er et brudd i tidsserien i år 2009. Statistisk Sentralbyrå (2009) rapporterer at fra og med 1. januar 2009 ble det innført en ny versjon av norsk standard for næringsgruppering. Siden tallene jeg benytter er på et aggregert nivå og ikke inneholder næring direkte har ikke dette noe stort utslag, men omkodningen har fortsatt en effekt. For år 2008 inkluderer SSB både nye og gamle tall. For gammel standard var sykefraværet henholdsvis 6,2; 6,0; 6,1 og 6,0 prosent for 1. til 4. kvartal. For den nye standarden var de tilsvarende tallene henholdsvis 6,2; 6,0; 6,2 og 6,0 prosent. Den lave forskjellen indikerer at jeg trygt kan benytte disse tallene. Jeg velger å beholde de nyeste tallene.

Det siste punktet er overgangen fra Aa-registeret til A-ordningen fra og med 1. januar 2015. Statistisk Sentralbyrå (2017b) rapporterer selv at endringen har hatt en liten effekt på aggregert nivå, og en mindre effekt på det legemeldte enn det egenmeldte sykefraværet. De foreslår et sett med justeringer til tallmaterialet for å gjøre tallene fra før skiftet mest mulig sammenlignbare med tallene etter. Disse er derimot veldig små og basert på sesongjusterte tall. Derfor har jeg valgt å benytte meg av de ikke-justerte tallene i panelet.

Det høyeste sykefraværet var 9,3 prosent og ble registrert i Finnmark i 2003. Det laveste sykefraværet ble registrert i Rogaland, hvor det var så lavt som 4,2 prosent i 2004. Generelt ser det ut til at de grove trekkene i sykefraværet på fylkesbasis holder i hele perioden. Finnmark, som har det høyeste sykefraværet i begynnelsen av perioden, har fortsatt ett av de høyeste i siste periode, og Rogaland har det laveste sykefraværet i første periode og ett av de laveste i siste.

Alt i alt virker det som om mye av variasjonen i sykefraværet er delt mellom fylkene, men det er fortsatt perioder med relativt store avvik, som f.eks. i 2007 hvor sykefraværet økte betydelig mer i Finnmark enn i andre fylker. Det ser også ut til å være en tendens til at sykefraværet homogeniseres over perioden. Forskjellen i sykefraværet mellom høyest og lavest er 2,8 prosentpoeng i første periode, men bare 1,7 prosentpoeng i den siste. Jeg kan tenke meg to grunner til dette. Den første er at regelendringen for legemeldt sykefravær i

2004 samkjørte sykemeldingspraksis slik at sykefraværsbeslutninger i fylkene ble mer like. En annen forklaring kan være at sykefraværet har en form for nedre grense slik at et fall i sykefraværet vil bety at de fylkene som har høyest fravær før fallet faller ned mot de som har lavest og nærmer seg disse.

Generelt vil jeg argumentere for at det er vanskelig å tenke seg rene nivåeffekter på sykefraværet. Hvis et individ har ti sykefraværsdager i året mener jeg det er rimelig å tro at det er lettere å redusere det fraværet til fem dager enn det vil være å redusere det fra fem dager til null. Derfor mener jeg det er et logisk valg å benytte den naturlige logaritmen til sykefraværet som avhengig variabel. Dette følger bl.a. Dyrstad og Lysø (1998) og er vanlig i slike studier.

4.3.2 Arbeidsledighet

Det er to kilder til tall for arbeidsledighet. Den første er tall fra AKU, og den andre er registertall fra NAV. Jeg vil estimere modellen for begge tallene, da de har fordeler og ulemper assosiert med seg. Statistisk Sentralbyrå (2015) oppsummerer forskjellene. Siden registertallene baserer seg på faktisk registrerte arbeidsledige som hos NAV er det ingen utvalgsusikkerhet med disse tallene. Derimot er det visse krav for å kunne motta dagpenger, og spesielt unge som er ute etter sin første jobb faller utenfor i disse tallene. I dette tilfellet gir AKU-tallene et bedre innblikk i ledighetssituasjonen siden de også omfatter unge som ikke har hatt jobb før.

Et viktig valg når det kommer til modeller av denne typen er om man lagrer ledighetsvariablene eller ikke. Rasjonale bak å lagge ledigheten, typisk med ett år i årsdata, er at det kan ta tid fra arbeidsmarkedssituasjonen blir observert til individer faktisk endrer sin atferd. Det samme gjelder for komposisjonseffekter, hvor dårlige tider i én periode ikke nødvendigvis fører til avsettelser før i en senere periode. I mange tilfeller er det kanskje færre ansettelser og ikke avsettinger som endrer komposisjonen. Dette gir mening med årsdata, hvor arbeidsledighet og sykefravær dekker en bred periode. Jeg mener derimot at det er vanskelig å forsvare en slik lagging av variablene når tallene kun gjelder for fjerde kvartal hvert år. Det er en mer ekstrem antakelse at arbeidsledigheten i oktober til desember i fjor predikerer arbeidsledigheten i samme periode i år, enn at arbeidsledigheten

i fjor som helhet predikerer sykefraværet som helet i år. Jeg har derfor ikke lagget arbeidsledighetsvariablene.

4.3.2.1 Arbeidsledighet i arbeidskraftundersøkelsen

Figur 4: Arbeidsledighet for 4. kvartal (AKU)



År

Tallene for arbeidsledighet i AKU er regnet ut fra tall for personer som er i arbeidsstyrken og personer som er sysselsatte, oppgitt i 1000 personer. Fra dette kommer jeg frem til arbeidsledigheten som:

$$arb_{jt}^{AKU} = \left[1 - \left(\frac{Sysselsatte_{jt}}{Arbeidsstyrken_{jt}} \right) \right] * 100 \quad (4.11)$$

Fra januar 2006 skjedde det en større omlegging av arbeidskraftundersøkelsen. Blant annet ble aldersgrensen for å bli tatt med i AKU senket fra 16 til 15 år, slik at undersøkelsen i dag omfatter aldersgruppen 15-74 år. Ifølge Statistisk Sentralbyrå (2007) har definisjonen for sysselsetting også blitt endret slik at de som er ufrivillig permitterte regnes som sysselsatte istedenfor arbeidsledige.

Direkte sammenligning av gamle og nye tall for 2006¹¹ viser at for andre og fjerde kvartal er det ingen endring, mens det i første og tredje kvartal er en endring på henholdsvis minus 0,2 og pluss 0,1 prosentpoeng fra de gamle til de nye tallene. Disse forskjellene er forholdsvis små, og kan trekke i begge retninger. Ifølge Statistisk Sentralbyrå (2007) vil inkluderingen av 15-åringer øke arbeidsledighetstallene noe, mens definisjonsendringen for ufrivillig permitterte vil senke den igjen. Det er derfor vanskelig å si noe konkret om hvilket utfall dette vil ha for en økonometrisk analyse. På tross av dette har jeg valgt å benytte hele tallserien, med tall fra før og etter endringene i 2006, fordi jeg veier viktigheten av en lengre tallserie tyngre enn eventuelle problemer endringen i AKU fører til. Jeg har igjen valgt å beholde de nye tallene.

4.3.2.2 Registerbasert arbeidsledighet

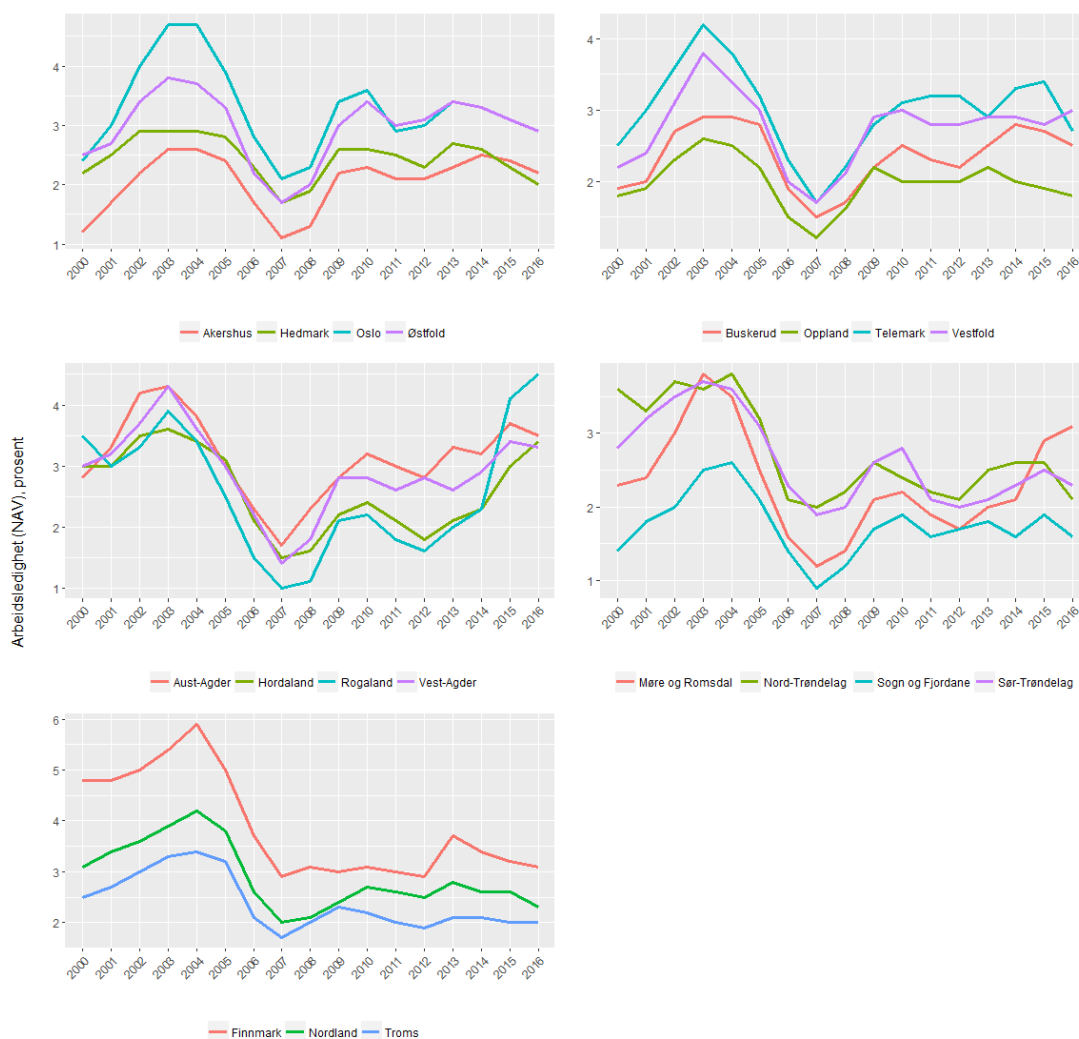
Tallene for registrerte arbeidsledige ble frem til og med 2014 publisert månedlig. Fra og med 2015 har jeg derimot kun tilgang til tall for november. Det viser seg at månedlige tall fortsatt blir publisert av NAV, og det hadde vært mulig å anvende disse, men tall for andre variabler er meg bekjent ikke tilgjengelig med like høy frekvens. Derfor har jeg holdt meg til SSB sine tall, noe som begrenser meg til november hvert år i perioden 2000-2016.

