

Arbeidsinnvandring og sykefravær

Virkingen av EU-utvidelsen i 2004

av

Guri Natalie Jordbakke

Masteroppgave

Masteroppgaven er levert for å fullføre graden

Master i samfunnsøkonomi

Universitetet i Bergen, Institutt for økonomi

Juni 2015

Forord

Arbeidet med masteroppgaven har både vært utfordrende og lærerikt, nå som siste ord er skrevet gjenstår det kun å takke. Jeg vil begynne med å rette en spesiell takk til min veileder Kjell Vaage, for forslag til oppgave, tilrettelegging av data og både faglige og språklige tilbakemeldinger underveis. Jeg ønsker også å takke min bror for gjennomlesning og språklige kommentarer.

Sist men ikke minst vil jeg takke min kjære Martin for både korrekturlesning, god støtte gjennom hele prosessen og for at du alltid har tro på meg.

Data som er benyttet i denne oppgaven er hentet fra Statistisk Sentralbyrå (SSB) sin database FD-trygd. Registerdata fra SSB med informasjon om utdanning og inntekt er påkoblet datasettet. Den overnevnte institusjonen er ikke ansvarlig for analyser eller tolkninger som er gjort i denne oppgaven.

Guri Natalie Jordbakke

Guri Natalie Jordbakke, Bergen 31. mai 2015

Sammendrag

Arbeidsinnvandring og sykefravær

Virkningen av EU-utvidelsen i 2004

av:

Guri Natalie Jordbakke, Master i samfunnsøkonomi

Universitetet i Bergen, 2015

Veileder: Kjell Vaage

Norge opplevde fra 2004 en betydelig reduksjon i det totale sykefraværet, en reduksjon som ofte er sett i sammenheng med reformen om en styrket portvaktrolle i 2004. Men på samme tidspunkt skjer det også en endring som skal ha betydning for det norske arbeidsmarkedet. I 2004 utvides EU med 10 nye land, noe som resulterte i en enorm økning i arbeidsinnvandrere til Norge, og som har bidratt til å øke den totale sysselsettingen. Samtidig viser det seg at arbeidsinnvandrere fra nye EU-land har en relativt liten andel som mottar helserelaterte trygdeytelser. På bakgrunn av dette stilles følgende spørsmål i denne oppgaven: Kan den store strømmen av ny arbeidskraft ha bidratt til den reduksjonen en så i sykefraværet i 2004?

For å undersøke dette er det i oppgaven delt inn i fire landgrupper. Norge som referansegruppen, ikke-vestlig og vestlig innvandrere og tilslutt nye EU land som har vært av særlig interesse. Den deskriptive delen viste at den nye arbeidsinnvandringen har hatt en betydelig større reduksjon i sykefraværet enn de andre gruppene. Denne observasjon ble fulgt opp i den empiriske analysen som har sett på både tilbøyeligheten for og varigheten av sykefravær. Resultatene viste at sannsynligheten for sykefravær er redusert for alle grupper, dette gjelder særlig gruppen med nye EU-land. Etter 2004 så har nye EU-land hatt en betydelig lavere sannsynlighet for sykefravær enn majoritetsbefolkningen, noe som har stryket troen på oppgavens hypotese. For øvrig er resultatene for de andre gruppene at ikke-vestlige har høyere, mens vestlige har lavere tilbøyelighet for sykefravær enn majoritetsbefolkningen. Forskjellene i sykefravær mellom gruppene og majoritetsbefolkningen kan delvis forklares av demografiske og arbeidsmarkedsforskjeller, men det gjenstår fortsatt en signifikant forskjell etter å ha kontrollert for disse.

Statistikkprogrammet Stata/IC 13.1 og Microsoft Excel 2010 er benyttet i oppgaven.

Innholdsfortegnelse

Forord	ii
Sammendrag	iii
Innholdsfortegnelse	iv
Tabeller.....	v
Figurer	vi
1.0 Innledning.....	1
2.0 Bakgrunn	4
2.1 Utvikling i og sammensetningen av sysselsettingen	4
2.2 Innvandring og arbeidsinnvandring etter EU-utvidelsen	6
2.3 Sykepengeordningen i Norge.....	9
2.4 Sykefraværet i Norge	12
2.5 Oppsummering	16
3.0 Teori og tidligere studier	17
3.1 Neoklassisk arbeidstilbudsmodell	17
3.2 Tidligere studier	24
3.3 Oppsummering	27
4.0 Økonometrisk metode	29
4.1 Paneldata	29
4.2 Lineære paneldatamodeller	29
4.3 Binære valg modeller	34
4.4 Paneldatamodeller med binær avhengig variabel.....	37
4.5 utfordringer ved estimering	38
5.0 Data og deskriptiv - statistikk og statistisk analyse.....	40
5.1 Datasettet.....	40
5.2 Sentrale variabler.....	41
5.3 Deskriptiv statistikk.....	44
5.4 Sykefravær	52
5.5 Oppsummerende kommentarer	55
6.0 Empirisk analyse	57
6.1 Introduksjon	57
6.2 Landbakgrunn og sannsynligheten for sykefravær	58
6.3 Landbakgrunn og varigheten på sykefravær	68
6.4 Kritisk vurdering	74
7.0 Oppsummering og avsluttende kommentarer	77
8.0 Litteraturliste	79
9.0 Vedlegg	84

Tabeller

Tabell 6.1 Gjennomsnitt for avhengige- og forklaringsvariabler for de fire landgruppene og det totale utvalget.	57
Tabell 6.2: Modeller på sykefravær (binær variabel), for det totale utvalget (alle grupper)...	59
Tabell 6.3 Modeller på sykefravær (binær variabel), for de fire landgruppene.....	66
Tabell 6.4 Modeller på sykefravær, ln(antall sykefraværsdager), for det totale utvalget (alle grupper).....	69
Tabell 6.5 Modeller på sykefravær, ln(antall sykefraværsdager), for de fire landgruppene...	73

Figurer

Figur 2.1 Sysselsetting i prosent, Arbeidskraftundersøkelsen (AKU), SSB.	5
Figur 2.2 Innvandring etter innvandringsgrunn 1990-2012.	6
Figur 2.3 Sykefravær i Norge for perioden: 2001-2014, totalt og etter kjønn.	14
Figur 5.1 Aldersutvikling i perioden 1998-2008 for fire landgrupper, etter sortering ned til 16-67 år.	45
Figur 5.2 Utviklingen i andel menn i perioden 1998-2008 for de fire landgrupper.	46
Figur 5.3 Utviklingen i sysselsetting (prosent) i perioden 1998-2008 for de fire landgruppene.	47
Figur 5.4 Utviklingen i andelen som jobber fulltid i perioden 1998-2008 for de fire landgruppene.	49
Figur 5.5 Reallønnsutviklingen i perioden 1998-2008 for de fire landgruppene.	50
Figur 5.6 Gjennomsnittlig utdanningsnivå, målt i antall år (høyeste grad) i perioden 1998-2008, for tre av landgruppene.	51
Figur 5.7 Utviklingen i prosenandelen med ett eller flere sykefraværstilfeller, for perioden 1998-2008.	53
Figur 5.8 Utviklingen i prosenandelen med ett eller flere sykefraværstilfeller i perioden 1998-2008, for menn.	54
Figur 5.9 Utviklingen i prosenandelen med et eller flere sykefraværstilfeller for perioden 1998-2008, kvinner.	55

1.0 Innledning

I en velferdsstat som Norge er sykefravær et viktig samfunnsøkonomisk tema, da et høyt sykefravær er forbundet med kostnader både for individet, virksomheter og samfunnet som helhet. I 2015 var statens direkte utgifter til sykepenger for arbeidstakere anslått til å være på rundt 35,2 milliarder kroner (Arbeids- og sosialdepartementet 2014). I tillegg kommer arbeidsgivers utgifter i form av utbetalte sykepenger de første 16 fraværsdagene. De vel 35 milliarder kronene inkluderer heller ikke andre indirekte kostnader, som utgifter til tiltak for å redusere sykefraværet. Bedriftene vil også oppleve produksjonstap knyttet til fravær. Endelig vil det, selv med 100 prosents lønnskompensasjon, være individuelle kostnader ved sykefravær i form av en lavere lønnsutvikling og lavere sannsynlighet for at bli ansatt og/eller forfremmet (Markussen 2012).

Utover 1990-tallet hadde Norge en betydelig vekst i sykefraværet, og det ble av den grunn i 2001 opprettet et tre-part samarbeid om et inkluderende arbeidsliv som hadde som mål å redusere fraværet. Evaluering av avtalen noen år senere viste en ytterligere økning i sykefravær. Dette var en medvirkende årsak til en reform i 2004 (Sosialdepartementet 2004a) som hadde som viktigste virkemiddel å styrke legens portvaktrolle.

Etter 2004 har det vært en betydelig reduksjon i sykefraværet, og reduksjonen er ofte sett i sammenheng med reformen innført i 2004 som Markussen (2009a) karakteriserer som «closing the gate». Mens en port lukkes, åpnes en annen, da EU ble utvidet med 10 nye land i 2004 og ytterligere to i 2007, åpnet dette porten for fri flyt av arbeidskraft, tjenester og fri etableringsrett fra nye EU-land til Norge. Dette ble starten på en enorm økning i arbeidsinnvandring til Norge som har ført til innvandringsrekorder flere år på rad. Generelt sett har disse innvandrerne hatt en høy sysselsetting som har bidratt til sysselsettingsvekst i Norge, samtidig som at det er en klart lavere andel av dem som mottar helserelaterte ytelser (Bratberg og Vaage 2015). Siden dette har vært en gruppe som har bidratt sterkt til vekst i sysselsettingen, vil det være rimelig å anta at denne gruppen også er stor nok til å kunne påvirke andre arbeidsmarkedstørrelser, slik som det samlede sykefraværet. På denne bakgrunn spørres det i denne oppgaven om den store strømmen av ny arbeidskraft kan ha bidratt til den reduksjonen en så i sykefraværet fra 2004.

Det er en rekke studier som undersøker det generelle sykefraværet og hva som kan forklare dette, både internasjonalt og nasjonalt. Men nasjonalt er det heller få empiriske bidrag som belyser innvanderens tilpasning av sykefravær relativt til norskfødte. Av de relativt få studiene som eksisterer, er fokuset på å forklare forskjellene en finner i sykefraværet mellom Norge og

innvandrere med ulike landbakgrunn, der det ofte er skilt mellom vestlig og ikke-vestlige innvandrere. En finner vanligvis at vestlige innvandrere har lavere sykefravær enn majoritetsbefolkningen, mens ikke-vestlige innvandrere har noe høyere sykefravær (Dahl, Hansen og Olsen 2010 og Hansen, Holmås, Islam og Naz 2014). Av studier som blant annet ser på den «nye» arbeidsinnvandringen etter 2004 kan en nevne Bratsberg, Rauum og Røed (2014b) og Kann, Bragstad og Thune (2013). Bratsberg et al (2014b) ser på innvandrere i arbeidslivet og sosialforsikring under et, mens Kann et al (2013) ser på forbruket av helserelaterte ytelser og finner at reduksjonen i forbruket hadde vært mindre uten innvandringen.

Likevel er det ingen studier, som undertegnede har kjennskap til, som har sett på/prøvd å belyse sammenhengen mellom den økte arbeidsinnvandringen siden 2004 og reduksjonen i sykefraværet en så på samme tid. Denne oppgaven fokuserer derfor på sammenheng mellom arbeidsinnvandring og sykefraværet. Ved å kombinere disse aspektene vil studien kunne bidra med ny og viktig empirisk kunnskap om sammenhengen. Oppgavens datagrunnlag strekker seg over perioden 1998-2008, en begrensning er derfor at en mangler observasjoner for flere år der innvandringen har vært rekordhøy. Likevel vil oppgaven kunne belyse utviklingen, da jeg har observasjoner for en viktig periode, både årene rett før og etter 2004. Interessen rundt innvandringens påvirkning av arbeidsmarkedsstørrelser, slik som sykefravær, skyldes ofte finansielle konsekvenser og bekymringen rundt såkalt «velferdsinnvandring» som gir et økt press på de allerede eksisterende, og generøse, tryde- og sykelønnsordningen. Det kan imidlertid tenkes at arbeidsinnvandringen i stedet letter presset på slike ordninger, særlig på kortsikt, på grunn av at de har lavere forbruk av disse ordningene, men også øker sysselsettingen slik at sykefraværsprosenten vil reduseres.

For å kunne si noe om eventuelle sammenhenger mellom det reduserte sykefraværet og økt arbeidsinnvandringen, er det i denne oppgaven laget fire landgrupper, med Norge som referansegruppe. Av disse fire gruppene vil den viktigste være gruppen som samler alle de landene EU ble utvidet med i 2004 og 2007. En vil videre se på de relative tilbøyelighetene for sykefravær og varighet av sykefravær, der en skiller mellom før og etter 2004 med en step-dummy.

