

Sosial interaksjon og sykefravær

En empirisk analyse av sosiale normer og sykefravær i et
norsk forsikringsselskap i perioden 2003-2007

Av

Linda Østergren

Masteroppgave

Masteroppgaven er levert for å fullføre graden

Masterstudiet i samfunnsøkonomi

Universitetet i Bergen, Institutt for økonomi

Vår 2016

UNIVERSITETET I BERGEN



Forord

Denne masteroppgaven markerer slutten på en fin studietid ved Universitetet i Bergen. Arbeidet med oppgaven har vært spennende og en svært lærerik prosess. Jeg er takknemlig for at jeg har fått mulighet til å vie et helt år til å jobbe med et fagområde som jeg synes er veldig interessant. Det har derimot vært utfordrende og litt frustrerende til tider.

Jeg ønsker først og fremst takke min hovedveileder Kjell Vaage for både forslag til oppgave og nyttige kommentarer underveis. Hans råd har vært verdifulle gjennom hele skriveprosessen. Jeg vil også takke min biveileder Gaute Torsvik som har gitt meg tilgang til dataene som benyttes i oppgaven.

For hjelp med korrekturlesing vil jeg takke min mor, Katarina Kaarbøe. Jeg vil også rette en takk til Oddvar Kaarbøe, både for at han har vært en fantastisk mentor gjennom hele studieløpet og for hans nyttige innspill til innhold og struktur i masteroppgaven. Til slutt ønsker jeg takke Sindre som gitt meg motivasjon og bidratt til å holde humøret oppe under hele prosessen.

Alle analyser og tolkninger i oppgaven er mine egne, og jeg står selv ansvarlig for eventuelle feil.

Linda Østergren

Linda Østergren, Bergen, 30.05.2016

Sammendrag

Sosial interaksjon og sykefravær

En empirisk analyse av sosiale normer og sykefravær i et norsk forsikringsselskap i perioden 2003-2007

av

Linda Østergren, Masterstudiet i samfunnsøkonomi

Universitetet i Bergen, 2016

Veiledere: Kjell Vaage og Gaute Torsvik

Vi benytter data fra et kundeservicesenter i et norsk forsikringsselskap til å analysere om det eksisterer sosiale interaksjonseffekter for sykefravær. Datasettet inneholder detaljert informasjon om de ansattes sykefravær med ukentlige observasjoner som løper fra første kvartal i 2003 til og med andre kvartal i 2007. De ansatte på kundeservicesenteret jobber i definerte team på ca. 16 medarbeidere. Vi har informasjon om 23 ulike team som sitter spredt i Norge. I tillegg til å undersøke om det faktisk eksisterer sosiale interaksjonseffekter, undersøker vi om noen typer sykefravær er mer påvirket enn andre, om det er noen forskjeller mellom kvinner og menn, og om effektene endres over tid. Fokuset i oppgaven er å identifisere den såkalte endogene sosiale interaksjonseffekten, dvs. effekten av teammedlemmenes sykefravær på et individs sykefravær.

Det er flere utfordringer knyttet til identifisering av sosiale interaksjonseffekter. De største utfordringene er knyttet til problemer med simultanitet, korrelerte effekter og seleksjon. Vi formulerer to modeller som forsøker å isolere den kausale effekten av teammedlemmenes sykefravær på et individs sykefravær. I den første modellen studerer vi sykefraværsadferd blant nyansatte. Denne omtales som «nyansatt-modellen». Vi undersøker om det er slik at nyansatte tenderer å tilpasse sitt sykefravær til det gjennomsnittlige sykefraværet i teamet som de tilhører. Vi benytter en fasteffekt modell og instrumentvariabelmetoden som identifikasjonsstrategi. I den andre modellen studerer vi sykefraværsadferd blant ansatte som flytter fra et team til et annet. Denne omtales som «flytte-modellen». Ved å sammenligne sykefravær i nytt team og

gammelt team, kan vi se om disse individene endrer adferd som følge av flyttingen. Vi benytter instrumentvariabelmetoden som identifikasjonsstrategi i denne modellen.

I nyansatt-modellen finner vi ingen indikasjon på at det eksisterer sosiale interaksjonseffekter for sykefravær i kundeservicesenteret. I flytte-modellen finner vi derimot signifikante effekter når vi studerer menn over en observasjonsperiode på to kvartal, og når vi studerer flytterne (kvinner og menn) over en observasjonsperiode på fire kvartal. Vi finner at menn reduserer sitt sykefravær når det gjennomsnittlige sykefraværet i teamet øker marginalt. Når vi studerer flytterne i fire kvartal finner vi en positiv effekt, dvs. sykefraværet øker når det gjennomsnittlige sykefraværet i teamet øker marginalt. Når vi gjennomfører placebotester for å teste robustheten til våre resultat finner vi at effektene trolig skyldes adferd hos teamet. Vi konkluderer med at det ser ut til å eksistere sosiale interaksjonseffekter for sykefravær i kundeservicesenteret som vi analyserer. Det er imidlertid noen svakheter ved analysen som må belyses. Vi har relativt få observasjoner og forklaringsvariabler, noe som potensielt skaper problemer med ekstern validitet.

Stata/IC 14 er benyttet til de statistiske beregningene i den empiriske analysen.

Innholdsfortegnelse

Forord.....	i
Sammendrag	ii
Innholdsfortegnelse.....	iv
Tabeller	vi
Figurer.....	vii
1 Innledning	1
2 Bakgrunn og institusjonelle forhold.....	6
2.1 Utviklingen i sykefraværet i Norge og dagens situasjon	6
2.2 Sykelønnsordningen i Norge: Regler for sykepenger.....	11
2.3 Internasjonal sammenligning av sykefravær og sykelønnsordninger	13
2.4 Er det aggregerte bildet preget av sosiale normer?	16
3 Teoretisk rammeverk	17
3.1 Optimalt bruk av sykepenger ved eksistens av sosial interaksjon	17
3.2 Lokal variasjon i sykefravær ved eksistens av sosial interaksjon	19
4 Tidligere empiri.....	23
5 Identifikasjon	28
5.1 Identifikasjon av sosiale interaksjonseffekter: en generell modell	28
5.2 Metodiske utfordringer	29
5.3 Identifikasjonsstrategier.....	32
6 Kundeservicesenteret	36
6.1 Arbeidsoppgaver.....	36
6.2 Organisering	37
6.3 IA-virksomhet.....	37
7 Data.....	39
7.1 Datasettet	39
7.1.1 Styrker ved datasettet	39
7.1.2 Svakheter ved datasettet	41
7.1.3 Paneldata	41
7.2 Variabler	41
7.2.1 Utfallsvariabler.....	41
7.2.2 Forklaringsvariabler	42
7.3 Deskriptiv statistikk.....	43
7.3.1 Gjennomsnittlige verdier til våre variabler.....	43
7.3.2 Utvikling i sykefravær over tid	47
7.3.3 Variasjon mellom teamene.....	49

8	Empiriske modellspesifikasjoner	52
8.1	Nyansatt-modellen.....	52
8.2	Flytte-modellen.....	53
8.3	Forutsetninger i fasteffekt modellen og krav til konsistens	54
8.4	Begrensninger i datasettet.....	57
8.4.1	Begrensninger i nyansatt-modellen	57
8.4.2	Begrensninger i flytte-modellen	58
9	Empirisk analyse	59
9.1	Resultat fra den generelle modellen.....	59
9.2	Resultat fra nyansatt-modellen	62
9.3	Resultat fra flytte-modellen	65
9.4	Heterogenitet over kjønn og tid	67
9.4.1	Heterogenitet i nyansatt-modellen.....	68
9.4.2	Heterogenitet i flytte-modellen	69
9.5	Oppsummering av resultat	71
9.6	Robusthet: placebo tester	71
9.7	Mekanismer	73
10	Svakheter ved analysen	76
10.1	Få observasjoner	76
10.2	Få forklaringsvariabler	77
10.3	Målefeil.....	79
11	Konklusjon	81
	Referanseliste	83
	Appendiks	86
A.1	Figurer	86
A.2	Tabeller	88

Tabeller

Tabell 1: Oversikt over sykelønnsordninger i utvalgte land.....	15
Tabell 2: Utfallsvariabler	44
Tabell 3: Forklaringsvariabler.....	45
Tabell 4: Kjønnforskjeller	46
Tabell 5: Resultat fra den generelle modellen.....	60
Tabell 6: Resultat fra nyansatt-modellen.....	63
Tabell 7: Resultat for ulike typer sykefravær i nyansatt-modellen	64
Tabell 8: Resultat fra flytte-modellen	66
Tabell 9: Resultat for ulike typer sykefravær i flytte-modellen	67
Tabell 10: Resultat for kvinner og menn i nyansatt-modellen	68
Tabell 11: Resultat for ulike observasjonsperioder i nyansatt-modellen.....	69
Tabell 12: Resultat for kvinner og menn i flytte-modellen	69
Tabell 13: Resultat for ulike observasjonsperioder i flytte-modellen	70
Tabell 14: Resultat fra placebotestene	73
Tabell A.1: Utfallsvariabler (nyansatte).....	88
Tabell A.2: Forklaringsvariabler (nyansatte)	88
Tabell A.3: Utfallsvariabler (flytterne)	89
Tabell A.4: Forklaringsvariabler (flytterne).....	89

Figurer

Figur 1: Sykefraværsprosenten fra 1980-2014 ifølge AKU	7
Figur 2: Andel arbeidstakere med legemeldt sykefravær 3. kvartal 2015	8
Figur 3: Andel arbeidstakere med legemeldt sykefravær etter alder 3. kvartal 2015	9
Figur 4: Andel arbeidstakere med legemeldt sykefravær etter bosted 3. kvartal 2015.....	10
Figur 5: Utvikling i sykefraværsprosenten i utvalgte land, 1996-2012	14
Figur 6: Multiple likevekter for fraværssraten	21
Figur 7: Utvikling i samlet sykefravær over tid	48
Figur 8: Utvikling i samlet sykefravær for kvinner og menn	49
Figur 9: Teamvariasjon i samlet sykefravær, kjønn, alder og teamstørrelse	50
Figur A.1: Utvikling i ulike typer sykefravær over tid.....	86
Figur A.2: Utvikling i ulike typer sykefravær over tid for kvinner og menn	87
Figur A.3: Sykefravær etter alder.....	87

1 Innledning

Sykefraværet i Norge har siden 1970-tallet vært et tilbakevendende tema i den offentlige debatten. Særlig den langvarige økningen i sykefraværet på 1990-tallet var bakgrunnen for at Stortinget vedtok den såkalte arbeidslinjen, og at partene i arbeidslivet gikk sammen om Inkluderende Arbeidsliv (IA-avtalen). Sykefraværet har etter den tid gått både opp og ned, men IA-avtalens målsetting fra 2001 er ennå ikke nådd. Proba samfunnsanalyse (2014a) påpeker at sykefraværet i Norge er høyt i internasjonal sammenheng. Det er for eksempel omtrent dobbelt så høyt i Norge som i Sverige, til tross for at landene er relativt like. I rapporten påpekes det at en av årsakene til at sykefraværet er så høyt i Norge sammenlignet med andre nordeuropeiske land, trolig er den sjenerøse sykelønnsordningen.

Høyt sykefravær er en stor kostnad for bedrifter og samfunnet, både på grunn av økte utgifter og tapt produksjon. Hem (2011) anslår at kostnaden av sykefravær for norske arbeidsgivere er ca. 13 000 kroner i uken. Myndighetene har anslått å bruke 40,8 milliarder kroner på sykepenger i 2016. Dette er en økning på 1,1 milliarder kroner fra saldert budsjett i 2015 (Regjeringen, 2015). En økt forståelse av hvilke faktorer som påvirker utviklingen i sykefraværet vil kunne bidra til økt innsikt i sykemeldingsproblematikken. Det er en forutsetning for å kunne redusere samfunnets kostnader til sykepenger, slik et samlet Storting har gått inn for.

Det er flere faktorer som påvirker et individs sykefravær. En opplagt faktor er individets helsetilstand. Et individ med dårlig helse vil ha høyere sannsynlighet for å bli sykemeldt enn et individ med god helse. Arbeidsmiljøet kan også påvirke en ansatt sitt sykefravær. Høye krav og press på arbeidsplassen kan føre til at ansatte blir sykemeldt som følge av stress, utbrenthet og psykiske lidelser. Videre viser flere studier at et lands konjunktursituasjon kan ha betydning for sykefraværet¹. Forskningen viser at det er en negativ korrelasjon mellom arbeidsledighet og sykefravær. Det finnes flere teorier som forsøker å forklare denne sammenhengen, blant annet marginaliseringshypotesen² og disiplineringshypotesen³ (NOU 2000:27).

¹ Se for eksempel Nordberg og Røed (2003), Askildsen et al. (2004) og Bjørnstad (2006)

² Marginaliseringshypotesen hevder at økt etterspørsel etter arbeidskraft gjør at flere med helseproblemer (marginale arbeidstakere) kommer inn i arbeidsstyrken, og da går sykefraværet opp.

³ Disiplineringshypotesen tar utgangspunkt i at det er mindre sannsynlig å miste jobben ved redusert arbeidsledighet. Arbeiderne har derfor svakere insentiver for å møte opp på jobb i «gode tider» og derfor går sykefraværet opp.

I tillegg til disse faktorene, kan sosial interaksjon være en årsak til sykefravær. Sosiale normer og holdninger knyttet til adferd vedrørende sykemeldinger vil kunne påvirke et individs sykefravær. Hvis mange på arbeidsplassen er sykemeldt, kan det tenkes å være lettere å sykemelde seg enn dersom det er få på arbeidsplassen som er sykemeldt. Vi påvirkes av våre kolleger både når det gjelder terskel for hva som oppfattes som «syk nok» til å være borte fra jobb, og terskel for regelrett skoft. Med andre ord, det er trolig lettere å være borte fra jobb dersom den sosiale normen på arbeidsplassen tilsier at det er greit å være borte fra jobb. Flere studier finner sosiale interaksjonseffekter knyttet til sykefravær⁴.

Sykelønnsordningen skal kompensere for den inntekten en arbeidstaker mister ved sykdom eller skade, og kan derfor ses på som en forsikringsordning⁵. Konseptet moralsk hasard blir naturlig å diskutere når vi snakker om en forsikringsordning. Moralsk hasard handler om at individ endrer adferd som følge av en forsikringskontrakt. Individ tar typisk høyere risiko når potensielle kostnader bæres delvis eller helt av andre⁶. I Norge, hvor erstatningsgraden for sykefravær er høy, reduseres individets kostnader ved å være borte fra jobb. Det er derfor ikke urealistisk at systemet overforbrukes til en viss grad. Overforbruket omfatter regelrett misbruk (skoft), men også at individet setter terskelen for sykefravær lavere enn det forsikringsfellesskapet legger til grunn. I tillegg er graden av kontroll lav i Norge, særlig for egenmeldt sykefravær men også for legemeldt sykefravær.

For en gitt helsetilstand er det mer sannsynlig at en person med sjenerøs sykelønnsordning er borte fra jobb enn en person med samme helsetilstand, men som har en mindre sjenerøs sykelønnsordning. Dette gjelder ikke ekstremtilfeller (de som er veldig syk eller skadet), men vil påvirke de arbeidstakerne som står på marginen til å være på jobb eller ikke. I denne gråsonen kan det tenkes at sosial interaksjon spiller en avgjørende rolle. Dersom den sosiale normen på arbeidsplassen er kjennetegnet ved en lav terskel for sykefravær kan det tenkes at mange av disse marginale arbeidstakerne blir sykemeldt. Kostnaden, både for bedrifter og samfunnet, av en kombinasjon av et sjenerøst system og positive sosiale interaksjonseffekter for sykefravær kan potensielt være veldig høy.

⁴ I kapittel 4 presenteres en rekke studier som ser på sammenhengen mellom sosial interaksjon og sykefravær. Se blant annet Ichino og Maggi (2000), Dale-Olsen et al. (2015) og Lindbeck et al. (2016).

⁵ Sykepenger er en type sosialforsikring. I dette tilfellet forsikres det mot risikoen for uforutsett inntektsbortfall ved arbeidsuførhet.

⁶ Se Eckhoudt et al. (2005) for mer om moralsk hasard.

Innen forskningen har det de senere år vært en økende interesse for betydningen av sosial interaksjon i ulike kontekster⁷. Sosiale interaksjonseffekter er viktig for utforming og analyser av politiske tiltak. Den norske sykelønnsordningen bygger på tillit, og kan i sin ytterste konsekvens bryte sammen dersom systemet misbrukes tilstrekkelig. Det er viktig å forstå betydningen av sosiale normer og holdninger til uttak av sykepenger. Sosial interaksjon kan skape sosiale multiplikatorer⁸. Dette betyr at en endring eller innføring av politiske tiltak ikke bare påvirker adferden til målsettingsgruppen, men også adferden til andre individ på grunn av avhengighet mellom individene gjennom sosial interaksjon. Sosial interaksjon har lenge blitt analysert fra et teoretisk perspektiv, men på grunn av metodiske utfordringer og dårlig tilgang på gode data er det utført få empiriske analyser. Likevel ser vi nå at den empiriske litteraturen vokser ettersom dette er et viktig fagområde.

Formålet med denne oppgaven er å undersøke om det eksisterer sosiale interaksjonseffekter for sykefravær, og i så fall hvilken grad. Vi benytter data fra et kundeservicesenter i et norsk forsikringsselskap der vi har detaljert informasjon om de ansattes sykefravær. Vi har ukentlige observasjoner som løper fra første kvartal i 2003 til andre kvartal i 2007. Det er 362 ansatte i kundeservicesenteret og de ansatte jobber i definerte team på ca. 16 medarbeidere. I tillegg til data om mange ulike typer sykefravær, har vi også bakgrunnsopplysninger om hvert enkelt individ. For vårt formål er likevel den viktigste egenskapen ved data at vi kjenner teamtilhørigheten til hvert individ, og kan derfor konstruere ulike mål på kollegenes – definert som teammedarbeidere – sykefravær. Vi observerer også når en ny medarbeider kommer inn i bedriften (og dermed teamet), og når en medarbeider skifter fra et team til et annet. Som vi skal gjøre detaljert rede for i kapittel 8, er dette potensielt svært viktige egenskaper ved data når det gjelder å isolere sosiale interaksjonseffekter.

Vi undersøker også om noen typer sykefravær er mer påvirket av sosiale interaksjonseffekter enn andre. En fordel med våre data i forhold til det som er blitt brukt i tidligere analyser, er at vi har informasjon om ulike typer sykefravær (legemeldt, egenmeldt, gradert, m.m.). Datasituasjonen i Norge og i noen grad de øvrige skandinaviske landene, er at det finnes til dels svært gode data på den typen sykefravær som blir betalt av det offentlige (i Norge gjennom NAV). Kortvarig sykefravær, spesielt det egenmeldte men også det legemeldte som finansieres

⁷ Se for eksempel Sacerdote (2001) om utdanning, Mas og Moretti (2009) om produktivitet og Rege et al. (2012) om sosialforsikring.

⁸ Se Glaeser et al. (2003) for mer om sosiale multiplikatorer.

av arbeidsgiver, er derimot mye dårligere representert. På grunn av datasituasjonen i Norge er det relativt lite forskning som benytter data om korttidssykefravær. Ettersom vi har tilgang til denne type data for bedriften vi studerer vil det være interessant å undersøke om det er noen forskjeller i sosiale interaksjonseffekter mellom fraværstyper som typisk er kortere (egenmeldt og sykefravær pga. sykt barn) og fraværstyper som typisk er lengre (legemeldt og gradert). Vi kan komme med argumenter for to motstridende hypoteser. Siden sosiale normer ikke direkte påvirker et individs helsetilstand, kan det tenkes at korttidssykefraværet er mer påvirket enn landtidssykefraværet. Det er trolig lettere å skrive en egenmelding enn å få en legeerklæring dersom man står i gråsonen til å være syk eller skadet, eller dersom man faktisk ikke er syk eller skadet. Noen ser på leger som en «portvakt» som beskytter sykemeldinger mot å misbrukes. På den andre siden er mesteparten av langtidssykefraværet knyttet til muskel- og skjelettlidelser og psykiske lidelser. Disse lidelsene er ikke lett å identifisere og mye av diagnostiseringen blir subjektiv. Det kan derfor tenkes at disse diagnosene brukes til å kamuflere forhold som ikke gir rett til sykepenger. Korttidssykefraværet antas ofte å være preget av luftveisinfeksjoner og omgangssyke, noe som enkelt kan identifiseres (Mastekaasa, 2015).

Videre lurer vi på om vi kan forklare de observerte forskjellene i sykefravær blant kvinner og menn, som i stor grad er uforklart, med sosiale interaksjonseffekter⁹. Ettersom kvinner observeres med høyere sykefravær enn menn, er en hypotese at kvinner er mer påvirket av slike effekter enn menn. Til slutt undersøker vi om det er slik at sosiale interaksjonseffekter endres over tid. En hypotese er at effektene blir sterkere over tid ettersom individ blir bedre kjent og får mer tid til å tilpasse seg hverandre.

Det er flere utfordringer knyttet til identifisering av sosiale interaksjonseffekter, blant annet problemer med simultanitet, korrelerte effekter og seleksjon. Vi formulerer to modeller som forsøker å isolere den kausale effekten av teammedlemmenes sykefravær på et individs sykefravær. Vi benytter en fasteffekt modell og instrumentvariabelmetoden i begge modellspesifikasjonene som identifiseringsstrategi. I den første modellen studerer vi sykefraværsadferd blant nyansatte (nyansatt-modellen). Vi undersøker om det er slik at nyansatte tenderer å tilpasse sitt sykefravær til det gjennomsnittlige sykefraværet i teamet som de tilhører. I nyansatt-modellen finner vi imidlertid ingen indikasjon på at det eksisterer sosiale

⁹ Se figur 1 og figur 2 i avsnitt 2.1.

interaksjonseffekter for sykefravær i kundeservicesenteret. I den andre modellen studerer vi sykefraværsadferd blant ansatte som flytter fra et team til et annet (flytte-modellen). Ved å sammenligne sykefravær i nytt team og gammelt team, kan vi se om disse individene endrer adferd som følge av flyttingen. Når vi studerer flytterne over en observasjonsperiode på fire kvartal finner vi en positiv effekt, dvs. sykefraværet øker når det gjennomsnittlige sykefraværet i teamet øker marginalt. I flytte-modellen finner vi også at menn reduserer sitt sykefravær når det gjennomsnittlige sykefraværet i teamet øker marginalt.

Oppbyggingen av oppgaven er som følger. Kapittel 2 tar for seg bakgrunn og institusjonelle forhold som angår sykefravær. Kapittel 3 presenterer det teoretiske rammeverket for oppgaven. Kapittel 4 gir oversikt over tidligere empiri. Kapittel 5 tar for seg metodiske utfordringer og forslag til løsninger knyttet til identifisering av effektene. Kapittel 6 gir informasjon om kundeservicesenteret vi skal studere, og kapittel 7 gir en oversikt over den deskriptive statistikken. Kapittel 8 presenterer de empiriske modellspesifikasjonene og kapittel 9 presenterer resultatene. I kapittel 10 belyses svakheter ved analysen, og i kapittel 11 er konklusjonen presentert.

2 Bakgrunn og institusjonelle forhold

I dette kapittelet vil vi ta for oss bakgrunn og institusjonelle forhold som angår sykefravær. Vi starter med å se på utviklingen i sykefraværet i Norge og dagens sykefraværssituasjon. Videre gis en presentasjon av sykelønnsordningen i Norge, og deretter gjennomføres en sammenligning av sykefravær og sykelønnsordninger i utvalgte nordeuropeiske land. Til slutt drøfter vi hvorvidt det aggregerte bildet ser ut til å være preget av sosiale normer. Hensikten med dette kapittelet er å få en bedre forståelse av sykefraværssituasjonen i Norge, et innblikk i den norske sykelønnsordningen og forståelse av hvordan Norge ligger an i en internasjonal sammenheng.

2.1 Utviklingen i sykefraværet i Norge og dagens situasjon

På 1990-tallet opplevde vi en langvarig økning i sykefraværet i Norge. Andelen uførepensjonister økte også dramatisk utover 1990-tallet. Denne økningen var en stor bekymring og er en av årsakene til dagens fokus på sykefraværet. Prognosene som ble utarbeidet på slutten av 1990-tallet viste at både sykefraværet og tilgangen til uførepensjon ville fortsette å øke dersom det ikke ble satt i gang nødvendige tiltak. Utviklingen ble vurdert til å skape velferdsmessige utfordringer i form av økte kostnader og tapt produksjon (Ose, 2010). I oktober 2001 ble det inngått en avtale mellom regjeringen og partene i arbeidslivet for å styrke arbeidslinjen i arbeidslivet; intensjonsavtalen om et inkluderende arbeidsliv (IA-avtalen)¹⁰. Det overordnede målet i IA-avtalen er å forebygge og redusere sykefravær, styrke jobbnærværet og bedre arbeidsmiljøet, samt hindre utstøting og frafall fra arbeidslivet.

IA-avtalen har tre delmål. Det første er å redusere sykefraværet med 20 prosent i forhold til nivået i andre kvartal i 2001. Dette innebærer at samlet sykefravær ikke skal overstige 5,6 prosent på nasjonalt nivå¹¹. Det andre er å hindre frafall og øke sysselsettingen blant personer med nedsatt funksjonsevne. Det tredje er å forlenge yrkesaktivitet etter fylte 50 år med tolv måneder. Med dette menes en økning sammenlignet med 2009 i gjennomsnittlig periode med yrkesaktivitet for personer over 50 år. I IA-virksomheter forplikter både arbeidsgiver og arbeidstaker seg til et systematisk samarbeid for å oppnå mer inkluderende arbeidsplasser. IA-

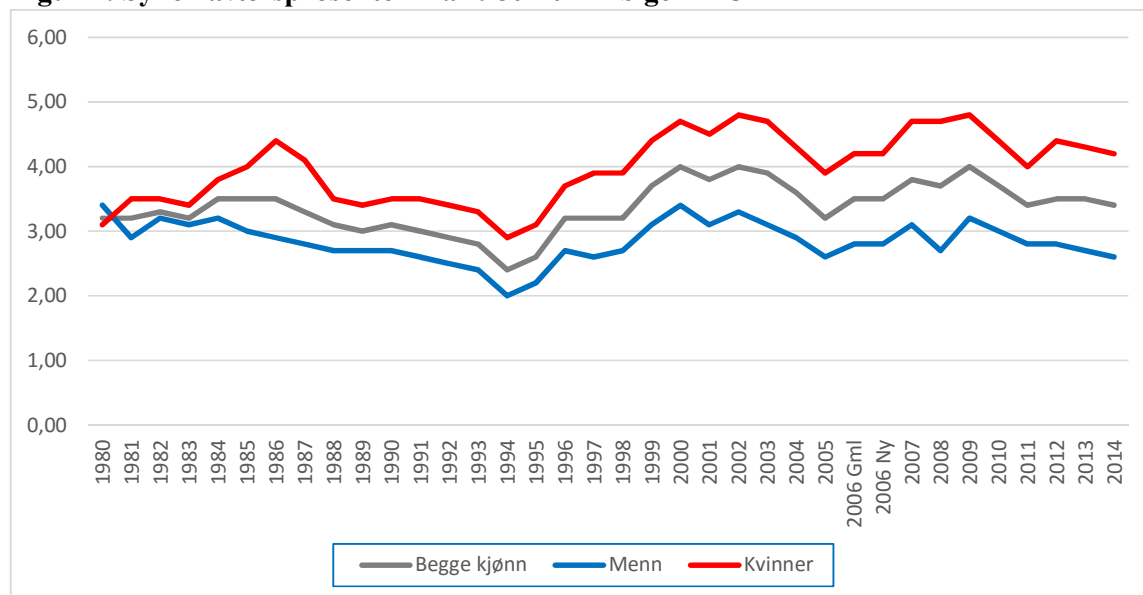
¹⁰ Nåværende avtale ble signert i 2014.

¹¹ Den prosentsatsen som ligger til grunn for IA-avtalen, er definert som andel syke i prosent av avtalte dagsverk. Dette forutsetter at vi kjenner sysselsettingsprosenten til de som inngår i statistikken. Dessverre ble denne statistikken lagt om fra og med 1. januar 2015. Fremdeles jobbes det med å finne en arbeidstidsvariabel som kan fungere i forhold til den tradisjonelle sykefraværstatistikken. Per 2. kvartal 2016 er det ikke funnet noen løsning, slik at verken NAV eller SSB for tiden publiserer akkurat denne sykefraværstatistikken.

virksomheter er virksomheter som har inngått en avtale med NAV, og disse virksomhetene får ekstra ressurser fra NAV for å arbeide mot å nå de tre delmålene. Nesten alle virksomheter i offentlig sektor i Norge er IA-virksomheter, mens tilslutningen i privat sektor er noe lavere (Arbeids- og sosialdepartementet, 2014).

IA-avtalen omfatter flere tiltak med fokus på å redusere sykefraværet, blant annet økt vektlegging av gradert sykemelding, «endring i sykmelders rolle» (første gang implementert våren 2004), tilretteleggingstilskuddet, «Raskere tilbake» (implementert i 2007) og endret oppfølgingsregime¹². Merk at både «endring i sykmelders rolle» og «Raskere tilbake» er implementert uavhengig om man jobber i IA-virksomhet eller ikke. Økt bruk av graderte sykemeldinger har blitt grundig evaluert og har vist seg å være et effektivt tiltak. I perioder hvor graderingsandelen går opp, har totalt sykefravær falt. Ellers foreligger det generelt lite effektforskning av disse tiltakene og dermed lite kunnskap om hva som faktisk virker. Begrenset måloppnåelse med tiltakene under IA-avtalen skaper mye usikkerhet omkring hva som fungerer, hva som potensielt er skadelig og hva som er helt uvirksomt. Målsettingen om en nedgang i sykefravær på 20 prosent fra andre kvartal i 2001 er ennå ikke nådd.

Figur 1: Sykefraværprosenten fra 1980-2014 ifølge AKU

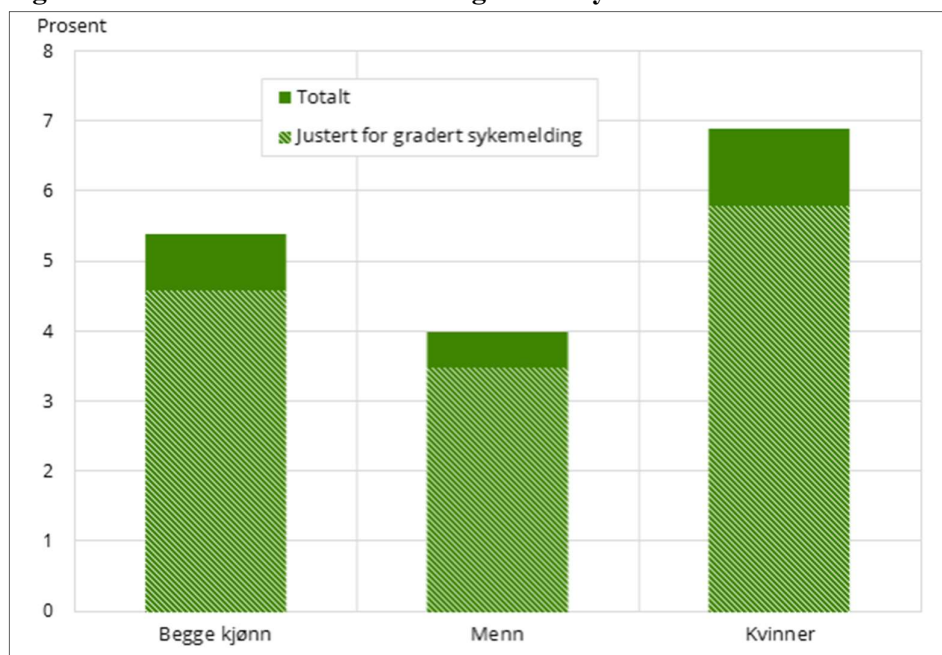


Kilde: SSB (2016)

¹² Se Ose et al. (2009), Markussen et al. (2012) og Mykletun og Brinchman (2013) for mer om tiltak under IA-avtalen og evaluering av disse.