Informasjon om ledighetstallene er hentet fra Statistisk Sentralbyrå (2017a), og jeg oppsummerer kort noen relevante punkter. Statistikken omfatter helt arbeidsledige, som er definert som «arbeidsføre personer som søker inntektsgivende arbeid ved NAV og ellers er

¹¹ 2006 er det eneste året hvor SSB oppgir tall basert på både gammel og ny utregningsmetode.

disponible for det arbeidet som søkes.» Personer som har vært arbeidsledige i mindre enn to uker er holdt ute fra statistikken. Disse tallene kommer fra NAV sine egne systemer. For å finne et prosenttall av arbeidsledige kombinerer ledighetstallene med arbeidsstyrketall fra AKU, og tallene på kommunalt nivå er satt sammen ved hjelp av Aa-registeret frem til 2015, og a-ordningen siden. I denne overgangen har det også vært en endring vekk fra å bruke tall fra AKU, noe som gjør at endringer i sysselsettingen fra 2014 til 2015 ikke nødvendigvis representerer en faktisk endring. Jeg har ikke noe tallgrunnlag for å vurdere viktigheten av denne endringen, og hvilket utslag det har for variablene mine, utover å lene meg på Statistisk Sentralbyrå (2016) som poengterer at «i forhold til totalt antall sysselsatte er forskjellen beskjeden». Jeg har derfor beholdt de nye tallene uendret.

Figur 5: Arbeidsledighet etter referanseuke i november (NAV)



Ar

4.3.2.3 Sammenligning av målene

Generelt sett er det store forskjeller i sykefravær mellom fylkene. For AKU-tallene var den laveste arbeidsledigheten som er målt i perioden null prosent. Dette var tilfellet i Sogn og Fjordane i 2007, 2009, og 2012, i Aust-Agder i 2007, samt i Finnmark i 2006 og 2013. Den høyeste arbeidsledigheten var 7,9 prosent i Finnmark i 2004. I gjennomsnitt var arbeidsledigheten 3,3 prosent. For NAV-tallene var den laveste arbeidsledigheten registrert til 0,9 prosent i Sogn og Fjordane i 2007, mens den høyeste ledigheten var registrert til 5,9 prosent i Finnmark i 2004. I snitt var denne ledigheten målt til 2,7 prosent.

4.3.3 Kjønn

Kjønnsforskjeller i sykefraværsadferd har vært et gjennomgående tema i litteraturen. Typisk finner man at kvinner har høyere sykefravær enn menn. Mastekaasa (2012) argumenterer for at en økning i kvinners sykefravær relativt til menn har skyltes at kvinner også trer inn i mer sykefraværsfremmende jobber i senere tid. I dag har derimot kvinner og menns yrkesmessige plassering ingen betydning på forskjellene i sykefraværet ifølge forfatteren. Dette vil i så tilfelle bety at de observerte kjønnsforskjellene i sykefraværet er reelle og må tas høyde for. Jeg forventer derfor en betydelig positiv sammenheng mellom sykefraværet og andelen kvinner i arbeidsstyrken.

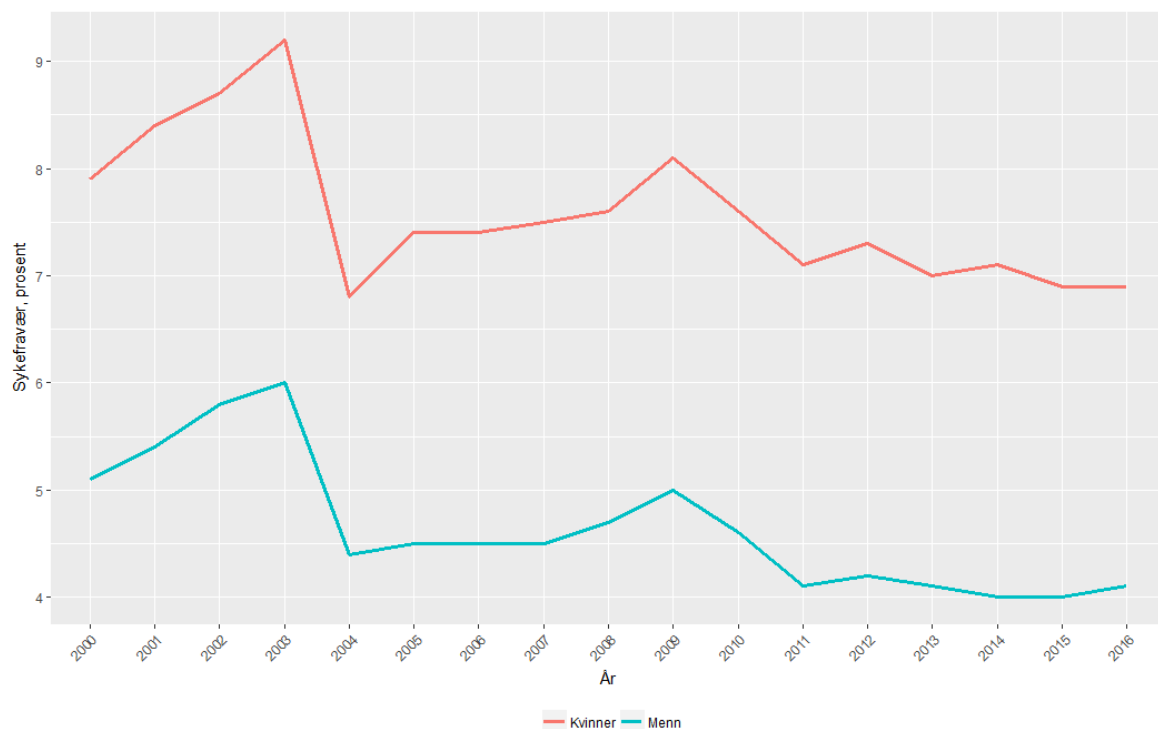
I perioden jeg har sett på var det gjennomsnittlige sykefraværet 4,6 prosent for menn, og hele 7,6 prosent for kvinner. Relativt få personer er fraværende fra jobb på grunn av sykdom til enhver tid, slik at prosenttallene kan virke misvisende med tanke på hvor stor forskjellen er mellom kjønnene. Forskjellen på 2,8 prosentpoeng mellom kvinner og menn betyr at kvinner har en sykefraværsprosent som er 63,2 prosent høyere enn menn. Det laveste sykefraværet for kvinner ble registrert som 5,4 prosent i Rogaland i 2004, og det høyeste ble registrert som 11,0 prosent i Finnmark 2003.

For å kontrollere for dette har jeg inkludert en variabel for kvinneandelen blant de sysselsatte. Variabelen er kodet som:

$$kvinner_{jt} = \frac{Sysselsatte\ kvinner_{jt}}{Sysselsatte_{jt}} * 100 \quad (4.12)$$

Den laveste verdien ble observert i Møre og Romsdal i 2000 med 45,2 prosent, og den høyeste i Oslo i 2003 med 48,5 prosent. Generelt er det en tendens til at kvinneandelen stiger betydelig frem mot 2004, holer seg ganske stabil frem til ca. 2010 og faller igjen mot 2013. Mot slutten av perioden har det vært en oppgang i kvinneandelen. Oppgangen i 2004 bidrar kanskje til økningen i sykefraværet i samme periode.

Figur 6: Sykefraværet til kvinner og menn for 4. kvartal



4.3.4 Alder

Det er klare aldersforskjeller i sykefraværet, og jeg vil derfor inkludere kontrollvariabler for alderskomposisjonen i befolkningen på fylkesbasis. Figur 7 viser ikke overraskende at sykefraværet er høyest i den eldste alderskategorien og lavest i den yngste. Den mest åpenbare årsaken til dette er at eldre har større helseproblemer enn yngre av naturlige årsaker. En annen faktor påpekt av Yaniv (1991) er at kostnaden av fravær går ned med tiden, siden det er færre resterende perioder man kan bli straffet i. Alderskategoriene i figuren samsvarer ellers ikke helt med aldersandelsvariablene jeg har tilgjengelig, men

illustrerer likevel hvor betydelige forskjellene kan være. Ikke inkludert er sykefraværet for de over 66 år, som jeg har med i panelet mitt. I tillegg mangler data for 2009.

Figur 7: Sykefraværet for 4. kvartal etter alder



Alderskategoriene jeg har tilgjengelig er 16-19, 20-24, 25-39, 40-54, 55-66 og 67-74 år.