I kapittel 2, bakgrunnskapittelet, presenteres utviklingen i innvandringen til Norge, med et spesielt fokus på arbeidsinnvandringen etter 2004. Videre i kapittelet ser en på de institusjonelle forholdene i sykepengeordningen, samt utviklingen og reformer av betydning for det norske

sykefraværet. I det neste kapitlet, kapittel 3, belyses den økonomiske teorien om sykefravær, som vil være av relevans for oppgavens problemstilling. Kapittel 4 vil gi en gjennomgang av oppgavens økonometriske metoder. Her legges det spesiell vekt på paneldatamodeller og binære valgmodeller. Kapittel 5 er todelt, der den første delen ser på oppgavens datasett med fokus på sorteringer som er gjort og sentrale variabler. Deretter vil sentral deskriptiv statistikk bli presentert og diskutert. Den empiriske analysen i kapittel 6 er todelt, med først fokus på sannsynligheten for sykefravær før en går videre å ser på varigheten av sykefraværet. Avslutningsvis blir det en diskusjon og oppsummering i kapittel 7.

2.0 Bakgrunn

Som vi skal komme nærmere tilbake til senere i oppgaven, er det mange ting som tyder på at sysselsetting og sykefravær henger sammen. I dette kapittelet skal vi derfor se på både sysselsettingsgraden og sammensetningen av ulike grupper i arbeidsstyrken. Videre skal en se nærmere på innvandringen til Norge, med fokus på EU-utvidelsens betydning. Avslutningsvis ses det på sykefraværet i Norge, med fokus på utviklingen og sentrale reformer.

2.1 Utvikling i og sammensetningen av sysselsettingen

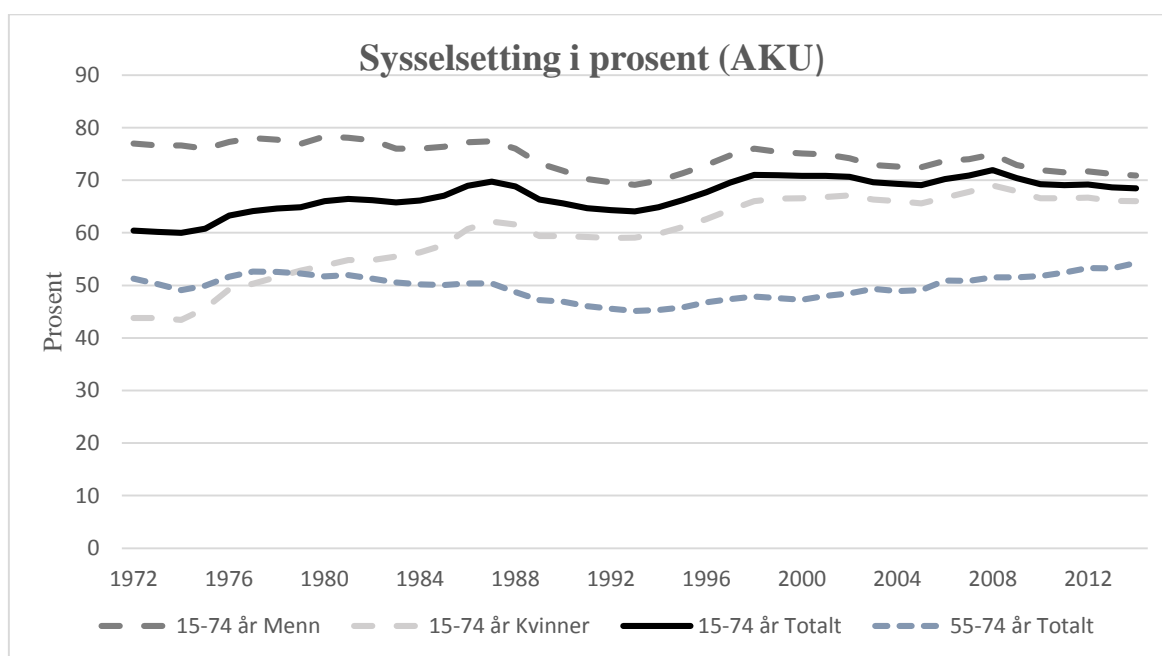
Flest mulig i arbeid er et sentralt mål i oppbygningen av den norske velferdsstaten, da trygdeordningen i stor grad blir finansiert av allmenn beskatning på inntekt, formue og forbruk. I 1992 ble betegnelsen arbeidslinja introdusert¹, dette kom som en følge av økende offentlige utgifter til trygdeordningene sammen med en bekymring for at mange ble ekskludert fra arbeidslivet. Underliggende her er at en var urolig for at tilgang til arbeidsfrie inntekter gav mindre incentiv til å forsørge seg selv. Med en noe strammere sosialpolitikk, betegnet som arbeidslinja, skulle en ha et sterkere fokus på å forene høy sysselsetting med et sikkerhetsnett for de som ikke kan forsørge seg selv i arbeidslivet (Hatland og Terum 2013). Norge har i flere år vært blant de landene i Europa med høyest sysselsetting. I 2013, for eksempel, var sysselsettingen i Norge på 68,65 prosent av befolkningen mellom 15-74 år, med henholdsvis 66,1 prosent og 71,2 prosent for kvinner og menn (SSB). Sysselsatte blir i AKU-statistikken definert som «personer i alderen 16-74 år som utfører inntektsgivende arbeid av minst en times varighet i undersøkelsesuka, samt personer som har et slikt arbeid, men som var midlertidig fraværene pga. sykdom, ferie, lønnet permisjon e.l.» (Bø og Håland 2002, s: 7). EU og OECD opererer med lignende tall, men da som andel av befolkningen mellom 15-64 år. Med dette utgangspunktet blir sysselsettingen enda høyere. Fra OECD-statistikk henholdsvis: 75,4 totalt, 73,5 for kvinner og 77,3 for menn i 2013 (OECD 2015).

En av hovedårsakene til at Norge har så høy sysselsetting internasjonalt sett, er at det er relativt mange kvinner som er yrkesaktive, selv om en stor andel av disse jobber deltid. Fra figur 2.1 ser en økningen i sysselsettingen fra midten av 70-tallet, denne økningen skyldes i stor grad kvinners inntog i arbeidslivet. Kvinners sysselsetting har økt jevnt fra 1972 med nær 44 prosent til toppåret 2008 med 69 prosent. Den største økningen for kvinners yrkesdeltakelse finner en i løpet av 80-tallet (Bjørnstad 2006). Når det gjelder sysselsettingen blant eldre arbeidstakere, som i figur 2.1 er definert som arbeidstakere mellom 55-74 år, har denne økt siden 1992 frem

¹ Men det er enighet om at den har røtter helt tilbake til 1930-tallet.

mot 2014. I 2014 var sysselsettingen for eldre arbeidstakere vel 4 prosentpoeng høyere enn i 1972. Sammensetningen av de yrkesaktive er viktig da det vil påvirke nivået på sykefraværet. Både kvinner og eldre arbeidstakere har typisk et høyere sykefravær enn menn, og en vil derfor forvente et høyere sykefravær når deres andel i sysselsettingen øker, alt annet likt. Bjørnstad (2006) går langt i forklare det økte sykefraværet med økt kvinnelig og eldre andeler i sysselsettingen. Det er i midlertid uenighet i hvilken grad dette kan forklare det økte sykefraværet, Ekspertgrupperapport (2010, s:7) rapporterer at «Noe økt yrkesdeltakelse blant kvinner og eldre kan bare forklare en liten del av endringen i sykefraværet siste år». De viser også til at Norge og Sverige har svært lik kvinnelig yrkesdeltakelse, samtidig som Sverige har et lavere sykefravær enn Norge.

Figur 2.1 Sysselsetting i prosent, Arbeidskraftundersøkelsen (AKU), SSB.



Kilde: SSB

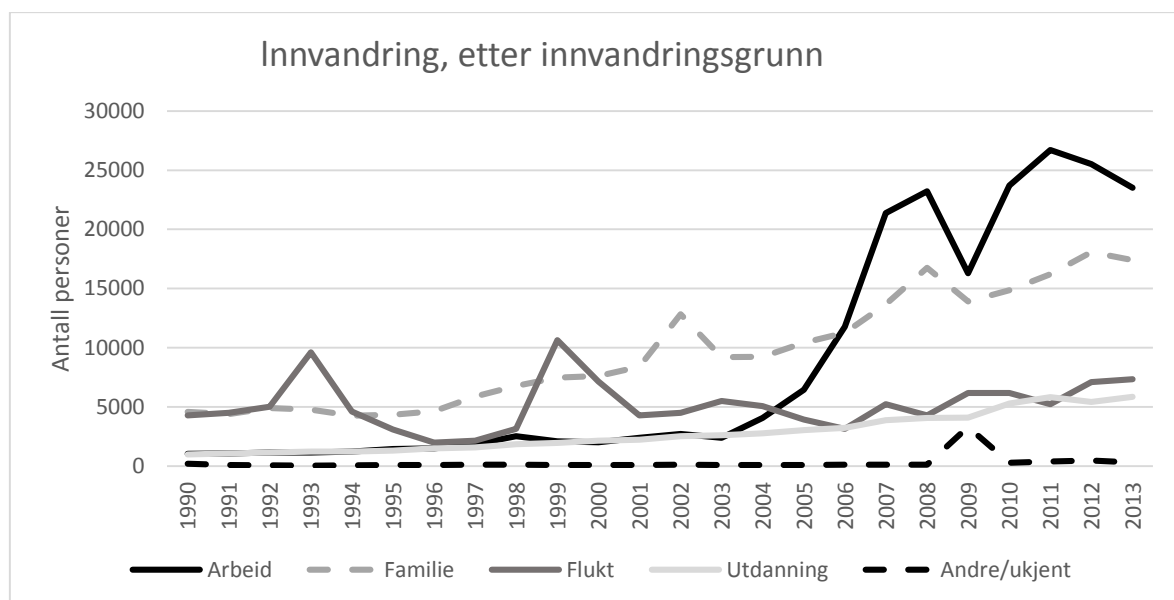
På 2000-tallet har det også vært vekst i sysselsetting (Olsen 2013), i figur 2.1 synes denne veksten å være varierende, men med en vekst fra 2003/2004 og frem til finanskrisen. En årsak til at denne rapporterte sysselsettingsveksten ikke vises i figuren, er at figuren tar utgangspunkt i AKU-statistikk. Innvandringsgruppene i AKU-tallene har relativt høy frafall og frafallsskjevhet, på grunn av dette er det stor usikkerhet knyttet til denne gruppen (Villund 2012). Ser en i stede på registerdata, tabell A1 i vedlegg, ser en en tydelig vekst i sysselsettingen, spesielt fra 2005 til 2008. I denne perioden var en vesentlig del av økningen i den totale sysselsettingen grunnet økt sysselsetting blant innvandrere. Fra 2002-2012 var 60

prosent av økningen i sysselsettingen knyttet til økningen av sysselsatte innvandrere. Den største økningen skyldes sysselsatte fra EU-landene i øst (Olsen 2013).

2.2 Innvandring og arbeidsinnvandring etter EU-utvidelsen

Stort innvandringsoverskudd til Norge er et relativt nytt fenomen. Før 1960-tallet har nettoinnvandringen stort sett vært negativ eller svakt positiv. Men i 1967 kom det første varselet om et nytt flyttemønster, gjerne kalt den «nye innvandringen»². Fra dette året har Norge stort sett hatt et innvandringsoverskudd³ (Vassenden 2012). På 70 og 80-tallet var det en økning i arbeidsinnvandrere helt til innvandringsstoppen. Fra da av har innvandringen hovedsakelig bestått av familie og flukt- innvandrere (Brochmann 2010).

Figur 2.2 Innvandring etter innvandringsgrunn 1990-2012.



Kilde: SSB

Av figur 2.2 ser en at frem til 2005/2006 har kategoriene familie og flukt vekselvis vært de største innvandringsgruppene i Norge. Figur 2.2 viser også at innvandringen har økt relativt kraftig siden 2004, dette gjelder for alle innvandringsgrupper, men en ser spesielt en økning i arbeid som innvandringsårsak. Dette skiftet i hovedårsak til innvandring og økningen av innvandrere, kommer i kjølvannet av EU-utvidelsen.

EU har en målsetting om å skape et felles arbeidsmarked for hele EU-området, og gjennom EØS-avtalen er reglene for de tre friheter: fri bevegelse for arbeidstakere, fri bevegelse av

² Ny, i form av at innvandrerne nå kom fra «nye» /andre land enn tidligere innvandrere.