Figur 1 viser langsiktig utvikling i sykefraværprosenten¹³ fra 1980 til 2014 ifølge Arbeidskraftundersøkelsen (AKU)¹⁴. Fra figur 1 ser vi at det var en langsiktig økning i sykefraværprosenten på 1990-tallet. Vi ser også at sykefraværet falt noe i perioden 2003-2005, for å deretter stige frem mot 2010. Siden den tid har det vært ganske stabilt. Nedgangen i sykefraværet kan skyldes «endring i sykmelders rolle». Sykemelder sitt fokus skulle dreies fra sykemelding til vurdering av arbeidstakernes «restarbeidsevne», og flere skulle oppfordres til å kombinere arbeid og sykemelding gjennom gradert sykemelding. Økt bruk av graderte sykemeldinger har som sagt vært et effektivt tiltak for å redusere sykefraværet, men effekten ser ut til å være noe forbigående. Videre observerer vi fra figur 1 at menn og kvinners sykefravær var temmelig likt i begynnelsen av denne observasjonsperioden, men at kvinners sykefravær økte betraktelig i forhold til menns utover 1980-tallet. Fra da av har det såkalte kjønnsgapet i sykefravær vært konstant eller svakt økende.

Figur 2: Andel arbeidstakere med legemeldt sykefravær 3. kvartal 2015



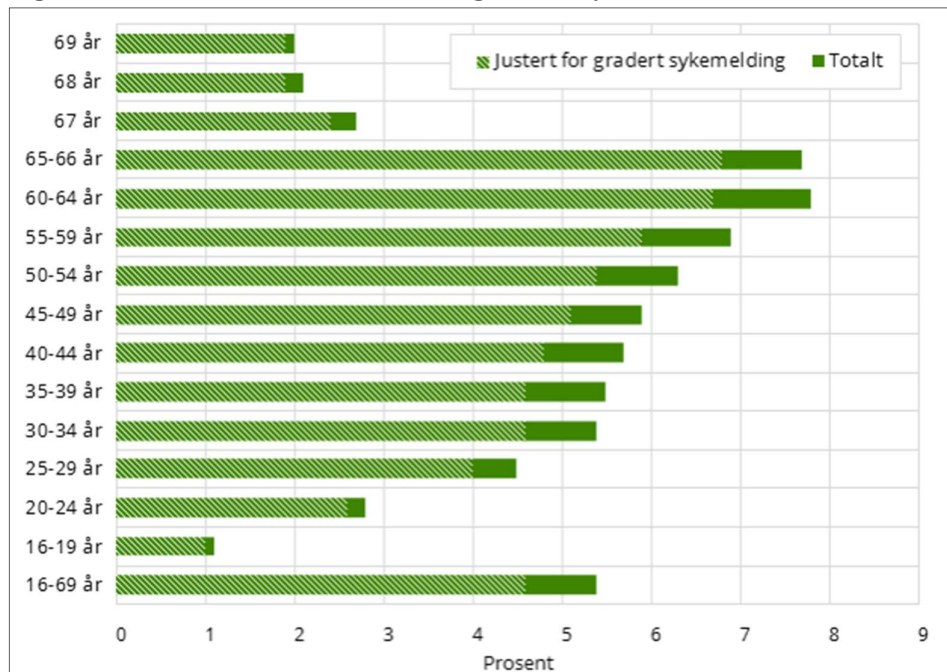
Kilde: SSB (2015)

¹³ Sykefraværprosenten er beregnet som antall sysselsatte 15-74 år som var fraværende fra arbeid hele referanseuka pga. egen sykdom i prosent av antall sysselsatte i alt.

¹⁴ AKU er en spørreundersøkelse der man i første rekke er ute etter å registrere sysselsetting og arbeidsledighet, men der også sykefravær inngår. Nivå tallene fra AKU skiller seg noe fra sykefraværstatistikken ettersom AKU ikke dekker sykefravær mindre enn en uke. Merk at i januar 2006 skjedde en omlegging av AKU, og for å bedre kunne sammenligne med tallene for 2006 med året før så er 2006 også lagt inn med tall basert på det gamle opplegget. Fra 2006 ble aldersgrensen for å bli med i AKU senket fra 16 til 15 år. Samtidig ble aldersdefinisjonen endret fra alder ved utgangen av året til alder på referansetidspunktene for undersøkelsen.

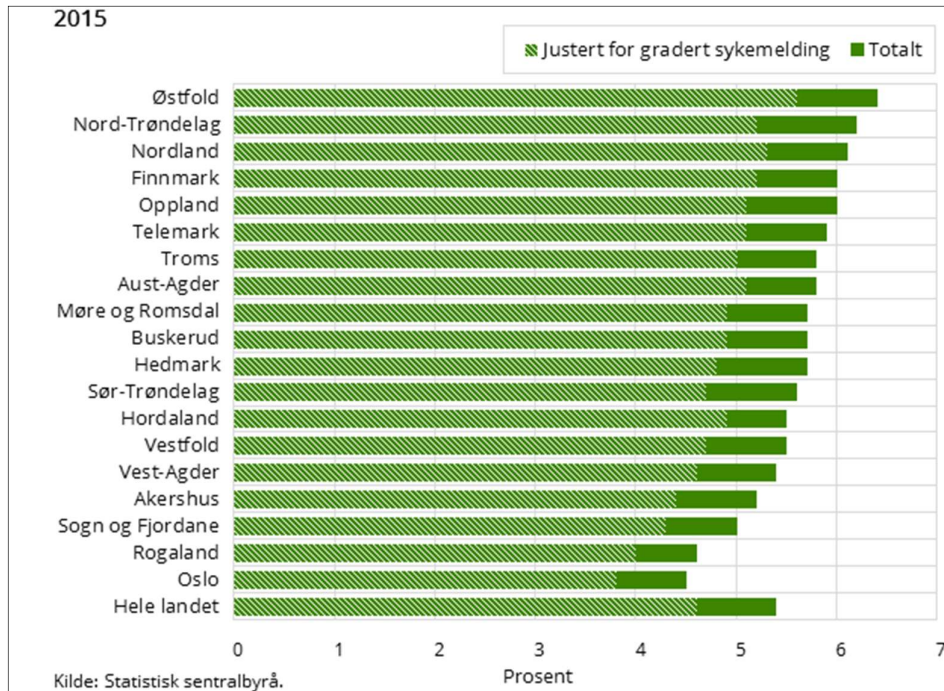
Figur 2 viser andel arbeidstakere med legemeldt sykefravær tredje kvartal i 2015. Det legemeldte sykefraværet utgjør den største andelen av samlet sykefravær, men egenmeldt sykefravær utgjør også en betydelig størrelse. Tredje kvartal i 2015 var andelen arbeidstakere med legemeldt sykefravær 5,4 prosent, der andelen kvinner var 6,9 prosent og andelen menn var 4 prosent. Det er mange arbeidstakere som har gradert sykefravær, dvs. ikke en 100 prosent sykemelding. Når vi tar hensyn til det graderte sykefraværet går andelen arbeidstakere med legemeldt sykefravær ned til 4,6 prosent. For kvinner går andelen ned til 5,8 prosent, mens for menn går andelen ned til 3,5 prosent. Menns legemeldte sykefravær utgjør omtrent 60 prosent av kvinners, både med og uten justering av sykemeldingsgrad (SSB, 2015).

Figur 3: Andel arbeidstakere med legemeldt sykefravær etter alder 3. kvartal 2015



Kilde: SSB (2015)

Figur 3 viser andel arbeidstakere med legemeldt sykefravær etter alder tredje kvartal i 2015. Vi observerer at eldre arbeidstakere har høyere legemeldt sykefravær enn yngre arbeidstakere, bortsett fra de aller eldste arbeidstakerne. Årsaken er at det er de friskeste eldre som fortsetter å arbeide. Det er størst andel fraværende blant arbeidstakere i aldersgruppene 60–64 år og 65–66 år, med henholdsvis 7,8 prosent og 7,7 prosent. Dersom vi justerer for graderte sykemeldinger reduseres andelen til 6,7 prosent og 6,8 prosent, men det er fortsatt disse to aldersgruppene som har størst andel legemeldt sykefravær (SSB, 2015).

Figur 4: Andel arbeidstakere med legemeldt sykefravær etter bosted 3. kvartal 2015

Kilde: SSB (2015)

Figur 4 viser andel arbeidstakere med legemeldt sykefravær etter bosted tredje kvartal i 2015. Vi ser at Østfold var det fylket med størst andel fraværende med 6,4 prosent. Det fylket med lavest andel fraværende var Oslo med 4,5 prosent. Dersom vi justerer for graderte sykemeldinger reduseres andelen fraværende arbeidstakere i alle fylker, men effekten er størst for Nord- og Sør-Trøndelag. Hordaland er fylket med minst nedgang når vi justerer for graderte sykemeldinger (SSB, 2015).

Det finnes lite statistikk over diagnoser for egenmeldt sykefravær siden arbeidstaker ikke er pliktig til å oppgi hvilken sykdom som har forårsaket sykefraværet. Det er likevel trolig at mye av dette sykefraværet skyldes sykdommer som forkjølelse, influensa og omgangssyke. For det legemeldte sykefraværet finnes bedre data, og det er diagnoser som muskel- og skjelettlidelser som utgjør den største andelen. Omtrent 40 prosent av sykepengedagene skyldes slike lidelser. Ellers er lettere psykiske lidelser en stor og økende andel av det legemeldte sykefraværet. Ose (2010) konkluderer imidlertid med at helsemessige forhold i liten grad kan forklare utviklingen i sykefraværet i Norge.

Både eldre og nyere sykefraværstatistikk viser at kvinner har høyere sykefravær enn menn. Kvinner er mer plaget av både muskel- og skjelettlidelser og psykiske lidelser enn menn

(Mastekaasa, 2015). Årsaker til kjønnsforskjeller i sykefravær er derimot i stor grad uforklart. Smeby et al. (2009) undersøker om kjønnsforskjeller i sykefravær kan forklares av yrke, arbeidsbetingelser, inntekt, helse og psykisk helse. Forfatterne finner ikke at noen av disse faktorene kan forklare de observerte kjønnsforskjellene i sykefraværet. Markussen et al. (2009) finner heller ikke noen forklarende faktorer for kjønnsforskjeller i sykefravær. Forfatterne gjennomfører en omfattende analyse av registerdata der de finner at kjønnsforskjellene i legemeldt sykefravær består når de kontrollerer for barn, graviditet, utdanning, yrke og inntekt.

2.2 Sykelønnsordningen i Norge: Regler for sykepenger

I dette avsnittet presenteres utformingen av den norske sykelønnsordningen. Vi vil ta for oss noen av de generelle bestemmelsene knyttet til sykepenger fastsatt i folketrygdloven (ftrl) (1997, §§ 8-1 – 8-55). Vi vil blant annet se på formålet med ordningen, hvem som har rett til å motta sykepenger, finansiering av sykepenger, lengde på sykefraværet, kompensasjonsgrad og ordninger for ulike typer sykefravær.

Formål og opptjeningstid

Formålet og regler rundt opptjeningstid og arbeidsuførhet presenteres i ftrl. (1997, §§ 8-1, 8-2 og 8-4). Formålet med sykepenger er å gi kompensasjon for bortfall av arbeidsinntekt for arbeidstakere som er arbeidsuføre på grunn av sykdom eller skade. For å få rett til sykepenger må arbeidstaker ha vært i arbeid minst fire uker før han/hun ble arbeidsufør (opptjeningstid). Sykepenger gis til den som er arbeidsufør på grunn av en funksjonsnedsettelse som opplagt skyldes sykdom eller skade. Arbeidsuførhet som skyldes for eksempel sosiale eller økonomiske problemer gir ikke rett til sykepenger.

Finansiering

Finansiering knyttet til sykepenger presenteres i ftrl. (1997, §§ 8-19 og 8-20). Arbeidsgiveren skal betale sykepenger opptil 16 kalenderdager. Disse 16 dagene omtales som arbeidsgiverperioden, og sykefravær utover disse 16 dagene betales av folketrygden som er finansiert over statsbudsjettet. Arbeidstaker eller arbeidsgiver kan søke om å få dekket utgifter til sykepenger i arbeidsgiverperioden dersom arbeidstaker har en langvarig eller kronisk sykdom som fører til at han/hun har spesielt høy risiko for sykefravær. Arbeidstaker eller arbeidsgiver kan også søke om å få dekket utgifter til sykepenger i arbeidsgiverperioden for en arbeidstaker som er sykemeldt grunnet svangerskapsrelatert sykdom, i tilfeller når det ikke er

mulig å finne alternative arbeidsoppgaver til vedkommende som gjør at hun kan fortsette å arbeide.

Dokumentasjon

Regler rundt dokumentasjon og bruk av egenmelding presenteres i ftrl. (1997, §§ 8-7, 8-23 og 8-24). For at en arbeidstaker skal ha rett til sykepenger må arbeidsuførheten dokumenteres med en legeerklæring, bortsett fra når arbeidstaker har rett til å benytte egenmelding. Ved en egenmelding melder arbeidstaker selv fra til arbeidsgiver om arbeidsuførheten uten å fremlegge legeerklæring. For å ha rett til å benytte en egenmelding må arbeidstaker ha arbeidet hos nåværende arbeidsgiver i minst to måneder. Vilklårene for bruk av egenmelding varierer mellom arbeidsgivere. I virksomheter som ikke har IA-avtale kan arbeidsgiver kreve sykemelding ved sykefravær utover 3 dager, mens i virksomheter som har IA-avtale er tilsvarende grense 8 dager. Ved sykefravær utover 16 dager skal det leveres legeerklæring uansett om virksomhet har IA-avtale eller ikke. I virksomheter uten IA-avtale kan ansatte ha fire perioder med egenmelding i løpet av ett år, mens i virksomheter med IA-avtale kan ansatte ha 24 dager i løpet av ett år (Proba samfunnsanalyse, 2014a).

Maksimal lengde på fraværet og beregning av sykepenger

Regler rundt lengde på sykefraværet, kompensasjonsgrad og beregning av sykepenger presenteres i ftrl. (1997, §§ 8-10, 8-12, 8-16, 8-28 og 8-30). I Norge er det ingen karensdager, dvs. ingen ventetid for å få utbetalt sykepenger. Videre er det ingen regler for maksimal lengde på sykefravær, men det betales sykepenger i maksimalt 12 måneder. Kompensasjonsgrad er 100 prosent av sykepengegrunnlaget, der sykepengegrunnlaget er den inntekten som sykepengene skal regnes ut etter. Sykepengegrunnlaget kan ikke overstige seks ganger grunnbeløpet (6G)¹⁵, men mange arbeidstakere – blant annet hele offentlig sektor - som tjener mer enn 6G får ekstra kompensasjon av sin arbeidsgiver. Ved beregning av sykepengegrunnlaget er det vanlig å basere det på arbeidstakers inntekt de siste 4 ukene. Dersom arbeidstakers inntekt de siste 4 ukene avviker mer enn 25 prosent fra deres inntekt det siste året bør andre mål vurderes¹⁶.

¹⁵ Grunnbeløpet tilsvarer 90 068 norske kroner, og beløpet justeres hvert år 1. mai (NAV, 2015).

¹⁶ Se ftrl. (1997, §§ 8-28 og 8-30) for mer om beregning av sykepengegrunnlaget.

Gradert sykemelding

I henhold til ftrl. (1997, § 8-13) kan det ytes graderte sykepenger dersom arbeidstaker er delvis arbeidsufør. Gradert sykemelding er en ordning som innebærer at arbeidstaker får redusert arbeidstid eller belastning og en kombinasjon av noe lønn og noe sykepenger. Sykepengeutbetalingene skal gjenspeile graden av funksjonsnedsettelse innenfor intervallet 20-100 prosent. Retningslinjer for leger og andre med sykemeldingsrett sier at gradert sykemelding skal vurderes før full sykemelding.

Aktiv sykemelding

Aktiv sykemelding er en ordning som tillater at arbeidstaker prøver å jobbe selv om han/hun er arbeidsufør. Arbeidsgiver skal finne andre arbeidsoppgaver til arbeidstaker enn de han/hun vanligvis utfører. En aktiv sykemelding kan inngås i en periode på 4 uker og den kan utvides til 8 uker ved behov. Ordningen ble derimot fjernet juli 2011 fordi den virket mot sin hensikt. Målet med ordningen var å få arbeidstakere til å komme raskt tilbake i jobb ved å opprettholde kontakt med arbeidsplassen. Mange endte derimot opp med å utføre hele eller deler av sitt vanlige arbeid, noe som gjorde at de gikk sykemeldt lengre (Høyskolen i Lillehammer, 2010).

Omsorgspenger

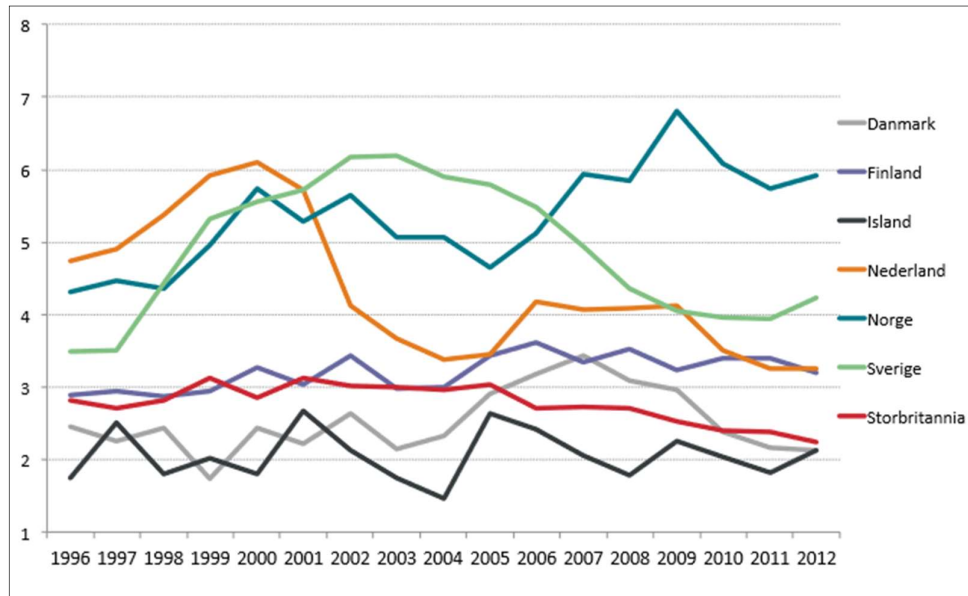
I henhold til ftrl. (1997, § 9-5) kan en arbeidstaker som er hjemme med et sykt barn motta omsorgspenger som erstatning for tapt arbeidsinntekt. Arbeidstaker kan også motta omsorgspenger dersom barnetilsyn er syk. Retten til omsorgspenger gjelder også fosterforeldre. Retten til omsorgspenger gjelder til og med det kalenderåret barnet fyller 12 år. Dersom barnet er kronisk sykt eller funksjonshemmet, gjelder retten til og med det året barnet fyller 18 år. Antall dager med omsorgspenger per år avhenger av hvor mange barn man har, og om man er alene om omsorg eller ikke. Dersom du har omsorg for ett eller to barn får du 10 dager per år. Dersom du har omsorg for mer enn to barn får du 15 dager per år. Dersom du er alene om omsorgen for ett eller to barn får du 20 dager per år. Dersom du er alene om omsorgen for mer enn to barn får du 30 dager per år.

2.3 Internasjonal sammenligning av sykefravær og sykelønnsordninger

I dette avsnittet vil vi sammenligne sykefraværet i Norge med sykefraværet i noen andre nordeuropeisk land; Sverige, Danmark, Island, Storbritannia, Finland og Nederland. Deretter vil vi foreta en kortfattet sammenligning av sykelønnsordningene i disse landene. Gjennom å

foreta disse sammenligningene ønsker vi å få en bedre forståelse av sykefraværssituasjonen i Norge.

Figur 5: Utvikling i sykefraværprosenten i utvalgte land, 1996-2012



Kilde: Proba samfunnsanalyse (2014b, s. 27)

Figur 5 viser utviklingen i sykefraværprosenten i utvalgte land fra 1996 til 2012, der tallene er hentet fra AKU¹⁷. Internasjonale sammenligninger av sykefravær tar ofte utgangspunkt i arbeidskraftundersøkelsene (AKU), ettersom det er vanskelig å sammenligne sykefravær mellom land når det er ulikheter i hvordan sykefravær registreres og defineres. I Norge har gjennomsnittlig sykefraværprosent steget fra omtrent 4 prosent i 1996 til 6 prosent i 2012. Sykefraværet var høyest i 2009 med 6,6 prosent, og lavest i 1996 med 4,3 prosent. Frem til 2006 hadde både Sverige og Nederland i perioder høyere sykefravær enn Norge, men etter 2006 har Norge hatt betydelig høyere sykefravær enn de andre landene. Det er vanskelig å tenke seg naturlige forklaringer til dette, som for eksempel dårligere folkehelse eller mer sykdomsbringende arbeidsliv. Den sjenerøse sykelønnsordningen i Norge peker seg derfor ut som en sentral forklaring.

¹⁷ Før 2006 er tall hentet fra 2. kvartal. Fra og med 2006 er tall fra alle fire kvartal inkludert i alle land, bortsett fra Finland og Storbritannia. I Finland er det bare data for 2. kvartal også i 2006, og i Storbritannia er det data fra alle kvartal først fra og med 2008.

Tabell 1: Oversikt over sykelønnsordninger i utvalgte land

Land	Karensdager	Maksimal lengde	Kompensasjonsgrad	Egenmeldingsperiode
Norge	0 dager	52 uker	100 %	3 eller 8 dager
Sverige	1 dag	52 uker	80 %	7 dager
Danmark	0 dager	52 uker	Maksimalt 4005 DKK per uke	30 dager
Island	0 dager hvis ansatt mer enn 2 mnd. 21 dager hvis ansatt mindre enn 2 mnd.	26 uker	80 %	Ingen formelle bestemmelser, men det krevers legemelding for å få sykepenger fra staten
Storbritannia	3 dager	28 uker	86,7 GBP per uke	8 dager
Finland	0 dager	52 uker	70 %	0 dager
Nederland	0 dager	104 uker	70 %	6 uker

Kilde: Egen tabell med informasjon fra Proba samfunnsanalyse (2014a)

Tabell 1 gir en oversikt over karensdager, maksimal lengde, kompensasjonsgrad og egenmeldingsperiode i Norge, Sverige, Danmark, Island, Storbritannia, Finland og Nederland¹⁸. Når vi sammenligner sykelønnsordninger på disse områdene ser vi at det norske systemet skiller seg ut ved å være veldig sjenerøst. I Norge får arbeidstakere sykepenger fra første fraværsdag med 100 prosent lønnskompensasjon (inntil 6G). Et par av de andre landene har karensdager og alle har en lavere kompensasjonsgrad. Vi legger derimot merke til at maksimal lengde på sykefraværet og egenmeldingsperioden er betydelig lengre i Nederland enn Norge. Videre ser vi at Storbritannia og Island har de minst sjenerøse sykelønnsordningene, og vi observerer også fra figur 5 at det er disse to landene som har lavest sykefravær. Tilsvarende har Norge, Sverige og Nederland de mest sjenerøse sykelønnsordningene, og er de landene med høyest sykefravær. Vi kan ikke konkludere med at det er en kausalitet, men en indikasjon på at det er en positivt korrelasjon mellom sykefravær og sjenerøsiteten i sykelønnsordningen.

I det norske systemet er det få eller ingen økonomiske insentiv til å arbeide ved sykdom eller skade, og utfordringen med moralsk hasard kan derfor tenkes å være større i Norge enn i andre land (særlig for marginale arbeidstakere/gråsoner til å være syk eller skadet). Det er viktig å presisere at sykefravær også kan skyldes skoft. Det er trolig mindre skoft i land med mindre sjenerøse sykelønnsordninger siden det er mindre attraktivt der. Den sjenerøse sykelønnsordningen i Norge bidrar trolig til å holde sykefraværet høyt i Norge sammenlignet med andre land. «Dårlige normer», svak kontroll og svake økonomiske insentiv til å arbeide ved sykdom eller skade kan potensielt være en dyr kombinasjon.

¹⁸ Se Proba samfunnsanalyse (2014a) for en grundigere gjennomgang av de ulike landenes sykelønnsordninger.

2.4 Er det aggregerte bildet preget av sosiale normer?

Vi har nå sett at Norge har høyt sykefravær i en internasjonal sammenheng, en svært sjenerøs sykelønnsordning, men også at det har vært en del variasjon i sykefraværet over tid. At sykefraværet er høyere i Norge enn andre nordeuropeiske land kan tenkes å være et resultat av et sjenerøst system, men også sosiale normer. Det norske systemet bygger på tillit, og vil kun fungere så lenge arbeidstakere har «gode normer». Den sosiale normen må fremme holdningen om at kun dem som har rett til sykepenger får sykemelde seg. Den langvarige, og i stor grad uforklarlige med tanke på helsemessige forhold, økningen i sykefraværet i Norge kan tyde på at sosiale normer knyttet til uttak av sykepenger har endret seg. Det kan se ut som at de «gode normene» forvitrer. Det kan tenkes å være både på grunn av at terskelen for hva som oppfattes som «syk nok» er blitt lavere samt høyere aksept for regelrett skoft. Med andre ord, det kan se ut som at stigma knyttet til «uførtjent» uttak av sykepenger er blitt mindre. Dette taler for at sosiale normer er av betydning for sykefravær i Norge. Likevel ser vi, både fra figur 1 og 5, at det har vært en del svingninger i sykefraværet i Norge over tid. Disse svingningene kan ikke trolig ikke forklares av sosiale normer, og noen vil derfor hevde at sosiale normer ikke er av betydning for sykefravær i Norge.

Vi har også sett at kvinner har høyere sykefravær enn menn. Fra figur 1 ser vi at kjønnsgapet i sykefravær har vært svakt økende siden 1980-tallet. Et argument for at sosiale normer ikke er av betydning for sykefravær er nettopp på grunn av denne forskjellen i sykefravær blant kvinner og menn. Ettersom kvinner og menn står ovenfor samme arbeidsmarked og samme sykelønnsordning burde de bli påvirket likt av sosiale normer knyttet til sykefraværsadferd. På den andre siden er den observerte forskjellen i sykefravær blant kvinner og menn i stor grad uforklart, og det kan derfor tenkes at sosiale normer påvirker kvinner og menn ulikt. Det kan altså se ut som at kvinner er mer påvirket av sosiale normer enn menn, i den grad normer drar i retning mot høyere sykefravær.

Et annet argument for at kvinner er mer påvirket av sosiale normer enn menn, er at kvinner i gjennomsnitt har svakere tilknytning til arbeidslivet enn menn. Arbeidskraftundersøkelsen for 2011 viste at 41 prosent av sysselsatte kvinner jobbet deltid, mens bare 13 prosent av sysselsatte menn jobbet deltid (SSB, 2013). I de fleste familier er det også kvinnene som har hovedansvaret for barna. Selv om husholdsarbeidet øker blant menn er det fortsatt kvinner som bruker mest tid til dette ifølge tidsbruksundersøkelsen fra 2010 (SSB, 2012). Begge deler kan bidra til lavere terskel for å være borte fra jobb, både ved egen og ved barns sykdom.

3 Teoretisk rammeverk

I dette kapitlet presenteres det teoretiske rammeverket for oppgaven. Målet med oppgaven er å undersøke om det eksisterer sosiale interaksjonseffekter for sykefravær, og vi trenger derfor en teoretisk modell som forklarer mekanismene som fører til at individ tilpasser sin adferd etter andre individ. Klassisk økonomisk teori fokuserer stort sett på individuell adferd knyttet til økonomiske insentiv. Analyser av sosial interaksjon har tidligere vært overlatt til sosiologer, men senere år har økonomilitteraturen innenfor dette området vokst og vist at sosiale normer spiller en viktig rolle i nyttemaksimering.

Sosiale interaksjonsmodeller kommer fra ideen om at preferansene til et individ over ulike alternativ avhenger av handlingene tatt av andre individ som individet relaterer til (referansegruppe)¹⁹. Hensikten med disse modellene er å gi en forklaring for gruppeadferd som oppstår ved avhengighet mellom individ. Denne avhengigheten mellom individ kan oppstå blant annet gjennom sosiale normer. Coleman (1990) definerer en sosial norm som «en regel for adferd som håndheves gjennom sosiale sanksjoner, som kan være i form av stigma». Sosiale interaksjonseffekter oppstår dersom stigma assosiert ved å ikke følge normen er sterkere når referansegruppen følger normen. Individ oppnår dermed høyere nytte ved å følge den sosiale normen/tilpasse sin adferd etter referansegruppen.

I dette kapitlet presenteres to teoretiske modeller. Den første modellen forsøker å forklare individuell tilpasning ved tilstedeværelse av sosiale normer, og hvordan sosiale interaksjonseffekter kan beregnes. Den andre modellen viser hvordan lokal variasjon i sykefravær kan oppstå og hvorfor vi kan få multiple likevekter.

3.1 Optimalt bruk av sykepenger ved eksistens av sosial interaksjon

Markussen og Røed (2015) presenter en teoretisk modell som beskriver hvordan sosial interaksjon påvirker økonomisk adferd. Denne modellen egner seg til å analysere hvordan sosial interaksjon (sosiale normer) påvirker sykefraværsadferd. Vi presenterer modellen slik den er fremstilt i Markussen og Røed (2015).

¹⁹ Se Brock og Durlauf (2001) og Cont og Löwe (2010) for generelle modeller.

La a_i indikere individ i 's bruk av sykepengar, og anta at nyttefunksjonen assosiert med denne handlingen kan dekomponeres til en privat og en sosial komponent. La a_i^0 være optimalt valg ved fravær av sosial interaksjon og la $j \in J$ være et sett av agenter som i relaterer til. Med kvadratisk nytte, kan vi skrive:

$$U_i(a_i; \{a_j, j \neq i\}) = -\pi(a_i^0 - a_i)^2 - \sum_{j \neq i} \gamma_{ij}(a_i - a_j)^2 \quad (3.1)$$

Den relevante delen av nyttefunksjonen er der hvor marginalnyttan er økende, dvs. $U'_i(a_i) > 0$. For å finne optimalt bruk av sykepengar for individ i må vi først finne førsteordensvilkårene til alle J agenter og sette dem lik null. Deretter kan vi løse for a_i , og finner da at optimalt bruk av sykepengar er gitt ved:

$$a_i^* = \frac{1}{\pi + \sum_{j \neq i} \gamma_{ij}} \left(\pi a_i^0 + \sum_{j \neq i} \gamma_{ij} a_j \right) \quad (3.2)$$

Her er π marginal disnytte fra å avvike fra det private optimum, γ_{ij} måler marginal gevinst i individ i 's nytte av å tilpasse seg handlingen til j . Individ i tilpasser seg den faktiske handlingen til j , ikke normer/holdninger som motiverer j 's adferd. Dette innebærer at γ_{ij} representerer det Manski (1993) kaller endogen interaksjon²⁰. Endogen interaksjon kan skape sosiale multiplikatoreffekter. Endogen interaksjon innebærer at optimale valg blir bestemt i et stort simultant ligningssystem med like mange ligninger som individ. Merk at i denne fremstillingen, med en kvadratisk nyttefunksjon, får vi kun en likevekt. Med mer generelle nyttefunksjoner kan man få multiple likevekter.