Regneoperasjonen er den samme som for kvinneandelen, f.eks. er aldersgruppen 20-24 år:

$$a_{24}_{jt} = \frac{Sysselsatte_{jt}^{20-24\text{år}}}{Sysselsatte_{jt}} * 100 \quad (4.13)$$

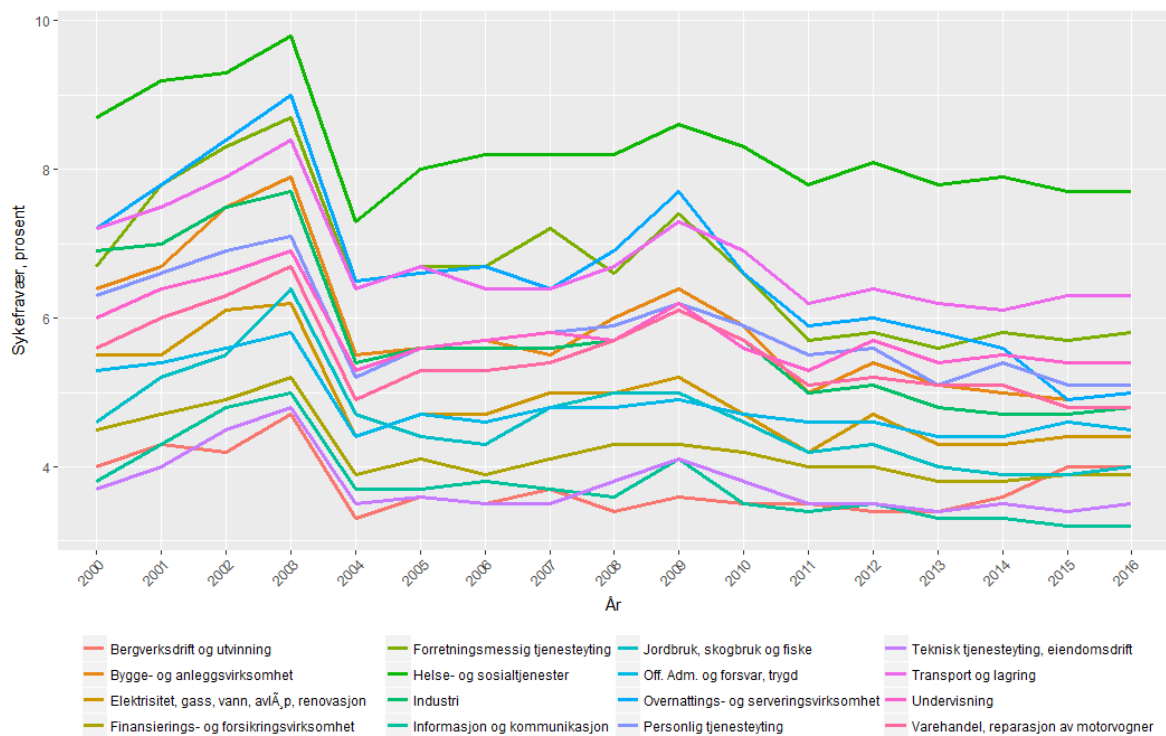
Resultatet er seks variabler som kanskje blir litt i overkant for å beskrive aldersutviklingen. Jeg har valgt å slå sammen de to yngste, a_{19} og a_{24} , til én variabel kalt *unge*. I tillegg vil jeg inkludere variabelen a_{74} . Det jeg håper å oppnå med dette valget er å fange opp den mest betydningsfulle forskjellen i alderssammensetning for fraværet, nemlig de aller yngste, hvor det er store endringer spesielt rundt 2008. Disse endringene er trolig er assosiert med finanskrisen. For den eldste gruppen har jeg observert at en økning for denne kategorien ofte estimeres til å ha en negativ effekt på sykefraværet. Det virker rimelig å tro at dette

kommer av at de som er syke i denne gruppen vil benytte seg av muligheten for å gå av med pensjon. Dermed blir det en seleksjonseffekt jeg ønsker å ta høyde for. Jeg merker meg at denne inndelingen gir ganske like estimater som å inkludere alle aldersvariablene.

4.3.5 Næring

Det er vanlig å kontrollere for forskjeller i arbeidsforhold i studier av sykefraværet, og jeg har forsøkt å gjøre dette med et sett næringsvariabler. På en side ønsker jeg å kontrollere for næringene direkte, da noen næringer nok involverer en naturlig høyere risiko for skader eller generelt mer stress. Samtidig håper jeg at næringsvariablene sammen med variablene for utdanning, kjønn og alder takler variasjonen i lønninger jeg ikke har tall for. Det er også mulig at forskjellige sektorer har forskjeller i overtid, hvor Ose og Dyrstad (2003) finner at flere overtidstimer fører til høyere sykefravær.

Figur 8: Sykefravær for 4. kvartal etter næring



Jeg har ikke direkte prosenttall for sykefraværet på samme næringsnivå som variablene jeg inkluderer i panelet mitt, men figur 8 inneholder noen av de samme kategoriene og gir et inntrykk av hvor betydelige disse forskjellene kan være.

SSB oppgir antall sysselsatte per næringsgruppe, samt totalt antall sysselsatte. Fra dette har jeg regnet ut andelen av de sysselsatte i hver næringssektor, som f.eks. for variabelen *jord*:

$$jord_{jt} = \frac{Sysselsatte\ i\ kategorien_{jt}}{Sysselsatte_{jt}} * 100 \quad (4.14)$$

Det største bruddet i tallseriene jeg har hentet inn skjer i disse næringsvariablene. Som nevnt i diskusjonen om tallene for sykefravær skjedde det en endring i standarden for næringsgruppering i 2009. Etter dette publiseres tall hovedsakelig etter SN2007 i motsetning til SN2002. SSB har publisert omkodingstabeller mellom de to standardene, men disse bryter forandringene ned på et mye mer detaljert nivå enn jeg har mulighet til med de tallene som er publisert. Næringskodene er fem siffer, og ser ut som f.eks. koden for svinehold som er 01.230 i SN2002 og 01.460 i SN2007. Tallene jeg har benyttet oppgir kategorier som «01-05 Jordbruk, skogbruk og fiske» for SN2002 og «01-03 Jodbruk, skogbruk og fiske» for SN2007. De to første sifrene her korresponderer til de to første sifrene i de detaljerte næringskodene. For å koble sammen tallene så godt som mulig har jeg gått igjennom de 1232 linjene i omkodingsdokumentet for å se hovedtrekkene i omgrupperingene. På grunnlag av dette har jeg kommet frem til en nedgruppering som ser ut til å stemme ganske godt for både SN2002 og SN2007 i år 2008 hvor tall basert på begge grupperingene er tilgjengelig. Jeg har oppsummert denne omkodingen i figurene 9 og 10 under, med alternerende bakgrunn som samsvarer med næringsvariablene etter sammenslåing. Figur 11 viser andelen sysselsatte i hver kategori på landsbasis etter både SN2002 og SN2007 for 2008.

For å redusere mengden variabler i modellen og spare inn frihetsgrader, spesielt med tanke på hvor mange dummyvariabler jeg har, har jeg slått sammen noen variabler. Her er det ikke helt åpenbart hvordan man skal gå frem, og jeg har derfor tenkt etter tre kriterier. For det første ønsker jeg å slå sammen variabler med lignende observert sykefravær. For det andre er det en fordel om de beskriver tilsvarende trender. For det tredje virker det fornuftig å eliminere noen av de minste kategoriene. Jeg har endt opp med å slå sammen variablene *fin* og *tjen* til samlevariabelen *fintek* fordi det er naturlig overlapp mellom disse, og fordi de begge har litt under gjennomsnittlig sykefravær. Jeg startet med *vare* som referansekategori da denne har sykefravær rundt gjennomsnittet. Jeg flyttet også *adm*, *und* og *and* til

referansekategoriene, da disse er relativt små og igjen har lignende sykefravær. Dermed har jeg med *jord, ind, fintek og hel* som variabler i regresjonene.

Figur 9: Næringer i SN2002

Kategorier SN2002
01-05 Jordbruk, skogbruk og fiske
10-37 Industri, bergv., olje- og gassutv.
40-41 Kraft- og vannforsyning
45 Bygge- og anleggsvirksomhet
50-55 Varehandel, hotell- og restaurantvirksomhet
60-64 Transport og kommunikasjon
65-67 Finansiell tjenesteyting
70-74 Forretningsmessig tjenesteyting, eiendomsdrift
75 Off.adm. og forsvar, sosialforsikr.
80 Undervisning
85 Helse- og sosialtjenester
90-99 Andre sosiale og personlige tjenester

Figur 10: Næringer i SN2007

Kategorier SN2007
01-03 Jordbruk, skogbruk og fiske
05-09 Bergverksdrift og utvinning
10-33 Industri
35-39 Elektrisitet, vann og renovasjon
41-43 Bygge- og anleggsvirksomhet
45-47 Varehandel, reparasjon av motorvogner
49-53 Transport og lagring
55-56 Overnattings- og serveringsvirksomhet
58-63 Informasjon og kommunikasjon
64-66 Finansiering og forsikring
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

Figur 11: Personer i samlede næringskategorier SN2002 og SN2007

Kategorier	SN2002	SN2007	02 i %	07 i %
jord	77969	76794	3.1%	3.1%
ind	521680	516274	20.8%	20.5%
vare	623724	689519	24.8%	27.4%
fin	50747	50646	2.0%	2.0%
tjen	302773	278057	12.0%	11.1%
adm	156755	141292	6.2%	5.6%
und	192982	194584	7.7%	7.7%
hel	480832	477682	19.1%	19.0%
and	106290	88249	4.2%	3.5%

4.3.6 Utdanningsnivå

En annen variabel som kan tenkes å bidra til å forklare sykefraværet er utdanningsnivå.

Primært håper jeg at disse variablene skal bidra til å korrigere for variasjon i lønninger, men andre utfall som f.eks. typen arbeid som utføres (om arbeidet er fysisk krevende eller ikke)

kan også tenkes å fanges opp her. Utdanning er oppgitt på fire nivåer, samt en samlekategori

for de som har uoppgitt utdanning. De fire nivåene er: grunnskole (nivå 1-2), videregående (nivå 3-5), universitet og høyskole (nivå 6, 1-4 år) og universitet og høyskole (nivå 7-8, over 4 år). Jeg har holdt de med uoppgitt utdanning utenfor og kodet tre variabler som korresponderer til de tre nivåene *videre* for videregående, *uni* for universitet og høyskole og *uno* for universitet og høyskole over 4 år. Andelen med grunnskoleutdanning er referansekategori. Som eksempel er andelen med videregående som høyeste fullførte utdanning kodet:

$$vidre_{jt} = \frac{Sysselsatte_{jt}^{videregående}}{Sysselsatte_{jt}} * 100 \quad (4.15)$$

Det er generelt mye samvariasjon i utdanningsvariablene. Jeg har derfor vært aggressiv med sammenslåingen her og definert variabelen *utd* som andelen som har fullført videregående (eller høyere) utdanning. Dette er en sterk trendvariabel, hvor andelen øker i samtlige fylker over perioden, men med et fall rundt finanskrisen. Trolig skyldes dette at så mange unge var en del av arbeidsstyrken.