³ Innvandringsoverskudd eller positiv nettoinnvandring er når det innvandrer flere personer til landet enn det utvandrer fra landet.

tjenester og fri etableringsrett, også gjort gjeldende for Norge (Utenriksdepartementet 2012). I prinsippet betyr dette at en arbeidstaker har rett til å reise til andre EU-land for å søke arbeid i inntil 3 måneder. Om en får arbeid, har en rett til opphold og familiegjenforening. Er en ansatt i innenlandsk virksomhet, har en plikt til å betale skatt og trygdeavgifter samtidig som en da har lik tilgang til velferdsytelsene, deriblant sykepenger⁴. I tilknytning til fri bevegelse av tjenester så kan utenlandske leverandører med tidsbegrensede tjenester fritt medbringe egen arbeidskraft, enten utstasjonert eller utsendte arbeidstakere for å utføre oppdrag. Disse arbeidstakere omfattes ikke av vertslandets trygde- og velferdsordninger, med mindre oppholdet eller kontrakten varer lengre enn 24 måneder⁵. Den siste friheten, fri etablering som også omfatter enmannsbedrifter, vil være underlagt de samme lover og regler som nasjonale virksomheter, da de driver permanent virksomhet i vertslandet (Friberg, Dølvik og Eldring 2013).

Disse og mange andre EU-regler som kan anvendes, gir rom for norske oppdragsgivere å rekruttere arbeidere fra EU gjennom flere kanaler. De fem viktigste er: Ansettelse av EU-borgere i en norsk etablert bedrift, tjenestekjøp av EU-firma, altså utsendte eller utstasjonerte arbeidstakere. En tredje måte er å leie arbeidskraft fra EU-firma, eventuelt kan en leie EU-arbeidskraft fra en nasjonal bedrift, og til slutt kan en foreta tjenestekjøp fra EU-enmannsfirma (ibid). Disse ulike kanalene omfattes av forskjellige regler knyttet til sosiale avgifter, skatt, arbeidsvilkår, lønnsdannelse og ansvarsforhold, som gir oppdragsgiver og oppdragstaker incentiv til å velge en tilpasning som er mest økonomisk gunstig for begge parter. Her finnes det derfor også flere gråsoner.

I 2004 ble EU utvidet med ti nye land, Polen, Estland, Latvia, Litauen, Slovenia, Slovakia, Tsjekkia, Ungarn, Malta og Kypros, og ytterligere to land i 2007, nemlig Romania og Bulgaria. Denne utvidelsen har betydd mye for tilstrømming av arbeidsinnvandrere i Norge de siste ti årene. Både i 2006, 2007, 2008 og senest i 2011 ble det rapportert om nye innvandringsrekorder. Som nevnt tidligere har den totale innvandring for alle innvandringsgrunner økt siden 2004, der arbeid som innvandringsårsak har hatt den desidert største veksten⁶, og er nå den største gruppen innvandrere, etterfulgt av familiegjenforening.

⁴ Mer om dette under sykepengeordningen.

⁵ De omfattes imidlertid av utstasjoneringsdirektivet, etter dette gjelder en kjerne av lovfestede arbeidsvilkår deriblant minstelønn/allmenngjort tariffavtale (Friberg et al 2013:14)

⁶ I tillegg så er dette data som ikke får med seg korte arbeidsopphold, så tallene for arbeidsinnvandrere antas å være enda høyere.

Så hvorfor har Norge blitt et attraktivt mottakerland for denne type innvandring? Klassisk økonomisk teori legger vekt på de makro-økonomiske forskjellene mellom mottakerland og avsenderland, men også alternative mottakerland. Individet som beslutningstaker velger det mottakerlandet (eller å bli værende i hjemlandet) som gir mest nytte, der inntektsforskjeller sammen med sannsynlighet for å få arbeid står sentralt (Borjas 1989)⁷. Friberg, Tronstad og Dølvik (2012) finner i sin survey-studie bevis på slike økonomiske incentiv til å immigrere. Funnene går ut på at Norge har blitt en stor og attraktiv arbeidsgiver for nye EU-land som i stor grad skyldes faktorer på etterspørselssiden. For eksempel har sterk økonomisk vekst som førte til stor etterspørsel etter arbeidskraft, spesielt i konjunkturutsatte segmenter (eks: bygg og anlegg), ført til større mulighet for arbeid, som i teorien vil tilsi større sannsynlighet for å få arbeid. I tillegg til høy etterspørsel, nevnes de relativt høye lønningene, spesielt blant fagarbeidere («less-skilled workers») i forhold til avsenderlandet, men også alternative mottakerland, som en viktig faktor.

Som en så av figur 2.2, så fikk innvandringen, spesielt arbeidsinnvandringen, et kraftig fall rundt 2008/2009, men kom raskt tilbake i 2010. Fallet ses i sammenheng med finanskrisen, et sjokk Norge i mindre grad enn andre Europeiske land⁸ ble rammet av. Dette har bidratt til at Norge etter 2009 har pekt seg ut som et relativt attraktivt destinasjonsland (ibid). I tillegg til relative konjunktur og lønnsforskjeller så vil andre faktorer som påvirker omfanget av arbeidsinnvandringen være offentlig velferdsordninger, nærhet i geografi, kultur og språk og nettverk (Bakken, Bråthen, Handal og Aarethun 2009).

De største avsenderlandene er Polen og Sverige, mens Litauen kommer på en tredjeplass og har hatt en formidabel økning siden 2003. Men en slik økning i innvandringen vil kunne påvirke den økonomiske bærekraften til velferdsstaten. En har sett at de kortsiktige økonomiske virkningene har vært positive for velferdsstaten, da arbeidsinnvandringen har bidratt til sysselsetting og produksjon i en situasjon med allerede sterk vekst; dette har gitt økte skatteinntekter og har dempet inflasjonspresset i den norske økonomien. Kann et al (2013) finner at nedgangen i bruken av helserelevante ytelser hadde vært mindre uten innvandringen, også når en justerer for alderssammensetningen. På utgiftssiden har arbeidsinnvandrere så langt hatt et relativt lavt forbruk av velferdsgodene (Friberg et al 2013). Norge ser derfor ut til å ha

⁷ Annen økonomiske teori ser på husholdningen som beslutningstaker som ønsker å spre sin inntektskilde og risiko. Annen ikke-økonomiske teori nevner faktorer som for eksempel: Nettverk, geografi, språk etc.

⁸ Alternative mottakerland Norge «konkurrerer» mot.

hatt en nettogevinst av den høye arbeidsinnvandringen frem til nå. Spørsmålet er om denne gevinsten vil opprettholdes på lengre sikt.

Arbeidsinnvandrere er overrepresentert i jobber som har fysiske belastninger, lav status og lav lønn og dårlig arbeidsmiljø, og dette er gjerne forbundet med større bruk og avhengighet av velferdsgoder (ibid). Friberg et al (2013) argumenter for at det er usikkerhet knyttet til hvordan arbeidsinnvandringen vil gi utslag på velferdsstaten økonomiske bærekraft på lengre sikt. Det vil blant annet avhenge av om arbeidsinnvandrere greier å opprettholde det høye sysselsettingsnivået eller ikke.

2.3 Sykepengeordningen i Norge.

Sykepengeordningen blir i Norge regulert av Folketrygden. Folketrygden kan ses på som et nasjonalt sosialt forsikringssystem, som skal forsikre medlemmer mot tap av arbeidsinntekt. Dette forsikringssystemet dekker spesielt for utgifter ved arbeidsløshet, svangerskap og fødsel, aleneforsørgere for barn, sykdom og skade, uførhet, alderdom og dødsfall (Folketrygdloven 1997). Så lenge det er mål om å komme tilbake i arbeid, skal medlemmene sikres for inntektstap ved sykdom og skade ved å motta sykepenger eller arbeidsavklaringspenger, avhengig av årsak og lengde på sykefraværet. Om det er snakk om mer varig uførhet, skal en overføres til uførepensjon.

Medlemskap i folketrygden er automatisk og pliktig for de som er bosatt eller arbeider i Norge eller på norsk kontinentalsokkel⁹. Som medlem gir sykepengeordningen full økonomisk kompensasjon (100 prosent) under sykdom eller skade¹⁰, i inntil et år, og opptil seks ganger folketrygdens grunnbeløp (G)¹¹. Rettigheter til sykepenger er betinget på en rekke krav, blant annet må en ha vært i arbeid i fire uker, også kalt opptjeningstid; i tillegg er det kun pensjonsgivende inntekt som er grunnlaget for sykepenger. Dessuten må inntektsgrunnlaget utgjøre mer enn 50 prosent av folketrygdens grunnbeløp (G) (Sosialdepartementet 2004a).

Som arbeidstaker har en rett til en egenmeldingsperiode på tre dager (åtte hvis arbeidsplassen har skrevet under Inkluderende Arbeidsliv (IA)-avtalen¹²), som kan benyttes opptil 4 ganger i løpet av et år. Etter egenmeldingsperioden må en ha legemelding for gyldig sykefravær.

⁹ Unntak ved noen typer arbeidsinnvandring som nevnt ovenfor.

¹⁰ Her regnes ukens 5 virkedager som sykepengedager.

¹¹ På en rekke arbeidsteder, blant annet hele offentlig sektor, er det avtalt at en får utbetalt 100 prosent kompensasjon også for lønn over 6G.

¹² Mer om IA-avtalen senere i kapittelet.

Arbeidsgiver kompensere arbeidstakeren med sykepenger de 16 første dagene i sykdomstilfelle, deretter tar NAV over utbetalingen av sykelønnen.

For arbeidsinnvandrere som bosetter seg i Norge¹³ vil de generelle ordningene gjelde, men ved kortvarig arbeidsopphold vil de i hovedsak ikke være dekket av trygdeytelsene. Generelt sett vil arbeidsinnvandrere som faller innunder EU-reglene ha lettere adgang til de aller fleste trygdeytelser enn andre innvandrere.

For eksempel vil en arbeidsinnvandrere fra EU-land få sykepenger etter en dag i arbeid, betinget på at de har arbeidet opptjeningstiden i et annet EU-land. I tillegg vil de som omfattes av EØS-reglene kunne ta med ytelsene til hjemlandet da bostedsvilkåret til ytelsene oppheves av EØS-reglene, dette for å forhindre en såkalt «lock in» effekt. For arbeidsinnvandrere som ikke omfattes av EØS-reglene, så vil reglene være strengere, minimum fire ukers opptjeningstid i Norge og ytelsen kan i alle hovedsak ikke tas med under opphold utenfor Norge (Arbeids- og inkluderingsdepartementet 2008).

Sykepengeordningens opprinnelse og endringer/reformer¹⁴

Loven om sykeforsikringen ble vedtatt 18. september 1909¹⁵, der retten på sykepenger i all hovedsak gjaldt lavtlønnede arbeidstakere (industriarbeidere). Siden den tid har ordningen blitt endret en rekke ganger, og i 1970 ble sykestrygden en del av folketrygden. Arbeidstakere var ikke lenger medlem av ulike trygdekasser med forskjellige betingelser. I 1978 fikk Norge en ny sykelønnsordning, som i hovedtrekk også gjelder i dag. Alle arbeidstakere hadde rett på 100% kompensasjon ved sykdom i inntil ett år¹⁶, i tillegg fikk alle rett til å bruke egenmelding de tre første fraværsdagene. Arbeidsgiver betalte de 14 første dagene før staten tok over betalingsansvaret. Dette gjorde at alle arbeidere hadde fått et felles regelverk med svært generøse ytelser (Hagelund 2014 og Sosialdepartementet 1977).

I løpet av 80-tallet fikk sykepengeordningen flere innskjerpinger, noen av disse var at minste inntektsgrunnlaget ble hevet først til 10 000 så ytterligere til ½ G, mens den øvre inntektsgrensen ble redusert fra 8G til 6G. I tillegg ble antall sykepengedager redusert fra 6 til

¹³ «Personer som kommer til Norge blir regnet som bosatt eller folketrygdens regler dersom oppholdet er ment å vare i minst ett år.» (Arbeids- og inkluderingsdepartementet 2008).

¹⁴ For en utdypende oversikt over endringer se Hagelund (2014, s:190)

¹⁵ Loven inkludert da også helsetjenester.

¹⁶ Ingen skulle ha karensdag. Karensdag er dag(er) uten støtte, før en har rett på sykepenger.

5 dager i uken og fra 52 til 50 uker i året. Til slutt ble ordningen som sykemelding II innført¹⁷ (Hagelund 2014).

Lovendringer i 2004, en strengere portvaktrolle.