Ved å parametrisere γ_{ij} på ulike måter, får man ulike modeller. For eksempel, ved å velge $\gamma_{ij} = \gamma/N$, der N er størrelsen på populasjonen ekskludert i , får man den globale interaksjonsmodellen. I den globale interaksjonsmodellen er hver agents preferanser påvirket av den gjennomsnittlige handlingen til alle andre. Ved å velge $\gamma_{ij} = \gamma(d_{ij})$, der d_{ij} er et mål på relasjonsavstand mellom individ i og j , får man den lokale interaksjonsmodellen. Den lokale interaksjonsmodellen antar at sosiale innflytelser oppstår i begrensede grupper.

²⁰ Mer om dette i avsnitt 5.1.

I denne oppgaven vil vi fokusere på lokal interaksjon. Vi ønsker å se hvordan gruppemedlemmer (teammедlemmer) påvirker hverandre. Vi vil studere endogene effekter på gruppenivå, og gruppegjennomsnitt blir hovedsakelig brukt som forklaringsvariabler. Den direkte innflytelsen av et individ på et annet er modellert homogent innen definerte grupper (team) og inverst relatert til gruppestørrelse. Vi definerer $\gamma_{ij} = \gamma_g/N_g$, der g indikerer gruppe og N_g er antall medlemmer i en gruppe ekskludert individ i . En viktig antakelse er at gjennomsnittlig relasjonsavstand øker med gruppestørrelse, ceteris paribus, slik at jo større gruppe, jo mindre påvirker individene hverandre. Omformulerer uttrykk (3.2) slik at vi får en lokal interaksjonsmodell:

$$a_i^* = \frac{1}{\pi + \sum_g \gamma_g} \left(\pi a_i^0 + \sum_g \gamma_g \bar{a}_{g,-i} \right) \quad (3.3)$$

Her er γ_g nytten av å tilpasse seg til den gjennomsnittlige adferden i gruppen g ($\bar{a}_{g,-i}$). Vi antar typisk at $\gamma_g \geq 0$, men det kan også være slik at $\gamma_g < 0$. Negative sosiale interaksjonseffekter kan oppstå når individ oppnår nytte av å opptre annerledes. For eksempel, et individ ønsker kanskje å signalisere avstand til grupper den ikke ønsker å bli assosiert med.

3.2 Lokal variasjon i sykefravær ved eksistens av sosial interaksjon

Lindbeck et al. (2016) presenterer en forenklet versjon av forsikringsmodellen som blir utledet i Lindbeck og Persson (2010) med ett representativt individ. Denne modellen egner seg til å studere hvordan sosiale normer påvirker individuell adferd knyttet til inntektssikringsordninger. Ved å bruke lokal variasjon i sykefravær for å studere betydningen av sosiale normer må man forstå hvordan slik variasjon kan oppstå. Først og fremst kan det oppstå variasjon mellom ulike grupper dersom individ er forskjellige og plasseres tilfeldig i de ulike gruppene, så lenge det er relativt få individ i hver gruppe. Dette er rimelig opplagt, men med sosiale normer kan det oppstå variasjon mellom de ulike gruppene selv om alle individ er ex ante homogene. Vi vil nå vise hvordan slik variasjon kan oppstå. Anta at en persons nytte av arbeid kan uttrykkes:

$$u^w = u(c^w) + \theta \quad (3.4)$$

der θ er en kontinuerlig fordelt tilfeldig variabel som representerer nytten (når $\theta > 0$) eller disnyten (når $\theta < 0$) av arbeid. Parameteren θ kan avhenge av flere ting, men for enkelthetskyld

ser vi på den som en helsevariabel som påvirker smerte eller glede ved arbeid, men ikke ved fritid. I virkeligheten, der et individs helsetilstand i stor grad er uobserverbar for forsikringsselskapet (arbeidsgiver/staten), er individets beslutning om å bli hjemme fra jobb påvirket av flere faktorer som ikke direkte er relatert til helse (arbeidsmiljø, sjefens humør, osv.). Vi bruker c^w til å betegne konsum tilgjengelig til individet dersom han/hun jobber (nettolønn). Vi bruker c^A til å betegne konsummulighetene dersom individet er borte fra jobb²¹. Nytte ved fravær fra jobb kan da uttrykkes:

$$u^A = u(c^A) - \phi \quad (3.5)$$

der ϕ er sosial stigma knyttet til å motta sykepenger/leve av velferdsytelser. La oss først behandle ϕ som en eksogen konstant. Individet er indifferent mellom arbeid og å bli hjemme dersom θ tar en verdi, $\hat{\theta}$, slik at $u(c^w) + \hat{\theta} = u(c^A) - \phi$. Individet blir hjemme dersom $u^A > u^w$, og realisering av den tilfeldige helsevariabelen er slik at:

$$\theta < \hat{\theta} \equiv u(c^A) - u(c^w) - \phi \quad (3.6)$$

La $F(\theta)$ være den kumulative sannsynlighetsfordelingen av den tilfeldige helsevariabelen θ . Siden c^A og c^w er gitt ved forsikringssystemet ser vi fra (3.6) at totalt sykefravær i samfunnet, π , er gitt ved:

$$\pi \equiv F(u(c^A) - u(c^w) - \phi) \quad (3.7)$$

Det totale sykefraværet i samfunnet er altså bestemt av forsikringssystemet (c^A , c^w). Et forsikringssystem som tilfredsstillers forsikringsselskapets budsjettbetingelse (utgifter er lik inntekter) vil ha en unik løsning for π ; $\pi c^A + (1 - \pi)c^w = (1 - \pi) * 1$. Vi antar at et individ som jobber produserer én enhet av konsumgodet. Med en fraværsrate på $\pi = F(\hat{\theta})$ så vil $(1 - \pi)$ enheter produseres.

Anta nå at sosial stigma knyttet til å være borte fra jobb (ϕ) ikke lengre er eksogent gitt, men en avtakende funksjon av totalt sykefravær i samfunnet:

$$\phi = \phi(\pi^{total}), \quad \phi'(\pi^{total}) < 0 \quad (3.8)$$

²¹ I Norge er sykelønnsordningen utformet slik at $c^A = c^w$.

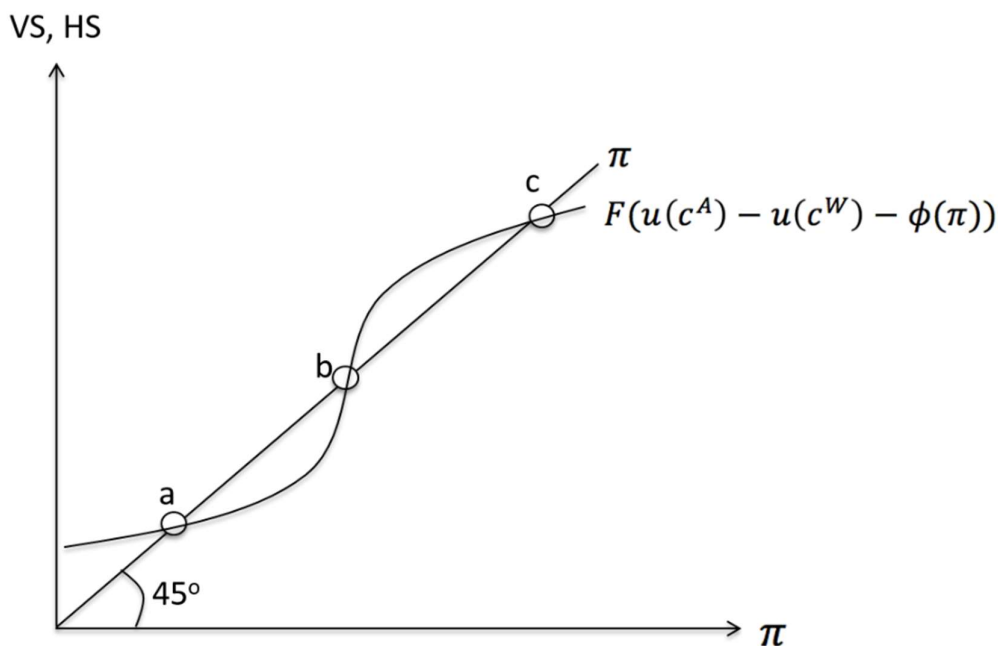
Det er flere mulige mekanismer som kan føre til dette. For eksempel, sosial stigma ved å være borte fra jobb kan synke dersom mange av ens venner, kolleger og naboer er borte fra jobb. Det kan også tenkes at man har større glede av fritidsaktiviteter dersom det er flere mennesker som kan delta, dvs. andre som også er borte fra jobb.

Dersom vi setter (3.8) inn i (3.7) og $\pi^{total} = \pi$ i en rasjonell forventinger likevekt med ex ante identiske individ får vi:

$$\pi = F(u(c^A) - u(c^W) - \phi(\pi)) \quad (3.9)$$

Siden $F(\cdot)$ generelt ikke er lineær, kan (3.9) ha multiple løsninger selv om $\phi(\cdot)$ funksjonen er lineær. Når begge funksjonene er ikke-lineære, er sannsynligheten for multiple løsninger stor.

Figur 6: Multiple likevekter for fraværsraten



Kilde: Egen figur basert på figur fra Lindbeck et al. (2016, s. 55)

I figur 6 er det illustrert at det kan oppstå multiple likevekter for fraværsraten. Venstre side av (3.9) er representert ved den rette 45-graderslinjen, og høyre side er representert ved den ikke-lineære kurven. Langs x-aksen måles fraværsraten π , og langs y-aksen måles verdien av venstre- og høyre side av (3.9). Slik som kurvene er tegnet her, er det tre mulige løsninger i modellen (a, b og c), dvs. π kan ta tre verdier som tilfredsstillende (3.9). Den midterste likevekten,

b, er derimot ikke stabil. Anta at vi befinner oss i likevekt b. Dersom fraværssraten (π) øker så vil sosial stigma ved å være borte fra jobb i samfunnet (ϕ) falle, og fraværet vil øke enda mer – helt til vi når likevekten i punkt c. Tilsvarende vil en reduksjon i fraværssraten (π) føre til at sosial stigma ved å være borte fra jobb (ϕ) øker, og fraværet vil falle enda mer – helt til vi befinner oss i likevekten i punkt a. I tilfeller med multiple likevekter er det typisk at man havner i en «god» eller en «dårlig likevekt». I dette tilfellet vil a være en god likevekt med lavt sykefravær, og c vil være en dårlig likevekt med høyt sykefravær.

For et gitt forsikringssystem (c^A, c^W), er fraværssraten π generelt ikke unik. Et nasjonalt forsikringssystem som er obligatorisk i hele landet vil en typisk få ulike fraværssrater ulike steder. På grunn av sosiale normer kan man få ulike fraværssrater i ulike deler av landet selv om individene er homogene ex ante og har samme sannsynlighetsfordeling av helsevariabelen, $F(\theta)$. Fra figur 4 så vi at det er geografisk variasjon i sykefravær på landsbasis. Variasjonen skyldes nok først og fremst heterogenitet mellom individene og industristrukturen i de ulike fylkene, men det kan også tenkes at sosiale normer forklarer noe av variasjonen.

I denne oppgaven skal vi studere betydningen av sosiale normer ved å bruke lokal variasjon i sykefravær. Vi vil benytte variasjon i sykefravær i de ulike teamene i kundeservicesenteret for å undersøke om det eksisterer sosiale interaksjonseffekter. Vi har nå sett, fra et teoretisk perspektiv, at det kan oppstå lokal variasjon i sykefravær (variasjon mellom team i vårt tilfelle) grunnet sosiale normer selv om individene man studerer er ex ante homogene. I avsnitt 7.3.3 ser vi nærmere på variasjon mellom de ulike teamene som vi studerer.

4 Tidligere empiri

Teorien antar at det eksisterer sosiale interaksjonseffekter, men hvorvidt slike effekter faktisk eksisterer er et empirisk spørsmål. I dette kapittelet vil vi derfor se nærmere på noen av resultatene som er funnet i forskningslitteraturen innenfor dette fagområdet. I empiriske studier er fokuset typisk å estimere en lokal interaksjonsmodell, ikke en global interaksjonsmodell. Vi tar først for oss studier som ser på sosial interaksjon i forbindelse med sykefravær og andre velferdsytelser, og deretter studier som ser på sosial interaksjon i andre kontekster.

Ichino og Maggi (2000) studerer sosial interaksjon i relasjon til fraværsadferd. Forfatterne bruker personelldata om fravær og skoft fra en stor italiensk bank til å studere ansatte som flytter mellom avdelinger. De undersøker om et individs skoft blir påvirket av kollegers skoft. I perioden de studerer, 1975-1995, har det vært 28 642 ansatte på 442 ulike avdelinger. For å studere sosiale interaksjonseffekter bruker forfatterne den såkalte flytte-modellen, hvor målsettingen er å estimere hva som bestemmer et individs skoft som en funksjon av det lokale gjennomsnittlige skoftet, individuelle egenskaper og lokale attributter (for eks, områdevariabler som arbeidsledighetsraten, avdelingsspesifikke variabler om kvalitet på ledelse, osv.). Ved å fokusere på flytterne og bruke ulike instrumenter for kollegers skoft kan de identifisere sosiale interaksjonseffekter. Forfatterne finner at ansatte som flytter mellom avdeling tenderer å tilpasse seg til det gjennomsnittlige nivået skoft i sin nye avdeling, og at ansatte i sør har høyere skoft enn ansatte i nord. Forskjellene mellom sør og nord kan forklares av ulik individuell bakgrunn, seleksjon til beliggenhet, lokale attributter og – særlig av interesse for vår analyse - sosiale interaksjonseffekter.

Hesselius et al. (2009) utnytter et randomisert sosialt eksperiment for å se hvordan kollegers innsats påvirker hverandres, der innsats er målt som sykefravær. Forfatterne bruker administrativ data fra 3000 arbeidsplasser. De ser på en reform som utvidet egenmeldingen for sykefravær for halvparten av befolkningen i Göteborg kommune i Sverige i 1988. Eksperimentet varte et halvt år og alle ansatte som var født på oddetallsdager inngikk i behandlingsgruppen. Ansatte som var omfattet av ordningen (behandlingsgruppen) fikk utvidet sin egenmelding til 14 dager, mens de som ikke var omfattet av ordningen (kontrollgruppen) fikk 7 dager egenmelding som før. I bedrifter der mange ansatte var omfattet av ordningen finner de at sykefraværet øker betraktelig. I studien finner de også sterke effekter for ansatte som ikke var omfattet av ordningen med utvidet sykemelding på arbeidsplasser med høy andel

ansatte som var omfattet av ordningen. Ansatte som ikke var omfattet av ordningen ser ut til å tilpasse sin adferd etter ansatte som var omfattet av ordningen med utvidet egenmelding. Forfatterne konkluderer med at sosial interaksjon på arbeidsplassen er en avgjørende faktor for ansatte sin innsats i bedriften, målt via sykefravær.

Rieck og Vaage (2010) studerer sosial interaksjon på norske arbeidsplasser. De undersøker om et individs sykefravær påvirkes av dens kollegers. De bruker registerdata for norske lærere fra Kommunenes Sentralforbund. Deres datasett omfatter lærere som er ansatt i offentlig sektor på barneskoler, ungdomsskoler og videregående skoler i de fleste kommuner i perioden 2002-2006. De har blant annet informasjon om sykefravær, utdanning, arbeidsmarkedsstatus, inntekt, alder og kjønn. Disse dataene er slått sammen med flere andre administrative registre for å få informasjon om skoler og kommuner. Forfatterne bruker to ulike tilnærminger for å studere om det eksisterer sosiale interaksjonseffekter for sykefravær. Først undersøker de om det er slik at lærere som bytter arbeidsplass tenderer å tilpasse sitt sykefravær til sine nye kollegers. Ved å fokusere på flytterne og bruke instrumentvariabelmetoden kan forfatterne identifisere sosiale interaksjonseffekter. Deretter undersøker de om nyansatte lærere tenderer å tilpasse sitt sykefravær til sine kollegers. Her benytter de en fasteffekt modell samt instrumentvariabelmetoden for å identifisere sosiale interaksjonseffekter. Forfatterne finner positive effekter for begge tilnærmingene.

Rege et al. (2012) undersøker om tilgang til uførepensjon påvirkes av sosial interaksjon. Forfatternes hypotese er at et individs valg om uførepensjonering avhenger av personer i dens referansegruppe. Det finnes flere mulige forklaringer til det. Hvis det er slik at trygd i utgangspunktet oppleves stigmatiserende kan det tenkes at det oppleves mindre stigmatiserende hvis man kjenner mange som er uføretrygdet, og mer stigmatiserende hvis få eller ingen man kjenner er uføretrygdet. Det kan også tenkes at det er lettere å søke om uførepensjon hvis man kjenner noen som er uføretrygdet ettersom de kan gi deg informasjon om søkeprosessen. Mer tid til fritidsaktiviteter sammen med personer i ens referansegruppe kan også gjøre det mer attraktivt å bli uføretrygdet. Forfatterne har et 11 år langt paneldatasett som inneholder sosioøkonomisk informasjon, ansattdata og uførepensjon status for hele den norske populasjonen. Referansegruppe er definert som jevnaldrende personer bosatt i samme nabolag. For å identifisere sosiale interaksjonseffekter bruker forfatterne instrumentvariabelmetoden. Forfatterne finner først at nedbemanning øker sannsynligheten for uførhet blant de som rammes, og denne eksogene variasjonen i uførhet benyttes derfor som instrument når de ser på

andre i nabolaget som ikke ble berørt av nedbemanning. Sosiale normer må derfor være forklaringen til at naboene som ikke ble berørt av nedbemanning påvirkes. I analysen finner de bevis for at sosial interaksjon spiller en rolle for trygdeytelser.

Dale-Olsen et al. (2015) analyserer hvordan sosial interaksjon påvirker sykefraværsadferd på arbeidsplassen. Forfatterne bruker registerdata fra perioden 2004-2006 med detaljert informasjon om alle menn i arbeidsstyrken i Norge. De har informasjon om blant annet lengde på ansettelse, lønn, arbeidstimer, legemeldt sykefravær, individuelle egenskaper, samt sosioøkonomiske variabler. For å identifisere sosiale interaksjonseffekter brukes en fasteffekt modell og instrumentvariabelmetoden. Forfatterne utnytter variasjon i sykefravær grunnet en skattereform som ble innført i 2006 som identifikasjonsstrategi. Skattereformen innførte blant annet en lavere marginal inntektsskatt for individ med høy inntekt. Endringene i marginalsattene påvirket sykefraværsadferd, og forfatterne benytter derfor gjennomsnittlig skatteendring blant gruppe medlemmene som instrument. Gruppe medlemmene er definert som mannlige kolleger ansatt på samme arbeidsplass med samme yrke. Forfatterne finner at et individs sykefravær er positivt korrelert med sine kollegers sykefravær.

Markussen og Røed (2015) bruker administrativ paneldata fra Norge for å studere hvordan sosial forsikringsavhengighet spres innen nabolag, familier, etniske minoriteter og mellom tidligere klassekamerater. Deres datasett dekker hele arbeidsstyrken som er over 17 år. Forfatterne undersøker også hvordan sosiale interaksjonseffekter varierer med relasjonsavstand, dvs. hvordan sosiale interaksjonseffekter varierer med «nære» og «fjerne» nettverksmedlemmer. Forfatterne bruker en fasteffekt modell og antakelsen om at sosiale interaksjonseffekter virker med tidsetterslep som identifikasjonsstrategi. De finner at et individ sitt nivå av sosiale ytelser er sterkt påvirket av nivået til personer i samme nettverk, og at effekten blir større jo sterkere relasjon individene har.

Lindbeck et al. (2016) bruker data fra hele den svenske populasjonen til å analysere hvordan gruppeeffekter påvirker sykefraværsadferd. Deres datasett kombinerer individuell sykefraværsdata fra Försäkringskassan med et stort antall sosioøkonomiske variabler fra LISA databasen, drevet av Statistiska centralbyrån. Forfatterne kan identifisere hvert individ sitt nabolag og arbeidsplass i tillegg til individuelle egenskaper. Datasettet er panel i tidsperioden 1996-2002. Deres sykefraværsdata dekker kun sykefravær som overstiger 14 dager. I artikkelen tar de for seg flere ulike tilnærminger for å studere gruppeeffekter. En tilnærming er å studere

om individ som flytter fra et nabolag i Sverige til et annet tenderer å tilpasse sitt sykefravær til det som er normalt i det nye nabolaget. En annen tilnærming er å studere om immigranter som flytter til Sverige tilpasser sitt sykefravær til de innfødtes sykefravær i nabolaget de har flyttet til. Forfatterne undersøker også hvorvidt et individs sykefravær er påvirket gjennom sosial interaksjon med personer den møter både i nabolaget og på arbeidsplassen. For å estimere de ulike tilnærmingerne bruker forfatterne ulike økonometriske modeller og identifiseringsantakelser. Forfatterne finner at det eksisterer sosiale interaksjonseffekter for sykefravær med alle strategiene.

Sacerdote (2001) studerer gruppeeffekter blant studenter på Dartmouth College i USA. Han undersøker om det eksisterer gruppeeffekter for studenters karaktergjennomsnitt, for deres beslutninger om å delta i studentorganisasjoner og for deres valg av studieretning senere i studieløpet. Forfatteren har informasjon om studentenes studieretning, hvilke aktiviteter de er med på, hvor på campus de bor, samt karakterer før og på college. Forfatteren ser på studenter som avsluttet sin grad i 1997 og 1998. Nye studenter får romkamerat tildelt tilfeldig, men samme kjønn. Dette innebærer at romkameratens egenskaper er ukorrelert med hverandre. Forfatter kan dermed estimere redusert form effekter av student i sin bakgrunn på romkameraten j sitt utfall. Sacerdote (2001) finner gruppeeffekter for romkameratens valg av studentorganisasjoner og for deres karaktergjennomsnitt. I analysen brukes en akademisk indeks som en proxyvariabel for forklaringsvariabelen *evner*. Effekten på karakterer når begge studentene har lav akademisk indeks er -0.331, mens den er -0.16 når studenten selv har lav akademisk indeks og romkameraten har høy akademisk indeks. Dette kan tyde på at studenter med høy akademisk indeks kan hjelpe studenter med lav akademisk indeks. Studenters valg av studieretning ser derimot ikke ut til å være påvirket av romkameraten sitt valg av studieretning.

Mas og Moretti (2009) studerer gruppeeffekter på arbeidsplassen. De undersøker hvordan produktiviteten til arbeidere som jobber i samme team påvirkes av hverandres. Ved å se på produksjonsprosesser der et produkt produseres av flere personer samtidig kan det tenkes to utfall på arbeidernes produktivitet. Det kan tenkes at gratispassasjerproblemet oppstår dersom individuell innsats ikke kan observeres, dvs. noen arbeidere blir mindre produktive dersom de jobber sammen med andre som er høyproduktive siden jobben blir gjort uansett. Det kan derimot også tenkes at det er stigma knyttet til å prestere dårlig, noe som gjør at den enkelte arbeideren har insentiv til å prestere bedre dersom den jobber sammen med andre som er mer produktiv enn dem selv. Dataene som benyttes i analysen inneholder informasjon om

produktiviteten til kasseoperatører og er tatt fra en supermarkedkjede. Kundene på supermarkedet er ikke pålagt å gå til en spesiell kassebetjent, men man observerer typisk at kunder foretrekker å stå i den køen som går raskest. Dette innebærer at en effektiv kassebetjent får mer å gjøre enn de som er mindre effektiv. Forfatterne finner sterke gruppeeffekter assosiert med å introdusere en høyproduktiv arbeider til en gruppe medarbeidere. De finner at en økning på 10 prosent i en kollega sin produktivitet fører til en økning på 1,5 prosent i individuell produktivitet, noe som indikerer at positive gruppeeffekter dominerer gratispassasjerproblemet. Videre finner forfatterne at lavproduktive arbeidere er mer følsom for sammensetningen av medarbeidere enn arbeidere som er høyproduktive. Lavproduktive arbeidere blir mer produktiv når høyproduktive arbeidere er til stedet, men høyproduktive arbeidere blir ikke mindre produktiv når lavproduktive arbeidere er tilstede. Sammensetningen av medarbeidere som maksimerer produktivitet er den hvor medarbeider mangfold i hvert skift maksimeres.

Resultatene fra disse studiene tyder på at det eksisterer lokale sosiale interaksjonseffekter i ulike kontekster. I dette kapitlet har vi sett at det finnes ulike metoder og strategier for å estimere sosiale interaksjonseffekter. I neste kapittel skal vi se nærmere på utfordringer knyttet til identifikasjon av sosiale interaksjonseffekter, samt ulike identifikasjonsstrategier man kan benytte i empiriske arbeid. Hvilke type data man har tilgjengelig er avgjørende for hvilke identifikasjonsstrategier man kan benytte og hvilke type analyser man kan gjennomføre. Vi har tilgang til bedriftsdata med informasjon om de ansattes sykefravær samt noen individuelle egenskaper, der de ansatte jobber i definerte team. Vi vil derfor hente inspirasjon om identifikasjon fra studier som har lignede data og variasjon.

5 Identifikasjon

I dette kapitlet skal vi se nærmere på estimering av sosial interaksjon i en generell sammenheng. Først vil vi gjennomgå en generell modell for estimering av sosiale interaksjonseffekter, deretter vil vi gjennomgå metodiske utfordringer ved estimering og til slutt vil vi se på hvordan man kan løse disse metodiske utfordringene.

5.1 Identifikasjon av sosiale interaksjonseffekter: en generell modell

Manski (1993) skiller mellom ulike mekanismer som forklarer hvorfor individ i samme gruppe oppfører seg likt. Han skiller mellom endogene, eksogene og korrelerte effekter. Endogene effekter er når individuell adferd er påvirket av gruppens adferd. Eksogene (kontekstuelle) effekter er når individuell adferd er påvirket av gruppens egenskaper. Korrelerte effekter er når gruppe-medlemmer har lik adferd fordi de har like egenskaper eller like institusjonelle rammer. Vi ser på et eksempel om skoleprestasjoner for å tydeliggjøre skillet mellom de tre effektene. I dette tilfellet vil en endogen effekt være når den enkelte elevens prestasjoner er påvirket av medelevers prestasjoner. En eksogen effekt er når medelevers sosiale bakgrunn påvirker enkeltelevens prestasjoner. Korrelerte effekter er når elevenes prestasjoner samvarierer fordi at de har samme sosial bakgrunn, samme lærere, osv. Endogen sosial interaksjon gir sosiale multiplikatoreffekter og er derfor viktige policyimplikasjoner.

For å identifisere sosiale interaksjonseffekter mellom individ trenger vi en modell som kan identifisere hvordan utfallet/adferden til et individ påvirkes av utfallet/adferden til andre individ i den samme referansegruppen. Vi tar utgangspunkt i en enkel modell presentert i Moffitt (2001). Dersom vi antar linearitet kan modellen formuleres slik:

$$y_{i,g} = \theta_0 + \theta_1 x_{i,g} + \theta_2 \bar{y}_{-i,g} + \theta_3 \bar{x}_{-i,g} + \varepsilon_{i,g} \quad (5.1)$$

Her er $y_{i,g}$ utfallet til individ $i=1, \dots, N$ i gruppe $g=1, \dots, G$. Videre er $x_{i,g}$ en eksogen forklaringsvariabel, $\bar{y}_{-i,g}$ er det gjennomsnittlige utfallet til de andre gruppe-medlemmene og $\bar{x}_{-i,g}$ er en gjennomsnittlig eksogen forklaringsvariabel til de andre gruppe-medlemmene. Parameteren θ_0 er konstantleddet og $\varepsilon_{i,g}$ er et stokastisk feilledd med forventning lik null og konstant varians. Parameteren θ_2 refereres til som den endogene sosiale interaksjonseffekten og er parameteren av interesse som vi ønsker å estimere. Parameteren θ_3 refereres til som en eksogen sosial interaksjonseffekt.

Estimering av (5.1) med OLS vil gi forventingskjevne og inkonsistente estimatorer. For at OLS skal være både forventingsrett²² og konsistent²³ må følgende fire Gauss Markov antakelser være oppfylt (Verbeek, 2012, s. 15):

1. $E(\varepsilon_i) = 0$. Forventingen til feilleddene må være lik 0. Det betyr at regresjonslikningen i gjennomsnitt er riktig.
2. $E(\varepsilon_i|x_i) = 0$. Feilleddene og forklaringsvariablene er uavhengige.
3. $Var(\varepsilon_i) = \sigma^2$. Variansen til feilleddene er konstant, dvs. alle feilledd har samme varians. Dette kalles homoskedastisitet. Dersom denne ikke gjelder sier vi at feilleddene er heteroskedastiske.
4. $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$. Kovariansen mellom feilleddene må være 0. Dette utelukker autokorrelasjon.

En konsistent estimator kan utledes under svakere betingelser enn en forventingsrett estimator. For at en estimator skal være konsistent holder det at forventingen til feilleddene er lik 0 og at feilleddene er ukorrelert med alle forklaringsvariablene, dvs. $E(\varepsilon_i x_i) = 0$. Det kan være flere årsaker til at OLS er inkonsistent. Noen av de vanligste årsakene er simultane ligninger, målefeil, utelatte variabler og dynamiske modeller med seriekorrelasjon. I neste avsnitt vil vi ta for oss hva som kan føre til at OLS er inkonsistent i vårt tilfelle, der målsettingen er å estimere sosiale interaksjonseffekter.

5.2 Metodiske utfordringer

Det er flere utfordringer når man skal estimere modell (5.1). I dette avsnittet skal vi ta for oss de største problemene som man møter ved empirisk identifisering av sosiale interaksjonseffekter. Disse problemene kan deles inn i tre kategorier: simultanitet, korrelerte effekter og endogent gruppedlemskap (Moffitt, 2001).

Simultanitetsproblemet oppstår når person A sin handling påvirker person B sin handling og vice versa. Når person A og person B påvirker hverandre simultant (samtidig) er det vanskelig å separere den kausale effekten som person A sitt utfall har på person B sitt utfall, eller omvendt. Manski (1993) refererer til dette som refleksjonsproblemet. Han stiller spørsmålet; «er det speilbildet som forårsaker at personen beveger seg eller reflekterer det dem?». Dette impliserer

²² OLS er forventingsrett når forventningen til OLS estimatoren er lik den sanne populasjonsverdien; $E(\hat{\beta}) = \beta$.