4.3.7 Dummyvariabler

Den siste variabelen i panelet er en tidsdummy for endringer i fastlegenes sykemeldingspraksis som et resultat av regelendringer. Disse endringene tredde i kraft fra og med 1. juli 2004. Jeg har derfor kodet variabelen *r04* som én i perioden [2004, 2016] og null ellers. Jeg har også inkludert en interaksjon mellom denne variabelen og arbeidsledigheten.

Jeg inkluderer også et fullt sett med årsummier. Ideen er å kontrollere for effekter unike for hvert år som jeg ikke ellers kan ta høyde for. Et eksempel på en slik effekt er influensatilfeller som varierer fra år til år. Dette er spesielt relevant siden tallene mine er fra 4. kvartal som markerer starten på influensasesongen.

4.4 KVASIEKSPERIMENT

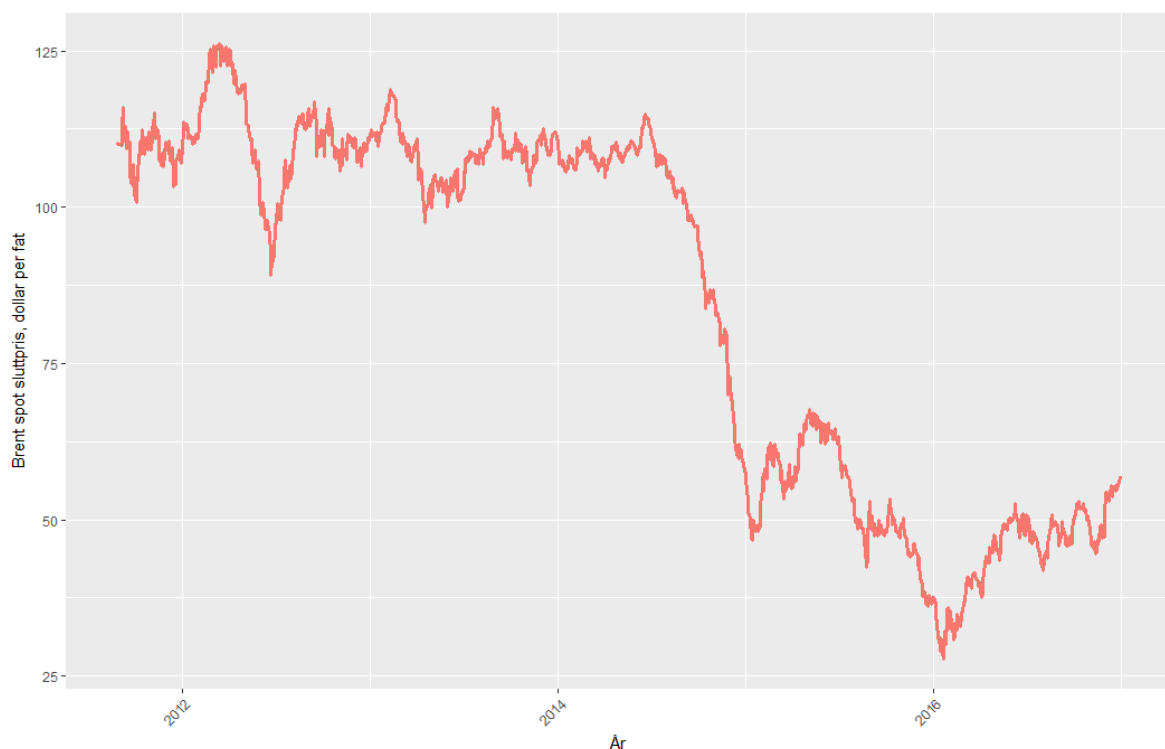
Estimering av makroøkonomiske effekter er problematisk av mange grunner. Én kilde til problemer er hvilke data som er tilgjengelig. Som samfunnsøkonom skulle man kanskje ønske at man kunne utføre eksperimenter, men dette er lite realistisk med tanke på etiske

og praktiske hensyn. Som svar på dette har man kvasieksperimenter, også kalt naturlige eksperimenter.

En vanlig kilde til denne typen analyse er gjennomføring av ny politikk. Hvis man f.eks. introduserer et nytt program ved en gruppe skoler, kan man studere effekten på de skolene opp mot effekten i en gruppe skoler som blir valgt som kontrollgruppe. En metode som kan anvendes i et slikt tilfelle er den såkalte «difference-in-differences» metoden.

Når det kommer til forholdet mellom arbeidsledighet og sykefravær er det kanskje ikke så åpenbart hvordan dette skal være nyttig. Jeg har ikke muligheten til å kontrollere arbeidsledighet som en variabel, og politikk som har som mål å endre sykefraværet er kanskje best undersøkt på sine egne meritter, som f.eks. ved introduksjonen av nye regler i 2004. Jeg vil allikevel forsøke å estimere en slik modell basert på en annen form for sjokk, nemlig det plutselige og dramatiske fallet i oljeprisene i 2014, som illustrert i figur 12.

Figur 12: Utviklingen i oljeprisen



For å bruke dette som et eksperiment har jeg kodet vestlands- og agderfylkene som behandlingsgrupper. Resten av landet blir brukt som kontrollgruppe. Grunnlaget for dette er andelen oljejobber i disse fylkene. For å introdusere så lite bias som mulig har jeg benyttet en rapport fra Blomgren et al. (2013) som bl.a. inneholder informasjon om andelen oljejobber i hvert fylke. Bare vestlands- og agderfylkene har en andel oljejobber som er lik eller høyere enn landsgjennomsnittet og jeg har derfor valgt disse som behandlingsgruppe.

Den andre faktoren i en «diff-in-diff» analyse er koding av en tidsvariabel som tar verdien én for perioden etter behandlingen, i dette tilfellet prisfallet, og null ellers. Dette valget er litt vanskeligere å rasjonalisere. En åpenbar kandidat er 2014. Panelet jeg har satt sammen har kun data fra siste kvartal hvert år og dette gir både rom for at oljeprisfallet har vært oppfattet som noe signifikant og for at adferd muligens kan endres av dette. En annen mulighet er å ta høyde for at responsen til prisfallet kan være noe forsinket, slik som man ofte gjør i sykefraværsanalyser i Norge ellers ved å lagge arbeidsledigheten. I dette tilfellet ville 2015 være en åpenbar kandidat. Jeg har valgt å inkludere en versjon av estimatene for begge alternativer.

«Diff-in-diff» estimeres ved bruk av vanlig MKM. Estimatet forklarer effekten av å motta behandlingen, og siden den estimeres med MKM kan det kontrolleres for andre faktorer. Metoden er et kvasieksperiment siden behandlingsgruppen strengt tatt ikke er tilfeldig valgt, men så lenge hendelsen er eksogen, f.eks. et uforutsett økonomisk sjokk, kan den anvendes.

Det er verdt å merke seg at denne variabelen tar plassen til arbeidsledigheten i modellen. Jeg estimerer således ikke effekten av arbeidsledigheten på sykefraværet i seg selv, men bruker variasjonen i oljeprisen som en slags proxyvariabel eller konjunkturindikator. Spørsmålet denne variabelen svarer på er egentlig «økte sykefraværet i vestlands- og agderfylkene spesielt som et resultat av fallet i oljeprisene»? Det blir utover det spekulasjon hva som er mekanismen bak. Begrunnet i teori kan man tenke seg to klare muligheter. Begge bygger på at fallet i oljeprisen kan tolkes som en lavkonjunktur som spesielt rammer disse fylkene. Med det i mente forventer jeg at sykefraværet vil falle, hovedsakelig fordi det vil ha en disiplinerende effekt på arbeidstakerne. Det er også mulig at en eventuell nedgang vil være et resultat av endret sammensetning i arbeidstakere, men andre Norske studier,

som f.eks. Askildsen et al. (2000), har sådd tvil om dette er spesielt relevant i Norge. Dyrstad og Ose (2002) finner at sammensetningshypotesen kan ha forklaringskraft for korttidsfraværet til menn, men for langtidsfraværet, som min analyse bruker, finner også de kun disiplineringseffekter. Ligningen jeg estimerer er:

$$\begin{aligned}
 syk_{jt} = & \alpha_j + did_{jt}\delta_1 + treat_j\delta_2 + post_t\delta_3 + kvinner_{jt}\beta_1 + unge_{jt}\beta_2 + a74_{jt}\beta_3 \\
 & + utd_{jt}\beta_4 + jord_{jt}\beta_5 + ind_{jt}\beta_6 + fintek_{jt}\beta_7 + hel_{jt}\beta_8 \\
 & + r04_t\delta_4 + arb_{jt} * r04_t\beta_9 + årsummier_t + \varepsilon_{jt}
 \end{aligned}
 \tag{4.16}$$

Koeffisienten δ_1 til variabelen $did = treat * post$ er den jeg er ute etter. Denne beskriver effekten av «behandlingen».

5 RESULTATER

Flere empiriske studier om forholdet mellom arbeidsledighet og sykefravær har benyttet seg av paneldatametoder, men det ikke meg bekjent at det finnes noen andre studier hvor enheten det er snakk om er fylker i Norge. Jeg har sånn sett ingen studier å direkte sammenligne resultatene mine med. Jeg vil derfor sammenligne resultatene mine med Norske studier som behandler ledigheten og sykefraværet på lavere nivå som bransjer, bedrifter og individer.

Jeg estimerer en versjon av modellen min med ledighetstall fra AKU og en med ledighetstall fra NAV. For hver av disse igjen vil jeg utføre regresjonen med både pooled, fixed og random effects. Dette resulterer i seks paneldataregresjoner totalt. I tillegg vil jeg presentere et kvasiekperiment i sin egen seksjon.

Jeg utfører tre spesifikasjonstester. Jeg kan ikke avvise nullhypotesen i Durbin-Watson testen om ingen seriekorrelasjon i restleddene, og har derfor ikke gjort noen justeringer av standardfeilene. Testen for uobserverte effekter er kjørt på fixed effects estimatene, og her avvises nullhypotesen om ingen seriekorrelasjon i restleddene. Dette indikerer at en pooled regresjon vil ha problemer med utelatte uobserverte effekter, og jeg har uansett grunn til å tro at resultatene her er upålitelige på bakgrunn av de observerte fylkesforskjellene i sykefraværet. For valget mellom fixed- og random effects har jeg anvendt Hausman-testen. For både regresjonen med AKU-ledighet og NAV-ledighet avvises nullhypotesen på et arbitrært lavt signifikansnivå. Det vil si at Hausman-testen indikerer at jeg ikke kan benytte estimatene med random effects. Alt i alt foretrekker jeg da som forventet estimatene med fylkeseffektene inkludert, og vil fokusere på disse.