Bakgrunnen for lovendringen i 2004 (sosialdepartementet 2004b) var at evalueringen av avtalen om et mer inkluderende arbeidsliv (IA) viste til en økning av sykefraværet. Økningen var primært i det legemeldte fraværet og spesielt fraværet utover arbeidsgiverperioden. Regelendringen i folketrygdloven bestod av fem punkter, med et overordnet tema om å styrke legens portvaktrolle. Første punkt var å innføre et aktivitetskrav. Dette innebærer at en arbeidstaker så tidlig som mulig skal prøve seg i arbeidsrelaterte aktiviteter, der inaktivitet over 8 uker skal medisinsk begrunnes. Videre ble det lagt føringer om at legen alltid skulle vurdere gradert sykemelding fremfor aktiv sykemelding. Gradert sykemelding er når arbeidstakeren kan utføre deler av sine vanlige arbeidsoppgaver, mens aktiv sykemelding at arbeidstakere er til stede og utfører alternative arbeidsoppgaver¹⁸. Det primære finansieringsansvaret vil da henholdsvis være delt på NAV og arbeidsgiver og NAV. I tillegg ble det en eksplisering av legens ansvar om funksjonsvurdering til den sykemeldte samt at sanksjonene overfor leger som ikke fulgte de nye reglene, ble skjerpet (ibid). Tidlig i 2004 ble også arbeidsgiverperioden utvidet fra 14 til 16 dager.

Inkluderende arbeidsliv (IA)

Den første intensjonsavtalen om et mer inkluderende arbeidsliv (IA-avtalen) ble inngått mellom regjeringen, arbeidstakere og arbeidsgivere (trepartssamarbeid) 3.oktober 2001 og varte vel fire år frem til 2005. Avtalen har blitt videreført tre ganger, der gjeldende avtale ble underskrevet 1. mars 2014 og gjelder til utgangen av 2018. Formålet med avtalen er å forebygge og redusere sykefraværet, hindre utstøtning samt frafall fra arbeidslivet ved å styrke jobbnærværet til den enkelte arbeidstaker. Dette skal gjøres ved å ha bedre og tidligere oppfølging av sykemeldte med større grad av aktiv dialog mellom arbeidsgiver og arbeidstaker for å tilrettelegge. Fokuset skulle rettes over på sykemeldtes funksjonsevnen med tanke på hans/hennes arbeidsoppgaver-hvilke oppgaver var mulig å utføre, og ikke diagnose/sykdom.

¹⁷ Sykemelding II anvendes når sykefraværet har vart i 8 uker, da skal et utvidet spørreskjema med undersøkelser og prognose rapporteres til trygdekontoret.

¹⁸ Ordningen med aktiv sykemelding er i dag avskaffet.

Aktivisering og nærværreformen¹⁹

I november 2009 oppnevnte arbeids- og inkluderingsdepartementet en ekspertgruppe som skulle komme med forslag om administrative tiltak for å redusere sykefraværet. Årsakene for opprettelsen var at den tredje IA-avtalene snart utløp (1.mars 2010) og det økende sykefraværet i 2009. Ekspertgruppen tok utgangspunkt i reformen som ble innført i 2004 som ga betydelig resultat ved reduksjon i sykefraværet. Her stod gradert sykemelding og aktivisering sentralt, og ekspertgruppens forslag er bygget videre på dette.

Ekspertgruppene la frem en 6-delt forslag på tiltak, som de mente ville bidra til reduksjon i sykefraværet uten å øke utstøtningen fra arbeidsmarkedet. Det første og mest sentrale elementet var Aktivisering- og nærværreformen, i tillegg ble 5 andre punkter nevnt²⁰ som mer eller mindre skulle støtte opp under denne foreslåtte reformen på ulike arenaer. Blant de viktigste elementene i reformen var en reel endring der gradert sykemelding skulle være hovedregel ved sykefravær.

2.4 Sykefraværet i Norge

Det finnes en rekke kilder til sykefraværstatistikk og Norge har fire følgende hovedkilder til sykefraværstatistikk: *Den sentrale sykefraværstatistikken*, *NAV's sykefraværstatistikk*, *Arbeidskraftundersøkelsen (AKU)* og *NHO's fraværstatistikk*. *Den sentrale sykefraværstatistikken* ble for første gang publisert i 2001, og publiseres kvartalsvis av NAV og SSB. Statistikken rapporterer både det egenmeldte og det legemeldte sykefraværet. Det sist nevnte er sykefravær f.o.m 4. dag (8 for IA bedrifter), og er basert på NAV's registreringer av sykemeldingsattester. Mens statistikkproduksjonen av egenmeldt sykefraværet skjer ved hjelp av en bedriftsbasert utvalgsundersøkelse, en undersøkelse SSB har ansvar for (Olsen 1997). Den sentrale sykefraværstatistikken rapporteres i prosent og måler antall tapte dagsverk av avtalt dagsverk. NAV's sykefraværstatistikk er et register over alle legemeldte sykefravær, det vil si sykefravær utover 16 dager, og omfatter variablene antall avsluttede sykefraværstilfeller og antall erstattede sykepengedager. På lik linje med den sentrale sykefraværstatistikken rapporterer *NHO's fraværstatistikk* fraværprosent på 1-3 dager samt alt fravær over 3 dager, altså både egen- og legemeldt fravær. Statistikken samles inn av et utvalg av medlemsbedriftene og var frem til opprettelsen av den sentrale sykefraværstatistikken den eneste kilden til

¹⁹Det vil bli en kort gjennomgang av reformen, da den er sentralt i sykefraværdiskusjonene, men utenfor oppgavens datamaterial (som strekker seg fra 1998-2008).

²⁰ Disse 5 punktene var: kunnskapsbaserte tiltak for å redusere sykefraværet, kompetanseløft, sykemelding ved opphør av arbeidsforhold og permittering og styrke NAV's kontroll og innkrevingsmuligheter.

egenmeldte sykefravær. Fraværstatistikken ble avvirket i 2002 etter at den sentrale sykefraværstatistikken da også inkluderer data på alle næringer. Til slutt har en AKU sin fraværstatistikk som dekker alle individ i alderen 15-74 som er registrert bosatt i Norge²¹. Statistikkgrunnlaget innhentes fra en utvalgsundersøkelse, som gjennomføres hvert kvartal, der et tilfeldig utvalg intervjues angående arbeid og fravær i referanseuken (Berge 2008). Sykefraværprosenten beregnes ut ifra antall sysselsatte som var borte hele referanseuken, grunnet egen sykdom, som prosent av alle ansatte.

AKU²² er utformet primært for å få informasjon om sysselsetting og arbeidsledighet, mens sykefraværstatistikken anses mer som et bi-produkt av dette. Sykefraværstatistikken blir likevel ofte brukt til internasjonal sammenligning i mangel på annen god sammenlignbar statistikk (Berge, Johannessen og Næsheim 2012)²³.

Ut ifra sammenligninger på tvers av land, med AKU data, viser det seg at Norge har et relativt høyt sykefravær. OECD rapporter at Norge har det høyeste sykefraværet (og uføretrygdede) av OECD-land og peker på den generøse sykepengeordningen som gir incentiv til å trekke seg ut fra arbeidsmarkedet (OECD 2014). Som en så tidligere, har andre forklaringer på det relativt høye sykefraværet vært sammensetningen av arbeidsstyrken, men flere studier viser at dette (sammen med flere andre bakgrunnsfaktorer) kun kan forklare små deler av forskjellen mellom sykefraværet i Norge og andre land. I en studie av Gleinsvik, Klingberg og Mastekaasa (2014) finner de blant annet at Norge har et vesentlig høyere sykefravær enn sammenligningslandene og at demografiske kjennetegn, kontraktsforhold, arbeidstid eller kjennetegn ved virksomheten kun forklarer denne forskjellen i noen grad²⁴. Analysen sammenligner Norge med land det er naturlig å sammenligne seg med, spesielt nordiske land der Sverige har flest likhetstrekk med utformingen av den norske sykepengeordningen. Ekspertgruppen, i tilknytning aktivisering- og nærværreformen, har også trukket frem Sverige når de vurderte størrelsen på det norske sykefraværet.

Nasjonalt sett har det vært et stort fokus på kjønnsforskjellene i sykefraværet, men rapporten fra ekspertmøte om kjønnsforskjeller i 2011 understreker at det høye sykefraværet er en større

²¹ Individ som er sysselsatt med opphold mindre enn seks måneder eller pendler til Norge med bosted i utlandet omfattes ikke av denne undersøkelsen. Som nevnt tidligere vil dette gi underrapportering av en gruppe innvandrere som oppgaven er interessert i.

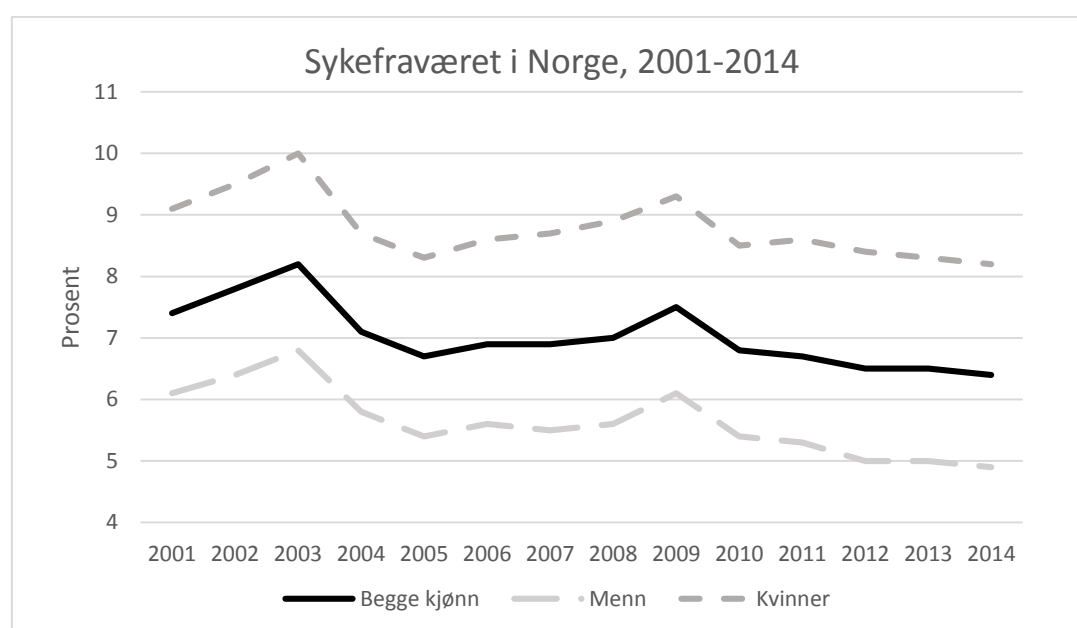
²² AKU, eller LFS («Labour Force Survey») gjennomføres i alle EU-land, undersøkelsen er standardisert for at produktene skal være mest mulig sammenlignbare.

²³ I deres rapport konkluderer de med at på tross av forskjeller i trygdesystemet i de landene de har undersøkt, så synes måten AKU måler sykefravær å være sammenlignbar (Berge et al 2012).

²⁴ Landene Norge ble undersøkt oppimot var Danmark, Finland, Island, Nederland, Sverige og Storbritannia.

«bekymring» enn kjønnsforskjellen. «Det var i møte bred enighet om at det høye totale sykefraværet er mer påfallende i Norge enn kjønnsforskjellen i sykefravær» (Mykletun og Vaage 2012). Ser en på det samlede sykefraværet i Norge i figur 2.3, som er både egenmeldt og legemeldt ²⁵ for begge kjønn, ligger sykefraværsprosenten mellom 6-9 prosent for perioden 2001 til 2014. Figuren illustrerer også godt nivåforskjellene mellom kjønn der differansen er på rundt 3 prosentpoeng hele den angitte perioden. Sykefraværet er stigende frem til 2003 før det får et relativt kraftig fall rundt 2003/2004, for både kvinner og menn. Sykefraværet får igjen en topp rundt finanskrisen 2008/2009 før det igjen sank frem til 2014.

Figur 2.3 Sykefravær i Norge for perioden: 2001-2014, totalt og etter kjønn.



Kilde: SSB

Siden datamaterialet i denne oppgaven er begrenset til perioden 1992-2008 vil det inneholde hendelsene rundt 2003-2004, men bare så vidt de store bevegelsene rundt finanskrisen i 2008-2009. Som tidligere nevnt ble det i 2004 innført en lovendring i sykepengeordningen med et fokus på å styrke legens portvaktrolle. Denne reformen, omtalt som «da sykefraværet falt som en stein» i Markussen (2010), har blitt gjenstand for en rekke studier. Markussen (2010) finner betydelig effekt av lovendringen i 2004 på korttids- og langtidsfraværet, både når det gjelder insidens og lengde på sykefravær. Forfatteren tolker funnene som at reformen har vært med på redusere sykefraværet fra 2004.