²³ OLS er konsistent når den predikerte estimatoren konvergerer mot den sanne verdien når antallet observasjoner øker; $plim(\hat{\beta}) = \beta$.

at $\bar{y}_{-i,g}$ er endogen²⁴ i modellen, noe som fører til forventingsskjevne og inkonsistente estimatorer. Vi kan illustrere simultanitetsproblemet ved å anta at vi har $g=1, \dots, G$ grupper med to individ ($i=1,2$) per gruppe. Modellen kan da formuleres på følgende måte:

$$y_{1,g} = \theta_0 + \theta_1 x_{1,g} + \theta_2 y_{2,g} + \theta_3 x_{2,g} + \varepsilon_{1,g} \quad (5.2)$$

$$y_{2,g} = \theta_0 + \theta_1 x_{2,g} + \theta_2 y_{1,g} + \theta_3 x_{1,g} + \varepsilon_{2,g} \quad (5.3)$$

Vi antar at $x_{1,g}$ og $x_{2,g}$ er strengt eksogen, [$E(\varepsilon_{1,g} | x_{1,g}, x_{2,g}) = E(\varepsilon_{2,g} | x_{1,g}, x_{2,g}) = 0$], og at feilleddene er uavhengig av hverandre, [$cov(\varepsilon_{1,g}, \varepsilon_{2,g}) = 0$]. Vi kan vise at $\varepsilon_{1,g}$ er korrelert med $y_{2,g}$ og at $\varepsilon_{2,g}$ er korrelert med $y_{1,g}$, ved å skrive (5.2) og (5.3) på redusert form. Dette gjøres ved innsetting av (5.3) i (5.2) og omvendt. Vi får da følgende ligninger:

$$y_{1,g} = \alpha + \beta x_{1,g} + \gamma x_{2,g} + v_{1,g} \quad (5.4)$$

$$y_{2,g} = \alpha + \beta x_{2,g} + \gamma x_{1,g} + v_{2,g} \quad (5.5)$$

hvor

$$\alpha = \frac{\theta_0(1 + \theta_2)}{1 - \theta_2^2} \quad (5.6)$$

$$\beta = \frac{\theta_1 + \theta_2 \theta_3}{1 - \theta_2^2} x_{2,g} \quad (5.7)$$

$$\gamma = \frac{\theta_3 + \theta_1 \theta_2}{1 - \theta_2^2} x_{1,g} \quad (5.8)$$

$$v_{1,g} = \frac{\varepsilon_{1,g} + \theta_2 \varepsilon_{2,g}}{1 - \theta_2^2} \quad (5.9)$$

$$v_{2,g} = \frac{\varepsilon_{2,g} + \theta_2 \varepsilon_{1,g}}{1 - \theta_2^2} \quad (5.10)$$

²⁴ En endogen variabel er en variabel som er korrelert med feilleddet i modellen, noe som bryter med Gauss Markov antakelse (2).

Ved å skrive ligningene på redusert form forsvinner de endogene variablene. Parameterne α , β og γ er funksjoner av θ_1 , θ_2 og θ_3 . Ligningene kan estimeres, men vi kan ikke skille mellom korrelerte, eksogene og endogene sosiale interaksjonseffekter. Hvis det ikke er sosiale interaksjonseffekter ($\theta_2 = \theta_3 = 0$) blir $\beta = \theta_1$ og $\gamma = 0$.

Fra våre antakelser om forklaringsvariablene og feilleddene følger det at kovariansen mellom $\varepsilon_{1,g}$ og $y_{2,g}$ er gitt ved:

$$\text{cov}(\varepsilon_{1,g}, y_{2,g}) = \text{cov}\left(\varepsilon_{1,g}, \frac{\varepsilon_{2,g} + \theta_2 \varepsilon_{1,g}}{1 - \theta_2^2}\right) = \text{var}(\varepsilon_{1,g}) \frac{\theta_2}{1 - \theta_2^2} \neq 0 \quad (5.11)$$

Den første overgangen følger av at $E[\varepsilon_{1,g} | x_{1,g}, x_{2,g}] = 0$, og den andre overgangen følger av at $\text{cov}(\varepsilon_{1,g}, \varepsilon_{2,g}) = 0$. Tilsvarende gjelder kovariansen mellom $\varepsilon_{2,g}$ og $y_{1,g}$. Dersom det eksisterer en endogen sosial interaksjonseffekt, dvs. $\theta_2 \neq 0$, ser vi fra (5.11) at feilleddet ($\varepsilon_{1,g}$) vil være korrelert med forklaringsvariabelen ($y_{2,g}$). Estimering med OSL vil dermed gi forventingsskjevne og inkonsistente estimatorene. Dette gjør at det blir vanskelig å identifisere den endogene sosiale interaksjonseffekten.

Korrelerte effekter kan føre til problemer med utelatte variabler dersom individspesifikke eller gruppespesifikke komponenter i feilleddet varierer mellom individ og grupper, og samtidig er korrelert med egenskaper til individene. Gruppemedlemmene kan ha lik adferd fordi de har like egenskaper som ikke kan observeres. For å illustrere dette modifierer vi den reduserte modellen ved å tillate $i=1, \dots, N_g$ individ per gruppe:

$$y_{i,g} = \alpha + \beta x_{i,g} + \gamma \bar{x}_{-i,g} + \mu_g + \mu_i + \eta_{i,g} \quad (5.12)$$

Komponenten μ_g er en uobservert gruppeeffekt, for eksempel egenskaper ved teamleder som varierer mellom teamene og samtidig påvirker sykefraværet til de ansatte. Komponentene μ_i er en uobservert individeffekt, for eksempel motivasjon, preferanser eller evner. Hvis μ_g eller μ_i er korrelert med $x_{i,g}$ har vi et utelatt variabelproblem²⁵. Estimering med OLS vil gi inkonsistente estimatorene.

²⁵ I likhet med simultanitetetsproblemet innebærer dette at Gauss Markov antakelse (2) ikke er oppfylt. Vi får et endogenitetsproblem.

Ved estimering av sosiale interaksjonseffekter kan man også få problemer med endogent gruppemedlemskap (seleksjon). Endogent gruppemedlemskap betyr at det finnes egenskaper ved individ som har betydning for hvilken gruppe de er medlem av. Det er ofte slik at individ velger referansegruppe selv, ikke plasseres i dem tilfeldig. Det kan tenkes at individ velger gruppe utfra observerbare egenskaper ved andre gruppemedlemmer. Individ kan for eksempel selektere seg inn i en gruppe basert på kjønns sammensetning, utdanningsnivå, alder, osv. En slik systematisk seleksjon er en spesifikk form for korrelerte effekter og fører til at estimatorene blir inkonsistent ved OLS estimering. For å studere utfall i definerte grupper er det viktig å vite hvordan initial gruppefordeling ble gjennomført, dvs. om gruppefordeling er tilfeldig eller potensielt utsatt for seleksjon.

5.3 Identifikasjonsstrategier

Det er ofte vanskelig å identifisere sosiale interaksjonseffekter, både på grunn av de metodiske utfordringene som vi har diskutert og typisk dårlig/lite passende data. Det finnes derimot noen metoder man kan benytte for å identifisere slike effekter. Soetevent (2006) diskuterer ulike identifikasjonsstrategier for å løse problemer med simultanitet, korrelerte effekter og endogent gruppemedlemskap ved estimering av sosiale interaksjonseffekter.

For å løse problemet med simultanitet kan man bruke sekvensielle lab-eksperiment som identifikasjonsstrategi. En fordel med lab-eksperiment er at forsker har full kontroll over individene som studeres. En strategi som kan benyttes for å identifisere sosiale interaksjonseffekter i lab-eksperiment er å gi behandlingsgruppen og kontrollgruppen ulik informasjon og la dem handle på ulike tidspunkt. Ved å gi mer informasjon til for eksempel behandlingsgruppen og la dem handle først kan det tenkes at denne gruppen vil signalisere noe til kontrollgruppen. Kontrollgruppen får observere handlingen til behandlingsgruppen før de handler selv. Det kan da tenkes at kontrollgruppen vil tilpasse sin adferd etter behandlingsgruppen, og ettersom handling foregår sekvensielt unngår man problemet med simultanitet. I lab-eksperiment er det viktig at gruppesammensetningen er randomisert. Potters et al. (2005) benytter en slik identifikasjonsstrategi når de studerer adferd knyttet til private bidrag til kollektive goder.

En annen identifikasjonsstrategi man kan benytte for å løse simultanitetsproblemet er instrumentvariabelmetoden. Identifikasjon oppnås ved å finne en eller flere instrumentvariabler

for den endogene variabelen. En instrumentvariabel er en variabel som (1) er korrelert med den endogene variabelen, (2) er ukorrelert med feilledet og (3) ikke er relevant for å forklare variasjon i utfallsvariabelen. (3) impliserer at det ikke eksisterer eksogen sosial interaksjon for denne variabelen. Det kan derimot være vanskelig å finne et slikt instrument. Dersom vi tar utgangspunkt i modell (5.2-5.3) og ønsker å estimere hvordan individ 1 påvirker individ 2, innebærer dette at vi må finne en variabel som inngår i (5.2) men ikke i (5.3). Flere studier analyserer gruppeeffekter ved å bruke instrumentvariabelmetoden²⁶.

En tredje måte å løse problemet med simultanitet er ved å manipulere gruppedlemskap. Løsningen kan for eksempel være å flytte individ mellom grupper. Den identifiserende antakelsen her er at individet som blir flyttet ikke påvirker de andre i gruppen/nabolaget/arbeidsplassen som den blir flyttet til. Individet som blir flyttet vil derimot kunne bli påvirket av de andre i gruppen. For eksempel, anta at en familie flytter til et nabolag med N familier, og at hver familie tilskriver en referansevekt lik $1/N$ til de andre familiene. Når $N \rightarrow \infty$ betyr dette at adferden til individene som allerede er bosatt i dette nabolaget ikke vil endres av at én ny familie flytter dit. Den nye familien vil derimot bli påvirket av alle sine nye naboer med en total referansevekt lik $N \times (1/N) = 1$. Dette kan imidlertid være vanskelig å gjennomføre i praksis. Ichino og Maggi (2000) benytter blant annet denne metoden for å studere sosiale interaksjonseffekter. Alternativt kan manipulering av gruppedlemskap bestå i å observere utfall for individ som er nye i gruppen, for eksempel nyansatte i en bedrift eller tilflyttere til et nabolag. Rieck og Vaage (2010) og Lindbeck et al. (2016) benytter henholdsvis disse metodene.

En metode for å løse problemet med korrelerte effekter er å bruke en funksjonsform som inkluderer individspesifikke og gruppespesifikke komponenter. Gitt at de individspesifikke og gruppespesifikke komponentene er faste over tid kan vi benytte en fasteffekt modell. Dette innebærer at vi inkluderer en dummyvariabel α_i for hvert individ i ($\alpha_i = 1$ for $i=j$, 0 ellers), og en dummyvariabel α_g for hver gruppe g ($\alpha_g = 1$ for $k=g$, 0 ellers). Disse dummyene vil fange opp nivåforskjeller mellom individ og grupper, og kontrollerer dermed for uobservert heterogenitet. Modell (5.12) over kan da utvides på følgende måte:

$$y_{i,g,t} = \theta_0 + \theta_1 x_{i,g,t} + \theta_2 \bar{y}_{-i,g,t} + \theta_3 \bar{x}_{-i,g,t} + \alpha_i + \alpha_g + \varepsilon_{i,g,t} \quad (5.13)$$

²⁶ Se for eksempel Rege et al. (2012) og Dale-Olsen et al. (2015).

Indeksen $t=1, \dots, T$ angir tid. Ved å kontrollere for uobservert heterogenitet mellom individ og grupper reduseres problemet med korrelerte effekter/utelatte variabler. Flere studier benytter denne metoden for å redusere problemet med korrelerte effekter²⁷.

Eksperiment som omfatter deler av populasjonen er en annen identifikasjonsstrategi man kan benytte for å løse problemet med korrelerte effekter. I slike eksperiment treffer et tiltak deler av populasjonen eller eventuelt noen av gruppemedlemmene, men ikke alle. Ta utgangspunkt i modell (5.2-5.3) og anta at kun individ 1 blir utsatt for tiltak $p_{1,g}$. Det kan for eksempel være en subsidie eller et stipend. Vi får da følgende uttrykk:

$$y_{1,g} = \theta_0 + \theta_1 x_{1,g} + \theta_2 y_{2,g} + \theta_3 x_{2,g} + \theta_4 p_{1,g} + \varepsilon_{1,g} \quad (5.14)$$

$$y_{2,g} = \theta_0 + \theta_1 x_{2,g} + \theta_2 y_{1,g} + \theta_3 x_{1,g} + \varepsilon_{2,g} \quad (5.15)$$

$p_{1,g}$ vil her fungere som en instrumentvariabel og identifisere parameterne i (5.15). Hesselius et al. (2009) bruker denne identifikasjonsstrategien. Naturlige eksperiment kan også ha denne egenskapen. I naturlige eksperiment ser man ofte på reformer som eksogent endrer forutsetningene for noen individ i gruppen eller populasjonen som studeres. Dersom noen i kontrollgruppen blir påvirket er ideen at dette følger av sosiale interaksjonseffekter. Denne identifikasjonsstrategien kan også benyttes for å løse simultanitetsproblemet ettersom den fungerer som instrumentvariabelmetoden.

Det er flere metoder man kan bruke for å løse problemet med endogent gruppemedlemsskap (seleksjon). Lab-eksperiment, instrumentvariabelmetoden, fasteffekt modellen og manipulering av gruppemedlemsskap kan brukes som identifikasjonsstrategi. Naturlig nok er spesielt manipulering av gruppemedlemsskap en god identifikasjonsstrategi ved problemer med endogent gruppemedlemsskap.

Vi har nå diskutert ulike identifikasjonsstrategier som kan benyttes for å løse problemer med simultanitet, korrelerte effekter og endogent gruppemedlemsskap. Som tidligere nevnt, vil tilgjengelig data være avgjørende for hvilke identifikasjonsstrategier man kan benytte. Ettersom vi har tilgang til bedriftsdata, er det lite aktuelt for oss å benytte lab-eksperiment og eksperiment

²⁷ Se blant annet Rieck og Vaage (2010) og Markussen og Røed (2015).

som omfatter deler av populasjonen som identifikasjonsstrategier. For å løse problemet med simultanitet vil det derimot være aktuelt for oss å benytte instrumentvariabelmetoden eller å manipulere gruppedlemsskap. Når det gjelder å manipulere gruppedlemsskap, kan vi benytte to metoder; den første er å se på utfall for nyansatte og den andre er å se på utfall for ansatte som flytter mellom team. I bedriften i denne analysen er det tilfeldig hvilket team de nyansatte plasseres i, dvs. at det er bemanningsbehovet i teamene, og ikke de nyansattes egenskaper, som bestemmer plasseringen. Dette innebærer at de ikke kan selektere seg inn i team basert på observerbare egenskaper. At ansatte flytter mellom team, antas også å skyldes bemanningsbehovene fremfor systematiske egenskaper ved individene og/eller teamene. For å redusere problemet med korrelerte effekter er det aktuelt for oss å benytte en fasteffekt modell. Denne metoden er relativt enkel å implementere i regresjonsanalysen, så lenge man benytter paneldata.

6 Kundeservicesenteret

Som tidligere nevnt skal vi studere sosiale interaksjonseffekter i et kundeservicesenter i et norsk forsikringsselskap. Vi skal undersøke om det er gjensidig avhengighet mellom de ansattes sykefravær. I dette kapittelet vil vi presentere de ansattes arbeidsoppgaver, organisering og mulige konsekvenser av at forsikringsselskapet ikke er en IA-virksomhet. Det er viktig med et innblikk i kundeservicesenteret for å forstå og kunne tolke våre resultat bedre. Informasjonen som presenteres i dette kapittelet er informasjon som vi har fått fra bedriften, og informasjon som er hentet fra Aarbu og Torsvik (2007).

6.1 Arbeidsoppgaver

Kundeservicesenterets arbeidsoppgaver er hovedsakelig å svare på innkommende samtaler fra eksisterende og nye kunder. De ansatte skal først og fremst være behjelpelige med svare på kundenes spørsmål, men de skal også prøve å selge nye forsikringsprodukter. Det eksisterer en bonusordning knyttet til salg av forsikringsproduktene for å øke insentiv for salg av disse. Innringende kunder skal informeres om deres gjeldende forsikringer, eventuelle endringer og nye produkter. De ansatte må ha god kunnskap om selskapets produkter, og de besvarer i gjennomsnitt 112 innkommende samtaler per uke (gjennomsnitt i perioden 2003-2008). De ansatte skal både sørge for at innringende kunder blir fornøyd samtidig som de skal selge nye produkter, men de driver altså ikke ordinært telefonsalg av forsikringsprodukter i form av oppringing til potensielle kunder.

I tillegg må ansatte ta stilling til problemer med ugunstig utvalg og moralsk hasard ved salg av forsikringsprodukter. Med ugunstig utvalg menes det at forsikringsselskapet ikke nødvendigvis vet hvilken type kunden er, dvs. om det er en kunde med høy eller lav risiko for skade. Moralsk hasard innebærer at kunder potensielt endrer adferd på grunn av forsikringen, for eks. ved å ta høyere risiko eller feilrapportere skader. Det er ikke nødvendigvis et mål for forsikringsselskapet å selge flest mulige forsikringer. Dersom forsikringsselskapet kun har «dårlige» kunder vil de gå konkurs. Ansatte må derfor ta hensyn til disse usikkerhetene gjennom å hente og tolke historisk informasjon om kunden. Videre skal de ansatte utføre sine arbeidsoppgaver mest mulig effektivt for å minimere ventetiden for innringende kunder. Det kan være en krevende jobb, og stress er en faktor som kan påvirke et individs sykefravær.

6.2 Organisering

De ansatte i kundeservicesenteret jobber i definerte team som er spredt i Norge, hvor hvert team består av ca. 16 medarbeidere. Den initiale fordelingen av team er tilfeldig, noe som innebærer det ikke er systematisk seleksjon inn til teamene. Nyansatte har ikke anledning til å selekttere seg inn i team basert på observerbare egenskaper som alder, kjønn, utdanning og lignende. Ansatte som flytter mellom ulike team antas å skyldes tilfeldigheter fremfor systematiske årsaker. Det er typisk bemanningsbehovet som er avgjørende for hvilket team nyansatte plasseres i, og grunnen til at noen ansatte bytter team. Dette er fordelaktig når vi skal studere sosiale interaksjonseffekter, ettersom metodene vi bruker er basert på at det ikke foregår systematisk seleksjon inn til teamene.

Teamene jobber i et åpent kontorlandskap, noe som gjør det enkelt for de ansatte å observere om dens medarbeidere er borte fra jobb eller ikke. Hvert team har en teamleder som er plassert sammen med teamet som vedkommende er ansvarlig for. Teamlederen har flere arbeidsoppgaver. Teamlederen skal først og fremst organisere arbeidet til teamet. Dette innebærer blant annet å holde de ansatte oppdatert ved å informere dem om nye forsikringsprodukter. En annen oppgave som teamlederen har er å skape et godt arbeidsmiljø samt motivere de ansatte. Teamlederen skal også kvalitetssikre arbeidet til de ansatte og føre statistikk over individuelt salg ukentlig. Videre skal teamlederen gi tilbakemeldinger på telefonsamtaler de ansatte har med kunder, og komme med forslag til forbedringer. Teamlederen kan ha stor betydning for teammedlemmenes trivsel på arbeidsplassen, noe som kan påvirke de ansattes sykefravær. Ansatte som er fornøyd med sin teamleder og opplever at de har et godt arbeidsmiljø vil trolig ha lavere sykefravær enn ansatte som er misfornøyd med sin teamleder. Vi vet dessverre ingenting om teamlederne - en faktor som potensielt påvirker utfallet vi studerer.

6.3 IA-virksomhet

Under perioden vi studerer, fra 2003 til midten av 2007, er forsikringsselskapet ikke en IA-virksomhet. Det innebærer blant annet at ansatte har rett til å skrive en egenmelding på opptil 3 dager sammenhengende, og maks 4 ganger per år. IA-virksomheter har derimot rett til å skrive en egenmelding på opptil 8 dager sammenhengende, og totalt 24 dager per år. I analysen vi vil blant annet undersøke om det er slik at egenmeldt sykefravær er mer eller mindre påvirket av sosiale interaksjonseffekter enn andre typer sykefravær. I kapittel 1 drøftet vi to motstridende

hypoteser når det gjelder hvilke typer sykefravær som potensielt er mest påvirket av sosiale interaksjonseffekter. Den første hypotesen var at egenmeldt sykefravær, eller de typisk kortere fraværstypene, er mer påvirket av sosiale interaksjonseffekter ettersom sosiale normer ikke direkte påvirker et individs helsetilstand. Siden det er lav grad av kontroll for egenmeldt sykefravær er kanskje denne type sykefravær mest påvirket av slike effekter. Den andre hypotesen var at det kan tenkes at sosiale interaksjonseffekter spiller en større rolle for det legemeldte sykefraværet, eller de typisk lengre fraværstypene, ettersom mesteparten utgjøres av mer diffuse diagnoser som muskel- og skjelettlidelser og lettere psykiske plager. Det er dermed mer subjektivitet knyttet til det legemeldte sykefraværet. Ettersom de ansatte i forsikringselskapet har relativt få egenmeldingsdager, kan det tenkes at effektene vil være større for det legemeldte sykefraværet. Med andre ord, ettersom forsikringselskapet ikke er en IA-virksomhet kan det tenkes at den andre hypotesen dominerer i vårt tilfelle.

7 Data

Formålet med dette kapitlet er å få et generelt innblikk i dataene som benyttes i den empiriske analysen. Avsnitt 7.1 gir generell informasjon om datasettet og hva som er dets sterke og svake sider i forhold til andre tilgjengelige data. Vi belyser også fordelene ved paneldatasett. Avsnitt 7.2 gir oversikt over aktuelle variabler og hvilke som brukes i analysen, mens avsnitt 7.3 presenterer den deskriptive statistikken.

7.1 Datasettet

Dataene som benyttes i den empiriske analysen er hentet fra et kundeservicesenter i et norsk forsikringsselskap. Datasettet er et panel, dvs. vi har flere observasjoner for hvert individ. Vårt datasett inneholder detaljert informasjon om ansattes sykefravær med ukentlige observasjoner som løper fra første kvartal i 2003 til og med andre kvartal i 2007. Siden vi ikke kan regne med å oppfatte normendringer fra uke til uke, har vi valgt å aggregere dataene til kvartalsintervaller. Vi har dermed 18 perioder med observasjoner (4,5 år). Det er totalt 5038 observasjoner med 362 ulike individ, spredt over 23 team. Datasettet er ubalansert, noe som innebærer at vi ikke observerer alle individ i alle perioder. Det kan være flere årsaker til dette, blant annet fødselspermisjon, ferier og at individ starter og slutter i bedriften på ulike tidspunkt.

7.1.1 Styrker ved datasettet

En styrke ved datasettet er at det er svært detaljert når det kommer til type sykefravær. Vi observerer legemeldt sykefravær, men også egenmeldt, gradert eller aktiv sykemelding, om den tilsatte er hjemme med sykt barn, eller om sykefraværet rett og slett er ubegrunnet. Vi kan dermed undersøke om det er slik at noen typer sykefravær er mer påvirket av sosiale interaksjonseffekter enn andre. De fleste sykefravær analysene som er utført i Norge, har kun data for det sykefraværet som overstiger 16 dager, siden det er denne informasjonen som er lettest tilgjengelig via offentlige registre²⁸. Det vil derfor være interessant å undersøke gruppeeffekter for kortere fraværstyper, særlig egenmeldt sykefravær, men også sykefravær som skyldes syke barn. Vi vil potensielt kunne få et bedre bilde av sosiale interaksjonseffekter knyttet til sykefravær. For eksempel, dersom det er slik at en eller begge disse kortere fraværstypene er mest påvirket av sosiale interaksjonseffekter, bør myndighetene vurdere hvordan disse påvirkes av sykelønnsordningen, og om det finnes forebyggende tiltak for spesielt disse typene sykefravær. Vi har tidligere påpekt at en sjenerøs sykelønnsordning og

²⁸ Sykefravær utover 16 dager dekkes av folketrygden og registreres hos NAV.

positive sosiale interaksjonseffekter kan være en dyr kombinasjon. Vi kan potensielt havne i en «dårlig likevekt» med høyt sykefravær.

En annen styrke for vår analyse er at vi har bedriftsdata. Mange andre studier ser på sosial interaksjon i nabolag, men når det gjelder sykefravær, vil vi hevde at bedriftsdata passer bedre. For det første er det mange som ikke kjenner sine naboer, og for det andre er det mange som vet lite om sine naboer sitt sykefravær. I en bedrift av den typen vi studerer, treffes man derimot hver dag og har umiddelbar tilgang til informasjon om sine kollegers sykefravær. Etersom teamene vi studerer jobber i et åpent kontorlandskap, er vi sikker på at de har god informasjon om hverandres sykefravær.

En tredje styrke ved dataene er at de ansatte jobber i relativt små og definerte team. Vi har informasjon om 23 ulike team hvor hvert team består av ca. 16 medarbeidere, og vi har full oversikt over hvilke individ som tilhører hvilke team, hvor lenge de har befunnet seg i sine respektive team og når/hvis individene flytter fra et team til et annet. For å studere gruppeeffekter er vi avhengig av å studere flere grupper samtidig, noe som er mulig her. Strukturen med full identifikasjon av alle individ som inngår i hvert team gjør det mulig å konstruere variabler som gir informasjon om egenskaper ved de ulike teamene, blant annet teamets gjennomsnittlige sykefravær, som er nøkkelvariabelen i vår analyse. For å kunne estimere sosiale interaksjonseffekter, er det viktig at vi kontrollerer for både observerbare og uobserverbare teamegenskaper.

Det er attraktivt at teamene består av relativt få individ. Både teorien og empirien viser at gruppestørrelse og relasjonsavstand er viktig for sosiale interaksjonseffekter. Teorien antar at et individ vil tilpasse seg etter andre individ som individet relaterer til, dvs. individ som inngår i dens referansegruppe. En viktig antakelse fra vårt teoretiske rammeverk er at gjennomsnittlig relasjonsavstand øker med gruppestørrelse, ceteris paribus. Dette innebærer at jo større gruppe, jo mindre påvirker individene hverandre. Dersom gruppen er veldig stor kan individ miste oversikt over andre i gruppen, og det kan danne seg flere små grupper innad i gruppen, slik at det blir meningsløst å studere gruppeeffekter i den store gruppen. Markussen and Røed (2015) finner sterkere sosiale interaksjonseffekter blant individ som er nære nettverksmedlemmer (liten relasjonsavstand).

7.1.2 Svakheter ved datasettet

Det er også noen svakheter ved datasettet. For det første har vi relativt få observasjoner sammenlignet med andre studier innenfor dette temaområdet. Vi har kun informasjon om ansatte i én bedrift, mens flere andre studier har informasjon om store deler av befolkningen. Dette kan potensielt føre til problemer med ekstern validitet. En annen svakhet er at vi har relativt få forklaringsvariabler. Vi mangler mye av den bakgrunnsinformasjonen som finnes i offentlige registre og som kan tenkes å påvirke individenes sykefravær, for eksempel utdanning, familiesituasjon, tidligere arbeids- og trygdehistorie, m.m. I kapittel 10 gir vi en grundigere gjennomgang av svakhetene ved datasettet og analysen.

7.1.3 Paneldata

Det er flere fordeler med å bruke paneldata. En fordel er at det øker antall observasjoner, og en annen er at det gir mulighet til å identifisere økonomiske fenomen via dynamiske justeringer som ikke kan identifiseres i rene tverrsnitt eller tidsserier. Den største fordelen er imidlertid at paneldata gjør det mulig å kontrollere for uobservert individuell heterogenitet. Når det gjelder sykefravær, finnes det en lang rekke forhold som må antas å ha betydning for individets sykefravær, men som ikke kan observeres i våre data; helse, familiesituasjon, ansvarsfølelse og utholdenhet, motivasjon for arbeidet, osv. Hvis dette er variabler som ikke er observerbare/ikke tilgjengelige, men påvirker adferd til hvert individ, vil resultatet være forventingsskjeve og inkonsistente estimatorer (Verbeek, 2012, s. 374). Paneldata gir oss muligheten til å korrigere for slike effekter, selv om de ikke kan inkluderes/observeres. Vi drøfter mer om dette i avsnitt 8.3.

7.2 Variabler

I dette avsnittet tar vi for oss variablene som benyttes i den empiriske analysen. Først tar vi for oss utfallsvariabler og deretter aktuelle forklaringsvariabler. Navnene i kursiv og parentes er slik variablene benevnes i tabeller og figurer.

7.2.1 Utfallsvariabler

Hovedformålet med denne oppgaven er å avdekke hvorvidt det eksisterer en gjensidig avhengighet i sykefravær mellom de ansatte på kundeservicesenteret. En naturlig utfallsvariabel er samlet sykefravær (*Sykefravær*). Denne variabelen angir antall timer sykemeldt per kvartal og består av flere ulike typer sykefravær; antall timer legemeldt sykefravær per kvartal

(*Legemeldt*), antall timer egenmelding per kvartal (*Egenmeldt*), antall timer gradert sykemelding per kvartal (*Gradert*), antall timer sykefravær pga. sykt barn per kvartal (*Syktbarn*), antall timer aktiv sykemelding per kvartal (*Aktiv*) og antall timer ubegrunnet sykefravær per kvartal (*Ubegrunnet*). Vi vil også dele opp disse variablene for å se om noen typer sykefravær er mer påvirket av sosiale interaksjonseffekter enn andre.

7.2.2 Forklaringsvariabler

I analysen benyttes både forklaringsvariabler som inneholder informasjon om individene, og forklaringsvariabler som inneholder informasjon om teamene. Som tidligere forklart, er en av styrkene ved vårt datasett nettopp at det er mulig å generere informasjon om teamene basert på kunnskap om individene som inngår, noe vi har gjort i stort omfang i denne oppgaven. I dette avsnittet vil vi gi en presentasjon av aktuelle forklaringsvariabler og presisere hvilke vi velger å benytte i analysen.

Datasettet inneholder informasjon om den tilsattes kjønn (*Mann*). Vi følger konvensjonen og representerer dette med en indikatorvariabel som tar verdier 1 dersom individet er mann og 0 dersom individet er kvinne. Vi har også informasjon om individets alder (*Alder*), og hvor lenge vedkommende har vært ansatt på tidspunkt t (*Ansiennitet*). Årsinntekt (*Lønn*) er oppgitt, men bare for året 2003, og har ikke blitt oppdatert årlig i datasettet. Vi har også informasjon om individenes produktivitet (*Salg*). Denne variabelen måler individuelt salg per kvartal.

Basert på de individuelle egenskapene benytter vi datastrukturen til å konstruere teamegenskaper i form av teamgjennomsnitt²⁹. Vi generer først og fremst en variabel for det gjennomsnittlige sykefraværet i team g ekskludert individ i per kvartal (*Team_sykefravær*). Det kan sies å være hovedforklaringsvariabelen, da vi er ute etter å identifisere hvordan et individs sykefravær blir påvirket av sykefraværet til sine teammedlemmer. Tilsvarende konstrueres gjennomsnittlig andel menn (*Team_mann*), gjennomsnittsalder (*Team_alder*), gjennomsnittslønn (*Team_lønn*), gjennomsnittsansiennitet (*Team_ansiennitet*) og gjennomsnittssalg (*Team_salg*). Til slutt har vi en variabel for antall ansatte per team per kvartal (*Team_størrelse*).

²⁹ Merk at alle teamgjennomsnittene er korrigerert for individets eget bidrag; i motsatt fall ville de vært endogene per konstruksjon.