Figur 13: Regresjonsresultater

	Sykefravær (log)					
	(1)	(2)	(3)	(4)	(5)	(6)
arbaku	0.004 (0.008)	0.008** (0.004)	0.007 (0.004)			
arbnav				-0.002 (0.010)	0.017** (0.007)	0.011 (0.008)
kvinner	-0.019 (0.013)	0.043*** (0.009)	0.029*** (0.010)	-0.039*** (0.013)	0.026*** (0.009)	0.015 (0.010)
r04	-0.174*** (0.042)	-0.109*** (0.038)	-0.072* (0.037)	-0.318*** (0.046)	-0.144*** (0.041)	-0.116*** (0.040)
unge	-0.025*** (0.005)	-0.025*** (0.004)	-0.015*** (0.004)	-0.022*** (0.005)	-0.018*** (0.005)	-0.011** (0.005)
a74	-0.041*** (0.014)	-0.019 (0.012)	-0.023* (0.013)	-0.016 (0.015)	-0.014 (0.012)	-0.018 (0.013)
utd	-0.017*** (0.002)	-0.007 (0.005)	-0.023*** (0.004)	-0.017*** (0.002)	-0.012** (0.005)	-0.025*** (0.004)
jord	-0.006* (0.003)	-0.030*** (0.005)	-0.015*** (0.005)	-0.006* (0.003)	-0.037*** (0.005)	-0.019*** (0.005)
ind	-0.004** (0.002)	0.004 (0.003)	0.005** (0.002)	-0.004** (0.002)	0.004 (0.003)	0.005** (0.002)
fintek	0.001 (0.004)	0.008*** (0.003)	0.005 (0.003)	0.004 (0.004)	0.010*** (0.003)	0.007** (0.003)
hel	0.040*** (0.004)	-0.016*** (0.006)	0.009* (0.005)	0.042*** (0.004)	-0.017*** (0.006)	0.009* (0.005)
arbaku:r04	-0.006 (0.009)	-0.007 (0.004)	-0.006 (0.005)			
arbnav:r04				0.033*** (0.012)	0.016*** (0.006)	0.014** (0.006)
FE	Nei	Ja	Nei	Nei	Ja	Nei
RE	Nei	Nei	Ja	Nei	Nei	Ja
DW	0.761	0.761	0.761	0.761	0.761	0.761
Wooldridge		0.000			0.000	
Hausman			0.000			0.000
Obs.	323	323	323	323	323	323
Justert R ²	0.839	0.930	0.915	0.846	0.936	0.918

* p<0.1; ** p<0.05; *** p<0.01

Effekten av ledigheten på sykefraværet er signifikant både for AKU-ledigheten og NAV-ledigheten. I begge tilfeller finner jeg overraskende en positiv sammenheng. Dette går imot tidligere norske studier som for eksempel Dyrstad og Lysø (1998). Den sterkeste effekten finner jeg med den registerbaserte ledigheten. Her er koeffisienten 0.017 som betyr at en økning i arbeidsledigheten på ett prosentpoeng vil øke sykefraværsprosenten med 1,7% – en veldig beskjeden effekt. For å illustrere dette tar jeg utgangspunkt i gjennomsnittet for sykefraværet i panelet på 6,18 prosent. NAV-ledigheten har et standardavvik på 0,8 og en økning på ett standardavvik vil dermed øke sykefraværet til 6,26 prosent. For AKU-ledigheten er effekten enda mer beskjeden. Denne ledigheten har litt større variasjon med et standardavvik på 1,22 som gir en predikert økning til 6,24 prosent. Selv om jeg finner en statistisk signifikant effekt kan man spørre seg om denne i praksis er veldig relevant.

Hvis jeg holder fast ved funnene til Dyrstad og Ose (2002) om at sammensettingshypotesen ikke er relevant for langtidsfraværet kan jeg tolke dette som at de empiriske funnene går imot de teoretiske prediksjonene om disiplineringshypotesen, og dermed også mange tidligere norske studier. En forklaring av det positive forholdet kan være at i nedgangstider blir deler av arbeidet til de som mister jobben gitt til de som beholder jobben, slik at deres arbeidsmengde øker. Dette kan føre til utslitte arbeidere som får høyere sannsynlighet for sykefravær.

Et annet alternativ er at sammensetning er av betydning, men at effekten er motsatt av den som er predikert av sammensetningshypotesen. Dette ville i så fall samsvare med funnene til Askildsen et al. (2000). Forfatterne finner her at marginale arbeidere har lavere fravær enn de ikke-marginale. Noe av dette skyldes trolig at de er yngre og tjener mindre, noe som passer godt overens med både teorimodellen og mine empiriske funn, men assosiasjonen overlever også kontroll for disse faktorene. Jeg tenker meg at en grunn til dette kan være at marginale arbeidere vet at de er først i køen for oppsigelse. Dette er spesielt sant i den grad det er vanlig å si opp de med laveste ansenitet først. I så fall kan det tenkes at straffen for fravær er forskjellig mellom individene. Dette gir en naturlig utvidelse av teorimodellen hvor straffen for eksempel kan betegnes som $P(A, U; Z)$ hvor Z representerer individspesifikke faktorer som ansenitet.

Et annet interessant resultat er interaksjonen mellom reformvariabelen og NAV-ledigheten. Denne er estimert til 0,016. Det vil si at effekten av arbeidsledigheten på sykefraværet har nesten doblet til 0,033. Jeg merker meg her at sykefraværet generelt er lavere etter reformen, og det er også variasjonen i arbeidsledigheten. Sykefraværet før og etter reformen er i snitt henholdsvis 7,09 og 5,90 prosent. Standardavviket til arbeidsledigheten er henholdsvis 0,85 og 0,73. Dette gir en predikert økning i sykefraværet av en ett standardavvik økning i arbeidsledigheten på 1,45 og 2,40 prosent. Dette tilsvarer en økning fra 7,09 til 7,19 og 5,90 til 6,04 prosent sykefravær. Det kvantitative utslaget er dermed ikke så mye større i praksis.

Med det i mente er det allikevel interessant å tenke på hvorfor utslaget skulle være større etter reformen. En mulighet jeg kan tenke meg er at disiplinering faktisk er relevant i sykefraværssammenheng, på tross av at jeg observerer en positiv effekt. Argumentet er at etter reformen ble implementert ble det vanskeligere å få legemeldt sykefravær, slik at disiplineringseffekten ble mindre. Dette har en direkte dempende effekt på sykefraværet som estimert ved reformvariabelen, siden det er vanskeligere å få fri selv om man skulle ønske det. Man kan si at for et gitt nivå på arbeidsledigheten vil sykefraværet være mindre. Reformen har også en effekt på forholdet mellom arbeidsledigheten og sykefraværet. Hvis arbeidsledigheten går ned forventer vi at folk vil være mer villige til å ta sykefravær. Hvis flere blir stoppet er derimot utslaget mindre. Samtidig er effekten av den økende andelen marginale arbeidere uendret slik at denne blir mer dominerende. Da øker den totale effekten som jeg har målt. Dessverre har jeg ikke muligheten til å dekomponere effektene her og sånn sett kan jeg heller ikke trekke noen sterke konklusjoner. Men dette ville stemme overens med argumentet om marginale arbeidere over. Det er meg bekjent ingen studier her jeg kan sammenligne resultatene mine med direkte.

På mange måter er jeg egentlig ute etter å svare på spørsmålet om hva effekten av arbeidsledigheten er på sykefraværet på et generelt nivå, men det er ikke nødvendigvis sant at en modell for sykefraværet på fylkesnivå vil gi lignende estimerte effekter som en på lands, bedrifts eller individnivå. Og det er ikke alltid klart at motsettende konklusjoner fra disse modellene er i direkte konflikt med hverandre. Det er mulig at en modell for individers

sykefravær finner at arbeidsledigheten har en negativ effekt på sykefraværet, mens en modell for sykefravær på fylkesnivå finner en positiv effekt. Begge disse resultatene kan være «riktige», f.eks. fordi sammensettingshypotesen ikke er relevant for å predikere sykefraværet til en enkelt person. Effekten i bedrifter kan igjen være annerledes av samme grunn. Hvis sammensettingseffekten i realiteten handler om typen arbeid som utføres, og typen arbeid som er etterspurt innad i hver bedrift ikke endres ved endringer i arbeidsledigheten, vil det i dette tilfellet ikke være sammensettingseffekter innad i hver bedrift, mens de kan finnes på f.eks. fylkesnivå hvor noen typer bedrifter legger ned mens andre opererer som normalt. På bakgrunn av dette resonnerer jeg ikke påstå at mine funn, at arbeidsledigheten har en positiv effekt på sykefraværet på fylkesbasis, er i direkte konflikt med tidligere funn. Derimot vil jeg trekke frem en gammel studie av Markham (1985) som fant at det negative forholdet holdt på alle nivå i USA: organisasjoner, regioner og for nasjonen som helhet. Det eneste unntaket var i nordøst, hvor han ikke fant et forhold. Studien er noe svak med tanke på perioder og variasjon i arbeidsledigheten i denne regionen, men forfatteren bemerker at regionen har sterkere arbeidslover og større konsentrasjon av fagforeninger enn de andre, noe som kanskje gjør regionen mer sammenlignbar med Norge.

Variablene for kvinneandelen er signifikante og positive. Dette samsvarer blant annet med Mastekaasa (2012). Det er interessant at kvinneandelen har betydelig større koeffisient i regresjonen med AKU-ledighet enn med NAV-ledighet, samtidig som effekten av arbeidsledigheten er mindre. Jeg lurer på om det kan være en sammenheng her mellom andelen av kvinner som fanges opp i fraværstatistikkene og resultatene mine. For eksempel var kvinners ledighet målt ved AKU 74% av menns ledighet mot 70% for den registerbaserte.¹² Uansett er effektene statistisk signifikante, men i praksis beskjedne. Kvinneandelen har et standardavvik på 0,59 noe som betyr at koeffisientene på 0,043 og 0,026 gir et utslag av en ett standardavvik økning i kvinneandelen på sykefraværsprosenten på 2,52 og 1,52 prosent.

¹² Tallene er fra fjerde kvartal 2016.