Men det vil være svært vanskelig å skille effekten av reformen fra andre mulige faktorer som kan påvirke sykefraværet samme periode, spesielt når en se på langtidseffekten, noe også

²⁵ Både egen- og legemeldt sykefravær fra den sentrale sykefraværstatistikken.

Markussen (2009b, s: 24) påpeker «Yet, shocks occurring near reform implementation will also here be misinterpreted as *reform effect*». Et aspekt som ble belyst tidligere, som kan ses på som et sjokk i arbeidsmarkedet, er den kraftige økningen av arbeidsinnvandrere fra 2004, spesielt fra nye EU-land. Herunder så vi også at denne innvandringsgruppen har hatt høy sysselsetting og et relativt lav forbruk av trygdeordningene i forhold til majoritetsbefolkningen. Dette gir grunnlag for å stille spørsmål om reduksjonen i sykefraværet fra 2004 også kan ses i sammenheng med økt arbeidsinnvandring fra EU-utvidede land. Kann et al (2013) studerer bruken av helserelaterte ytelser²⁶ og finner at nedgangen fra 2005 gjelder generelt i befolkningen, men er større for arbeidsinnvandringen. Dette vil si at arbeidsinnvandringen har bidratt til en ytterligere nedgang i helserelaterte ytelser enn en ville hatt ellers. Her er det derfor interessant å se om en kan finne lignende sammenhenger for sykefraværet isolert for ulike landgrupperinger.

Dersom en finner lignende sammenheng med landtilhørighet og sykefravær kan det tenkes at dette er reformeffekt (2004), men jeg vil i mindre grad anta at reformen kan forklare en eventuell reduksjon i deres sykefravær. Det er flere årsaker til dette, for det første vil en rekke av disse individene fra EU-utvidede land ha innvandret på eller etter reform-tidspunktet. De fleste har derfor ikke opplevd noe endring i sykepengeordningen som følge av reformen i 2004²⁷. Slik at dersom nedgangen i sykefraværet for individ fra EU-utvidede land har redusert det totale sykefraværet, kan det argumenteres for at dette i en mindre grad er forårsaket av reformen i 2004.

Den høye arbeidsinnvandringen i siste halvdel av 2000-tallet gjør at vi må være ekstra nøye med hvordan vi måler og tolker sykefraværstallene. Det er rimelig at antall sykemeldte går opp når antall sysselsatte går opp. Men siden sykefraværet oftest måles som andel sykefraværende i forhold til antall sysselsatte, vil et konstant eller økende antall sykefraværstilfeller kunne ende opp som en reduksjon i sykefraværsprosenten, så lenge sysselsettingen har økt tilstrekkelig. Et annet aspekt er derfor at den store sysselsettingsøkningen blant nye EU-land i Norge kan i seg selv gi en reduksjon i sykefraværet.

²⁶ I helserelaterte ytelser inngår sykepengemottakere, uførepensjonister, mottakere av attføringspenger og rehabiliteringspenger, arbeidsavklaringspenger og tidsbegrenset uførestønad (Kann, Bragstad og Thune 2013).

²⁷ Selv om de ikke har opplevd endringen, vil de likevel stå ovenfor samme sykemeldingprosedyre som andre etter reformen.

2.5 Oppsummering

I dette kapitlet har en sett at det etter utvidelsen av EU i 2004 kom en stor strøm av arbeidsinnvandrere og disse har hatt en høy sysselsetting og har bidratt til en økning i den generelle sysselsettingen i Norge. Arbeidsinnvandrerne fra nye EU-land har hatt lettere tilgang til norske trygdeytelser enn innvandrere utenfra EU, likevel har de hatt et relativt lavt forbruk av disse ytelsene i forhold til majoritetsbefolkningen. Rundt samme tid, fra og med 2004, har det vært en relativt kraftig reduksjon i sykefraværet og denne reduksjonen knyttes gjerne til 2004-reformen som ble innført for å styrke legens portvaktrolle. Men denne reform-effekten kan for eksempel ha fanget opp tilbudssjokket av ny arbeidskraft fra nye EU-land. Jeg argumenterte videre for at en eventuell reduksjon i sykefraværet, for individ fra nye EU-land, i mindre grad kunne være forårsaket av reformen i 2004. Forbruket av helserelaterte trygdeytelser har vært relativt lav for denne gruppen. Dette gir grunnlag for å spørre om økningen i arbeidsinnvandringen kan ha bidratt til reduksjonen i sykefraværet etter 2004.

Før en skal gå nærmere inn på dette og undersøke oppgavens datasett, skal en se nærmere på teori om sykefravær som kan være med på å belyse problemstilling og økonometrisk bakgrunn.

3.0 Teori og tidligere studier

Den enkeltes helsetilstand vil være den viktigste årsaken til sykefravær. Men det er ikke å utelukke at sykefraværet også blir påvirket av både psykologiske, sosiale og økonomiske forhold, enten direkte eller indirekte. Innen økonomiske teoretiske modeller forklarer en i hovedsak sykefraværet med aktørens atferd (incentiv) og beslutningstaking påvirket av gitte ytre forhold. Men hva sier teorien om sykefraværet til innvandrere og hva tyder de empiriske resultatene på? I dette kapitlet skal en se på teoretiske perspektiv og empiriske funn som kan være med å belyse problemstillingen. Dette gjøres ved en todeling av kapitlet. I den første delen av kapitlet ser en nærmere på relevant teori med hensyn til sykefravær og innvandring. Siden det eksisterer få teoretiske bidrag som kombinerer disse to aspektene, vil jeg ta utgangspunkt i det teoretiske rammeverket rundt sykefravær som grunnlag for hypoteser rundt innvandrers tilpasning av sykefravær i Norge. I den andre delen av kapitlet gis en presentasjon av studier som har studert sykefravær (eller samlet trykgeforbruk) blant innvandrere i Norge. Studiene som blir presentert er derfor avgrenset til å være norske studier.

3.1 Neoklassisk arbeidstilbudsmodell

Den klassiske arbeidstilbudsmodellen tar utgangspunkt i individ som maksimere sin nytte ved å velge en optimal tilpasning av konsum og fritid. For å konsumere, er individet avhengig av inntekt som består av både arbeidsinntekt og eventuell arbeidsfrie inntekter. Problemet kan derfor ses på som en avveining av tidsbruk på arbeid og fritid eller godene konsum og fritid²⁸. Individene i modellen tar lønnen for gitt og tilpasser tilbudet av arbeidskraft deretter. Den optimale tilpasningen for individet vil være når lønn er lik det marginale bytteforholdet mellom fritid og konsum. I dette punktet vil individet på marginen være indifferent til litt mer fritid eller konsum. Effekten av økt lønn vil som vanlig bestå av en inntektseffekt som taler for mer fritid og substitusjonseffekten som tilsier mer arbeid, altså mindre fritid. Vanligvis antar en at substitusjonseffekten dominerer, og at økt lønn får individene til å arbeide mer. Om individet har inntekter utenom sin arbeidsinntekt vil individet velge å konsumere mer fritid (altså arbeide mindre). Fritid blir av den grunn sett på som et normalt gode i modellen.

Allen (1981) tar utgangspunkt i denne modellen når han ser på årsaker til sykefraværet på tilbudssiden. Han definerer fravær som «Absence results when an individual decides to engage in nonwork activity throughout a scheduled work period» (Allen 1981, s:78).

²⁸ Der prisen på konsum er normalisert til 1.

I motsetning til den klassiske modellen, ser Allen for seg at et individ som tilbys arbeid, da også en arbeidskontrakt, som har oppgitt både lønn og arbeidstimer. Individet kan derfor ikke justere arbeidstimene til sin optimale tilpasning av konsum og fritid. Men det vil likevel være slik at individet vil godta et jobbtillbud selv om antall arbeidstimer overstiger arbeidstimene i en optimal tilpasning. Dette gir individet incentiv til å konsumere mer fritid, og en kan justere denne mistilpasningen ved å ta ut sykefravær. I tillegg ser han på muligheten for at individ kan være fraværende fra jobb selv om de opprinnelig har for få arbeidstimer eller er optimalt tilpasset. Dette begrunner Allen med at faste arbeidstider gjør at individet hindres fra deltakelse i andre aktiviteter som kun er tilgjengelig i arbeidstiden. Elementene i disse andre aktivitetene vil variere fra dag til dag, som gjør at alternativkostnaden til arbeid varierer. Noen dager vil alternativkostnaden være så høy at en foretrekker fritid, til en annen aktivitet, enn arbeidsinntekt til konsum, og individet er derfor fraværende fra jobb.

Når arbeidsplassen opplever sykefravær/fravær så vil dette resultere i produksjonstap så lenge arbeideren blir erstattet av en som generelt er mindre effektiv, eller ikke blir erstattet i det hele tatt. Om fraværet er uforventet, vil dette også krever ressurser for å tilpasse arbeidsplassen til den «nye» situasjonen. For å kompensere for dette tapet, vil en i tillegg til å minste inntekten for de/den fraværende periodene bli straffet med et engangsbeløp (D) for hver fraværende periode. I Norge gis det 100 % lønnskompensasjon ved sykefravær. En kan likevel tenke seg en straff i form av redusert sannsynlighet for å bli forfremmet eller få lønnsforhøyelse, samt en økt sannsynlighet for å bli oppsagt.

Lar en t^A være tid med fravær (for eksempel antall timer eller dager) fra jobb, så kan en skrive:

$$D = D(t^A), \quad D' \geq 0, D'' \geq 0, \quad D(0) = 0$$

Det vil si at arbeidstakeren med flest fraværsdager betaler den største straffen, denne kostnaden antas å være ikke-avtakende, med en konstant eller økende straffestruktur. En kan nå analysere arbeidsoppmøteavgjørelsen innenfor det vanlige arbeid- og fritid-valg rammeverket.

Arbeidere maksimerer en nyttefunksjon som inneholder konsum (x) og fritid (L):

$$U = U(x, L) \quad (3.1)$$

Lar vi R være inntekt fra andre kilder enn arbeid, så har vi følgende budsjettbetingelse:

$$R + w(t^C - t^A) - D(t^A) - x = 0 \quad (3.2)^{29}$$

²⁹ Her er t^C tiden brukt på arbeid (C for konsum) og t^L er tidsbruk på fritid.

Budsjettbetingelsen viser at konsumet ikke kan være større enn inntekten, der den totale inntekten avhenger av hvor mye fravær en har og hvor mye en er på jobb. Dersom en ikke har noe fravær: $t^A = 0$ så vil inntekten være summen av arbeidsuavhengig inntekt og arbeidsinntekt (timelønn*antall timer).

Arbeideren står ovenfor følgende tidsbegrensning, der t er den totale tiden (for eksempel døgnet 24 timer):

$$t - t^C - t^L = 0 \quad \rightarrow \quad t = t^C + t^L \quad (3.3)$$

Som en ser av ligning (3.3), fordeler individet tiden mellom konsum (arbeid) og fritid, slik at ligningen holder ved likhet. Setter vi (3.2) og (3.3) inn i (3.1) og differensierer med hensyn på t^A , gir det oss følgende førsteorden/likevektsbetingelse:

$$U_L - (w - D)U_x = 0 \quad \rightarrow \quad U_L = (w - D)U_x \quad (3.4)$$

Ligning (3.4) illustrerer punktet der individet er indifferent mellom det å jobbe litt mer eller å ha litt mer fritid. I dette punktet er marginalnyten av fritid like stor som marginalnyten av konsumet som avhenger av antall arbeidstimer og straffen ved fravær D ³⁰. Effekten av en økning (reduksjon) i lønn er ikke entydig, da inntektseffekten og substitusjonseffekten drar i forskjellig retning³¹, men en antar vanligvis at substitusjonseffekten dominerer slik at en økt lønn tilsier økt arbeidstilbud og mindre fravær.

Ser en på tilfelle der en får inntektskompensasjon ved fravær, i form av en sykepengeordning, vil inntektseffekten helt eller delvis forsvinne, ettersom i hvor stor grad en blir kompensert. Ved 100% kompensasjon som i Norge, vil nytten av arbeid kun avhenge av straffen. Har en i tillegg ingen straff ($D=0$), vil modellen predikere at ingen ønsker å jobbe. Dette oppfattes å være en av svakheten ved modellen. Bildet justeres imidlertid når en inkluderer straff ved fravær ($D > 0$). Så lenge straffen er økende med fravær, så vil økt straff ($D \uparrow$) både redusere antall tilfeller av sykefravær, men også lengden på sykefraværet, alt annet likt. Straff kan som nevnt reflektere sannsynligheten for å bli oppsagt og kostanden til individet ved oppsigelse vil avhenge av

³⁰ Om individet ikke har noe fravær vil $D=0$, har en den klassiske løsningen at individet tilpasser sitt konsum der marginalnyten er lik hverandre.