I den deskriptive statistikken er det av interesse å presentere lønn og salgsinnsats, både på individ- og teamnivå. Både lønn og salgsinnsats kan tenkes å påvirke individuelt sykefravær og kunne derfor vært aktuelle å benytte i analysen. Vi har imidlertid ikke inkludert disse variablene i regresjonsmodellene ettersom at de potensielt er endogene grunnet simultanitet. Individuelt sykefravær vil trolig påvirke individuelt salg gjennom at individ som er borte fra jobb typisk selger mindre enn de som er på jobb. Ettersom de ansatte mottar bonus avhengig av salg av forsikringsprodukter kan det tenkes at lønn er utsatt for simultanitet. Individ som er borte fra jobb vil trolig selge mindre og får dermed mindre bonus. Dessuten er ikke lønn oppdatert årlig, vi har kun lønn fra 2003, så det kan være noe misvisende å inkludere denne som en forklaringsvariabel.

Teamets gjennomsnittlige salg kan også være utsatt for simultanitet. Ved å være færre på jobb vil trolig teamets gjennomsnittlige salg reduseres. Det kan også tenkes at andre teammedlemmer tar lengre/kortere kaffepauser når et bestemt individ er borte fra jobb slik at sykefraværet til dette individet påvirker teamets gjennomsnittlige salg. Dersom et individs sykefravær påvirker de andre ansattes produktivitet, vil det også påvirke de andre ansattes lønn ettersom lønn påvirkes av salg gjennom bonus. Dermed kan også teamets gjennomsnittlige lønn være utsatt for simultanitet. Ved å droppe å inkludere disse variablene i analysen kan vi derimot få problemer med utelatte variabler. Denne avveiningen drøftes grundigere i avsnitt 10.2.

7.3 Deskriptiv statistikk

Vi vil nå presentere deskriptiv statistikk for variablene vi har omtalt så langt. Vi starter med å se på gjennomsnittlige verdier til variablene, deretter ser vi på utviklingen i sykefravær over tid og til slutt ser vi nærmere på egenskaper ved teamene som individene tilhører.

7.3.1 Gjennomsnittlige verdier til våre variabler

I dette avsnittet presenteres gjennomsnittlige verdier til våre variabler. Gjennomsnittene er kalkulert for alle 362 individ i utvalget, over alle 18 kvartal. Tabell 2 viser gjennomsnittlige verdier til ulike sykefraværsvariabler som benyttes som utfallsvariabler i analysen. Tabell 3 viser gjennomsnittlige verdier til aktuelle forklaringsvariabler, og tabell 4 viser gjennomsnittlige verdier til enkelte variabler separat for kvinner og menn.

Tabell 2: Utfallsvariabler

Variabel	Gjennomsnitt	Standardavvik	Forklaring
Sykefravær	24.414	59.505 (50.746)	Antall timer samlet sykefravær per kvartal
Legemeldt	14.485	50.767 (44.595)	Antall timer legemeldt sykefravær per kvartal
Egenmeldt	4.628	8.924 (8.093)	Antall timer egenmeldt sykefravær per kvartal
Gradert	3.039	21.496 (19.538)	Antall timer gradert sykefravær per kvartal
Syktbarn	2.034	6.637 (5.340)	Antall timer sykefravær pga. sykt barn per kvartal
Aktiv	.195	6.373 (6.091)	Antall timer aktiv sykefravær per kvartal
Ubegrunnet	.031	1.338 (1.301)	Antall timer ubegrunnet sykefravær per kvartal
<i>Observasjoner</i>	<i>5038</i>		
<i>Antall individ</i>	<i>362</i>		

Merk: Within standardavvik i parentes.

Fra tabell 2 ser vi at gjennomsnittlig samlet sykefravær per kvartal er 24,4 timer, eller ca. 3 arbeidsdager per kvartal. Legemeldt sykefravær utgjør som ventet den største andelen av samlet sykefravær med 14,48 timer per kvartal, eller ca. 2 arbeidsdager per kvartal. Egenmeldt sykefravær utgjør i underkant av 1 arbeidsdag per kvartal av samlet sykefravær. Gradert sykefravær og sykefravær grunnet sykt barn utgjør begge ca. ½ arbeidsdag per kvartal av samlet sykefravær. Aktiv og ubegrunnet sykefravær utgjør en veldig liten andel av samlet sykefravær.

Fra tabell 2 observerer vi også at standardavvikene³⁰ er relativt store, spesielt for legemeldt og gradert sykefravær dersom vi ser på standardavvikene i termer av antall timer og ikke andeler. Store standardavvik tyder på stor variasjon både mellom individene og fra kvartal til kvartal for hvert individ (within standardavvik). Fra tabell 2 opplever vi altså at de typisk lengre fraværestypene har høyere standardavvik enn de typisk kortere fraværestypene. Noe som skiller legemeldt og gradert sykefravær fra egenmeldt sykefravær og sykefravær pga. sykt barn, er at det kreves en sykemelding fra lege/sykmelder. Det kan derfor tenkes at noe av denne variasjonen skyldes sykmelder, men det kan også tenkes at variasjonen skyldes sosiale interaksjonseffekter. Det vil derfor bli interessant å se om vi finner sterkere effekter for legemeldt og gradert sykefravær.

³⁰ Within standardavvik viser standardavvik over tid for individ *i*.

Tabell 3: Forklaringsvariabler

Variabel	Gjennomsnitt	Standardavvik	Forklaring
Mann	.465	.498 (0)	Andel menn per kvartal
Alder	31.922	7.676 (1.194)	Alder målt per kvartal
Lønn	293139	45782.86 (0)	Årsinntekt målt per kvartal
Ansiennitet	7.110	5.086 (4.647)	Lengde på ansettelse på tidspunkt t
Salg	93.728	117.883 (85.227)	Salg per kvartal
Teamvariabel			
Team_sykefravær	20.895	20.241 (16.915)	Gjennomsnittlig teamsykefravær ekskludert individ i
Team_mann	.467	.128 (.075)	Gjennomsnittlig andel menn ekskludert individ i
Team_alder	31.917	5.037 (2.258)	Gjennomsnittlig teamalder ekskludert individ i
Team_lønn	292453.5	22338.86 (12987.83)	Gjennomsnittlig teamlønn ekskludert individ i
Team_ansiennitet	7.178	3.979 (3.705)	Gjennomsnittlig teamansiennitet ekskludert individ i
Team_salg	61.795	63.218 (61.795)	Gjennomsnittlig teamsalg ekskludert individ i
Team_størrelse	16.116	6.634 (4.053)	Teamstørrelse per kvartal
Observasjoner	5038		
Antall individ	362		

Merk: Within standardavvik i parentes.

Tabell 3 viser gjennomsnittlige verdier til aktuelle forklaringsvariabler. Vi ser at det er en relativt jevn fordeling av kvinner og menn i utvalget, hvorav 46,5 prosent er menn. Gjennomsnittlig alder i utvalget er ca. 32 år og gjennomsnittlig årsinntekt er 293 139 kroner. Det gir ikke mye mening å tolke gjennomsnittet til ansiennitetsvariabelen ettersom denne er løpende, men den er aktuell å benytte i analysen fordi adferd kan tenkes å endres avhengig av hvor lenge man har vært ansatt på et hvert tidspunkt. Gjennomsnittlig salg per kvartal er ca. 94 produkter. I gjennomsnitt består hvert team av 16 individ per kvartal. Gjennomsnittlig teamsykefravær er forklaringsvariabelen av interesse ettersom vi ønsker å undersøke hvordan teamets sykefravær påvirker individ i 's sykefravær. Det er naturlig at denne størrelsen er litt mindre enn gjennomsnittsverdien til samlet sykefravær, ettersom vi har trukket ut individ i 's sykefravær i hvert team. I de resterende teamvariablene har vi også trukket ut individ i 's utfall, bortsett fra for variabelen for teamets størrelse.

Tabell 4: Kjønnforskjeller

Variabel	Kvinner		Menn	
	Gjennomsnitt	Standardavvik	Gjennomsnitt	Standardavvik
Sykefravær	30.580	68.375 (58.845)	17.328	46.308 (39.445)
Legemeldt	18.766	57.932 (51.272)	9.565	40.472 (35.407)
Egenmeldt	4.535	8.639 (7.897)	4.735	9.241 (8.314)
Gradert	4.756	26.798 (24.424)	1.065	12.675 (11.619)
Syktbarn	2.382	7.508 (6.201)	1.635	5.442 (4.137)
Aktiv	.115	3.761 (3.616)	.287	8.428 (8.046)
Ubegrunnet	.023	.807 (.784)	.039	1.762 (1.712)
Alder	32.900	8.631 (1.190)	30.795	6.209 (1.198)
Lønn	291486.7	42174.36 (13626.9)	295066.4	49555.91 (17648.67)
Salg	86.118	109.747 (79.928)	102.475	126.046 (90.955)
<i>Observasjoner</i>	<i>2693</i>		<i>2343</i>	
<i>Antall individ</i>	<i>195</i>		<i>167</i>	

Merk: Within standardavvik i parentes.

I tabell 4 presenteres gjennomsnittlige verdier til utvalgte variabler separat for kvinner og menn. Vi vil se om det er noen merkbare forskjeller i variablene for kvinner og menn. Vi ser at kvinners gjennomsnittlige samlede sykefravær tilsvarer ca. 4 dager per kvartal, mens menns gjennomsnittlige samlede sykefravær tilsvarer ca. 2 dager per kvartal. Kvinner har altså nesten dobbelt så høyt samlet sykefravær per kvartal som menn. Vi ser at det som først og fremst er utslagsgivende, er at kvinner har nesten dobbelt så høyt legemeldt sykefravær som menn. De har også høyere gradert sykefravær, og litt høyere sykefravær grunnet sykt barn. Egenmeldt sykefravær er omtrent det samme i antall timer, noe som innebærer at menn har større *andel* egenmeldt sykefravær enn kvinner (vel 27 prosent for menn mot knappe 15 prosent for kvinner). Alders- og inntektsfordelingen mellom kvinner og menn er relativt lik. Videre observerer vi at menn har noe høyere gjennomsnittlig salg enn kvinner, noe som virker naturlig ettersom kvinner har høyere gjennomsnittlig samlet sykefravær.

Det at kvinner har høyere legemeldt sykefravær enn menn i vårt utvalg er i tråd med hva som er gjeldende for resten av den norske befolkningen, se figur 1 og 2. På landsbasis er det muskel- og skjelettplager og psykiske lidelser som utgjør den største andelen av det legemeldte sykefraværet, og disse diagnosene er mest utbredt blant kvinner. Det kan tenkes at det er tilfellet

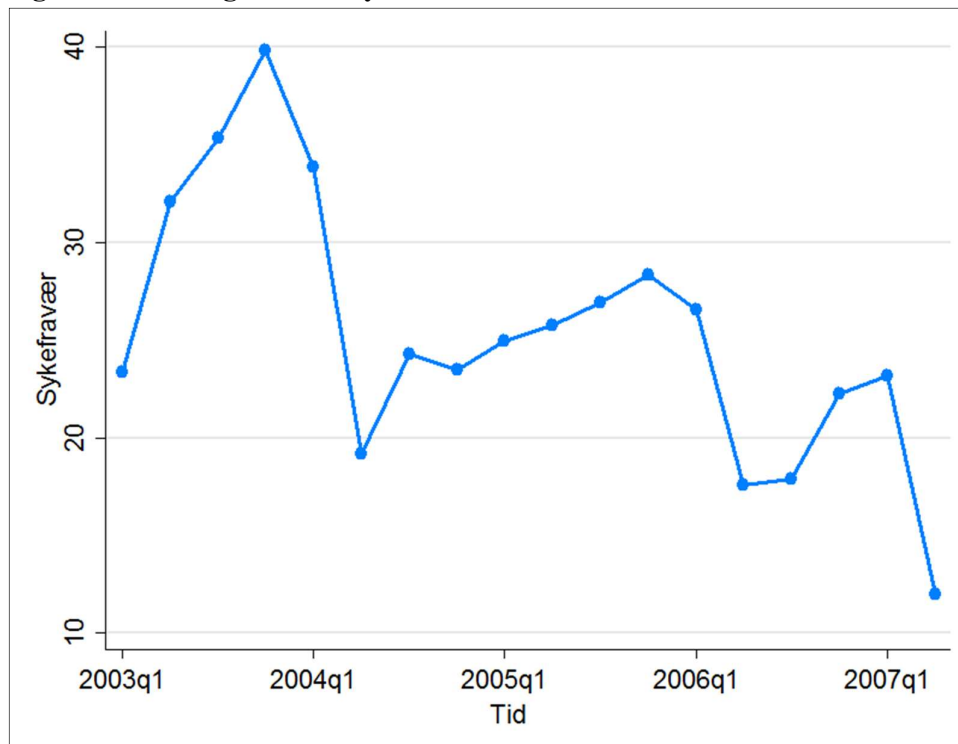
i våre data også, men vi har ikke data om diagnoser for å verifisere dette. Vi drøftet i kapittel 1 at dette potensielt er diagnoser som brukes for å kamuflere forhold som ikke gir rett til sykepenger, og at kjønnsforskjeller i sykefravær i stor grad er uforklart. Det kan derfor tenkes at vi vil finne sterkere sosiale interaksjonseffekter blant kvinner enn hos menn.

Det finnes også andre argumenter for at vi vil finne sterkere sosiale interaksjonseffekter blant kvinner enn menn. Vi påpekte i avsnitt 2.3 at kvinner i gjennomsnitt har svakere tilknytning til arbeidslivet enn menn, og at kvinner bruker mer tid til husholdsarbeid enn menn. Begge deler kan bidra til lavere terskel for å være borte fra jobb, både ved egen og ved barns sykdom. Fra våre data kan vi ikke se at kvinner har merkbart høyere sykefravær grunnet omsorg for barn, men det kan likevel tenkes at de tar ut sine egne fraværsdager for å være hjemme med syke barn.

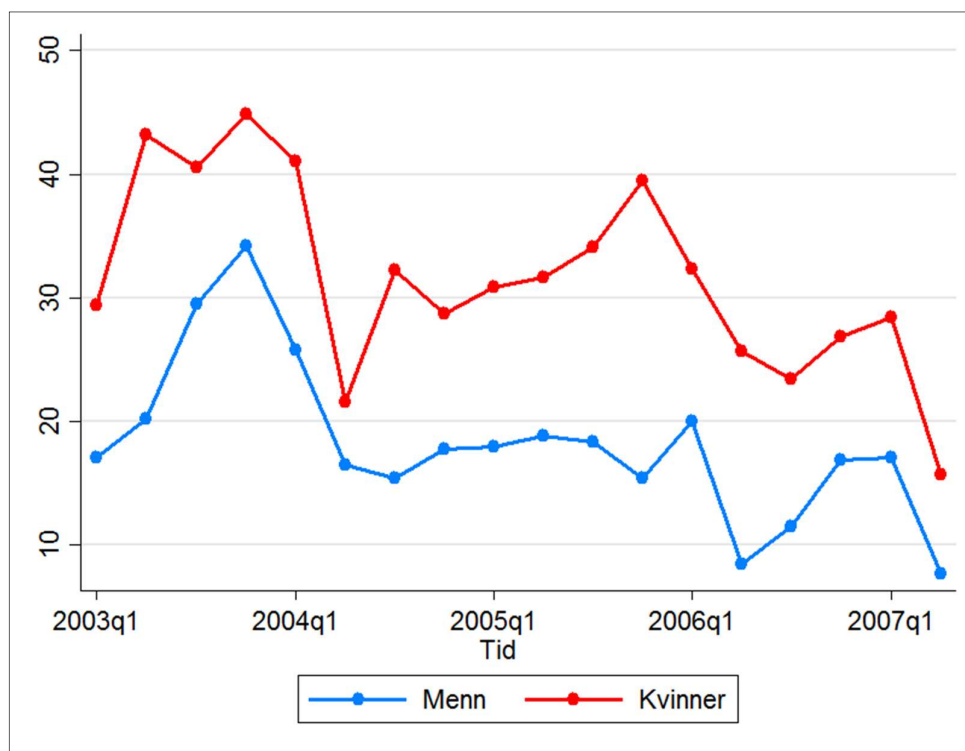
Vi gjør oppmerksom på at tilsvarende statistikk er presentert i appendiks, tabell A.1-A.4, for nyansatte og for ansatte som har flyttet mellom team. Se avsnitt 8.4 for definisjon av nyansatte og flyttere, samt kommentarer til denne statistikken.

7.3.2 Utvikling i sykefravær over tid

Vi vil nå se nærmere på hvordan enkelte variabler har utviklet seg under perioden vi studerer, dvs. fra første kvartal i 2003 til og med andre kvartal i 2007. Ved å se deskriptivt på hvordan variablene har utviklet seg over tid vil vi få et inntrykk om det er noe spesielt som har skjedd i løpet av perioden vi studerer, og om det eventuelt er noe vi må ta hensyn til i analysen. For eksempel, om ett år skiller seg veldig fra andre kan det være nyttig å prøve å finne årsak eller eventuelt kontrollere for dette. Vi ser først på utviklingen i samlet sykefravær for hele utvalget, og deretter separat for kvinner og menn. Alle figurene i dette kapittelet, samt appendiks, presenterer gjennomsnittlige verdier målt per kvartal.

Figur 7: Utvikling i samlet sykefravær over tid

Fra figur 7 observerer vi at samlet sykefravær, målt i timer per kvartal, er høyere i begynnelsen av 2000-tallet enn resten av perioden. Den bratte reduksjonen i sykefraværet i første halvdel av 2004 samsvarer med aggregert sykefraværstatistikk, og skyldes trolig «endring i sykmelders rolle». Dette var en reform som ble implementert våren 2004, som også gjaldt for ikke IA-virksomheter. Fastleger og sykmeldere ble oppfordret til å øke bruken av graderte sykemeldinger fremfor 100 prosent sykemeldinger. Figur A.1 i appendiks viser utviklingen i ulike typer sykefravær separat. Her ser vi at det er nettopp det legemeldte sykefraværet som har falt betraktelig, samtidig som gradert sykefravær øker. De andre fraværstypene utgjør en relativt liten andel av samlet sykefravær. Vi ser at det er en del svingninger i disse, men de er relativt små. Fra figur 7 legger vi også merke til at sykefraværet faller i begynnelsen av 2007. Dette kan tenkes å være relatert til finanskrisen. Som nevnt i kapittel 1 finner flere studier at sykefravær varierer med konjunktursituasjon, og at sykefraværet typisk er lavt under en lavkonjunktur.

Figur 8: Utvikling i samlet sykefravær for kvinner og menn

Figur 8 viser utviklingen i samlet sykefravær, målt i timer per kvartal, separat for kvinner og menn. Som ventet, på bakgrunn av deskriptiv statistikk for hele landet, ser vi at kvinner har høyere sykefravær enn menn under hele perioden vi studerer. Svingninger i sykefravær er derimot relativt lik for kvinner og menn. Figur A.2 i appendiks viser utviklingen i ulike typer sykefravær separat for kvinner og menn. Vi ser at kvinners legemeldte sykefravær ligger stabilt over menns legemeldte sykefravær. Når det gjelder egenmeldt sykefravær ser vi derimot at disse ligger ganske nært for kvinner og menn, og at kurvene krysser hverandre noen ganger. Enkelte tilfeller har altså menn høyere egenmeldt sykefravær enn kvinner. Vi legger også merke til at kvinner har noe høyere gradert sykefravær enn menn, samt litt høyere sykefravær grunnet sykt barn under stort sett hele perioden.

Se figur A.3 i appendiks for hvordan sykefravær varierer etter alder. Forskjellen i sykefravær etter aldersgruppe i bedriften vi studerer ser ut til å følge den på landsbasis, se figur 3. Eldre har høyere sykefravær enn yngre, bortsett fra de aller eldste som er veldig lite syke.

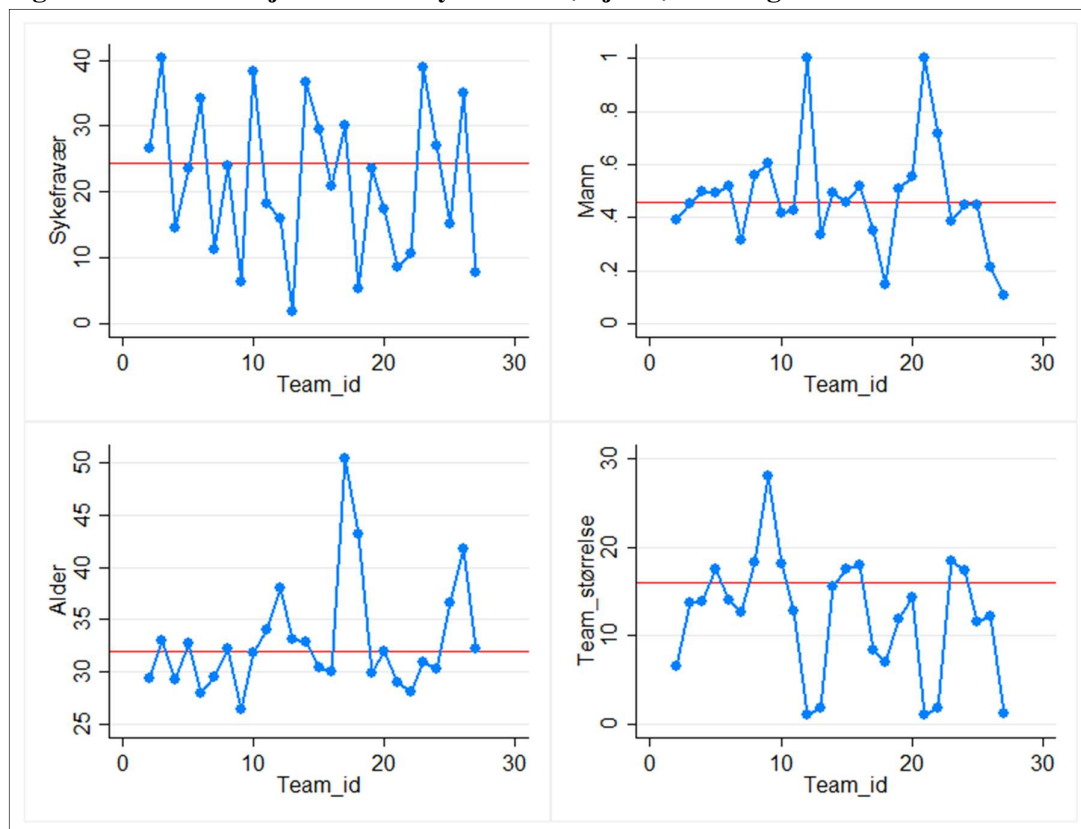
7.3.3 Variasjon mellom teamene

I dette avsnittet vil vi se om det er noen målbare forskjeller mellom de ulike teamene når det

gjelder samlet sykefravær, kjønn, alder og teamstørrelse. Fra avsnitt 3.2 så vi at det kan oppstå variasjon i sykefravær mellom ulike grupper selv om gruppene er ex ante identiske grunnet sosial interaksjon. Det kan føre til multiple likevekter, der noen typisk havner i en «god likevekt» med lavt sykefravær og noen havner i en «dårlig likevekt» med høyt sykefravær. Sosiale normer knyttet til sykefraværsadferd antas å spille en viktig rolle for hvilken likevekt man havner i gjennom stigma. I dette avsnittet vil vi derfor prøve å få en forståelse av hvor like de ulike teamene vi studerer er. Dersom teamene er ex ante veldig ulike vil det være mer naturlig at det er variasjon i sykefravær mellom teamene.

Vi vil se nærmere på hvordan kjønns- og aldersfordelingen ser ut mellom de ulike teamene ettersom vi har sett rent deskriptivt at det er variasjon i sykefravær avhengig av både kjønn og alder. Vi vil også se om det er stor variasjon i teamstørrelse mellom de ulike teamene. Fra vårt teoretiske rammeverk har vi antatt at det er sterkere sosiale interaksjonseffekter i mindre grupper. Ideen er at individ vil ha vanskeligheter med å indentifisere seg med andre i gruppen når gruppen blir stor.

Figur 9: Teamvariasjon i samlet sykefravær, kjønn, alder og teamstørrelse



Figur 9 viser teamvariasjon i samlet sykefravær, kjønn, alder og teamstørrelse. Den røde linjen viser gjennomsnittet for hele utvalget over alle perioder. Vi observerer først og fremst at det er variasjon i samlet sykefravær mellom teamene, noe som gjør det interessant å studere gruppeeffekter for sykefravær i dette datasettet. Vi ser at noen team har et gjennomsnittlig samlet sykefravær på rundt 0-10 timer per kvartal, mens andre team har et gjennomsnittlig samlet sykefravær på rundt 30-40 timer per kvartal. Det er altså ganske stor variasjon i samlet sykefravær mellom teamene.

Videre ser vi at de fleste teamene har en relativt jevn fordeling av kvinner og menn. Vi ser derimot at det er 4 team som skiller seg, der to av dem består kun av menn og to som har majoriteten kvinner. Vi ser også at det er relativt jevn aldersfordeling mellom teamene, men at noen få team skiller seg ut ved å ha en høy andel eldre. Når det gjelder teamstørrelse ser vi at mesteparten av teamene sirkulerer rundt gjennomsnittet som er 16 medlemmer, men at enkelte team har færre medlemmer. Alle teamene består altså av relativt få medlemmer, slik at det er enkelt for ansatte å ha god oversikt over alle sine medarbeidere.

Basert på disse figurene opplever vi at teamene er relativt like, slik at mye av variasjonen i sykefravær potensielt kan forklares av sosiale interaksjonseffekter. Det som måtte være av forskjeller, blir for øvrig kontrollert for i form av egne teamvariabler (kjønn, alder og størrelse). Fra figur 9 kan det dermed se ut som at noen team er i en «god likevekt», mens andre team er i en «dårlig likevekt».

8 Empiriske modellspesifikasjoner

I dette kapitlet presenterer vi to modellspesifikasjoner som benyttes i den empiriske analysen. Vi benytter to modeller som adresserer mange av de metodiske utfordringene nevnt i avsnitt 5.2. Vi benevner disse som henholdsvis nyansatt-modellen³¹ og flytte-modellen³². Hensikten med modellene er forklare hvordan utfallet/sykefraværet til et individ påvirkes av det gjennomsnittlige sykefraværet til de andre teammedlemmene. Ved å benytte disse modellene forsøker vi å isolere den kausale effekten av teamets sykefravær på individ i 's sykefravær.

I avsnitt 8.1 presenteres nyansatt-modellen og i avsnitt 8.2 presenteres flytte-modellen. I avsnitt 8.3 gjør vi greie for forutsetningene i fasteffekt modellen og hva som kreves for konsistent estimering. For å estimere disse modellene gjennomfører vi noen begrensinger i datasettet, disse spesifiseres i avsnitt 8.4

8.1 Nyansatt-modellen

Den første strategien vi benytter for å identifisere sosiale interaksjonseffekter er å studere individ som er nyansatte i kundeservicesenteret. Vi velger å begrense analysen til å studere dem det første halvåret (to kvartal) de jobber i kundeservicesenteret. Ved estimering bruker vi følgende lineære gjennomsnittsmoell:

$$y_{i,g,t} = \alpha_i + \theta_1 x_{i,g,t} + \theta_2 \bar{y}_{-i,g,t} + \theta_3 \bar{x}_{-i,g,t} + \theta_4 W_{g,t} + \alpha_g + \mu_t + \varepsilon_{i,g,t} \quad (8.1)$$

Her er $y_{i,g,t}$ sykefraværet til et nyansatt individ i som tilhører team g på tidspunkt t . $x_{i,g,t}$ er en vektor av eksogene variabler som kan tenkes å påvirke individuelt sykefravær og $\bar{x}_{-i,g,t}$ er en vektor av observerbare egenskaper til de andre teammedlemmene. $W_{g,t}$ er en vektor av observerbare lokale attributter (dvs. *teamspesifikke variabler* som for eksempel teamets størrelse og kvalitet på teamlederen, og *områdevariabler* som for eksempel arbeidsledighetsraten). Parameteren α_i er en uobservert individspesifikk komponent som er konstant over tid og kontrollerer for uobservert individuell heterogenitet. Tilsvarende er parameteren α_g en uobservert teamspesifikk komponent som er konstant over tid og kontrollerer for uobservert team heterogenitet. Parameteren μ_t er en tidsdummy (tidsfasteffekt)

³¹ Denne modellen er også brukt i Rieck og Vaage (2010), og minner om immigrant-modellen brukt i Lindbeck et al. (2016).

³² Denne modellen er også brukt i Ichino og Maggi (2000), Rieck og Vaage (2010) og Lindbeck et al. (2016).

som kontrollerer for kvartalseffekter, og $\varepsilon_{i,g,t}$ er et stokastisk feilledd. Variabelen $\bar{y}_{-i,g,t}$ angir gjennomsnittlig sykefravær til de andre teammedlemmene. Parameteren θ_2 er det vi kaller for endogen sosial interaksjon og er den som er av størst interesse i denne analysen. Parameteren θ_2 angir effekten av en marginal økning i gjennomsnittlig teamsykefravær på individuelt sykefravær. θ_2 kan være både negativ og positiv, eller 0. Dersom gjennomsnittlig sykefravær i teamet øker, kan det tenkes at sosiale normer vil gjøre det lettere for individ å være borte fra jobb og θ_2 vil da være positiv. θ_2 kan tenkes å være negativ dersom individ ønsker å ta avstand fra teamet (skille seg ut), selv om vi anser en slik effekt å være mindre sannsynlig enn positiv. Ved fravær av gjensidig avhengighet i sykefravær vil θ_2 være null.

Som nevnt i avsnitt 5.2 er det en del metodiske utfordringer ved identifikasjon av sosiale interaksjonseffekter. Ved å studere nyansatte kan det tenkes at vi løser problemet med simultanitet i modellen siden disse individene er «ferske» og ikke har blitt eksponert for gruppeeffekter enda, men det kan også tenkes at de nyansatte påvirker teamet det første halvåret vi bruker som observasjonsperiode. For å være sikker på at vi unngår problemet med simultanitet/refleksjon, bruker vi derfor lagget verdi av gjennomsnittlig sykefravær til de andre teammedlemmene, $\bar{y}_{-i,g,t-2}$, som instrument for gjennomsnittlig sykefravær til de andre teammedlemmene, $\bar{y}_{-i,g,t}$. Merk at ettersom vi bruker kvartalsdata og observerer de nyansatte i et halvt år, må vi lagge to perioder slik at vi bruker en verdi fra før individet startet i aktuelt team. Dette instrumentet er attraktivt å bruke fordi det trolig er korrelert med den endogene forklaringsvariabelen, $\bar{y}_{-i,g,t}$, samtidig som den i seg selv ikke er relevant for å forklare variasjon i utfallsvariabelen, $y_{i,g,t}$. For å redusere problemet med korrelerte effekter bruker vi fasteffekt modellen ved estimering. Vi legger til grunn at individene (de nyansatte) plasseres tilfeldig i team, slik at seleksjon ikke er et tema.