Reformvariabelen er kanskje av en viss interesse i seg selv. Denne er iallfall beregnet til å være negativ som forventet etter rapporten til Ose (2010). Effekten av reformen er også betydelig sammenlignet med variablene ellers. Koeffisientene på -0,109 og -0,144 betyr at reformen førte til henholdsvis 10,3 og 13,4 prosent reduksjon i sykefraværsprosenten.

Jeg anser aldersvariablene hovedsakelig som kontrollvariabler, men merker meg at sammenhengen mellom andelen unge og sykefraværet er negativ som forventet. Jeg finner ikke en signifikant effekt av andelen i den eldste aldersgruppen i denne spesifiseringen. Sammen med alder håper jeg at utdannings- og næringsvariablene skal kontrollere for lønn og type arbeid som utføres. Dette blir selvfølgelig ikke så presist med andeler på fylkesnivå, og jeg vil være litt forsiktig med å legge for mye vekt både på fortegn og signifikans her. Jeg finner en negativ sammenheng mellom utdanning og sykefravær, noe som går imot teorimodellens prediksjoner om effekten av lønn. Derimot gir det mening i den grad det måler effekten av utdanning i seg selv. Dette samsvarer f.eks. med resultatene til Askildsen et al. (2000) som finner en negativ effekt av utdanning på sykefraværet. Litt overraskende finner jeg en negativ effekt av andelen i helsesektoren, en gruppe som har høyt sykefravær. En faktor her er kanskje at denne sektoren har veldig stort antall kvinner i forhold til menn¹³ som er kontrollert for med en egen variabel.

5.1 KVASIEKSPERIMENT

Mange forsøk har blitt gjort på å spesifisere det nøyaktige forholdet mellom arbeidsledigheten og sykefraværet på et teoretisk nivå. Hele litteraturen bygger egentlig på en ganske enkel men generell idé: individer opplever gode eller dårlige økonomiske tider, noe som spiller inn på deres arbeidsmarkedsdeltakelse og sykefraværsbeslutninger. Et kvasieksperiment av den typen jeg har benyttet forholder seg mer agnostisk til nøyaktig hva som eventuelt forårsaker en endring i observert sykefravær, og retningen av denne effekten. I tillegg kan metoden gi videre støtte for at forholdet jeg observerer er kausalt. Modellen inneholder de samme industri, alder og utdanningsvariablene som de andre regresjonene, men disse er utelatt fra tabellen. I tillegg har jeg med både fylkes- og årsummier.

¹³ 81% av de ansatte i helse- og sosialsektoren var kvinner i 2016.

Variablene av interesse er did14 og did15. Effekten jeg estimerer for vestlandsfylkene etter oljeprisfallet er større enn effekten jeg har estimert for arbeidsledigheten ellers, med koeffisienter på 0,035 og 0,046. Dette tilsvarer en økning i sykefraværsprosenten på 3,56 og 4,71 prosent. Dette støtter opp under regresjonene over og indikerer sammen med disse at hvis en effekt av arbeidsledigheten på sykefraværet er en relevant forklaringsvariabel på fylkesbasis er det som en positiv effekt.

Videre er det interessant å tenke på hvorfor denne effekten er større enn den jeg finner ellers. Dette indikerer at noe relatert til konjunkturer påvirker fraværsbeslutningen utover arbeidsledigheten¹⁴. Jeg mener den mest naturlige forklaringen er et utslag i helsevariabelen fra teorimodellen. Spesielt tenker jeg at økt mentalt stress kan få utslag i sykefraværet. Dette er nok spesielt relevant når det skjer en plutselig endring i situasjonen i arbeidsmarkedet og for individets jobbsikkerhet. Dette er noe man kunne undersøkt direkte ved å studere tilfeller av forskjellige diagnoser i perioden.

Figur 14: Resultat av kvasiekspersiment

	Sykefravær (log)	
	(1)	(2)
treat	0.255*** (0.051)	0.250*** (0.051)
post14	-0.015 (0.012)	
did14	0.035*** (0.011)	
post15		-0.025** (0.012)
did15		0.046*** (0.012)
Behandlingsår	2014-16	2015-16
Obs.	323	323
Justert R ²	0.967	0.968

* p<0.1; ** p<0.05; *** p<0.01

¹⁴ En regresjon som inkluderer arbeidsledigheten i kvasiekspersimentet bekrefter dette.

Et mulig alternativ ville være at forventninger er viktig. Når oljeprisen faller svarer ikke arbeidsledigheten umiddelbart, men individer danner seg en forventning om det som kommer og kan endre adferd på bakgrunn av dette. Derimot er det ikke klart hvordan dette skulle føre til et større positivt utslag. Jeg ville tro at dette har en disiplinerende effekt på sykefraværet. Jeg har også vurdert dette opp mot forklaringen jeg presenterte over om lavere sykefravær blant marginale arbeidere. At effekten er større i kvasiekperimentet taler ikke imot denne forklaringen. Hvis andelen som oppfatter seg som marginale arbeidere går opp basert på forventninger til arbeidsmarkedet fremover vil jeg forvente en negativ effekt på sykefraværet, men jeg vil argumentere at konseptet marginale arbeidere er mer permanent. Arbeidere med lang ansenitet i en bedrift, og som har mye arbeidserfaring generelt, kan forvente å finne seg en ny jobb relativt fort hvis de først skulle miste den de har. Dermed er de ikke marginale arbeidere i ordets rette forstand.

6 OPPSUMMERING OG KONKLUSJON

Jeg har undersøkt sammenhengen mellom arbeidsledighet og det legemeldte sykefraværet på fylkesnivå med paneldata. I strid med forventningene jeg hadde basert på teori og tidligere studier fant jeg en positiv sammenheng. Disse funnene støttes opp av tilsvarende resultater for en kvasiexperimentell metode hvor effekten er beregnet til å være litt større. Alt i alt er ikke effektene store, men jeg mener de samlet sett er nok til å tale ikke bare imot at det er et negativt forhold mellom arbeidsledigheten og sykefraværet, men for at dette forholdet er positivt. Et viktig poeng her er at dette resultatet ikke nødvendigvis går imot tidligere funn som har vært på andre nivåer enn fylkesnivå.

Jeg finner også en effekt av endringene i fastlegenes sykemeldingspraksis gjennom reformen i 2004. Direkte er denne effekten negativ som både antatt og konkludert med i tidligere studier. Mer interessant er det at jeg finner en positiv effekt av interaksjonen mellom denne variabelen og arbeidsledigheten som målt med NAV-tallene. Jeg kan ikke dekomponere effektene formelt, men en mulig forklaring er at reformen har motvirket unødvendig fravær noe som har dempet eventuelle disiplineringseffekter av arbeidsledigheten. Da blir de resterende effektene mer gjeldene hvis disse er positive. Dette passer med en ide om at marginale arbeidere faktisk har mindre sannsynlighet for å være fraværende. Dette mener jeg ikke er urimelig siden de marginale arbeiderne kan tenkes å straffes hardere for fraværet sitt enn andre. Derfor har jeg foreslått at det er mulig å utvide straffefunksjonen med en individspesifikk variabel Z for å fange opp dette. Samtidig ville det være interessant å undersøke denne antakelsen nærmere og mer direkte.

7 LITTERATUR

- Allen, S. G. (1981). An Empirical Model of Work Attendance. *The Review of Economics and Statistics*, 63(1), 77-87.
- Askildsen, J. E., Bratberg, E. & Nilsen, Ø. A. (2000). *Sickness Absence over the Business Cycle*. University of Bergen,
- Blomgren, A., Quale, C., Bayer, S. B., Nyvold, C. E., Steffensen, T., Tovmo, P., . . . Hage, S. E. (2013). *Industribyggerne: Norsk olje- og gassnæring ut med havet og mellom bakkar og berg*: IRIS. Hentet fra [http://gammelweb.iris.no/internet/student.nsf/199f312efd2a0cacc125680e00635b85/561b24a55a1d7a61c12580ca0041a180/\\$FILE/IRIS%20rapport%202013-31%20Industribyggerne.pdf](http://gammelweb.iris.no/internet/student.nsf/199f312efd2a0cacc125680e00635b85/561b24a55a1d7a61c12580ca0041a180/$FILE/IRIS%20rapport%202013-31%20Industribyggerne.pdf)
- Brage, S., Nygård, J. F. & Tellnes, G. (1998). The gender gap in musculoskeletal-related long term sickness absence in Norway. *Scandinavian Journal of Social Medicine*, 26(1), 34-43.
- Brown, S. & Sessions, J. G. (1996). The Economics of Absence: Theory and Evidence. *Journal of Economic Surveys*, 10(1), 23-53.
- Dale-Olsen, H. & Markussen, S. (2010). Økende sykefravær over tid? Sykefravær, arbeid og trygd 1972–2008. *Søkelys på Arbeidslivet, Årgang 26 (1-2)*, 105-121.
- Dyrstad, J. M. & Lysø, N. (1998). Økonomiske faktorer bak sykefraværet. *Norsk Økonomisk Tidsskrift*, 112(2), 155-185.
- Dyrstad, J. M. & Ose, S. O. (2002). Worker absenteeism: Discipline or Composition Effects? *Essays on Worker Absenteeism*.
- Goddard, J. & Audas, R. (2001). Absenteeism, seasonality, and the business cycle. *Journal of Economics and Business*, 53, 405-419.
- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), 1251-1271.
- Hem, K.-G. (2011). *Bedriftenes kostnader ved sykefravær (SINTEF A19052)*: NHO. Hentet fra <https://www.sintef.no/contentassets/5c5963abef68468a9f22c5b5c3af201a/kostnader-sykefrav2011.pdf>