³¹ Substitusjonseffekten, ved økt lønn, sier at alternativkostnad ved fritid øker (fritid blir relativt dyrere) som taler for å arbeide mer. Inntektseffekten for normale goder vil innebære relativ økning i etterspørselen etter fritid som taler for å arbeide mindre.

muligheten for å få ny jobb³². En tolkning vil derfor være at høy arbeidsledighet genererer et lavere sykefravær, og denne tolkningen skal en se videre på under teorien om effektivitetslønn.

Som nevnt vil sykefraværet avhenge av individets helsetilstand, men også familiesituasjonen. Dette kan inkorporeres ved å anta at dårlig helse og familiestiftelse øker nytte av fritid. Gitt alt annet, så vil økt nytte av fritid føre til økt sykefravær, motsatt vil redusert nytte av fritid føre til redusert fravær. Her kan en for eksempel trekke inn den ny arbeidsinnvandringen, som har relativt kort fartstid i Norge. Det kan argumenteres for at deres nytte av fritid er mindre enn for majoritetsbefolkningen, da de har sin «vanlige» omgangskrets (familie og venner) i sitt opprinnelsesland samtidig som de i mindre grad er integrert i den norske kultur og samfunn. Samtidig kan det argumenteres for at de har større nytte av arbeid(sinntekten) enn andre, da mange forholder seg til priser i sitt opprinnelsesland, slik at for gitt lønn genereres et større konsum. Det vil si at kjøpekraften for gitt lønn er større for dem enn andre som bort fast i Norge. Et annet poeng er at lønningene i Norge er relativt høye i forhold til avsenderlandene, som gjør at individene som har innvandret opplever en betydelig lønnsøkning. Dette vil også øke nytten av arbeid for denne gruppen, men nå relativt til før de innvandret. Prisen på fritid har derfor økt, og som følge av substitusjonseffekten vil individene ifølge modellen ønske å arbeide mer.

En hypotese er derfor at den nye arbeidsinnvandringen fra nye EU-land ha mindre nytte av fritid som følge av mindre nettverk og kjennskap i Norge, i tillegg vil lønnsøkningen individet opplever ved å innvandre til Norge gjør fritid relativt dyrere. Samtidig har de en større nytte av arbeid, som gir en teoretisk prediksjon om at de vil ha et relativt lavere sykefravær enn majoritetsbefolkningen³³.

Effektivitetslønn³⁴

Teori for effektivitetslønn tar utgangspunkt i at arbeidsgiveren(e) ikke kan observere arbeiderens innsats til enhver tid. Derfor har de tidvis kontroller der de som blir tatt i unnasluntring sies opp. Når en skal se på sykefravær i dette teoretiske rammeverket vil unnasluntring tolkes som (syke)fravær, for en mer realistisk tolkning tolkes «hvit» sykefravær ikke som unnasluntring. Hvit sykefravær vil være klare sykefraværstilfeller, mens sort vil være ren skoft. Med bakgrunn i den neoklassiske arbeidsmarkedsmoellen der alle arbeidere får

³² I praksis vil denne kostanden også avhenge andre trygdeordninger som kan erstatte arbeidsinntekten, som arbeidsledighetstrygd.

³³ Et interessant poeng er at dette kan tenkes å endre seg om oppholdet blir mer permanent. En kan da tenke seg at nytten av fritid øker og at de forholder seg mer og mer til det norske prisnivået, slik at de ikke har høyere kjøpekraft for gitt lønn. Dette vil tilsi at fraværet øker.

³⁴ Avsnittet er basert på Shapiro og Stiglitz (1984)

utbetalt markedslønn og det i likevekt ikke er arbeidsledighet, så er det verste som kan skje ved unnasluntring å få en oppsigelse. Men individet vil umiddelbart få en ny jobb (siden det ikke er noe arbeidsledighet) og bærer derfor ingen kostnad ved å unnasluntre, siden en får tilsvarende lønn i den nye jobben. Det vil si at ved mangelfull kontroll og full sysselsetting så vil arbeiderne ha incentiv til å sluntre unna (være fraværende).

For å forhindre unnasluntring, vil bedriften sette lønnen noe høyere enn markedslønn. Dermed vil arbeiderene som blir tatt i unnasluntring som sies opp, oppleve en reduksjon i lønn når de blir ansatt i en annen bedrift. Dette lønnspåslaget vil være kostnaden eller straffen ved unnasluntring. Dersom dette er lønnsomt for én bedrift, er det lønnsomt for alle bedrifter, og dette fører til at det generelle lønnsnivået øker. Incentivene for å ikke unnasluntre er nå forsvunnet. Et generelt høyere lønnsnivå fører til lavere etterspørsel etter arbeidskraft, dette fører igjen til tilbudsoverskudd av arbeidskraft, altså arbeidsledighet. Med arbeidsledighet er ikke arbeideren lengre sikret ny jobb ved oppsigelse, og dette vil nå reflektere den individuelle kostnaden ved oppsigelse. Arbeiderne får nå mindre incentiv til å unnasluntre, siden oppsigelse medfører en kostnad³⁵. Teorien viser derfor at lønn over markedslønn kan være effektivt.

Shapiro og Stiglitz (1984) viser ved bruk av teorien om effektivitetslønn at informasjonsstrukturen i arbeidsforholdet mellom arbeidere og arbeidsgiver kan forklare den «naturlige» ledigheten i likevekt. Ser en effektivitetsteorien i sammenheng med sykefravær, er det to elementer som påvirker individets incentiv til unnasluntring. For det første vil fraværet reduseres med økende sannsynlighet for å bli tatt i unnasluntring, altså mer kontroll. I tillegg så vil økt straff ved unnasluntring redusere unnasluntring, altså fraværet. Økt straff i form av lavere sannsynlighet for å få nytt arbeid, altså større arbeidsledighet ved en lavkonjunktur vil kunne redusere fraværet. En vet at arbeidsinnvandrere typisk arbeider i konjunkturutsatt virksomhet (Bratsberg, Raaum og Røed 2014a), en kan derfor tenke seg at de i større grad utsettes for denne straffen. Det vil si at sykefraværet til individ fra nye EU land kan tenkes å påvirkes mer av konjunkturer enn for andre grupper.

Sykefravær og konjunkturer

Teorien om effektivitetslønn tilsier at sykefraværet varierer negativt med arbeidsledighet, noe som betyr at det varierer positivt med konjunkturerne. En forklaring på dette er at arbeidernes kostnad ved å sykemelde seg øker når arbeidsledigheten øker, slik en så i teorien om

³⁵ Merk at arbeidsledigheten må være tilstrekkelig stor for at det skal være mere lønnsomt å jobbe enn å ta sjansen på å sluntre unna.

effektivitetslønn. Kostanden vil ta form som økt risiko for å få sparken og/eller lavere sannsynlighet for å få nytt arbeid, og disse kostnadene øker med arbeiderens individuell sykefravær. Når kostnaden ved sykefravær øker, ønsker arbeiderne å redusere risikoen for oppsigelse. Dette har en disiplinerende effekt, derav *disiplineringshypotesen*, på arbeiderne som tar ut mindre sykefravær.

Den disiplinerende effekten vil antakeligvis være sterkest for arbeidere i privat sektor (spesielt konjunkturutsatte bransjer), midlertidig ansatte og unge med lav ansiennitet (Nossen 2009). Arbeidsinnvandring, og kanskje spesielt den nye arbeidsinnvandringen siden 2004, er som nevnt tidligere typisk ansatt i konjunkturutsatte bransjer, i tillegg er mange midlertidig ansatt. Dette betyr at de i større grad vil kunne være påvirket av denne disiplinerende effekten, enn andre. En implikasjon er da at deres sykefravær vil være lavere under lavkonjunkturer enn for andre som ikke utsettes for denne risikoen i like stor grad, for eksempel offentlige ansatte.

En annen årsak til negativ korrelasjon mellom sykefravær og konjunkturer er at yrkesdeltakelse for personer med marginal tilknytning til arbeidslivet³⁶, er særlig konjunkturavhengig. Det vil si at ved en høykonjunktur vil yrkesdeltakelsen til de marginale arbeiderne være relativt høy. Siden disse har dårligere helse og derav høyere sykefravær så vil sykefraværet øke som følge av *sammensetningen* av den yrkesaktive befolkningen er endret. I tillegg til denne *sammensetningshypotesen* og disiplineringshypotesen er det også formulert en tredje hypotesen rundt sammenhengen mellom sykefravær og konjunkturer. Hypotesen, kalt *stresshypotesen*, tar utgangspunkt i at det økende aktivitetsnivået ved høykonjunkturer får helserelaterte konsekvenser som øker sykefraværet i travle tider. De to sistnevnte vil i mindre grad være relevante for oppgavens fokus, da en kan tenke seg at arbeidsinnvandrere fra nye EU-land er en selektert gruppe av relativt unge (menn) med relativt god helse.

Moral hazard – Sykefravær som en sosialforsikring³⁷

Sykepengeordningen kan bli sett på som en sosialforsikring fra et økonomisk perspektiv, der individet forsikrer seg mot arbeidsuførhet på grunn av sykdom eller skade. Som økonomiske aktører med risiko (for uførhet), har individene et ønske om å spre risikoen ut på et forsikringskollektiv. Under perfekte forhold vil private aktører tilby denne type forsikring, men problemet oppstår imidlertid når det er asymmetrisk informasjon mellom forsikringstaker og

³⁶ Med marginal tilknytning går blant annet individ med redusert helse eller funksjonsevner og derfor også relativt høyere sykefravær,

³⁷ Basert på Bratberg og Risa (2000).

forsikringskollektivet/selskapet. Disse problemene går under samle betegnelsen «Moral Hazard» og deles gjerne inn i tre: rapporteringsrisiko («hidden information»), atferdsrisiko («hidden action») og ugunstig utvalg («adverse selection»).

Dersom forsikringsselskapet ikke kjenner til individenes risikoprofil, så vil typisk markedsprisen på forsikringen reflektere gjennomsnittsriskoen til den forsikrede populasjonen. Lavrisikokundene vil da vurdere forsikringen som for dyr og trekke seg som forsikringstaker, og dette øker gjennomsnittsriskoen og dermed også forsikringsprisen. Den nye prisen vurderes som for høy for individ med relativt lav risiko i den forsikrede populasjonen. Dette medfører at den forsikrede populasjonen stadig får en høyere forsikringspris som reflekterer den økte gjennomsnittsrisiko. Tilslutt vil kun de med relativt høy risiko ønske å forsikre seg, som for forsikringsselskapet er et *ugunstig utvalg* å tilby forsikring til. En løsning på problemet med ugunstig utvalg er en kollektiv obligatorisk forsikringsordning, slik som i Norge med sykelønnsordningen.

Et annet problem er når forsikringstaker har mulighet til å påvirke sin sannsynlighet for et forsikringstilfelle uten at forsikringsselskapet/kollektivet kan observere det. Dette blir kalt *atferdsrisiko* og omfatter alt fra ren svindel til mer indirekte tilfeller som øker sannsynligheten for et forsikringstilfelle. Ved forsikring er det også en risiko for feil-rapportering (*rapporteringsrisiko*). Dette omfatter både direkte feil-rapportering av for eksempel ikke observerbare skader eller sykdom, eller at det er ulike oppfatninger om hva som kreves for et forsikringstilfelle. I litteraturen skiller en gjerne mellom for «illness», «disease» og «sickness» for å beskrive dette (Alexanderson et al 2003 og Dahl et al 2010). Der «illness» er individets egen oppfattelse av sin helse, mens «disease» vil omfatte en medisinsk diagnose og «sickness» karakteriseres som en rolle individet inntar som følge av «illness» eller en «disease». Et viktig poeng er derfor at det ikke nødvendigvis er slik at sykdom og sykefravær er det samme, da det kan være individuelle forskjeller på hva som ses på som et forsikringstilfelle. Siden sykefravær på over 3 eller 8 dager må sertifiseres, som regel av fastlege, kan det tenkes å være en mindre utfordring ved lange sykefravær som er fokuset i denne oppgaven. Det er imidlertid funnet at det i stor grad er pasienten selv som bestemmer om en får sykemelding eller ikke og ikke legen (Carlsen og Nyberg 2009).