8.2 Flytte-modellen

Den andre strategien vi benytter for å identifisere sosiale interaksjonseffekter er å studere individ som flytter mellom team. Vi velger å observere flytterne et halvt år nytt team og et halvt år i gammelt team. Modellen estimeres ved å ta førstedifferansen mellom utfall i nytt team og utfall i gammelt team. Vi studerer ansatte som skifter fra team m på tidspunkt $t-1$ til team n på tidspunkt t . Vi tar utgangspunkt i notasjon fra modell (8.1) og får nå følgende modell:

$$y_{i,n,t} - y_{i,m,t-1} = \theta_1(x_{i,n,t} - x_{i,m,t-1}) + \theta_2(\bar{y}_{-i,n,t} - \bar{y}_{-i,m,t-1}) + \theta_3(\bar{x}_{-i,n,t} - \bar{x}_{-i,m,t-1}) + \theta_4(W_{n,t} - W_{m,t-1}) + \alpha_g + \mu_t + (\varepsilon_{i,n,t} - \varepsilon_{i,m,t-1}) \quad (8.2)$$

Ved å se på individ som bytter team, reduserer vi problemet med simultanitet i vår modell. Siden disse individene er nye i teamet på tidspunkt t , ble de ikke utsatt for gruppeeffekter i dette teamet på tidspunkt $t-1$. I likhet med nyansatt-modellen kan det likevel tenkes at sykefraværet til flytterne påvirker sykefraværet til teammedlemmene i løpet av vår observasjonsperiode på et halvt år. Igjen, for å være sikker på at vi unngår problemet med simultanitet, instrumenterer vi $\bar{y}_{-i,n,t}$ med sin laggede verdi, $\bar{y}_{-i,n,t-2}$. Vi forholder oss til problemet med korrelerte effekter ved å ta førstedifferansen, ettersom det er ekvivalent med en standard fasteffekt modell, som innebærer at den kontrollerer for uobservert individuell heterogenitet som er konstant over tid. Når det gjelder seleksjon i flytte-modellen, antas det at bytting mellom team skyldes tilfeldigheter fremfor systematisk seleksjon basert på observerbare og ikke-observerbare individ- og teamegenskaper.

8.3 Forutsetninger i fasteffekt modellen og krav til konsistens

Begge modellspesifikasjonene, (8.1) og (8.2), estimerer en fasteffekt modell, dvs. en modell som tar hensyn til uobservert individuell heterogenitet ved å inkludere en uobservert individspesifikk komponent. Nyansatt-modellen estimeres i utgangspunktet med within transformasjon og flytte-modellen estimeres i utgangspunktet med førstedifferanse transformasjon. Ettersom vi inkluderer et instrument i begge modellene estimeres nyansatt-modellen med 2SLS within estimatoren og flytte-modellen med 2SLS førstedifferanse estimatoren. I dette avsnittet vil vi ta for oss forutsetningene som må være oppfylt i fasteffekt modellen, hva som kreves for konsistent estimering, samt hva implementering av 2SLS (two-stage least squares) innebærer for våre estimatorene.

Fasteffekt modellen kontrollerer for tidskonstant individuell heterogenitet ved å inkludere individdummyer i modellen. Ved å inkludere teamdummyer i våre modeller kontrollerer vi på samme måte for tidskonstant uobservert team heterogenitet, men vi drøfter bare individuell heterogenitet her. En årsak til heterogenitet er utelatte variabler som er konstante over tid. Disse utelatte variablene er ikke observerbare eller ikke tilgjengelige, men påvirker individuell adferd. I vårt tilfelle kan dette tenkes å være en konstant helsevariabel. Et individs «opprinnelige» helsetilstand kan tenkes å være konstant over tid, noen individ er født friskere eller med bedre

gener enn andre. Vi får forventingskjevne og inkonsistente estimat dersom det ikke tas hensyn til. Fasteffekt modellen tar hensyn til dette, men filtrerer bort alt som ikke varierer over tid, noe som gjør at vi ikke kan identifisere tidskonstante parametre. Utgangspunktet for fasteffekt estimering er følgende modell, der vi har $i=1, \dots, N$ individ og $t=1, \dots, T$ tidsperioder:

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it} \quad (8.3)$$

I fasteffekt modellen er det 4 forutsetninger som må være oppfylt:

1. $\alpha_i \neq \alpha_j$. Dette innebærer at individene i modellen har forskjellige konstantledd, eller forskjellige uobserverbare individspesifikke komponenter. Dersom ($\alpha_1 = \alpha_2 \dots = \alpha_N$) eksisterer det ikke individuell heterogenitet, og vanlig OLS er den mest effisiente måten å estimere panelparametre.
2. α må være konstant over tid. Fasteffekt transformasjon eliminerer de uobserverte individ spesifikke komponentene ($\alpha_1, \dots, \alpha_N$) så lenge disse er tidskonstante. Dersom disse ikke er konstant over tid er det ikke mulig å estimere fasteffekt modellen.
3. $\beta_i = \beta_j$. Dette innebærer at individene har samme helningsparametre. Dersom individene har ulike helningsparametre, kreves det paneldatamodeller som er utenfor fokus for denne oppgaven.
4. Feilledet ε_{it} er uavhengig og identisk fordelt (i.i.d), med forventning lik 0 og varians lik σ^2 .

Within estimatoren og førstedifferanse estimatoren er gitt ved henholdsvis (8.4) og (8.5):

$$\widehat{\beta}_{FE} = \frac{\sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y}_i)(x_{it} - \bar{x}_i)}{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'} \quad (8.4)$$

$$\widehat{\beta}_{FD} = \frac{\sum_{i=1}^N \sum_{t=2}^T (y_{it} - y_{i,t-1})(x_{it} - x_{i,t-1})}{\sum_{i=1}^N \sum_{t=2}^T (x_{it} - x_{i,t-1})(x_{it} - x_{i,t-1})'} \quad (8.5)$$

I (8.4) er \bar{y}_i og \bar{x}_i gjennomsnitt over tid for individ i , også kalt innom gruppegjennomsnitt. Within estimatoren er konsistent når $(x_{it} - \bar{x}_i)$ er ukorrelert med $(\varepsilon_{it} - \varepsilon_i)$. Dette er oppfylt

under antakelse om streng eksogenitet³³, noe som innebærer at variabelen ikke avhenger av nåværende, fremtidige eller tidligere verdier av feilleddet. Within estimatoren vil være konsistent når N og/eller $T \rightarrow \infty$. Førstedifferanse estimatoren er konsistent når $(x_{it} - x_{i,t-t})$ er ukorrelert med $(\varepsilon_{it} - \varepsilon_{i,t-1})$. Dette er oppfylt under antakelse om svak eksogenitet³⁴. Denne betingelsen for konsistens er svakere enn for within estimatoren. Den tillater for eksempel korrelasjon mellom x_{it} og $\varepsilon_{i,t-2}$. Altså, den tillater fremtidige verdier av forklaringsvariabelen å være korrelert med feilleddet. Førstedifferanse estimatoren er derimot typisk mindre effisient enn within estimatoren så lenge ε_{it} er i.i.d. For $T=2$ er de to estimatorene identiske (Verbeek, 2012, s. 384), som er tilfellet når vi observerer individene i to kvartal. Vi kan derimot forvente å få ulike utfall i våre modeller siden vi observerer to ulike grupper individ – nyansatte versus flyttere.

Siden vi har en endogen forklaringsvariabel i begge modellspesifikasjonene, benytter vi instrumentvariabelmetoden. Dette gjøres ved å implementere 2SLS i within modellen og førstedifferanse modellen. Vi inkluderer et instrument som er korrelert med den endogene variabelen samtidig som den ikke er korrelert med feilleddet i modellen. Instrumentvariabelen skal heller ikke være relevant i seg selv for å forklare variasjon i utfallsvariabelen. Det er sjelden lett å finne passende instrumentvariabler. 2SLS er en to-steps prosedyre der det første steget er å estimere den reduserte formen av modellen. Det andre steget er å estimere opprinnelig modell der vi erstatter den endogene variabelen med predikerte verdier fra den reduserte formen. Den endogene variabelen blir altså erstattet med en lineær kombinasjon av de eksogene variablene³⁵.

For konsistent estimering gjelder samme betingelser som før, bortsett fra at instrumentvariabelen nå inngår som en forklaringsvariabel. For within modellen har vi konsistent estimering når $E(\varepsilon_{it} | \alpha_i, z_{i1}, \dots, z_{iT}) = 0$. Her er z_{it} er en vektor av alle eksogene variabler og instrument. Vi har dermed konsistent estimering når eksogene variabler og instrumentvariabel er ukorrelert med feilleddet i alle perioder. Vi antar sterk eksogenitet. For førstedifferanse modellen har vi konsistent estimering når $E(\varepsilon_{it} | \alpha_i, z_{i1}, \dots, z_{it}) = 0$. Det vil si, under antakelse om svak eksogenitet (Cameron et al. 2010, s. 287).

³³ Streng eksogenitet impliserer at $E(\varepsilon_{it} | \alpha_i, x_{i1}, \dots, x_{iT}) = 0$.

³⁴ Svak eksogenitet impliserer at $E(\varepsilon_{it} | \alpha_i, x_{i1}, \dots, x_{it}) = 0$.

³⁵ Se Verbeek (2012) s. 162 for mer om implementering av 2SLS.

Konsistent estimering krever dessuten relativt sterke instrument. For at et instrument skal være gyldig må (1) forventingen mellom instrument og feilledd være lik 0, og (2) forventingen mellom instrument og den endogene variabelen være ulik 0. Vi kan og bør teste (2), dvs. instrument relevans. For å teste instrument relevans estimerer vi førstesteget i 2SLS og regner ut en F-test for om instrument har forklaringskraft i den reduserte formen. En tommelfingerregel er at dersom $F > 10$ har vi et gyldig instrument. Hvis forventingen mellom instrument og den endogene variabelen er liten har vi et svakt instrument. I dette tilfellet vil estimatoren være forventingsskjev (Woolridge, 2014, s. 406).

8.4 Begrensninger i datasettet

I dette avsnittet vil vi ta for oss begrensningene som er gjort i datasettet for å estimere nyansatt- og flytte-modellen. I begge modellene observerer vi kun en begrenset del av utvalget over en begrenset tidsperiode. Dette gjøres i forsøk på å isolere den endogene sosiale interaksjonseffekten. Vi har valgt en tidsbegrensning på et halvt år ettersom vi tror at sosiale interaksjonseffekter vil spille en rolle da, men vi kunne alternativt ha valgt en annen observasjonsperiode. I avsnitt 9.4 undersøker vi om effektene endrer seg ved å benytte andre observasjonsperioder.

8.4.1 Begrensninger i nyansatt-modellen

I nyansatt-modellen observerer vi kun nyansatte. Vi har definert nyansatte som individ som har en ansettelsesdato tidligst tre kvartal etter første observasjon i datasettet, dvs. individ som har en ansettelsesdato senere eller lik tredje kvartal i 2003. Videre begrenser vi også denne analysen til å studere disse individene i et halvt år. Ved å innføre en slik restriksjon forsvinner individ som har ansettelsesdato senere enn siste kvartal i 2006. Dette er fordi at vi skal kunne observere dem i et halvt år. Vi blir da stående igjen med 159 individ og 2 observasjoner per individ, hvorav 86 er kvinner og 73 er menn. Totalt har vi 318 observasjoner i nyansatt-modellen.

Se tabell A.1 og A.2 i appendiks for deskriptiv statistikk for nyansatte. I disse tabellene inkluderes alle som har startdato senere eller lik tredje kvartal i 2003. Det er ingen større forskjeller når vi sammenligner denne med den deskriptive statistikken for hele utvalget, men vi observerer at de nyansatte har noe lavere gjennomsnittsverdier for sykefraværsvARIABLENE.

8.4.2 Begrensinger i flytte-modellen

I flytte-modellen observerer vi kun individ som på et eller annet tidspunkt bytter team. Vi har begrenset analysen til å kun studere individ som bytter team én gang. Dette tilsvarer ca. 35 prosent av utvalget. Videre har vi begrenset analysen til å studere disse individene i et halvt år. Ved å bruke en observasjonsperiode på et halvt år begrenser vi videre analysen til å kun inkludere individ som har vært minimum et halvt år i både gammelt og nytt team. Til slutt står vi igjen med 92 individ og 2 observasjoner per individ, hvorav 53 er kvinner og 39 er menn. Vi har totalt 184 observasjoner i flytte-modellen.

Se tabell A.3 og A.4 i appendiks for deskriptiv statistikk for flytterne. I disse tabellene inkluderes alle som har flyttet én gang. Den deskriptive statistikken for flytterne er temmelig lik den deskriptive statistikken for hele utvalget. Det er ingen merkverdige forskjeller i de kalkulererte gjennomsnittene.

9 Empirisk analyse

Formålet med denne oppgaven er å undersøke om et individs sykefravær påvirkes av sine kollegers sykefravær gjennom sosial interaksjon. I kapittel 5 presenterte vi en generell modell for estimering av sosiale interaksjonseffekter, metodiske utfordringer ved estimering samt ulike identifikasjonsstrategier man kan benytte i empiriske arbeid. I vårt tilfelle er det først og fremst to utfordringer knyttet til estimering av sosiale interaksjonseffekter; korrelerte effekter og simultanitet. I kapittel 8 presenterte vi to modeller, nyansatt- og flytte-modellen, som adresserer disse metodiske utfordringene.

I dette kapittelet vil vi presentere resultatene fra disse to modellene, men vi starter med å estimere den generelle modellen med vanlig OLS for hele utvalget. Som nevnt i avsnitt 5.1, vil estimering av den generelle modellen med OLS gi inkonsistente estimat. Modellen klarer ikke å skille mellom endogene, eksogene og korrelerte effekter, men den gir likevel et nyttig bilde av «bruttoeffekten». Det vil si, den gir oss et inntrykk om individene ser ut til å tilpasse seg kollegenes sykefravær, men uten at vi kan tolke dette som en endogen effekt. I avsnitt 9.2 og 9.3 presenteres resultat fra henholdsvis nyansatt- og flytte-modellen, hvor vi forsøker å isolere den endogene sosiale interaksjonseffekten. Til slutt studerer vi heterogenitet over kjønn og tid i avsnitt 9.4.

9.1 Resultat fra den generelle modellen

I dette avsnittet presenterer vi resultatene fra den generelle modellen (5.1). Vi estimerer først modellen med vanlig OLS, der vi inkluderer alle relevante forklaringsvariabler. Vi benytter samlet sykefravær som utfallsvariabel, og studerer hele utvalget over alle 18 perioder. Vi har 362 individ og totalt 4997 observasjoner. Deretter estimerer vi den generelle modellen med en fasteffekt (FE) modell, der vi i tillegg inkluderer fasteffekt komponenter (individ-, team-, og tidsdummyer). Dette gjør at vi kan se hvordan utfallet endres når vi kontrollerer for uobserverte effekter. Individ- og teamdummyer skal kontrollere for uobservert tidskonstant heterogenitet mellom individ og teamene, mens tidsdummyer skal kontrollere for eventuelle trender og sesongvariasjoner i dataene. Tidsdummyene er generert ved å inkludere en dummyvariabel for hvert kvartal. Vi benytter cluster-robuste standardfeil³⁶ i alle regresjonene.

³⁶ Vi bruker cluster-robuste standardfeil ettersom modellen har heteroskedastiske feilledd. Dette har vi undersøkt i Stata gjennom å utføre en såkalt «hettest». Se Verbeek (2012, s. 105) for mer om testing av heteroskedastisitet.

Tabell 5: Resultat fra den generelle modellen

Sykefravær	(1) OLS	(2) FE
Team_sykefravær	0.427*** [.061] (0.000)	.260*** [.061] (0.000)
Mann	-12.61*** [3.303] (0.000)	0 (omitted)
Alder	0.234 [.349] (0.503)	-.012 [2.460] (0.996)
Ansiennitet	0.311 [.422] (0.463)	-1.427 [1.290] (0.269)
Team_mann	-5.579 [10.556] (0.597)	16.563 [12.820] (0.197)
Team_alder	-0.478 [.526] (0.364)	1.497 [1.325] (0.259)
Team_ansiennitet	-0.356 [.547] (0.516)	.346 [.309] (0.263)
Team_størrelse	-0.268 [.214] (0.211)	.346 [.321] (0.281)
<i>Individdummyer</i>	<i>Nei</i>	<i>Ja</i>
<i>Teamdumymer</i>	<i>Nei</i>	<i>Ja</i>
<i>Tidsdummyer</i>	<i>Nei</i>	<i>Ja</i>
<i>Observasjoner</i>	4997	4997
<i>Antall individ</i>	362	362
<i>R²</i>	0.038	0.023
<i>Samlet F-test</i>	9.58	-

*Merknader: Cluster-robuste standardfeil i klammeparentes og p-verdi i parentes. Asterisker angir signifikans: *p<0.1, **p<0.05, ***p<0.01.*

Resultat fra den generelle modellen er presentert i tabell 5. Når vi estimerer modellen med OLS ser vi at gjennomsnittlig teamsykefravær har en klart signifikant effekt på individenes sykefravær, med en koeffisient på 0,427. Teammedlemmenes sykefravær ser altså ut til å ha en positiv «bruttoeffekt» på individets sykefravær. Vi kan derimot ikke si om effekten skyldes endogene, eksogene eller korrelerte effekter. Når vi estimerer modellen med en FE modell og inkluderer fasteffekt komponenter, ser vi at denne koeffisienten faller til 0,26. Dette er en relativt stor endring som viser viktigheten av å kontrollere for uobserverte effekter. Ved å inkludere fasteffekt komponenter forholder vi oss til problemet med korrelerte effekter i modellen, men vi kan fortsatt ikke skille mellom endogene og eksogene effekter. Koeffisienten er statistisk signifikant på et svært lavt signifikansnivå, og tyder fortsatt på at det finnes en positiv «bruttoeffekt» i våre data. Det kan dermed se ut som at sosiale normer spiller en rolle for individuelt sykefravær i denne bedriften.

Det er også interessant å se nærmere på koeffisientene til de andre forklaringsvariablene. Når vi estimerer modellen med OLS ser vi at koeffisienten til kjønn (*Mann*) er negativ og statistisk signifikant på et svært lavt nivå. En koeffisient på -12,61 betyr at menn i gjennomsnitt har 12,61 timer lavere sykefravær enn kvinner per kvartal, noe som stemmer overens med den deskriptive statistikken i tabell 4. Denne variabelen faller ut i FE modellen siden den er konstant over tid.

Fra tabell 5 ser vi at koeffisienten til alder er positiv når vi estimerer modellen med OLS og negativ med FE modellen. Den er derimot ikke statistisk signifikant i noen av tilfellene. Det virker fornuftig at denne koeffisienten er positiv ettersom vi har sett rent deskriptivt at de som er eldre typisk har høyere sykefravær. Koeffisienten er riktignok ikke statistisk signifikant, men det skyldes trolig få observasjoner og derav lite variasjon i data. Variabelen «overlever» i FE modellen, men uttrykker nå bare variasjon over tid, ikke mellom individ. Det er derfor som ventet at den blir enda mindre signifikant. At den i tillegg får negativt fortegn gir lite mening, men bør ikke tilskrives noen vekt når den er så langt unna å være signifikant.

Koeffisienten til ansiennitet er positiv i OLS, men negativ i FE modellen. Det er ikke opplagt hvilken retning denne variabelen vil dra. Et argument for at koeffisienten er positiv er at individ som er relativt ny på arbeidsplassen (lav ansiennitet) vil vise seg frem for arbeidsgiver ved å stille opp på arbeidsplassen, selv om det innebærer å gå på jobb om man er syk. Jo lengre man har vært ansatt (høy ansiennitet), jo mer trygg føler man seg ovenfor arbeidsgiver slik at man lettere kan sykemelde seg. Et argument for at koeffisienten er negativ er at individ som har vært ansatt lenge har mer ansvar slik at det blir vanskeligere å være borte fra jobb. Ettersom de har mer krevende arbeidsoppgaver, og kanskje andre på arbeidsplassen som er avhengig av dem for at ting skal fungere, går de på jobb selv om de faktisk er syke. Det kan også tenkes at det blir fanget opp en del uobserverte faktorer når vi inkluderer fasteffekt komponenter i modellen, for eksempel motivasjon og god helse. Det kan tenkes å være grunnen til at koeffisienten skifter fortegn. Det er kanskje slik at individ som er motivert og har god helse er de som blir lengst i bedriften, og nettopp derfor drar denne koeffisienten mot mindre sykefravær. Koeffisienten til ansiennitet er derimot ikke statistisk signifikant i noen av tilfellene.

Fra tabell 5 ser vi at ingen av teamvariablene er statistisk signifikant, verken i OLS eller FE modellen. Det innebærer at vi ikke kan forkaste nullhypotesen om at disse variablene ikke er av betydning for individuelt sykefravær. Vi synes derimot det virker rart at disse variablene ikke skulle være av betydning, men samtidig er det ikke helt opplagt hvilken retning disse

variablene vil dra. Vi synes likevel det virker mest fornuftig at koeffisienten til teamvariabelen for kjønn (*Team_mann*) er negativ ettersom et team med flere menn trolig vil ha lavere gjennomsnittlig teamsykefravær. Dette kan tenkes å påvirke holdninger og normer rundt sykefravær i teamet, og føre til at den sosiale normen for akseptert sykefravær blir lavere, som i sin tur vil dra i retning mot lavere individuelt sykefravær. Det er uklart hvilken retning koeffisienten til gjennomsnittlig teamalder eventuelt vil dra. Generelt i befolkningen, og i våre data, er det slik at sykefraværet øker med alder. Imidlertid er gjennomsnittsalderen 32 år i våre data, med et standardavvik på 5 år. Dermed er det kanskje ikke noen alderseffekt. I likhet med koeffisienten til ansiennitet synes vi det er uklart hvilken retning koeffisienten til gjennomsnittlig teamansiennitet vil dra.

Vi nevnte i avsnitt 3.1 antakelsen om at gjennomsnittlig relasjonsavstand øker med teamstørrelse, slik at jo større team, jo mindre påvirker individene hverandre. En skulle derfor forvente at koeffisienten til teamstørrelse er negativ når koeffisienten til gjennomsnittlig teamsykefravær er positiv, slik at samlet effekt blir mindre. Fra tabell 5 ser vi at den er negativ i OLS, men positiv i FE modellen. Vi ser derimot at ingen av dem er statistisk signifikant.

Tabell 5 viser også en samlet F-test som undersøker om modellen er statistisk signifikant³⁷. Nullhypotesen er at alle estimatorene er lik 0, mens alternativhypotesen er at minst én av estimatorene er ulik 0. En høy F-verdi fører til forkasting av nullhypotesen. Dersom vi velger signifikansnivå på 5 prosent vil vi, med frihetsgrader på 8 og 361, forkaste nullhypotesen om F-verdien er større enn 2,29. Med en F-verdi på 9,58 forkaster vi dermed nullhypotesen. Selv om ikke alle estimat i seg selv er statistisk signifikant ser modellen ut til å være det.

9.2 Resultat fra nyansatt-modellen

I dette avsnittet presenterer vi resultat fra nyansatt-modellen som ble presentert i avsnitt 8.1. Vi starter med å estimere nyansatt-modellen med OLS for å studere «bruttoeffektene» i våre data. Deretter estimerer vi modellen med en FE modell der vi inkluderer fasteffekt komponenter. Dette gjøres for å omgå problemer med korrelerte effekter. I forsøk på å unngå problemet med simultanitet inkluderer vi deretter et instrument i modellen. Vi instrumenterer gjennomsnittlig teamsykefravær med sin laggede verdi. Denne modellen estimeres også med en FE modell der vi inkluderer fasteffekt komponenter, og gis derfor notasjon FE/IV. Dette gjør at vi kan tolke

³⁷ Stata estimerer kun en samlet F-test for modellen uten fasteffekt komponenter.

koeffisienten til gjennomsnittlig teamsykefravær som en endogen sosial interaksjonseffekt. Til slutt presenteres resultat for ulike typer sykefravær, også estimert med en FE/IV modell.

Tabell 6: Resultat fra nyansatt-modellen

Sykefravær	(1) OLS	(2) FE	(3) FE/IV
Team_sykefravær	.276 [.194] (0.156)	-.057 [.460] (0.901)	-16.784 [50.306] (0.739)
Mann	-4.731 [4.146] (0.256)	0 (omitted)	0 (omitted)
Alder	.2452 [.278] (0.380)	0 (omitted)	0 (omitted)
Ansiennitet	9.481 [5.228] (0.072)	16.124 [11.254] (0.154)	207.080 [572.334] (0.717)
Team_mann	-6.748 [13.509] (0.618)	108.449 [128.371] (0.399)	-1041.314 [3515.962] (0.767)
Team_alder	-.226 [.943] (.943)	4.309 [5.290] (0.416)	61.920 [172.636] (0.720)
Team_ansiennitet	.943 [1.336] (0.826)	-1.245 [3.957] (0.753)	.558 [23.119] (0.981)
Team_størrelse	-.142 [.180] (0.432)	.565 [1.806] (0.755)	-27.784 [83.720] (0.740)
Individdummyer	Nei	Ja	Ja
Teamdumymer	Nei	Ja	Ja
Tidsdummyer	Nei	Ja	Ja
Observasjoner	318	318	318
Antall individ	159	159	159
R ²	0.069	0.067	-
Samlet F-test	3.43	-	-
F-test (first-stage)	-	-	1.84

Merknader: Cluster-robuste standardfeil i klammeparentes og p-verdi i parentes. Asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Fra tabell 6 ser vi resultat fra nyansatt-modellen estimert med ulike modeller; OLS, FE og FE/IV. Vi observerer nyansatte over en periode på to kvartal, og benytter samlet sykefravær som utfallsvariabel i alle modellene. I OLS modellen ser vi at koeffisienten til gjennomsnittlig teamsykefravær er 0,276, og at den er ikke statistisk signifikant med en p-verdi på 0,156. Når vi estimerer modellen med en FE modell, ser vi at koeffisienten endres til -0,057. Vi ser også at p-verdien har økt drastisk til 0,938. Med andre ord, den er langt fra å være statistisk signifikant. Variabelen for kjønn faller ut ettersom denne er konstant over tid. Variabelen for alder faller også ut ettersom den er konstant over tid i perioden vi studerer (to kvartal). Videre finner vi ingen signifikante effekter blant de andre forklaringsvariablene, verken i OLS eller FE modellen. Fra den samlede F-testen i OLS modellen ser vi derimot at vi forkaster nullhypotesen

om at ingen av estimatorene er statistisk signifikante. Samlet sett har denne modellen en viss forklaringskraft selv om ingen av koeffisientene er statistisk signifikant når vi ser på dem alene. Vi legger også merke til at begge disse modellene har en veldig lav R^2 -verdi.

FE/IV modellen er vår «endelige» modell som vi håper skal fange opp den endogene sosiale interaksjonseffekten. Kjønn og alder faller bort ettersom de er konstant over tid. Dette gjør at vi ikke kan estimere effekter for kjønn og alder. Fra p-verdiene ser vi at våre koeffisienter er langt fra å være statistisk signifikante. Vi observerer også at de fleste koeffisientene er veldig høye, noe som svekker troverdighetene til våre estimat. Tabell 6 viser også en F-test fra førstesteget, noe som viser at dette er et svakt instrument. Dette er opplagt en stor ulempe og kan føre til inkonsistent estimering. Andre studier av sosiale interaksjonseffekter finner ikke samme indikasjon på at dette instrumentet er ugyldig³⁸.

Tabell 7: Resultat for ulike typer sykefravær i nyansatt-modellen

Utfallsvariabel	(1) FE/IV Legemeldt	(2) FE/IV Egenmeldt	(3) FE/IV Gradert	(4) FE/IV Syktbarn
Team_sykefravær	-17.229 [51.287] (0.737)	.660 [1.954] (0.735)	-.115 [.387] (0.765)	.754 [2.356] (0.749)
<i>Individdummyer</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>
<i>Teamdummyer</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>
<i>Tidsdummyer</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>
<i>Forklaringsvariabler</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>
<i>Observasjoner</i>	318	318	318	318
<i>Antall individ</i>	159	159	159	159
R^2	-	-	-	-

Merknader: Cluster-robuste standardfeil i klammeparentes og p-verdi i parentes. Asterisker angir signifikans:

** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Tabell 7 viser utfall for ulike typer sykefravær i nyansatt-modellen estimert med en FE/IV modell. Vi ønsket å undersøke om det er noen forskjell for de fraværstypene som typisk er lengre (legemeldt og gradert) og de fraværstypene som typisk er kortere (egenmeldt og sykefravær pga. sykt barn). Vi kom med to motstridene hypoteser. Den første var at vi kunne forvente å finne sterkere effekter blant de kortere fraværstypene ettersom sosiale normer ikke direkte påvirker et individs helsetilstand og at det er lettere å skrive en egenmelding enn å gå til legen. Den andre var at vi kunne forvente å finne sterkere effekter blant lengre fraværstyper ettersom mesteparten av slikt sykefravær skyldes diagnoser som ikke er lett å identifisere, i motsetning til kortere sykefravær som typisk skyldes forkjølelse, influensa og omgangssyke, som er lett identifiserbart. Fra tabell 7 ser vi derimot at vi ikke finner noen målbare sosiale

³⁸ Se Ichino og Maggi (2000) og Rieck og Vaage (2010).

interaksjonseffekter for noen av fraværstypene. Merk at alle regresjonene inkluderer alle relevante forklaringsvariabler.

9.3 Resultat fra flytte-modellen

I dette avsnittet presenterer vi resultat fra flytte-modellen som ble presentert i avsnitt 8.2. Flytte-modellen er estimert ved å ta differansen mellom utfall fra de to første kvartalene i nytt team og de to siste kvartalene i gammelt team. Flytte-modellen er en FE modell, dvs. den inkluderer en individspesifikk komponent som kontrollerer for uobservert individuell heterogenitet. Dette er ekvivalent med å inkludere individdummyer.

Vi starter med å estimere flytte-modellen der vi kun inkluderer relevante forklaringsvariabler. Her ønsker vi å fange opp om det er noen «bruttoeffekter» i våre data. Dette er modell (1) i tabell 8. Deretter ser vi om utfallet endres når vi legger til team- og tidsdummyer, dvs. når vi kontrollerer for alle fasteffekt komponenter. Dette er modell (2) i tabell 8. I forsøk på å unngå problemet med simultanitet inkluderer vi deretter et instrument i modellen. Vi instrumenterer gjennomsnittlig teamsykefravær med sin laggede verdi. Denne modellen inkluderer alle fasteffekt komponentene, og gis notasjon FE/IV. Dette gjør at vi kan tolke koeffisienten for gjennomsnittlig teamsykefravær som en endogen sosial interaksjonseffekt. Til slutt presenteres resultat for ulike typer sykefravær, også estimert med en FE/IV modell.