- Knutsson, A. & Goine, H. (1998). Occupation and unemployment rates as predictors of long term sickness absence in two Swedish counties. *Social Science & Medicine*, 47(1), 25-31.
- Leigh, J. P. (1985). The Effects of Unemployment and the Business Cycle on Absenteeism. *Journal of Economics and Business*, 37(2), 159-170.
- Leigh, J. P. (1995). Smoking, Self-Selection and Absenteeism. *The Quarterly Review of Economics and Finance*, 35(4), 365-386.
- Markham, S. E. (1985). An Investigation of the Relationship between Unemployment and Absenteeism: A Multi-Level Approach. *The Academy of Management Journal*, 28(1), 228-234.
- Mastekaasa, A. (2012). Kvinners og menns sykefravær – en stadig økende forskjell? *Søkelys på Arbeidslivet, Årgang 29* (1-2), 21-32.
- O'Reilly, F. W. & Stevens, A. B. (2002). Sickness absence due to influenza. *Occupational Medicine*, 52(5), 265-269.
- Ose, S. O. (2010). *Kunnskap om sykefravær: nye norske bidrag. Rapport A14516*. SINTEF Helse. Hentet fra <https://www.sintef.no/globalassets/upload/konsern/media/sykerapport.pdf>
- Ose, S. O. & Dyrstad, J. M. (2003). Absenteeism and Overtime *Essays on Worker Absenteeism*.
- Ose, S. O., Jensberg, H., Reinertsen, R. E., Sandsund, M. & Dyrstad, J. M. (2006). *Sykefravær: Kunnskapsstatus og problemstillinger*: SINTEF helse Rapport A325. Hentet fra <https://www.sintef.no/globalassets/upload/helse/okt/pdf-filer/sykefravar.pdf>
- Paringer, L. (1983). Women and Absenteeism: Health or Economics? *The American Economic Review*, 73(2), 123-127.
- Shapiro, C. & Stiglitz, J. E. (1984). Equilibrium Unemployment as a Worker Discipline Device. *The American Economic Review*, 74(3), 433-444.
- Statistisk Sentralbyrå. (2007). Forbedringer i Arbeidskraftundersøkelsen (AKU) i 2006. Hentet 14. august 2017 fra <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/forbedringer-i-arbeidskraftundersokelsen-aku-i-2006>
- Statistisk Sentralbyrå. (2009). Ny standard for næringsgruppering. Hentet 15. august 2017 fra <https://www.ssb.no/energi-og-industri/ny-standard-for-naringsgruppering>

- Statistisk Sentralbyrå. (2015). Hvorfor ulike arbeidsledighetstall? Hentet 24/11/2017 fra <http://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/hvorfor-ulike-arbeidsledighetstall--236723>
- Statistisk Sentralbyrå. (2016). Sysselsetting, registerbasert, 2015, 4. kvartal. Hentet 11. November 2017 fra <https://www.ssb.no/arbeid-og-lonn/statistikker/regsys/aar/2016-05-27>
- Statistisk Sentralbyrå. (2017a). Registrerte arbeidsledige. Hentet 11. November 2017 fra <https://www.ssb.no/arbeid-og-lonn/statistikker/regledig>
- Statistisk Sentralbyrå. (2017b). Sykefravær. Hentet fra <https://www.ssb.no/sykefratot>
- Thomson, L., Griffiths, A. & Davidson, S. (2000). Employee absence, age, and tenure: a study of nonlinear effects and trivariate models. *Work & Stress*, 14(1), 16-34.
- Verbeek, M. (2004). A Guide to Modern Econometrics (s. 342-358).
- Vistnes, J. P. (1997). Gender Differences in Days Lost from Work Due to Illness. *Industrial and Labor Relations Review*, 50(2), 304-323.
- Wooldridge, J. M. (2002). Testing for the Presence of an Unobserved Effect *Econometric Analysis of Cross-Section and Panel Data* (s. 264-265): MIT Press.
- Wooldridge, J. M. (2013). Random Effects Models *Introductory Econometrics A Modern Approach* (6th. utg., s. 474-476).
- Yaniv, G. (1991). Absenteeism and the Risk of Involuntary Unemployment: A Dynamic Analysis. *The Journal of Socio-Economics*, 20(4), 359-372.

8 VEDLEGG

8.1 FYLKEEFFEKTER OG DESKRIPTIV STATISTIKK PÅ FYLKESNIVÅ

Fylkeseffektene er som jeg mistenkte før estimeringen ofte betydelige og signifikante. Sogn og Fjordane har lavere observert sykefravær enn referansefylket Akershus. Derfor er det litt interessant at den fylkesfaste effekten estimeres til å være høyere. Oppland, som har et betydelig lavere faktisk sykefravær enn Finnmark, skiller seg ut ved å ha en nesten like høy beregnet fylkesfast effekt. Sammen med Nord-Trøndelag, Nordland og Troms danner de en gruppe av fylker med store positive fylkesfaste effekter. Alle de tre nordligste fylkene er med i denne gruppen.

Figur 15: Fylkesfaste effekter

	AKU	NAV
Akershus	0.714	1.789***
Aust-Agder	0.249***	0.213***
Buskerud	0.189***	0.174***
Finnmark	0.575***	0.524***
Hedmark	0.404***	0.406***
Hordaland	0.204***	0.184***
Møre og Romsdal	0.314***	0.293***
Nord-Trøndelag	0.555***	0.574***
Nordland	0.493***	0.478***
Oppland	0.478***	0.498***
Oslo	-0.139***	-0.145***
Østfold	0.310***	0.266***
Rogaland	0.001	-0.036
Sogn og Fjordane	0.334***	0.356***
Sør-Trøndelag	0.254***	0.250***
Telemark	0.258***	0.231***
Troms	0.522***	0.524***
Vest-Agder	0.257***	0.219***
Vestfold	0.150***	0.130***

Figur 16: Deskriptiv statistikk på fylkesnivå

Akershus					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	5.740	0.590	5.100	7.100
arbnav	17	2.050	0.480	1.100	2.600
arbaku	17	2.910	0.750	1.560	4.080
kvinner	17	47.860	0.210	47.510	48.110
a74	17	2.120	0.540	1.400	2.990
jord	17	1.060	0.200	0.730	1.330
ind	17	14.090	0.530	13.390	15.380
hel	17	17.150	1.030	14.900	18.250
r04	17	0.760	0.440	0.000	1.000
fintek	17	17.360	0.740	16.430	18.590
utd	17	78.290	1.480	76.110	80.720
unge	17	11.960	0.630	10.930	13.030

Aust-Agder					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	6.040	0.560	5.600	7.400
arbnav	17	3.130	0.670	1.700	4.300
arbaku	17	3.650	2.020	0.000	7.020
kvinner	17	46.860	0.460	45.820	47.720
a74	17	2.110	0.720	1.100	3.260
jord	17	2.400	0.320	1.870	2.900
ind	17	24.460	1.000	22.500	26.550
hel	17	20.200	1.280	18.180	22.790
r04	17	0.760	0.440	0.000	1.000
fintek	17	10.050	0.390	8.980	10.640
utd	17	77.620	1.420	75.620	80.280
unge	17	14.190	0.830	12.820	15.830

Buskerud					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	6.290	0.720	5.500	8.000
arbnav	17	2.350	0.440	1.500	2.900
arbaku	17	3.460	0.960	2.170	5.480
kvinner	17	47.210	0.270	46.780	47.610
a74	17	2.310	0.680	1.360	3.400
jord	17	2.220	0.340	1.590	2.770
ind	17	23.070	0.660	22.200	24.750
hel	17	19.370	1.140	16.830	20.950
r04	17	0.760	0.440	0.000	1.000
fintek	17	12.530	0.500	11.610	13.260
utd	17	75.680	1.720	73.040	78.310
unge	17	12.500	0.650	11.400	13.770

Finnmark					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	7.360	0.900	6.100	9.300
arbnav	17	3.880	1.020	2.900	5.900
arbaku	17	3.120	1.930	0.000	7.890
kvinner	17	47.210	0.340	46.450	47.690
a74	17	2.230	0.660	1.250	3.150
jord	17	6.720	0.410	6.030	7.410
ind	17	17.240	1.150	15.390	19.190
hel	17	21.300	0.740	19.810	22.860
r04	17	0.760	0.440	0.000	1.000
fintek	17	7.290	0.930	5.780	8.900
utd	17	69.520	1.690	67.220	72.730
unge	17	14.270	1.110	12.620	15.630

Hedmark					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	6.590	0.850	5.500	8.400
arbnav	17	2.450	0.360	1.700	2.900
arbaku	17	3.380	1.080	2.060	5.500
kvinner	17	47.160	0.260	46.540	47.610
a74	17	2.640	0.870	1.460	3.930
jord	17	5.740	0.670	4.380	6.790
ind	17	20.280	1.350	18.600	22.610
hel	17	21.510	1.700	18.430	24.130
r04	17	0.760	0.440	0.000	1.000
fintek	17	9.450	0.510	8.450	10.260
utd	17	73.370	2.250	69.940	76.990
unge	17	12.520	0.440	11.990	13.590

Hordaland					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	5.960	0.520	5.400	7.400
arbnav	17	2.590	0.690	1.500	3.600
arbaku	17	3.250	1.020	1.930	4.980
kvinner	17	46.860	0.270	46.430	47.420
a74	17	1.860	0.450	1.180	2.560
jord	17	2.150	0.370	1.710	2.850
ind	17	23.490	0.560	22.290	24.430
hel	17	19.660	0.830	18.140	21.370
r04	17	0.760	0.440	0.000	1.000
fintek	17	12.720	0.630	11.400	13.540
utd	17	78.910	1.850	75.940	82.190
unge	17	14.000	0.630	13.030	15.290

Møre og Romsdal					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	5.860	0.480	5.300	7.000
arbnav	17	2.340	0.730	1.200	3.800
arbaku	17	3.230	0.930	1.550	4.760
kvinner	17	45.970	0.380	45.220	46.600
a74	17	2.340	0.620	1.510	3.270
jord	17	5.130	1.140	3.610	6.970
ind	17	27.300	0.870	26.190	29.270
hel	17	19.460	1.170	17.290	21.760
r04	17	0.760	0.440	0.000	1.000
fintek	17	8.790	0.870	7.270	9.870
utd	17	75.470	2.000	72.450	78.960
unge	17	13.920	0.820	12.450	15.600

Nordland					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	6.750	0.710	5.900	8.400
arbnav	17	2.890	0.660	2.000	4.200
arbaku	17	3.430	0.910	1.720	5.130
kvinner	17	47.010	0.260	46.390	47.410
a74	17	2.190	0.720	1.170	3.270
jord	17	5.960	0.890	4.600	7.210
ind	17	19.310	0.550	18.310	20.140
hel	17	22.130	1.080	19.970	23.890
r04	17	0.760	0.440	0.000	1.000
fintek	17	7.890	0.230	7.400	8.310
utd	17	72.960	1.670	70.470	75.860
unge	17	13.970	0.630	13.170	15.050