Nyere hypoteser

Av nyere hypoteser er det verdt å nevne *omorganisering og nedbemanning, sosiale normer og legens portvaktrolle*. *Omorganisering og nedbemanning* går ut på at økt globalisering øker omorganiseringen, flytting av bedrifter, raske teknologiutvikling osv. For at bedriftene skal klare å konkurrere i et slik marked, må de stadig omstruktureres og foreta nedbemanning. Dette kan virke belastende for arbeiderne og dermed føre til redusert helse i ulik grad, som igjen fører til et høyere sykefravær. Hesselius, Nilson og Johansson (2009) viser i sin studie at *sosiale normer* kan være med på å påvirke sykefraværet. I hypotesen antas det at sosiale normer og holdninger i samfunnet kan påvirke individets nytte og derfor incentiv til å ta ut sykefravær. Dersom disse sosiale normene eller holdningene tillater et høyere sykefravær hos majoritetsbefolkningen enn for nye arbeidsinnvandrere, nettopp fordi normene ikke er innarbeidet. Kan det tenkes at sykefraværet blant arbeidsinnvandringen øker med oppholdstiden i Norge ettersom disse normene blir integrert. Legen beskrives gjerne som helsevesenets *portvakt*, da de aller fleste tilfeller av (langtids)sykefravær går gjennom fastlegen. Etter reformen innført i 2004, som satte fokus på en strengere portvaktrolle, og følgende reduksjon i sykefravær fant Markussen (2009a) at 1/3 av denne reduksjonen bestod av redusert tilgang til sykefravær. Dette viste derfor at legen har en sentral portvaktrolle som kan påvirkes.

3.2 Tidligere studier

Dahl et al (2010) studerer det langsiktige sykefraværet blant innvandrere i Norge for perioden 1992-2003. De peker på to spørsmål de ønsker å belyse i studien. For det første ser de på i hvilke grad sykefraværnivået avhenger av opprinnelsesland. I tillegg ønsker de å undersøke hvilke effekt oppholdstid i vertslandet, målt i oppholdsår, og andregenerasjonsinnvandrere har på sykefraværet. Ved å bruke norske registerdata som inneholder detaljert informasjon om alle innbyggere i Norge mellom 16-67 år, estimerer de effekten av landopprinnelse på sykefraværet. 8 landgrupper blir konstruert med Norge som referansegruppe, og de kontrollerer samtidig for en rekke sosioøkonomiske, demografiske og arbeidsmarkedsfaktorer.

Studien til Dahl et al (2010) er en av de første som ser nærmere på langsiktig sykefravær blant innvandrere i en vertsnasjon der innvandring er et relativt nytt fenomen og som har universale helse/sosialtjenester. De oppsummerer studien med fire hovedfunn: det er ingen forskjell på sykefraværet mellom etnisk norske og individ fra andre nordiske land, Vest-Europa og Øst-Europa etter å ha kontrollert for demografiske, sosioøkonomiske og arbeidsmarkedsfaktorer.

Observert forskjell mellom Øst-Europa (høyere) og Vest-Europa (lavere) og etnisk norske kan bli forklart ved forskjellige karakteristikk på de overnevnte kontrollfaktorene.

De finner også at sykefraværet er høyere for individ fra Asia (både menn og kvinner) og Afrika (menn) enn etnisk norske, mens menn fra Nord-Amerika har lavere sykefravær enn etnisk norske. Disse forskjellene kan ikke forklares ved demografiske, sosioøkonomiske eller arbeidsmarkedsfaktor forskjeller. I tillegg finner de ingen forskjell i sykefraværet mellom etniske norske og andregenerasjonsinnvandrere. Avslutningsvis viser resultatene fra deres studie at oppholdstid har en kurvet-lineær effekt (motsatt U form) på sykefraværet, for alle innvandringsgruppene, med unntak av Nord-Amerika og Oseania. Som nevnt er analyseperioden fra 1992 til 2003, som gjør at den ikke kan si noe eksplisitt om arbeidsinnvandringen fra nye EU-land som denne oppgaven er interessert i. Den gir likevel et bilde av hvordan forholdene var før 2004, som også er sentralt for å sammenligne sykefraværet før og etter EU-utvidelsen i 2004.

Hansen et al (2014) undersøker hypotesen om «yrkesmessige forskjeller» kan forklare differansen mellom sykefraværet til vestlige innvandrere, norskfødte og ikke-vestlige innvandrere. Ved bruk av norsk registerdata og «matching» teknikk prøver de å finne ut om yrkesbestemte faktorer som er assosiert med ikke-vestlige innvandrere har sammenheng med det relativt høye sykefraværet. De ser på sykefravær over 16 dager og bruker data for perioden 2003-2006. På samme måte som Dahl et al (2010) deler de inn i åtte landgrupper.

Resultatene viser at før de kontrollerer for yrkesmessige forskjeller, så har vestlige innvandrere lavere sykefravær enn norskinnfødte mens ikke-vestlige innvandrere har høyere. Ved å bruke «exact matching» for å kontrollere for yrkesmessige forskjeller finner Hansen et al (2014) fortsatt at ikke-vestlige innvandrere har signifikant høyere sykefravær enn innfødte, dog noe redusert. Det vil si at det relativt høye sykefraværet til ikke-vestlige innvandrere bare i noen grad kan forklares fra yrkesmessige faktorer. Analysene gjøres også for hvert kjønn, og resultatene samsvarer med funnet på det totale utvalget, med unntak av at kvinner fra Øst-Europa nå har lavere sykefravær enn norskfødte. Det vil si at kvinner fra Øst-Europa, Vest-Europa, Nord-Amerika og Oseania har signifikant lavere sykefravær enn innfødte norske kvinner. Det er også gjennomgående at kvinner har et høyere sykefravær enn menn som er i samsvar med annen empiri. I samsvar med denne oppgaven så studerer også Hansen et al (2014) noe av perioden etter EU-utvidelsen i 2004, dette åpner opp muligheten for sammenligne selv

om gruppeinndelingene ikke er helt identiske. En betydelig forskjell er at de bygger sin analyse på forskjeller blant yrkesgrupper, noe denne oppgaven mangler pålitelige opplysninger på.

Kann, Bragstad og Thune (2013)

Ved bruk av deskriptiv analyse ser studien på utviklingen i bruken av helserelaterte ytelser, herunder sykefravær, uførepensjonister, rehabilitering- og attføring, tidsbegrenset uførestønad og arbeidsavklaringspenger, fra 1992 til utgangen av 2012. Studien fokuserer på endringen i den demografiske sammensetningen og peker på at Norge har en aldrende befolkning som i stor grad skyldes de store fødselskullene etter krigen. Samtidig har innvandringen etter EU-utvidelsen bidratt til befolkningsvekst, der innvandrerne hovedsakelig er menn under 45 år som har bidratt til å gi en yngre befolkning.

Utviklingen i ytelsene viser en sterk vekst i andelen av befolkningen som mottok helserelaterte ytelser frem til 2003, fra 2003 har det vært en utflating og en svak nedgang. Aldersfordelingen trekker i motsatt retning, slik at det er andre forhold enn alder som har bidratt til endring i bruken av helserelaterte ytelser. De finner en reell nedgang for norske statsborgere, men peker på at nedgangen ville vært svakere uten den nye innvandringen vi fikk etter 2004. Analysen fokusere i stor grad på samme grupper og periode som denne oppgaven. Siden deres analyse er deskriptiv og studien min også består av en deskriptiv analysen er studien likevel relevant for sammenligning.

Bratsberg et al (2014b)

I en omfattende studie basert på norske registerdata, undersøker Bratsberg et al (2014b) den langsiktige tilpasningen i arbeidsmarkedet og trygd for alle store kohorter med innvandrere til Norge siden 1970-tallet. De har et særlig fokus på arbeidsinnvandringen på 1970-tallet og den «nye» arbeidsinnvandringen fra 2004. Deres funn viser at effekten av arbeidsinnvandring avhenger kritisk av karakteristikken til innvandrerne og forskjellen i levestandarden mellom avsender- og mottakerlandet, som viser seg å være en svært viktig prediktor for presentasjon i arbeidsmarkedet. Innvandrere fra land med likt inntektsnivå som Norge, presterer godt og på lik linje med norskinnfødte i det norske arbeidsmarkedet. Motsatt så synes arbeidsinnvandrere fra lav-inntektsland å ha større risiko for arbeidsledighet og trygdeavhengighet.

For kohorter som kan studeres i periode lengre enn 10-15 år i Norge, finner de at avtakende deltakelse i arbeidsmarkedet som blir fulgt av økende bruk av uføreforsikring. Dette er gjerne trigget av en nedgangskonjunktur. De fastslår at det er for tidlig å si om den «nye»

arbeidsinnvandringen fra 2004 vil følge samme mønster. De viser at denne gruppen hadde økt ledighet og avhengighet av sosialeforsikringer under finanskrisen, men at sysselsettingen økte relativt raskt når arbeidsmarkedet tok seg opp igjen.

Når det gjelder andregenerasjonsinnvandrere, så viser studien at innvandrerbarn født i Norge presterer bedre enn de som er født utenlands og mye bedre enn deres foreldre. De finner imidlertid også at barn av ikke-europeisk arbeidsinnvandrere, tar mindre utdanning og har lavere sysselsetting og inntekt enn norsk innfødte på samme alder.

Brekke og Schøne (2013) undersøker om selvrapportert helse kan forklare de observerte forskjellene mellom ikke-vestlige innvandrere og majoritetsbefolkningen. Data består av helseundersøkelsen i Oslo (Hubro) foretatt i 2000-2001, som er koblet på FD-trygd databasen. De ser primært på den mannlige befolkningen. Ved bruk av Poisson regresjon finner de at egenrapportert helse er et viktig element for å forklare sykefraværet. Når de kontrollerer for egenrapportert helse med generell helsestatus og spesifikke helseproblemer, er det ikke lenger en forskjell i sykefravær mellom ikke-vestlige innvandrer menn og majoritetsmenn. De peker videre på at det er spesielt muskel- og skjelettproblemer som forklarer forskjellen i sykefraværet.

3.3 Oppsummering

Siden det eksisterer få teoretiske bidrag som sier noe om innvandrernes tilpasning av sykefravær i mottakerland, ble det i den teoretiske delen brukt teorier om sykefravær for å danne hypoteser som kunne si noe om sykefravær blant innvandrere. Gjennomgangen resulterte i all hovedsak i to hypoteser. Hypotesen som kommer ut av arbeidstilbudsmodellen var at nye arbeidsinnvandrere ville ha et lavere sykefravær enn majoritetsbefolkningen. En lignende tolkning kunne også komme av hypotesen om sosiale normer. Den andre hypotesen gikk ut på at arbeidsinnvandrere fra nye EU-land i stor grad arbeider i konjunkturutsatte bransjer og vil derfor i større grad være påvirket av konjunktursvingningen. En kunne derfor tenke seg at sykefraværet til arbeidsinnvandrerne svinger mer med konjunktorene enn for andre.

Av empiri om sykefravær blant innvandrere, har fokuset vært på å forklare forskjellene mellom sykefraværsnivåene mellom norskfødte og vestlig (lavere) og ikke-vestlig (høyere). Funnene i de overnevnte studiene viser at demografiske kjennetegn og arbeidsmarkedforhold kan forklare deler av forskjellene, men ikke alt. Det er også gjentakende at andregenerasjonsinnvandrere har et lavere sykefravær enn andre innvandrere, og at sykefraværet øker med oppholdstiden.

Gjennomgangen viser også at nye arbeidsinnvandrere har hatt et relativt lavt forbruk av helserelaterte ytelser. Men det er likevel ingen studier, som forfatteren per dagsdato vet om, som eksplisitt ser på sykefraværet hos arbeidsinnvandrere etter det store «frisippet» i 2004. Derfor vil en videre ha fokus på sykefraværet spesielt for nye arbeidsinnvandrere relativt til majoritetsbefolkningen etter 2004.

4.0 Økonometrisk metode

Som nevnt tidligere skal en estimere hvordan opprinnelsesland påvirker sannsynligheten for og varigheten av sykefraværet. Siden vi måler sykefraværet langs disse to dimensjonene ser vi i denne oppgaven på effekten på sykefravær i to ulike analyser/regresjoner. I dette kapittelet ser en på de økonometriske metodene som er benyttet samt deres begrensninger. Først beskrives paneldata og paneldatamodeller, herunder fast effekt og tilfeldig effekt. I tillegg ser en på binære valg modeller (Probit og Logit), før en tar for seg paneldatamodeller med binær avhengig variabel. Avslutningsvis blir det en gjennomgang av mulige estimeringsproblemer.

4.1 Paneldata

Longitudinal eller paneldata, er data som følger et gitt utvalg av enheter³⁸ over tid. Dette gjør at en får flere observasjoner per enhet i utvalget. Paneldata er dermed en kombinasjon av tverrsnitt- og tidsseriedata (Verbeek 2012, Hsiao 2003). En skiller gjerne mellom «vide og korte» og «lange og smale» panel, der førstnevnte har observasjoner på relativt mange individ (stor N) over en kortere periode (liten T), dette er typisk paneldatasett for mikrodata³⁹.