Tabell 8: Resultat fra flytte-modellen

Sykefravær	(1) FE	(2) FE	(3) FE/IV
Team_sykefravær	-1.02 [.204] (0.616)	-.245 [.240] (0.310)	-.499 [.690] (0.469)
Mann	0 (omitted)	0 (omitted)	0 (omitted)
Alder	4.660 [10.518] (0.659)	1.877 [2.330] (0.422)	5.548 [12.242] (0.650)
Ansiennitet	0 (omitted)	0 (omitted)	0 (omitted)
Team_mann	-19.834 [35.695] (0.580)	10.935 [55.945] (0.845)	4.102 [54.461] (0.940)
Team_alder	-1.066 [1.061] (0.318)	.267 [2.293] (0.907)	.029 [2.3199] (0.990)
Team_ansiennitet	2.345 [2.626] (0.374)	-1.045 [4.111] (0.800)	.119 [5.505] (0.983)
Team_størrelse	.029 [.801] (0.971)	-.680 [1.114] (0.543)	-.702 [1.047] (0.502)
<i>Individdummyer</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>
<i>Teamdummyer</i>	<i>Nei</i>	<i>Ja</i>	<i>Ja</i>
<i>Tidsdummyer</i>	<i>Nei</i>	<i>Ja</i>	<i>Ja</i>
<i>Observasjoner</i>	184	184	184
<i>Antall individ</i>	92	92	92
<i>R²</i>	0.006	0.068	0.063
<i>Samlet F-test</i>	0.33	-	-
<i>F-test (first-stage)</i>	-	-	35.33

*Merknader: Cluster-robuste standardfeil i klammeparentes og p-verdi i parentes. Asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Fra tabell 8 ser vi resultat fra flytte-modellen estimert med ulike modeller. Vi benytter samlet sykefravær som utfallsvariabel i alle modellene. Ettersom flytte-modellen er estimert med en FE modell, vil den transformere bort alt som ikke varierer over tid (observasjonsperioden), som i dette tilfellet er variabelen for kjønn. I nyansatt-modellen ble også variabelen for alder transformert bort fordi den var konstant over observasjonsperioden (to kvartal). I denne modellen benytter vi derimot informasjon fra fire perioder ved estimering, noe som medfører at vi får nok variasjon til at den «overlever» fasteffekt transformasjon. Modellen utelater variabelen for ansiennitet grunnet kolinearitet, ikke fordi den er konstant over tid³⁹. I modell (1) ser vi at koeffisienten til gjennomsnittlig teamsykefravær er -0,102, og at den er ikke statistisk signifikant med en p-verdi på 0,616. Når vi inkluderer alle fasteffekt komponentene,

³⁹ I vårt tilfelle hvor vi ser på differansen mellom ansiennitet de to første kvartalene i nytt team og de to siste kvartalene i gammelt team, vil alle observasjoner for denne variabelen være lik 2. Variabler som er konstant over tid vil være 0 i en slik differanse-modell.

i modell (2), ser vi at koeffisienten endres til -0,245. Vi ser også at p-verdien har falt til 0,310, men at den fortsatt ikke er statistisk signifikant. Vi finner heller ingen signifikante effekter blant de andre forklaringsvariablene som vi kan observere, verken i modell (1) eller modell (2). Fra den samlede F-testen i modell (1) ser vi at vi ikke kan forkaste nullhypotesen om at estimatorene til alle forklaringsvariablene er lik 0. Vi ser også at begge modellene har lav R^2 -verdi.

Fra vår «endelige» modellspesifikasjon i flytte-modellen, FE/IV modellen, ser vi at koeffisienten til gjennomsnittlig teamsykefravær lik -0,499 og har en p-verdi på 0,69. Den er altså ikke statistisk signifikant. Vi finner heller ingen signifikante effekter blant de andre forklaringsvariablene i denne modellspesifikasjonen. Tabell 8 viser også F-testen fra førstesteget, som viser at vårt instrument i denne forstand er gyldig i flytte-modellen.

Tabell 9: Resultat for ulike typer sykefravær i flytte-modellen

Utfallsvariabel	(1) FE/IV Legemeldt	(2) FE/IV Egenmeldt	(3) FE/IV Gradert	(4) FE/IV Syktbarn
Team_sykefravær	-0.131 [.651] (0.840)	-0.004 [.084] (0.954)	-0.360 [.248] (0.146)	-0.002 [.064] (0.974)
<i>Individdummyer</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>
<i>Teamdummyer</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>
<i>Tidsdummyer</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>
<i>Forklaringsvariabler</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>	<i>Ja</i>
<i>Observasjoner</i>	184	184	184	184
<i>Antall individ</i>	92	92	92	92
<i>R²</i>	0.057	0.243	0.106	0.201

*Merknader: Cluster-robuste standardfeil i klammeparentes og p-verdi i parentes. Asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Tabell 9 viser resultat for ulike typer sykefravær i flytte-modellen, estimert med en FE/IV modell. Merk at alle relevante forklaringsvariabler er inkludert i regresjonene. I likhet med resultat fra nyansatte-modellen, finner vi heller ikke noen signifikante effekter i flytte-modellen når vi studerer ulike typer sykefravær separat. Utfall for gradert sykefravær skiller seg derimot ut med å ha en betydelig lavere p-verdi enn de andre fraværstypene.

9.4 Heterogenitet over kjønn og tid

Vi har nå undersøkt om det eksisterer sosiale interaksjonseffekter i både nyansatt- og flytte-modellen. I begge modellene observerte vi både kvinner og menn over en periode på to kvartal. Vi vil nå undersøke om det er slik at vi finner effekter dersom vi ser på kvinner og menn separat, og om vi studerer individene over en lengre tidsperiode. Vi har sett fra den deskriptive

statistikken at kvinner har høyere sykefravær enn menn, og lurer på om sosiale normer og holdninger knyttet til sykefraværsadferd kan være en av årsakene. Videre lurer vi også på om vi vil oppdage noen effekter ved å estimere nyansatt- og flytte-modellen med observasjonsperioder på tre og fire kvartal. Det kan tenkes at effektene blir sterkere over tid ettersom individene får mer tid til å tilpasse seg hverandre⁴⁰. Vi studerer først heterogenitet i nyansatt-modellen og deretter flytte-modellen.

9.4.1 Heterogenitet i nyansatt-modellen

Tabell 10: Resultat for kvinner og menn i nyansatt-modellen

Sykefravær	(1) FE/IV Kvinner	(2) FE/IV Menn
Team_sykefravær	-13.484 [149.239] (0.928)	67.778 [416.927] (0.871)
Individdummyer	Ja	Ja
Teamdummyer	Ja	Ja
Tidsdummyer	Ja	Ja
Forklaringsvariabler	Ja	Ja
Observasjoner	172	146
Antall individ	86	73
R ²	-	-

*Merknader: Cluster-robuste standardfeil i klammeparentes og p-verdi i parentes. Asterisker angir signifikans:
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Tabell 10 viser resultat for nyansatte kvinner og menn separat, der utfallsvariabelen er samlet sykefravær. Vi observerer dem de to første kvartalene de er nye på arbeidsplassen. En hypotese var at vi skulle finne sterkere effekter for kvinner ettersom at kvinner har uforklarlig høyere sykefravær enn menn, svakere tilknytning til arbeidsmarkedet og mer ansvar i hjemmet – alle faktorer som kan dra i retning mot lavere terskel for å sykemelde seg. Vi finner derimot ingen signifikante effekter når vi studerer kvinner og menn separat. Vi observerer at koeffisientene og p-verdiene er veldig høye, både for kvinner og menn.

⁴⁰ Flere studier finner at sosiale interaksjonseffekter blir sterkere over tid, se blant annet Rieck og Vaage (2010).

Tabell 11: Resultat for ulike observasjonsperioder i nyansatt-modellen

Sykefravær	(1) FE/IV Tre kvartal	(2) FE/IV Fire kvartal
Team_sykefravær	.102 [1.128] (0.928)	6.178 [10.730] (0.565)
Individdummyer	Ja	Ja
Teamdummyer	Ja	Ja
Tidsdummyer	Ja	Ja
Forklaringsvariabler	Ja	Ja
Observasjoner	471	584
Antall individ	157	146
R ²	0.069	-

Merknader: Cluster-robuste standardfeil i klammeparentes og p-verdi i parentes. Asterisker angir signifikans:
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

For å studere nyansatte over en lengre tidsperiode endrer vi på definisjonen av nyansatte. For en observasjonsperiode på tre kvartal definerer vi nyansatte som individ som har en startdato tidligst tre kvartal etter første observasjon i datasettet, og senest tre kvartal før siste observasjon i datasettet. Dette gjør at vi blir stående igjen med 157 individ. Med tre observasjoner per individ får vi totalt 471 observasjoner. For en observasjonsperiode på fire kvartal definerer vi nyansatte som individ som har en startdato tidligst tre kvartal etter første observasjon i datasettet, og senest fire kvartal før siste observasjon i datasettet. Vi blir da stående igjen med 146 individ. Med fire observasjoner per individ får vi totalt 584 observasjoner. Vi har valgt følgende definisjoner slik at vi kan observere alle nyansatte i henholdsvis tre og fire kvartal. Fra tabell 11 ser vi derimot at vi ikke finner noen signifikante effekter for noen av observasjonsperiodene.

9.4.2 Heterogenitet i flytte-modellen

Tabell 12: Resultat for kvinner og menn i flytte-modellen

Sykefravær	(1) FE/IV Kvinner	(2) FE/IV Menn
Team_sykefravær	-.309 [.935] (0.741)	-1.202** [.485] (0.013)
Individdummyer	Ja	Ja
Teamdummyer	Ja	Ja
Tidsdummyer	Ja	Ja
Forklaringsvariabler	Ja	Ja
Observasjoner	106	78
Antall individ	53	39
R ²	0.098	0.244

Merknader: Cluster-robuste standardfeil i klammeparentes og p-verdi i parentes. Asterisker angir signifikans:
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Tabell 12 viser resultat for kvinner og menn som bytter team separat. Utfallsvariabelen er samlet sykefravær og vi benytter en observasjonsperiode på to kvartal. For kvinner finner vi ingen signifikant effekt, men for menn finner vi at koeffisienten til gjennomsnittlig teamsykefravær er lik -1,202 og at denne er statistisk signifikant på 1,3 prosent signifikansnivå. Vi finner altså en negativ endogen sosial interaksjonseffekt for menn. Tolkningen er at en marginal økning i teamets gjennomsnittlige sykefravær vil redusere menns sykefravær med ca. 1,2 timer per kvartal. Som tidligere nevnt hadde vi ikke forventet å finne negative sosiale interaksjonseffekter.

Tabell 13: Resultat for ulike observasjonsperioder i flytte-modellen

Sykefravær	(1) FE/IV Tre kvartal	(2) FE/IV Fire kvartal
Team_sykefravær	.105 [.416] (0.800)	.463** [.200] (0.021)
<i>Individdummyer</i>	<i>Ja</i>	<i>Ja</i>
<i>Teamdummyer</i>	<i>Ja</i>	<i>Ja</i>
<i>Tidsdummyer</i>	<i>Ja</i>	<i>Ja</i>
<i>Forklaringsvariabler</i>	<i>Ja</i>	<i>Ja</i>
<i>Observasjoner</i>	225	264
<i>Antall individ</i>	75	66
<i>R²</i>	0.086	0.104

Merknader: Cluster-robuste standardfeil i klammeparentes og p-verdi i parentes. Asterisker angir signifikans:

** p < 0.1, ** p < 0.05, *** p < 0.01.*

For å studere heterogenitet over tid i flytte-modellen utfører vi ytterligere noen begrensinger i datasettet. For å studere individene i tre kvartal begrenses modellen til å kun inkludere individ som har vært minst tre kvartal i både gammelt og nytt team. Vi står da igjen med 75 individ, og tre observasjoner per individ. Vi får dermed totalt 225 observasjoner. For å studere individene i fire kvartal begrenses modellen til å kun inkludere individ som har vært minst fire kvartal i både gammelt og nytt team. Vi blir stående igjen med 66 individ, og fire observasjoner per individ. Vi får dermed totalt 264 observasjoner. Fra tabell 13 ser vi at vi ikke finner noen signifikante effekter når vi observerer individene i tre kvartal. Når vi observerer individene i fire kvartal finner vi derimot at koeffisienten til gjennomsnittlig teamsykefravær er lik 0,463 og at denne er statistisk signifikant på 2,1 prosent signifikansnivå. Vi finner altså en positiv endogen sosial interaksjonseffekt når vi studerer individene i fire kvartal i flytte-modellen.

9.5 Oppsummering av resultat

Vi estimerte først den generelle modellen for hele utvalget med OLS, der utfallsvariabelen var samlet sykefravær. Denne ga en indikasjon på at det ser ut til å være en positiv «bruttoeffekt» i våre data. Modellen klarer imidlertid ikke å avgjøre om dette skyldes endogene, eksogene eller korrelerte effekter. Vi estimerte så den generelle modellen med en FE modell, hvor vi inkluderte team- og tidsdummyer, for å kontrollere for uobservert heterogenitet (korrelerte effekter). Vi finner fortsatt en positiv «bruttoeffekt» i dataene, men vi vet ikke om det skyldes endogene eller eksogene effekter.

Deretter estimerte vi nyansatt- og flytte-modellen i forsøk på å isolere den kausale effekten av teammedlemmenes sykefravær på et individs sykefravær (den endogene effekten). En fordel med å studere nyansatte og flytterne er at det kan tenkes å redusere problemet med simultanitet. Ved å dessuten instrumentere gjennomsnittlig teamsykefravær med sin laggende verdi blir problemet ytterligere redusert. I nyansatt-modellen finner vi imidlertid ingen signifikante effekter. Dessuten så vi at instrumentet vi har benyttet i nyansatt-modellen ikke er gyldig, noe som kan føre til inkonsistent estimering. I flytte-modellen finner vi derimot en signifikant effekt for menn når vi observerer dem i to kvartal. Her finner vi en negativ endogen sosial interaksjonseffekt. Videre finner vi en signifikant effekt i flytte-modellen når vi studerer flytterne (kvinner og menn) over en observasjonsperiode på fire kvartal. Her finner vi en positiv endogen sosial interaksjonseffekt, som er det forventede fortegnet. I flytte-modellen finner vi dessuten signifikant F-verdi når vi tester for instrumentets gyldighet.

9.6 Robusthet: placebotester

Før vi eventuelt drøfter mekanismer bak våre resultat vil vi undersøke robustheten til våre estimat. For å unngå feiltolkninger er det viktig å utføre robusthetstester. Vi vil undersøke robustheten til de koeffisientene som er statistisk signifikante. Vi har funnet to effekter som er statistisk signifikant i våre data, både når vi studerte menn i flytte-modellen og når vi observerte flytterne over en lengre observasjonsperiode på fire kvartal. I dette avsnittet vil vi gjennomføre en såkalt placebotest der vi undersøker hvordan de ansatte responderer til en tilfeldig gruppe medarbeidere. Ved å gjøre dette kan vi finne ut om de estimerte effektene skyldes egenskaper ved teamet som den ansatte tilhører, eller om det er mer generelle faktorer som forklarer effektene.

En generell faktor som kan være en forklaring til våre resultat er «regression toward the mean». Dette er et statistisk fenomen som kan føre til at naturlig variasjon i gjentatte data ser ut som ekte endringer. Dette kan føre til feiltolkninger gjennom at man tilskriver observert variasjon en årsak som egentlig bare skyldes naturlig variasjon. Fenomenet oppstår når uvanlig små eller store observasjoner blir fulgt av observasjoner som ligger nærmere gjennomsnittet. Årsaken til at det oppstår er fordi verdier observeres med tilfeldige feil. Disse tilfeldige feilene er ikke-systematisk variasjon i observerte verdier som ligger rundt det sanne gjennomsnittet, for eks. tilfeldige målefeil eller tilfeldig fluktasjoner hos den/de som observeres⁴¹.

Det kan være flere årsaker til at sykefravær er utsatt for «regression toward the mean». På kort sikt, kvartal til kvartal, kan stokastiske sjokk være avgjørende for om et individ har høyt eller lavt sykefravær. Smittsomme sykdommer kan være et slikt sjokk. Et individ som har vært utsatt for en smittsom sykdom et kvartal og følgelig har høyt sykefravær på grunn av det, vil trolig observeres med lavere sykefravær i neste periode ettersom sannsynligheten for å bli smittet igjen er lavere. Dette medfører at individets sykefravær beveger seg mot gjennomsnittet, og at vi observerer en sosial interaksjonseffekt. Dette er derimot resultat av «regression toward the mean», og ikke sosiale normer.

For å teste om resultatene vi finner skyldes egenskaper ved teamet eller simpelthen «regression toward the mean» utfører vi placebotester. Dette gjøres ved å kjøre de samme regresjonene som vi allerede har gjort, bortsett fra at vi plasserer de ansatte med en tilfeldig gruppe medarbeidere fra kundeservicesenteret⁴². Dette vil fange opp om det er noen spesielle egenskaper ved teamet som den ansatte tilhører som driver effektene, eller om det er mer generelle forklaringer. Dersom resultatene fra placebotestene ikke er statistisk signifikante er det trolig egenskaper ved teamet som driver de estimerte effektene, og ikke mer generelle forklaringer som «regression toward the mean». Da vil våre resultat fremstå som robuste.

⁴¹ Se Stigler (1997) og Barnett et al. (2005) for mer om «regression toward the mean» og eksempler rundt dette statistiske fenomenet.

⁴² Vi har implementert dette i Stata ved å tilskrive alle observasjoner et tilfeldig tall mellom 0 og 1. Vi sorterte dataene ut fra disse tallene og delte deretter dataene inn i 23 ulike grupper.

Tabell 14: Resultat fra placebotestene

Sykefravær	(1) FE/IV Menn – to kvartal	(2) FE/IV Flytterne – fire kvartal
Pseudo_sykefravær	-.124 [.079] (0.119)	-1.038 [.891] (0.244)
<i>Individdummyer</i>	<i>Ja</i>	<i>Ja</i>
<i>Teamdummyer</i>	<i>Ja</i>	<i>Ja</i>
<i>Tidsdummyer</i>	<i>Ja</i>	<i>Ja</i>
<i>Forklaringsvariabler</i>	<i>Ja</i>	<i>Ja</i>
<i>Observasjoner</i>	78	264
<i>Antall individ</i>	39	66
<i>R²</i>	-	-

*Merknader: Cluster-robuste standardfeil i klammeparentes og p-verdi i parentes. Asterisker angir signifikans:
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Tabell 14 viser resultatene fra placebotestene vi har utført. Vi har estimert placebotester for de estimerte effektene som er statistisk signifikante, dvs. når vi studerte menn i flytte-modellen og når vi studerte flytterne over en observasjonsperiode på fire kvartal. Basert på disse tallene ser vi at de estimerte placeboeffektene ikke er statistisk signifikante. Placeboteamet ser altså ikke ut til å ha noen effekt på sykefravær. Dette tyder på at det er egenskaper ved teamet som individet faktisk tilhører som driver effektene. Vi kan dermed konkludere med at resultatene ser ut til å være robuste. Dette innebærer at vi har funnet indikasjoner på sosiale interaksjonseffekter i våre data.

9.7 Mekanismer

Vi finner signifikante sosiale interaksjonseffekter blant menn i flytte-modellen, og når vi observerer flytterne over en lengre observasjonsperiode på fire kvartal. Fra placebotestene så vi at de estimerte effektene var robuste, dvs. at de estimerte effektene ser ut til å skyldes egenskaper ved teamet som den ansatte tilhører. Vi har dermed identifisert to endogene sosiale interaksjonseffekter for sykefravær i kundeservicesenteret som vi studerer. Placebotestene bringer oss imidlertid ikke noe nærmere et svar for hvilke mekanismer innad i teamene som driver effektene. Det er trolig flere mekanismer som spiller en rolle her, blant annet den sosiale normen for akseptert nivå sykefravær, terskel for hva som oppfattes som «syk nok» og stigma knyttet til skoft.

Når vi studerer menn i flytte-modellen finner vi en negativ endogen sosial interaksjonseffekt. Mer presist, vi finner at menn reduserer sitt sykefravær med 1,2 timer per kvartal når teamets gjennomsnittlige sykefravær øker med én time per kvartal. Dette var uventet ettersom vi

forventet å finne positive effekter, men vi kan likevel tenke oss et par mulige forklaringer til dette. En mulig forklaring kan være at menn tar mer ansvar, og dermed er mindre borte fra jobb, dersom det er flere som er sykemeldt i teamet. Menn har typisk en sterkere tilknytning til sin arbeidsplass enn kvinner, og er kanskje villig til å stå på litt ekstra når det trengs, selv om det innebærer å gå på jobb når man føler seg dårlig. Det kan også tenkes at menn har mindre behov for å tilpasse seg gruppen enn kvinner. Eventuelt ønsker menn å skille seg ut fra gruppen ved å være mer tilstede på jobb, og dermed fremstå som en sterk person. Dersom det er mange som er sykemeldt i teamet kan det tenkes også at teamlederen legger mer press på mennene enn på kvinnene om at de må stille opp på jobb. Det er ikke en uvanlig oppfatning at kvinner er mer «sårbar» enn menn.

Når vi studerer flytterne (kvinner og menn) over en periode på fire kvartal finner vi derimot en positiv endogen sosial interaksjonseffekt. Vi finner at individuelt sykefravær øker med 0,463 timer per kvartal når teamets gjennomsnittlige sykefravær øker med én time per kvartal. En mulig forklaring til at individene tilpasser seg hverandre kan være knyttet til den sosiale normen for akseptert nivå sykefravær. Det er trolig lettere å sykemelde seg dersom mange andre i teamet er sykemeldt. Dersom mange er sykemeldt kan stigma knyttet til å være borte fra jobb reduseres. På en arbeidsplass kan det danne seg en sosial norm både vedrørende terskel for hva som er «syk nok» og terskel for regelrett skoft. Slike normer kan også være knyttet til rettferdighet gjennom at individ føler at de fortjener å være borte fra jobb dersom andre er borte fra jobb.

Det er imidlertid litt forvirrende at vi finner negative effekter blant menn, ingen signifikante effekter blant kvinner, men positive effekter når vi studerer både kvinner og menn i fire kvartal i flytte-modellen. Det kan kanskje tyde på at det er kvinner som driver de estimerte effektene i flytte-modellen når vi studerer dem i fire kvartal ettersom vi finner en negativ effekt blant menn. Det er vanskelig å vite hva det er som faktisk driver de estimerte effektene, og det kan virke noe motsigende å argumentere for både positive og negative sosiale interaksjonseffekter for sykefravær. Vi klarer ikke å komme med noen konkrete svar, og denne analysen er dermed bare spekulativ.

For å få bedre forklaringer til mulige mekanismer som driver de estimerte effektene kunne det vært aktuelt å intervjuer ansatte i kundeservicesenteret, eller eventuelt gjennomført en spørreundersøkelse. Det ville vært interessant å få tilbakemeldinger på hva de ansatte opplever rundt sosiale normer knyttet til hva som er kvalifisert som «syk nok» til å være borte fra jobb.

For eksempel, er det slik at man er borte fra jobb dersom man er *litt* forkjølet? Vi har derimot ikke hatt anledning til å gjennomføre intervjuer eller spørreundersøkelser på kundeservicesenteret som vi analyserer.

10 Svakheter ved analysen

I dette kapittelet vil vi belyse noen svakheter ved vår analyse. Det er hovedsakelig tre bekymringer. Vi har relativt få observasjoner, få forklaringsvariabler og det er potensielt fare for målefeil.

10.1 Få observasjoner

En svakhet, trolig den største, ved vår analyse er at vi har få observasjoner. I utgangspunktet virker datasettet lovende med 362 individ og 18 observasjonsperioder, men i forsøk på å isolere den kausale effekten av teammedlemmenes sykefravær på et individs sykefravær foretar vi flere begrensninger i datasettet. I nyansatt-modellen observerer vi kun nyansatte, og i flytte-modellen observerer vi kun ansatte som bytter team. Dette gjøres for å redusere simultanitetsproblemet, men vi mister veldig mange individ i datasettet. Videre begrenser vi også observasjonsperioden vi studerer disse individene. I utgangspunktet studerer vi dem kun i to tidsperioder, dvs. to kvartal. Dette gjøres fordi det er mest interessant å studere individene i begynnelsen – når de er relativt ny på arbeidsplassen eller når de nettopp har byttet til et nytt team. Jo lengre vi observerer dem, jo mer problemer kan vi tenkes å få med simultanitet. Dette gjør at vi blir stående igjen med kun to observasjoner per individ. I nyansatt-modellen står vi igjen med 159 individ, og totalt 318 observasjoner. I flytte-modellen står vi kun igjen med 92 individ, og totalt 184 observasjoner.

Vi studerte også individene i tre og fire kvartal for å se om det er slik at effektene endres over tid. Dette medførte vi at vi måtte utføre ytterligere noen begrensninger i datasettet slik at vi ble stående igjen med enda færre individ, men flere observasjoner grunnet lengre observasjonsperioder. For å kunne studere nyansatte i tre og fire kvartal, endret vi definisjonen av nyansatte til å inkludere individ som henholdsvis hadde startdato senest tre og fire kvartal før siste observasjon i datasettet, slik at vi kunne observere alle like lenge. For å studere flytterne i tre og fire kvartal, inkluderte vi kun individ som henholdsvis har vært minst tre og fire kvartal i både gammelt og nytt team.

Fra avsnitt 8.3 så vi at T og/eller N må gå mot uendelig for at våre estimat skal være konsistent. Vi har hvert fall for få observasjonsperioder (T), så spørsmålet er om antall individ (N) er nok til at våre estimat er konsistent. Videre trenger man et sterkt instrument for konsistent

estimering, og vi så fra analysen at nyansatt-modellen ikke hadde et gyldig instrument. Dette gjør at resultatene fra denne modellen blir lite troverdig.

10.2 Få forklaringsvariabler

En annen svakhet er at vi har få forklaringsvariabler i vår analyse. Vi har relativt lite informasjon om individene vi studerer utover deres sykefravær. I analysen benyttes kun informasjon kjønn, alder og ansiennitet som gir oss forklaringsvariablene; *Mann*, *Alder*, *Ansiennitet*, *Team_mann*, *Team_alder* og *Team_ansiennitet*.

Det er trolig flere forhold som påvirker et individs sykefravær enn de som vi har inkludert i modellen. Fasteffekt modellen kontrollerer heldigvis for utelatte variabler som er konstant over tid. Utelatte variabler som ikke er konstante over tid og påvirker individuelt sykefravær kan derimot føre til inkonsistent estimering (utelatt variabelproblem). I dette tilfellet er en eller flere av forklaringsvariablene korrelert med feilledet, noe som fører til at vi får en eller flere endogene variabler i modellen. Sykefravær skyldes først og fremst forhold som påvirker et individs helse (sykdom/skade), men også mer uforklarlige forhold som ikke nødvendigvis gir rett til sykepenger.

Det er flere forhold som påvirker et individs helse, blant annet kosthold, fysisk aktivitet, utdanning og familiesituasjon. Alle disse forholdene kan tenkes å endres over tid, og vil dermed ikke bli fanget opp i modellen. Kosthold og fysisk aktivitet vil opplagt påvirke et individs helse, som igjen påvirker et individs sykefravær. Dersom individene vi studerer har vaner som er tidskonstante knyttet til kosthold og fysisk aktivitet vil fasteffekt modellen kontrollere for dette. Det er ikke urimelig å tenke seg at noen har «gode vaner» over tid, mens andre har «dårlige vaner» over tid. Det er derimot ikke sikkert at *alle* har slike faste vaner over tid. Videre finner flere studier at individ med høy utdanning har lavere sykefravær enn individ med lav utdanning⁴³. Dette er trolig også knyttet til helse, gjennom at individ med høy utdanning spiser bedre og trener mer enn individ med lav utdanning. Dersom individene vi studerer er ferdigutdannet vil fasteffekt modellen kontrollere for uobservert heterogenitet knyttet til utdanning, men dersom noen fortsatt er under utdanning vil ikke dette fanges opp i modellen. Et individs familiesituasjon, som typisk endres over tid, kan også påvirke et individs helse. Det

⁴³ Se for eksempel Markussen et al. (2009) og Mastekaasa (2015).

kan tenkes at individ med små barn er mer syk enn andre ettersom det ofte er mye smittsomme sykdommer i barnehager.

Det er som sagt ikke nødvendigvis bare et individs helse (sykdom/skade) som påvirker et individs sykefravær. Det kan også være mer uforklarlige forhold som påvirker individuelt sykefravær og som varierer over tid, som er utelatt fra modellen. For eksempel, individ kan oppleve endringer i deres livssituasjon som medfører at de sykemelder seg, til tross for at dette ikke nødvendigvis kvalifiseres som gyldig grunn til å være borte fra jobb og motta sykepenger. Dette kan for eksempel være konflikter på jobb eller i familie, sosiale problemer, fyllesyke og økonomiske vanskeligheter. Dette vil ikke fanges opp i modellen.

Man kan løse problemet med utelatte variabler ved å inkludere en proxyvariabel⁴⁴ i modellen for den variabelen man tror er utelatt. Det kan derimot være vanskelig å finne slike proxyvariabler og de er ikke alltid tilgjengelig. Alternativt kan man bruke instrumentvariabler for å løse problemet med utelatte variabler. Vi har allerede benyttet denne metoden for den endogene variabelen *Team_sykefravær* grunnet simultanitet, men det kunne vært fornuftig å instrumentert andre variabler også som potensielt er endogen grunnet utelatte variabler. Vi har derimot ikke noen fornuftige instrumenter tilgjengelig å benytte.

Som nevnt i avsnitt 7.2.2 har vi informasjon om individenes lønn og produktivitet (salg) i våre data, men har valgt å ikke benytte disse variablene i analysen ettersom de trolig er endogene grunnet simultanitet. Sykefravær og produktivitet er tett koblet sammen og vil trolig være korrelert, så ved å utelate informasjon om individenes produktivitet kan det tenkes at vi får problemer med utelatte variabler. Tilsvarende gjelder for informasjon om lønn. Det blir altså en avveining mellom problemer med simultanitet eller utelatte variabler. En vanlig måte å komme unna simultanitetsproblemet er å lagge forklaringsvariablene man tror er utsatt for simultanitet⁴⁵. Et argument for å gjøre dette er at man unngår å bruke svake instrumenter. Reed (2013) viser derimot at man ikke nødvendigvis løser simultanitetsproblemet ved å lagge. Det blir spesielt en utfordring dersom de endogene forklaringsvariablene er serie korrelert, dvs. verdien i periode t er korrelert med verdien i periode $t-1$. Dette vil typisk være tilfellet med både lønn og produktivitet. Ettersom vi ikke har noen fornuftige variabler å instrumentere disse

⁴⁴ Proxyvariabelen trenger ikke være den samme som den utelatte variabelen, men de må være korrelert. For eks. IQ for evner.

⁴⁵ Se for eksempel Bansak et al. (2007), Aschhoff og Schmidt (2008) og Clemens et al. (2012).

variablene med, har vi derfor valgt å utelate dem fra analysen, men det kan tenkes at vi får noe inkonsistent estimering som følge av dette.

Vi har også lite informasjon om lokale attributter (teamspesifikke- og områdevariabler). Den eneste variabelen vi har er teamets størrelse. Det er trolig flere forhold som varierer mellom team og over tid, som påvirker et individs sykefravær. For eksempel, arbeidsmiljø, teamlederen og arbeidsledighetsraten i området teamet befinner seg. Etter hvert som individ starter og slutter i et team kan det tenkes at arbeidsmiljøet i teamet endres, noe som potensielt kan påvirke et individs sykefravær. Vi har også sett at flere studier finner korrelasjon mellom arbeidsledighet og sykefravær. Vi vet derimot ingenting om beliggenheten til teamene, og dermed ingenting om arbeidsledigheten i de ulike områdene der teamene befinner seg. Dersom dette er forhold som er konstant over tid vil derimot teamdummyene kontrollere for dette.

10.3 Målefeil

I våre data er det en del sykefraværsobservasjoner som mangler. Vi mottok i utgangspunktet ukesdata fra forsikringsselskapet som vi analyserer. For alle fraværsvariabler i datasettet (*Legemeldt*, *Egenmeldt*, *Gradert*, *Syktbarn*, *Ubegrunnet*, *Sykefravær* og *Team_sykefravær*) er det 10544 ukesobservasjoner som mangler av totalt 63777 ukesobservasjoner. Dette innebærer at ca. 16,5 prosent av observasjonene for disse variablene mangler. Vi vet ikke årsaken til at vi mangler data for disse observasjonene, men det kan se ut som at forsikringsselskapet har «glemt» å føre data i disse periodene siden det er de samme periodene for alle fraværsvariabler. Vi kan ikke utelate at dette introduserer målefeil i sykefraværsvariablene.