Oslo					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	5.310	0.710	4.400	6.700
arbnav	17	3.260	0.750	2.100	4.700
arbaku	17	4.110	0.990	2.450	5.410
kvinner	17	48.210	0.190	47.860	48.450
a74	17	1.680	0.380	1.200	2.340
jord	17	0.200	0.050	0.130	0.250
ind	17	10.200	1.050	9.010	12.140
hel	17	16.810	1.010	14.430	17.660
r04	17	0.760	0.440	0.000	1.000
fintek	17	21.500	1.390	19.910	23.880
utd	17	83.540	1.530	80.870	85.800
unge	17	11.230	0.580	10.340	12.320

Rogaland					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	4.780	0.400	4.200	5.800
arbnav	17	2.580	1.070	1.000	4.500
arbaku	17	3.150	1.620	0.860	7.090
kvinner	17	46.290	0.360	45.900	47.250
a74	17	1.680	0.410	1.100	2.320
jord	17	3.550	0.930	2.410	5.000
ind	17	29.140	0.630	28.160	30.480
hel	17	17.040	1.130	14.710	19.310
r04	17	0.760	0.440	0.000	1.000
fintek	17	11.540	0.930	10.020	12.760
utd	17	76.950	1.850	74.310	80.490
unge	17	14.800	0.870	12.990	16.450

Nord-Trøndelag					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	6.490	0.600	5.500	8.000
arbnav	17	2.740	0.640	2.000	3.800
arbaku	17	2.850	1.050	1.470	4.920
kvinner	17	46.580	0.330	45.990	47.230
a74	17	2.310	0.710	1.250	3.270
jord	17	8.460	1.450	6.190	10.700
ind	17	21.840	0.530	21.010	22.720
hel	17	21.020	1.280	18.870	23.430
r04	17	0.760	0.440	0.000	1.000
fintek	17	8.120	1.030	6.650	9.820
utd	17	78.110	1.250	75.810	80.110
unge	17	13.680	0.690	12.710	14.900

Oppland					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	6.710	0.710	5.900	8.200
arbnav	17	1.980	0.350	1.200	2.600
arbaku	17	2.990	0.830	2.080	4.250
kvinner	17	47.010	0.200	46.520	47.270
a74	17	2.690	0.870	1.500	4.040
jord	17	6.200	0.900	4.610	7.660
ind	17	21.700	0.820	20.730	23.450
hel	17	21.650	1.250	19.190	23.500
r04	17	0.760	0.440	0.000	1.000
fintek	17	8.880	0.510	7.870	9.480
utd	17	75.380	1.930	71.820	78.340
unge	17	13.150	0.520	12.280	14.190

Østfold					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	6.960	0.720	6.200	8.700
arbnav	17	2.970	0.580	1.700	3.800
arbaku	17	3.780	0.820	2.360	5.070
kvinner	17	46.980	0.210	46.450	47.350
a74	17	2.110	0.650	1.190	3.150
jord	17	2.220	0.350	1.670	2.730
ind	17	23.970	1.810	21.650	27.280
hel	17	19.810	1.570	16.700	22.210
r04	17	0.760	0.440	0.000	1.000
fintek	17	11.100	0.430	10.210	11.710
utd	17	73.780	1.660	71.120	76.540
unge	17	12.910	0.550	12.110	13.960

Sogn og Fjordane					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	5.330	0.520	4.800	6.600
arbnav	17	1.750	0.420	0.900	2.600
arbaku	17	2.100	1.470	0.000	5.560
kvinner	17	46.530	0.280	45.890	46.970
a74	17	2.850	0.720	1.820	3.970
jord	17	7.430	1.400	5.240	9.730
ind	17	25.790	0.820	24.660	27.510
hel	17	20.290	1.770	16.050	22.790
r04	17	0.760	0.440	0.000	1.000
fintek	17	7.510	0.640	6.610	8.200
utd	17	77.310	1.710	74.680	80.110
unge	17	14.470	0.670	13.380	15.880

Sør-Trøndelag					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	6.060	0.660	5.400	7.600
arbnav	17	2.640	0.600	1.900	3.700
arbaku	17	3.310	0.690	2.170	4.960
kvinner	17	46.810	0.230	46.420	47.100
æ74	17	1.910	0.490	1.200	2.720
jord	17	3.650	0.730	2.570	4.820
ind	17	18.980	0.460	18.360	19.910
hel	17	19.190	1.170	16.620	20.640
r04	17	0.760	0.440	0.000	1.000
fintek	17	14.090	0.940	12.520	15.280
utd	17	79.950	1.850	76.850	82.900
unge	17	13.220	0.760	12.060	14.360

Telemark					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	6.310	0.770	5.500	8.100
arbnav	17	3.010	0.610	1.700	4.200
arbaku	17	3.770	1.140	2.350	6.100
kvinner	17	47.250	0.400	46.470	47.840
æ74	17	2.310	0.810	1.190	3.620
jord	17	2.330	0.250	1.800	2.700
ind	17	26.030	1.490	24.450	29.120
hel	17	21.380	1.410	18.800	23.350
r04	17	0.760	0.440	0.000	1.000
fintek	17	10.080	0.880	8.730	11.170
utd	17	76.340	1.760	73.700	79.420
unge	17	13.290	0.680	12.240	14.580

Troms					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	6.970	0.650	6.000	8.200
arbnav	17	2.380	0.540	1.700	3.400
arbaku	17	3.160	1.020	1.160	4.940
kvinner	17	47.220	0.290	46.410	47.490
æ74	17	2.160	0.740	1.080	3.170
jord	17	4.510	0.940	3.100	6.120
ind	17	15.060	0.420	14.490	15.750
hel	17	23.870	1.420	20.310	25.790
r04	17	0.760	0.440	0.000	1.000
fintek	17	9.140	0.440	8.400	9.950
utd	17	74.880	1.730	71.920	77.880
unge	17	14.050	0.580	13.170	14.810

Vest-Agder					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	5.970	0.790	5.200	7.600
arbnav	17	2.910	0.700	1.400	4.300
arbaku	17	3.740	1.160	2.100	5.430
kvinner	17	46.620	0.460	45.660	47.540
æ74	17	2.030	0.570	1.180	2.830
jord	17	2.140	0.410	1.570	2.730
ind	17	25.620	0.650	24.220	26.630
hel	17	19.880	1.420	17.150	22.410
r04	17	0.760	0.440	0.000	1.000
fintek	17	10.110	0.670	8.730	11.150
utd	17	78.520	1.460	76.060	81.310
unge	17	15.180	0.720	13.870	16.790

Vestfold					
Variabel	Obs.	Gj. Snitt	St. Avvik	Min	Maks
syk	17	5.870	0.610	5.300	7.300
arbnav	17	2.750	0.530	1.700	3.800
arbaku	17	3.660	0.860	2.520	5.600
kvinner	17	47.430	0.310	46.660	47.820
æ74	17	2.170	0.620	1.320	3.150
jord	17	2.000	0.390	1.460	2.620
ind	17	22.430	1.190	20.840	25.040
hel	17	19.650	1.380	16.880	21.770
r04	17	0.760	0.440	0.000	1.000
fintek	17	11.480	0.420	10.820	12.190
utd	17	77.180	1.450	74.920	79.430
unge	17	13.330	0.780	12.050	14.780

8.2 DATAKILDER

Tallene for sykefravær er publisert i «Tabell: 08320: Legemeldt sykefravær for arbeidstakere, etter bostedsfylke, kjønn og alder (F)».

Tallene for arbeidsledighet i AKU er hentet fra «Tabell 05615: Arbeidsstyrken og sysselsatte.

NAV-arbeidsledigheten er hentet fra tabell 04471 og 10593, begge med tittelen «Registrerte arbeidsledige 15-74 år, etter kjønn (prosent) (K)».

Næringsvariablene, kvinneandelen og aldersvariablene kommer fra tabell 03257, 06446 og 07984 som alle har tittelen «Sysselsatte per 4. kvartal, etter bosted, arbeidssted, kjønn, alder og næring».

Utdanningsvariablene for er hentet fra «Tabell: 11615: Sysselsatte per 4. kvartal, etter bosted, arbeidssted, kjønn, alder, fagfelt og utdanningsnivå (K)», som også er basert på registertall. Jeg har benyttet tall for bostedsfylke.

8.3 KOMPARATIVE STATISKE RESULTATER

De andre komparative statistiske resultatene følger på samme måte som for ledigheten, og jeg går fort igjennom de her. Dette gjør jeg ved å finne $\det(A_W)$, $\det(A_k)$, $\det(A_H)$ og $\det(A_S)$. $-f_{1,W} = H - A(1 - k)$ som ikke generelt kan bestemmes. Uttrykket er større enn null for A mindre enn H eller hvis k er lik én. $-f_{2,W} = (1 - k), > 0$. Dermed finner jeg at $\det(A_W) = (1 - k) - f_{2,C}[H - A(1 - k)]$. Siden $f_{2,C} > 0$ får jeg at

$$\frac{\delta A}{\delta W} = \frac{\det(A_W)}{\det(A)}, > 0 \text{ hvis } k = 1 \text{ eller } H > A, \text{ ellers ubestemt.}$$

Intuisjonen her er at hvis kompensasjonsgraden er hundre prosent vil det være en ren inntektseffekt av økningen i lønn. Siden både konsum og fritid er normale goder forventer jeg at etterspørselen etter begge øker.

Videre finner jeg at $-f_{1,k} = WA, > 0$ og $-f_{2,k} = -W, < 0$. Da får jeg at $\det(A_k) = -W - f_{2,C}WA, < 0$ slik at

$$\frac{\delta A}{\delta k} = \frac{\det(A_k)}{\det(A)}, > 0.$$

En høyere kompensasjonsgrad reduserer prisen på fritid og øker fraværet. For H finner jeg $-f_{1,H} = W, > 0$ og $-f_{2,H} = 0$ noe som gir $\det(A_H) = -f_{2,C}W, < 0$. Fra dette følger det at

$$\frac{\delta A}{\delta H} = \frac{\det(A_H)}{\det(A)}, > 0.$$

De siste resultatene er $-f_{1,S} = 0$ og dermed får jeg at

$$\det(A_S) = -f_{2,S} = -\frac{V_{LS}V_C - V_{CS}V_L}{V_C^2}, < 0.$$

Fra dette får jeg

$$\frac{\delta A}{\delta S} = \frac{\det(A_S)}{\det(A)}, > 0.$$

Oppsummert finner jeg dermed at A er en funksjon av W, k, H, U og S som:

$$A = A(W, k, H, U, S)$$

? + + - +