Ved å aggregere opp til kvartalsdata er det ikke lengre noen observasjoner som er spesifisert at de mangler i datasettet, men det innebærer at vi ikke har komplette data for alle kvartal. Altså, ukesobservasjoner som mangler blir ikke tatt med når vi aggregerer opp til kvartal. Dette kan føre til misvisende resultat da de sanne gjennomsnittsverdiene trolig er noe annet enn hva vi rapporterer. Dersom verdiene som mangler burde vært 0 rapporterer vi for høye gjennomsnittsverdier, mens dersom verdiene som mangler burde være noe annet enn 0 er det umulig å si om vi rapporterer for høye eller for lave gjennomsnittsverdier. utfordringen er først og fremst knyttet til forklaringsvariabelen *Team_sykefravær*. Mye manglede observasjoner kan tenkes å føre til målefeil i variabelen *Team_sykefravær*, noe som fører til inkonsistent

estimering⁴⁶. Mye manglende observasjoner i utfallsvariabelen, *Sykefravær*, er et mindre problem. Det fører ikke nødvendigvis til forventingsskjevne estimat, men kan skape en del ekstra støy. Det kan tenkes at denne ekstra støyen er med på å forklare den lave graden av signifikante sykefraværskoeffisienter i mange av våre modeller.

Til slutt vil vi også nevne et potensielt problem i flytte-modellen som oppstår som følge av at vi aggregerer ukedata til kvartalsdata. En av de største utfordringene med å estimere sosiale interaksjonseffekten er knyttet til simultanitet. Vi har tatt hensyn til simultanitet ved å se på flytterne samt instrumentere *Team_sykefravær* med sin laggede verdi. I prinsippet er dette en lovende prosedyre, men vi står overfor praktiske problemer som gjør at metoden likevel er usikker. Ettersom vi har aggregert ukedata til kvartalsdata oppstår et tolkningsproblem. Vi er ute etter å estimere sykefravær de to første kvartalene i nytt team i forhold til de to siste kvartalene i gammelt team. Fra ukedataene ser vi hvilken uke individene bytter team. Når vi aggregerer til kvartal innebærer dette at individene skifter team en eller annen gang i løpet av de 13 ukene som utgjør et kvartal. I kvartalsdataene ser vi ikke når i løpet av et kvartal et individ har byttet team, noe som innebærer at individ kan blir påvirket av både nytt og gammelt team et par uker i løpet av et kvartal. En måte å løse dette på er å droppe kvartalet der skiftet skjer, men ettersom vi har så få observasjoner i utgangspunktet er det lite attraktivt i vårt tilfelle. Dette svekker imidlertid resultatene fra flytte-modellen.

⁴⁶ Dersom variabelen er utsatt for målefeil vil den være korrelert med feilledet i modellen, dvs. $Cov(x, \varepsilon) \neq 0$. Se Verbeek (2012 s. 142) for mer om målefeil.

11 Konklusjon

Hovedformålet med denne oppgaven har vært å undersøke om det eksisterer sosiale interaksjonseffekter for sykefravær på et kundeservicesenter i et norsk forsikringsselskap. Vi har benyttet bedriftsdata med informasjon om de ansattes sykefravær samt individuelle- og teamegenskaper for å undersøke om det er slik at et individs sykefravær påvirkes av dens teammedlemmers sykefravær. Fokuset i oppgaven har vært å identifisere den såkalte endogene sosiale interaksjonseffekten. Vi ønsket først og fremst å undersøke om eksisterer sosiale interaksjonseffekter for sykefravær, men også om det er slik at disse effektene skiller seg for ulike typer sykefravær, for kvinner og menn, og eventuelt om effektene endres over tid.

Vi har benyttet to modellspesifikasjoner der vi forsøker å isolere den endogene sosiale interaksjonseffekten, dvs. den kausale effekten av teammedlemmenes sykefravær på et individs sykefravær. I nyansatt-modellen finner vi ingen indikasjon på at det eksisterer sosiale interaksjonseffekter for sykefravær, og dessuten er instrumentet vi har benyttet svakt. I flytte-modellen finner vi derimot signifikante effekter når vi studerer menn og når vi studerer flytterne over en lengre observasjonsperiode på fire kvartal. Vi finner at menn reduserer sitt sykefravær med 1,2 timer per kvartal når teammedlemmenes sykefravær øker med én time per kvartal. Når vi studerer flytterne i fire kvartal finner vi derimot en positiv effekt. Vi finner at flytternes sykefravær øker med 0,46 timer per kvartal når teammedlemmenes sykefravær øker med én time per kvartal. Instrumentet vi benytter i flytte-modellen viser seg dessuten å være gyldig. Vi gjennomførte placebotester for å undersøke robustheten til våre resultat, og kan konkludere med at effektene vi finner ser ut til skyldes adferden til teamet som den ansatte tilhører.

Vi har imidlertid sett at det er flere svakheter ved vår analyse. Den største svakheten er trolig at vi har relativt få observasjoner. Ved å studere nyansatte og flytterne ser vi kun på en begrenset del av utvalget, noe som gjør at antall individ og antall observasjoner dropper betydelig. Videre har vi relativt lite informasjon om individene vi studerer utover deres sykefravær, noe som skaper problemer med utelatte variabler. Til slutt møter vi også noen utfordringer ved å aggregere data fra ukedata til kvartalsdata. Ettersom det er noen observasjoner på ukensnivå som mangler, vil vi ikke få komplette data på kvartalsnivå. Dette kan potensielt føre til målefeil, men vi anser aggregeringen nødvendig da vi tror at sosiale interaksjonseffekter for sykefravær virker med litt tidsetterslep.

Til tross for en del svakheter ved analysen finner vi, i likhet med flere andre studier, indikasjon på at det eksisterer sosiale interaksjonseffekter for sykefravær på lokalt nivå. Vi har sett at Norge har et høyt sykefravær i en internasjonal sammenheng, og at det er vanskelig å finne naturlige årsaker til dette, som for eksempel dårligere folkehelse eller et mer sykdomsbringende arbeidsmarked. Den sjenerøse sykelønnsordningen og lav grad av kontroll nevnes derfor ofte som en mulig forklaring til dette. Dersom det er slik at sykefravær påvirkes av sosiale normer og man opplever en økning i sykefravær kan det være grunn til bekymring. Gruppeneffekter, for eksempel gjennom fraværskultur på arbeidsplassen, kan forsterke sosiale problemer og opprettholde dårlige likevekter. Sosiale normer i retning lavere terskel for sykefravær kan også oppstå på mindre lokalt nivå, for eksempel nasjonalt, noe som ligger utenfor rammen for den empiriske analysen i denne oppgaven. I et land som Norge hvor vi har en sjenerøs sykelønnsordning er det ekstra viktig at «gode normer» som fremmer jobbdeltakelse ikke forvitrer, da dette kan svekke velferdssystemet betraktelig. Vi mener derfor at det er viktig å tenke gjennom betydningen av sosiale interaksjonseffekter ved utforming og analyser av politiske tiltak.

På lang sikt vil det trolig ikke være bærekraftig med høyt sykefravær. Norge har i stor grad vært avhengig av oljeinntekter og aktivitet i petroleumsnæringen for å finansiere velferdssystemet som vi har i dag. Det er mye usikkerhet om fremtidige inntekter for offentlig sektor på grunn av den kraftige nedgangen i oljeinntekter samt økt fokus på fornybare energikilder. Som følge av dette vil velferdsstaten trolig bli mer avhengig av et fungerende arbeidsmarked fremover. Reduksjon i offentlige inntekter kan bety reduksjon i offentlige ytelser eller økte skatter. Dersom målsettingen er å beholde dagens sjenerøse sykelønnsordning, må skattenivået trolig økes på sikt for at myndighetene skal ha samme spillerom som tidligere når oljeinntektene var høye. Da er det viktig at folk jobber og genererer skatteinntekter. Det blir viktig for velferdsstaten å beholde normen om at kun de som faktisk er syk eller skadet har rett til sykepenger. Det er fortsatt relativt lite forskning om betydningen av sosiale normer for sykefravær. Mer forskning trengs for å forstå hvilken grad sosiale normer potensielt påvirker sykefravær.

Referanseliste

- Aarbu, K. O. & Torsvik, G. (2007) Pay and Performance in a Call Centre; Principals and Agents or Principally Angels? *NHH Department of Economics Discussion Paper*, Nr. 33
- Arbeids- og sosialdepartementet (2014) *Intensjonsavtalen om et mer inkluderende arbeidsliv*. [Internett] Tilgjengelig fra: https://www.regjeringen.no/globalassets/upload/asd/dokumenter/2014/ia_20142018/s_ignert_ia_avtale.pdf [Lest 15.04.2015]
- Aschhoff, B. & Schmidt, T. (2008) Empirical Evidence on the Success of R&D Cooperation – Happy Together? *Review of Industrial Organization*, 33(1) August, s. 41-62
- Askildsen, J. E., Bratberg, E. & Nilsen, Ø. A. (2004) Svingninger i sykefraværet: Er arbeidsledigheten avgjørende? *Søkelys på arbeidsmarkedet*, 21, s. 3-12.
- Bansak, C., Morin, N. & Starr, M. (2007) Technology, Capital Spending and Capacity Utilization. *Economic Inquiry*, 45(3) July, s. 631-645
- Barnett, A. G., Pols, J. C. & Dobson, A. J. (2005) Regression to the Mean: What it is and How to Deal With it. *International Journal of Epidemiology*, 34(1), s. 215-220
- Bjørnstad, R. (2006) Er det økte sykefraværet tegn på mer inkluderende eller ekskluderende arbeidsliv? *Økonomiske analyser*, 6, s. 48-55.
- Brock, W. & Durlauf, S. N. (2001) Interactions-Based Models. I: J. Heckman & E. Leamer, red. *Handbook of Econometrics*, Amsterdam (North-Holland): Elsevier Science B.V, s. 3299-3371
- Cameron, A. C. & Trivedi, P. K. (2010) *Microeconometrics Using Stata*, College Station (Texas): Stata Press
- Clemens, M. A., Radelet, S., Bhavnani, R. R. & Bazzi, S. (2012) Counting Chickens when they Hatch: Timing and the Effects of Aid on Growth. *The Economic Journal*, 122(561) June, s. 590-617
- Coleman, J. S. (1990) *Foundation of Social Theory*, Cambridge (Massachusetts): Harvard University Press
- Cont, R. & Löwe, M. (2010) Social Distance, Heterogeneity, and Social Interactions. *Journal of Mathematical Economics*, 46(4) March, s. 572-590
- Dale-Olsen, H., Nilsen, M. N. & Schøne, P. (2015) Imitation, Contagion, or Exertion? Using a Tax Reform to Reveal How Colleagues' Sick Leave Influence Work Behavior. *The Scandinavian Journal of Economics*, 117(1) January, s. 57-83
- Eckhoudt, L., Gollier, C. & Schlesinger, H. (2005) *Economic and Financial Decision under Risk*, Princeton (New Jersey): Princeton University Press
- Folketrygdloven. Lov 28. februar 1997 nr. 19 om folketrygd.
- Glaeser, E. L., Scheinkman, J. A. & Sacerdote, B. (2003) The Social Multiplier. *Journal of the European Economic Association*, 1, s. 345-353
- Hem, K. G. (2011) *Bedriftenes kostnader ved sykefravær*. Rapport nr. A19052. Oslo: SINTEF
- Hesselius, P. (2009) Is Leisure Contagious? The Relationship Between Sickness Absence and Spousal Retirement. *National Institute Economic Review*, 209(1) July, s. 104-115
- Høyskolen i Lillehammer (2010) *Sykefraværprosedyre: Aktiv sykemelding*. [Internett] Tilgjengelig fra: http://www.hil.no/HMS-system/sykefravaersprosedyre_aktiv_sykmelding [Lest 29.02.2016]
- Ichino, A. & Maggi, G. (2000) Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm. *The Quarterly Journal of Economics*, 115(3), s. 1057-1090
- Lindbeck, A., Palme, M. & Persson, M. (2016) Sickness Absence and Local Benefit Cultures. *The Scandinavian Journal of Economics*, 118(1) January, s. 49-78

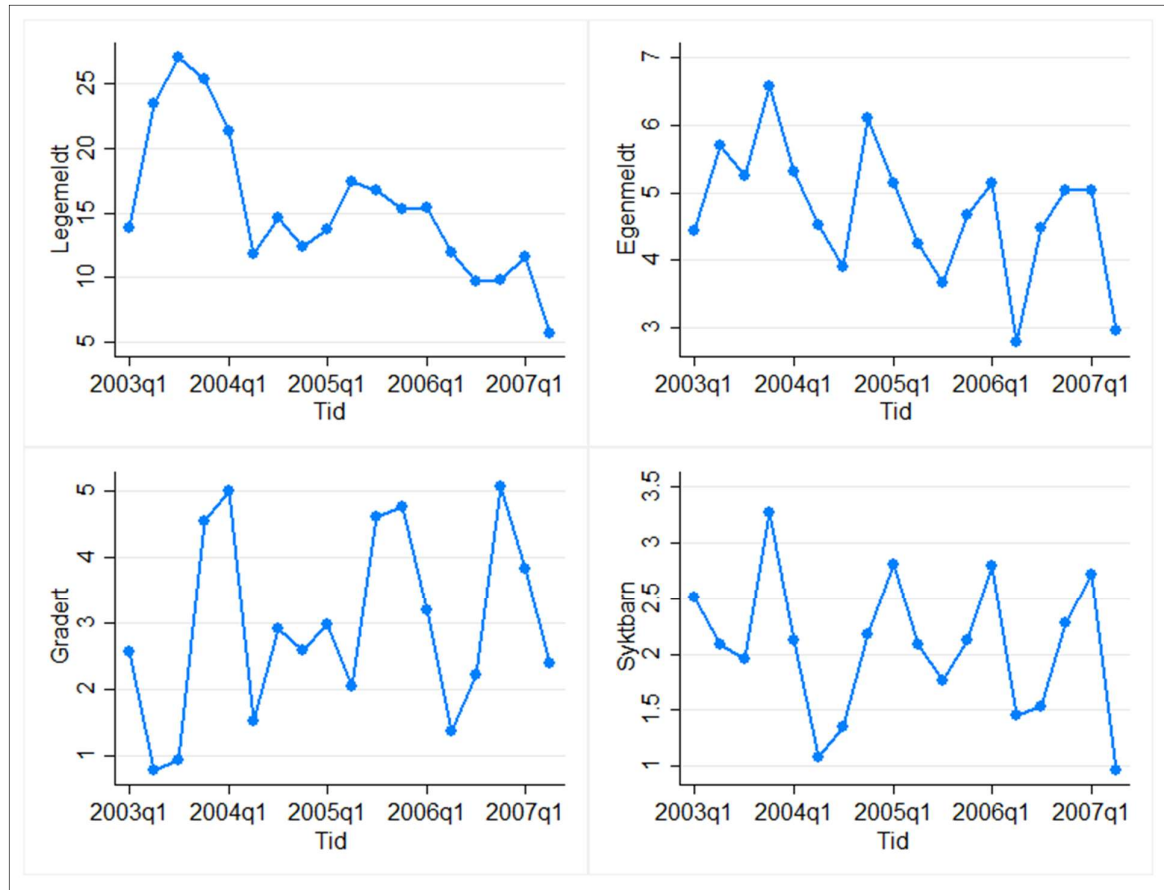
- Lindbeck, A. & Persson, M. (2010) A Continuous Theory of Income Insurance. *CESifo Working Paper*, Nr. 3097
- Manski, C. F. (1993) Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3) July, s. 531-542
- Markussen, S., Mykletun, A. & Røed, K. (2012) The Case of Presenteeism – Evidence from Norway's Sickness Absence Insurance Program, *Journal of Public Economics*, 96, s. 959-972
- Markussen, S. & Røed, K. (2015) Social Insurance Networks. *The Journal of Human Resources*, 50(4), s. 1081-1113
- Markussen, S., Røed, K., Røgeberg O. J. & Gaure S. (2009) The Anatomy of Absenteeism. *IZA Discussion Paper*, Nr. 4240
- Mas, A. & Moretti, E. (2009) Peers at Work. *American Economic Review*, 99(1) March, s. 112-145
- Mastekaasa, A. (2015) Sosiale og demografiske variasjoner i korttidssykefraværet. *Søkelys på Arbeidslivet*, 32, s. 3-20
- Moffitt, R. (2001) Policy Interventions, Low-Level Equilibria and Social Interactions. I: S. Durlauf & H.P. Young, red. *Social Dynamics*. Washington (D.C): Brooking Institution Press, s. 45-82
- Mykletun, A. & Brinchman, B. (2013) *Effekter av tiltak under IA-avtalen – Rapport fra forskermøte på oppdrag fra Arbeidsdepartementet*. Oslo: Nasjonalt folkehelseinstitutt.
- NAV (2015) *Grunnbeløpet i Folketrygden*. [Internett] Tilgjengelig fra: <https://www.nav.no/no/NAV+og+samfunn/Kontakt+NAV/Utbetalinger/Grunnbeloet+i+folketrygden> [Lest 05.01.2016]
- Nordberg, M. & Røed, K. (2003) Absenteeism, Health Insurance, and Business Cycles. *HERO Working Paper*, Nr. 17
- NOU 2000:27. *Sykefravær og uførepensjonering: Et inkluderende arbeidsliv*.
- Ose, S. O. (2010). *Kunnskap om sykefravær: nye norske bidrag*. Rapport nr. A14516. Trondheim: SINTEF
- Ose, S. O., Jensberg, A. M., Hem, K. G., Johnsen, A., Lippestad, J., Paulsen, B., Mo, T. O. & Saksvik, P. Ø. (2009). *Evaluering av IA-avtalen (2001-2009)*. Rapport nr. A11947. Trondheim: SINTEF
- Potters, J., Sefton, M. & Vesterlund, L. (2005) After you – Endogenous Sequencing in Voluntary Contribution Games, *Journal of Public Economics*, 89(8) August, s. 1399-1419
- Proba samfunnsanalyse (2014a) *Internasjonal sammenligning av sykefravær – del 1*. Rapport nr. 2014-03. Oslo: Proba samfunnsanalyse
- Proba samfunnsanalyse (2014b) *Internasjonal sammenligning av sykefravær – del 2*. Rapport nr. 2014-05. Oslo: Proba samfunnsanalyse
- Reed, W. R. (2013) A Note on the Practice of Lagging Variables to Avoid Simultaneity. *University of Canterbury Department of Economics and Finance Working Paper*, Nr. 32
- Rege, M., Telle, K. & Votruba, M. (2012) Social Interaction Effects in Disability Pension Participation: Evidence from Plant Downsizing. *The Scandinavian Journal of Economics*, 114(4) December, s. 1208-1239
- Regjeringen (2015) *Folketrygdens utgifter*. [Internett] Tilgjengelig fra: <https://www.regjeringen.no/no/aktuelt/folketrygdens-utgifter/id2456725/> [Lest 08.10.2015]

- Rieck, K. & Vaage, K. (2010) *Social Interactions at the Workplace: Exploring Sickness Absence Behavior*. [Internett] Tilgjengelig fra:
<<http://www.hha.dk/nat/peterj/gradinprowe/krieck.pdf>> [Lest 02.09.2015]
- Sacerdote, B. (2001) Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, 116(2) May, s. 681-704
- Smeby, L., Bruusegard, D. & Claussen, B. (2009) Sickness Absence: Could Gender Divide be Explained by Occupation, Income, Mental Distress and Health? *Scandinavian Journal of Public Health*, 37(7) September, s. 674-681
- Soetevent, A. R. (2006) Empirics of the Identification of Social Interactions; An Evaluation of the Approaches and Their Results. *Journal of Economic Surveys*, 20(2) April, s. 193-228
- SSB (2012) *Tidsbruksundersøkelsen, 2010*. [Internett] Tilgjengelig fra:
<<https://www.ssb.no/kultur-og-fritid/statistikker/tidsbruk/hvert-10-aar/2012-01-18>> [Lest 10.04.2016]
- SSB (2013) *Deltidsarbeid – blir forskjellene utjevnet?* [Internett] Tilgjengelig fra:
<https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/deltidsarbeid-blir-forskjellene-utjevnet> [Lest 13.04.2016]
- SSB (2015) *Sykefravær 3. kvartal 2015*. [Internett] Tilgjengelig fra:
<<https://www.ssb.no/arbeid-og-lonn/statistikker/sykefratot/kvartal/2015-12-10>> [Lest 23.02.2016]
- SSB (2016) *Sykefravær ifølge Arbeidskraftsundersøkelsen*. [Internett] Tilgjengelig fra:
<<https://www.ssb.no/statistikkbanken/selectvarval/Define.asp?subjectcode=&ProductId=&MainTable=AKUSykefravaer&nvl=&PLanguage=0&nyTmpVar=true&CMSSubjectArea=arbeid-og-lonn&KortNavnWeb=sykefratot&StatVariant=&checked=true>> [Lest 29.02.2016]
- Stigler, S. M. (1997) Regression Towards the Mean, Historically Considered. *Statistical Methods in Medical Research*, 6(2) April, s. 103-114
- Verbeek, M. (2012) *A Guide to Modern Econometrics*, Chichester (West Sussex), John Wiley & Sons Ltd
- Woolridge, J. M. (2014) *Introduction to Econometrics*, Andover (Hampshire), Cengage Learning EMEA

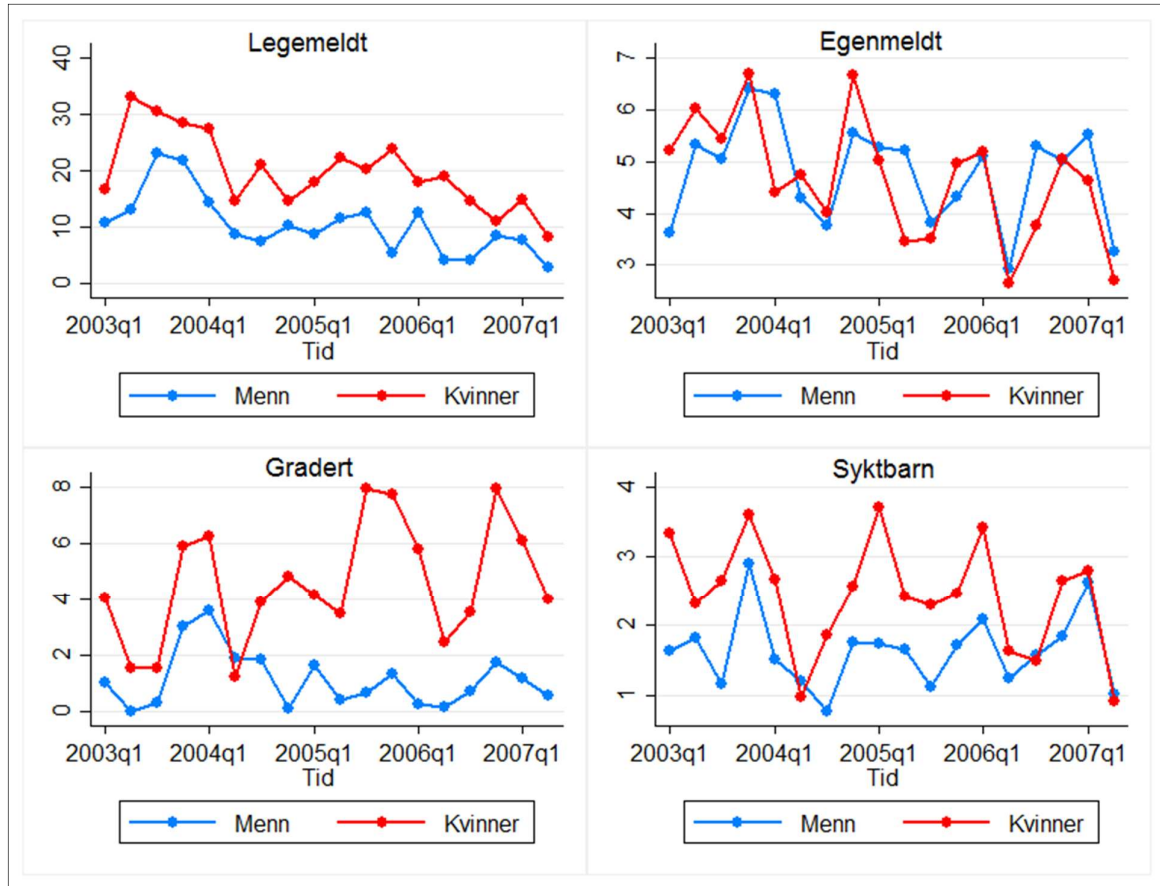
Appendiks

A.1 Figurer

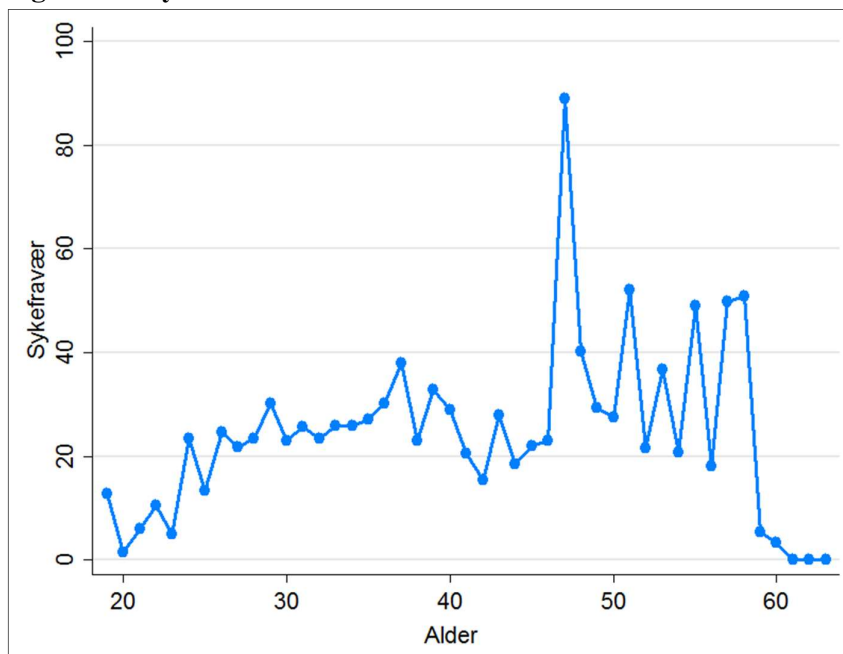
Figur A.1: Utvikling i ulike typer sykefravær over tid



Figur A.2: Utvikling i ulike typer sykefravær over tid for kvinner og menn



Figur A.3: Sykefravær etter alder



A.2 Tabeller

Tabell A.1: Utfallsvariabler (nyansatte)

Utfallsvariabel	Gjennomsnitt	Standardavvik	Forklaring
Sykefravær	20.543	60.328 (48.892)	Antall timer samlet sykefravær per kvartal
Legemeldt	11.858	49.813 (42.424)	Antall timer legemeldt sykefravær per kvartal
Egenmeldt	3.739	8.517 (7.698)	Antall timer egenmeldt sykefravær per kvartal
Gradert	2.992	24.307 (21.148)	Antall timer gradert sykefravær per kvartal
Syktbarn	1.561	5.784 (4.459)	Antall timer sykefravær pga. sykt barn per kvartal
Aktiv	.385	10.686 (10.206)	Antall timer aktiv sykefravær per kvartal
Ubegrunnet	.004	.181 (.174)	Antall timer ubegrunnet sykefravær per kvartal
<i>Observasjoner</i>	1528		
<i>Individ</i>	167		

Merk: Within standardavvik i parentes.

Tabell A.2: Forklaringsvariabler (nyansatte)

Variabel	Gjennomsnitt	Standardavvik	Forklaring
Mann	.465	.498 (0)	Andel menn per kvartal
Alder	31.165	8.941 (.884)	Alder målt per kvartal
Lønn	282109.4	39787.59 (12599.03)	Årsinntekt målt per kvartal
Ansiennitet	4.365	3.628 (3.245)	Lengde på ansettelse på tidspunkt t
Salg	94.760	131.304 (94.988)	Salg per kvartal
Teamvariabel			
Team_sykefravær	16.25026	16.556 (12.744)	Gjennomsnittlig teamsykefravær ekskludert individ i
Team_mann	.4531393	.139 (.059)	Gjennomsnittlig andel menn ekskludert individ i
Team_alder	31.38271	5.251 (1.628)	Gjennomsnittlig teamalder ekskludert individ i
Team_lønn	286926.2	21373.96 (9612.139)	Gjennomsnittlig teamlønn ekskludert individ i
Team_ansiennitet	7.863107	3.151 (2.523)	Gjennomsnittlig teamansiennitet ekskludert individ i
Team_salg	46.17262	54.042 (40.742)	Gjennomsnittlig teamsalg ekskludert individ i
Team_størrelse	19.6398	7.674 (3.830)	Teamstørrelse per kvartal
<i>Observasjoner</i>	1528		
<i>Individ</i>	167		

Merk: Within standardavvik i parentes.

Tabell A.3: Utfallsvariabler (flytterne)

Utfallsvariabel	Gjennomsnitt	Standardavvik	Forklaring
Sykefravær	24.028	51.855 (44.386)	Antall timer samlet sykefravær per kvartal
Legemeldt	13.324	43.737 (39.287)	Antall timer legemeldt sykefravær per kvartal
Egenmeldt	5.040	8.670 (7.863)	Antall timer egenmeldt sykefravær per kvartal
Gradert	2.915	17.916 (16.343)	Antall timer gradert sykefravær per kvartal
Syktbarn	2.723	7.607 (5.961)	Antall timer sykefravær pga. sykt barn per kvartal
Aktiv	.012	.505 (.491)	Antall timer aktiv sykefravær per kvartal
Ubegrunnet	.012	.505 (.491)	Antall timer ubegrunnet sykefravær per kvartal
<i>Observasjoner</i>	1769		
<i>Individ</i>	120		

Merk: Within standardavvik i parentes.

Tabell A.4: Forklaringsvariabler (flytterne)

Variabel	Gjennomsnitt	Standardavvik	Forklaring
Mann	.483	.499 (0)	Andel menn per kvartal
Alder	31.830	6.838 (1.229)	Alder målt per kvartal
Lønn	293931.6	43456.34 (15718.58)	Årsinntekt målt per kvartal
Ansiennitet	7.494	5.195 (4.813)	Lengde på ansettelse på tidspunkt t
Salg	109.480	116.307 (89.023)	Salg per kvartal
Teamvariabel			
Team_sykefravær	21.719	19.785 (16.834)	Gjennomsnittlig teamsykefravær ekskludert individ i
Team_mann	.488	.121 (.085)	Gjennomsnittlig andel menn ekskludert individ i
Team_alder	31.405	3.790 (2.457)	Gjennomsnittlig teamalder ekskludert individ i
Team_lønn	291869.4	20120.05 (13958.98)	Gjennomsnittlig teamlønn ekskludert individ i
Team_ansiennitet	7.082	4.181 (3.903)	Gjennomsnittlig teamansiennitet ekskludert individ i
Team_salg	71.505	61.995 (49.606)	Gjennomsnittlig teamsalg ekskludert individ i
Team_størrelse	16.024	5.948 (4.585)	Teamstørrelse per kvartal
<i>Observasjoner</i>	1769		
<i>Individ</i>	120		

Merk: Within standardavvik i parentes.