Bruk av paneldata kontra tverrsnitt og tidsserie, har en rekke fordeler. Som nevnt, så gir paneldata vanligvis et stort antall observasjoner (datapunkter), noe som øker frihetsgradene og variasjonen i data. I tillegg vil det kunne redusere kollineariteten mellom forklaringsvariablene. Disse egenskapene fører til mer effektive og presise økonomiske estimat (Hsiao 2003). Med paneldata vil det også være mulig å måle effekter som rett og slett er umulig å måle ved tverrsnitt eller tidsserie. Til slutt så kan en ved paneldata kontrollere for uobservert individuell heterogenitet⁴⁰ som kan skape skjevheter i tilsvarende analyse ved tverrsnitt og tidsserie (Baltagi 2008). Men med paneldata kommer også noen utfordring. Det vil ikke lenger være fornuftig å anta at observasjonene er uavhengig, siden en følger individuelle enheter over flere perioder. For eksempel vil et individets helsetilstand være avhengig av helsetilstanden i tidligere periode.

4.2 Lineære paneldatamodeller

Siden paneldata varierer både over tid og individ kan vi indeksere variablene med i for individ ($i=1,2,\dots,N$) og t for tid ($t=1,2,\dots,T$), med dette kan en sette opp en lineær paneldatamodell som følger:

³⁸ Individ, bedrift, land etc.

³⁹ For makrodata har en ofte lange (stor T) og smale (liten N) paneldatasett.

⁴⁰ Individuell heterogenitet vil si individuelle unike karakteristikk ved et individ.

$$y_{it} = \beta_0 + X_{it}'\beta + \varepsilon_{it} \quad (4.1)$$

Der x_{it}' er en K-dimensjonal vektor med alle forklaringsvariablene som varierer på tvers av individ og over tid. Koeffisientene β_0 og β er lik for alle individ og tidsperioder, derfor vil β være endringen i y når x_{it} endres som følge av tid eller individ (to dimensjoner). Feilledet ε_{it} varierer også over individ og tid, og fanger opp alle uobserverte faktorer som påvirker den avhengig variabelen y_{it} (Verbeek 2012).

En ser gjerne på feilledet bestående av to komponenter⁴¹, et individspesifikk komponent α_i som kun variere på tvers av individ, og det resterende leddet som ses på som en komponent som variere over tid og individ u_{it} . Dette kan uttrykkes som:

$$\varepsilon_{it} = \alpha_i + u_{it} \quad (4.2)$$

For å få forventningsrette, konsistente og effesiente estimat for (4.1) med MKM, så må de klassiske antagelsen for MKM være oppfylt. Men ofte vil det være urealistisk å anta at feilledene (ε_{it}) på ulike tidspunkt er ukorrelerte med hverandre. Individuell uobserverte heterogenitet som helsetilstand (som fanges opp i α_i) vil typisk være korrelert over tid. Det vil si at en har autokorrelerte feilledd, som fører til at standardavviket til de estimerte koeffisientene blir forventingsskjeve og inkonsistente (ibid). En estimeringsmetode som tar høyde for denne korrelasjonen vil være mer effesient enn MKM.

Et annet problem er at disse individuelle heterogene kjennetegnene kan være avhengig av/korrelerte med forklaringsvariablene. Om vi ser på feilledet definert i (4.2) så må følgende antagelser være oppfylt for at MKM, men også tilfeldig effekt estimatoren, skal gi konsistent estimat:

$$E[u_{it}, X_{it}'] = 0 \quad (4.3)$$

$$E[\alpha_i, X_{it}'] = 0 \quad (4.4)$$

Men antagelsen (4.4) om at de individuelle heterogene egenskapene α_i er uavhengig av forklaringsvariablene x_{it}' er sterk, og ofte urimelig å anta. I så tilfelle vil en få problemer med endogenitet (forventningsskjev og inkonsistente estimatorer) og en må se etter estimeringsmetoder som tar høyde for dette problemet, som for eksempel fast effektmodellen.

⁴¹ Noen ganger også tre, for et konstantledd som kun varierer over tid, et slikt konstantledd er mer relevant ved lange og smale data.

Som en skal se senere, så vil valget mellom fast effekt- og tilfeldig effekt-modellen ofte være basert på om (4.4) holder eller ikke.

MKM

Setter vi (4.2) inn i (4.1) får vi følgende modell:

$$y_{it} = \beta_0 + X_{it} \beta + (\alpha_i + u_{it}) \quad \alpha_i \sim IID(0, \sigma_\alpha^2) \quad \& \quad u_{it} \sim IID(0, \sigma_u^2) \quad (4.5)$$

Der $\varepsilon_{it} = \alpha_i + u_{it}$ og antar at α_i er en stokastisk variabel i feilleddet. Der u_{it} er antatt å være ukorrelert over tid. Derfor vil enhver korrelasjon mellom feilleddene over tid knyttes til den individspesifikke effekten α_i . Videre er det antatt at α_i og u_{it} er gjensidig uavhengig og uavhengig av $x_{js} \forall j$ og s . MKM estimatene vil da være forventningsrett og konsistent.

Problemet med å estimere paneldatamodeller med MKM kan vises ved å se på kovariansmatrisen til feilleddet ε_{it} når α_i er korrelert over tid. Først kan vi se på kovariansmatrisen dersom de klassiske antagelsene er oppfylt i tilfellet uten individuell heterogenitet, dvs når $\alpha_i = 0$:

$$V[\varepsilon_{it}] = \begin{bmatrix} \sigma_u^2 & 0 & \dots & 0 \\ 0 & \sigma_u^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_u^2 \end{bmatrix} = \sigma_u^2 I_k$$

Her er $E(\varepsilon_{it}, \varepsilon_{is}) = 0$ og $E(\varepsilon_{it}, \varepsilon_{it}) = \sigma_u^2$. Men med autokorrelasjon grunnet α_i , har vi følgende:

$$E(\varepsilon_{it}^2) = E((\alpha_i + u_{it})^2) = E(\alpha_i^2 + 2\alpha_i u_{it} + u_{it}^2) = E(\alpha_i^2) + E(u_{it}^2) = \sigma_\alpha^2 + \sigma_u^2 \quad \text{og}$$

$$\begin{aligned} E(\varepsilon_{it}, \varepsilon_{is}) &= E((\alpha_i + u_{it})(\alpha_i + u_{is})) = E(\alpha_i^2) + E(\alpha_i u_{is}) + E(u_{it} \alpha_i) + E(u_{it} u_{is}) \\ &= E(\alpha_i^2) = \sigma_\alpha^2 \end{aligned}$$

Da har kovariansmatrisen ε_{it} følgende form:

$$V[\varepsilon_{it}] = \Omega = \begin{bmatrix} \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_u^2 + \sigma_\alpha^2 \end{bmatrix} \neq \sigma_u^2 I_k \quad (4.6)$$

Her vil MKM være ineffisient, standardfeilene forventnings skeive og standardtestene ikke lenger gyldige.

Tilfeldig effekt

Tilfeldig effektmodellen antar at alle regressorer som ikke er inkludert i regresjonen, kan hensiktsmessig og passende plasseres i feilleddet. Dette vil bety at en antar at den individ spesifikke komponenten α_i er en tilfeldig faktor som er uavhengig og identisk fordelt (IID) over alle individ, og derfor behandles den som en del av feilleddet. Vi har da modellen (4.5):

$$y_{it} = \beta_0 + X_{it} \beta + (\alpha_i + u_{it}) \quad \alpha_i \sim IID(0, \sigma_\alpha^2) \quad \& \quad u_{it} \sim IID(0, \sigma_u^2)$$

Som tidligere er $\varepsilon_{it} = \alpha_i + u_{it}$ et feilleddet som inneholder to komponenter. Som vi så i (4.6), har feilleddet en komponentstruktur som gir autokorrelasjon med mindre $\sigma_\alpha^2 = 0$. I likhet med tverrsnitt så kan problemet med autokorrelerte feilledd løses ved å transformere feilleddet og bruke generalisert minste kvadraters metode (GMKM). Dette gjør at modellen tar høyde for autokorrelasjonen slik at ε_{it} blir IID $\sim(0, \sigma_\varepsilon^2)$.

Den transformerte modellen tar følgende form:

$$y_{it} - \vartheta \bar{y}_i = \beta_0(1 - \vartheta) + (x_{it} - \vartheta \bar{x}_i) \beta + v_{it} \quad , \quad v_{it} \sim IID(0, \sigma_v^2) \quad (4.7)$$

Hvor $\vartheta = 1 - \psi^{\frac{1}{2}}$ der $\psi = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$.

Fordelene med tilfeldig effektmodellen er at den er mer effesient enn fast effektmodellen under (4.5), samt at en kan estimere effekten av tidskonstante variabler. Dette er sentralt i oppgaven, da vi ønsker å estimere hvordan opprinnelsesland – som er tidskonstant- påvirker sykefravær.

Fast effekt

Fast effektmodellen er en lineær regresjon som inneholder et konstantledd som varierer for hvert individ, men er konstant over tid. Det vil si at den individspesifikke komponenten α_i anses som en fast effekt for hvert individ.

$$y_{it} = \alpha_i + X_{it} \beta + u_{it} \quad , \quad u_{it} \sim IID(0, \sigma_u^2) \quad (4.8)$$

Om (4.3) holder, vil fast effekt estimatoren være forventningsrett⁴² for β , og en merker seg at en ikke trenger antagelser om den individspesifikke komponenten.

En kan skrive denne regresjonen ved å inkludere en dummyvariabel for hvert individ i og estimere med MKM⁴³. Dette er imidlertid svært tungvint, spesielt dersom N er stor, derfor kan

⁴² For konsistens kreves det at: $E\{(x_{it} - \bar{x}_i)u_{it}\} = 0$. Tilstrekkelig for dette er at x_{it} og \bar{x}_i er ukorrelert med feilleddet u_{it} .

⁴³ Denne kalles gjerne LSV (Least sqaerd dummy variable).

en heldigvis få samme estimat for β ved å kjøre regresjonen på avviket fra de individuelle gjennomsnittene.

$$y_{it} - \bar{y}_i = (\alpha_i - \alpha_i) + (X_{it} - \bar{X}_i)\beta + (u_{it} - \bar{u}_i)$$

$$y_{it} - \bar{y}_i = (X_{it} - \bar{X}_i)\beta + (u_{it} - \bar{u}_i) \quad (4.9)^{44}$$

Der \bar{y}_i er det individuelle gjennomsnittet over tid av den avhengige variabelen, definert som: $\bar{y}_i = T^{-1} \sum_t y_{it}$. Og tilsvarende for \bar{X}_i , som er de individuelle gjennomsnittene over tid for forklaringsvablene.

En ser at den individspesifikke komponenten α_i forsvinner når en transformerer modellen, og derfor trenger en ikke restriktive antagelser om hvordan α_i forholder seg til seg selv og de andre variablene i modellen. Estimatet av β i denne modellen blir gjerne kalt «within» estimator eller fast effekt estimator. Estimatoren sier noe om hvor mye individet avviker fra sitt individuelle gjennomsnitt og er derav variasjon innen hvert individ. Under svake betingelser så er fast effekt estimatoren asymptotisk normalfordelt, slik at en kan bruke de vanlige inferensprosedyrene. En ulempe med modellen er at også tidskonstante kontrollvariabler vil forsvinne under transformasjonene. Av den grunn anses faste effektmodellen som lite aktuell i oppgavens analyse.

Fast effekt eller tilfeldig effekt?

Det er gjerne en utfordring å finne et godt fullstendig grunnlag for å velge mellom tilfeldig effekt og fast effekt estimatoren. Til nå har vi sett at forskjellene i disse estimatorene ligger i hvordan en behandler den individspesifikke variabelen α_i , enten som en tilfeldig stokastisk variabel eller en fast individuell effekt. I mange tilfeller, spesielt når T er liten, så vil det være store forskjeller mellom estimatorene.

Et vurderingsgrunnlag er om individene i utvalget er «one of a kind» og ikke kan ses på som tilfeldig trekt fra en underliggende populasjon. I så fall er fast effekt estimatoren passende. Dette gir mening hvis de observerte enhetene er et land, et stort firma eller industri. Ved tilfeldig effekt tilnærmingen som er ubetinget på individenes individuelle karakteristikk α_i , er en ofte ikke interessert i verdien til individenes α_i (Verbeek 2012).

⁴⁴ Alternativ til (4.9) er å ta første differansen for å eliminere den individ spesifikke effekten.

Om en er interessert i å estimere α_i , er derfor fast effekt best egnet, dette gir mening om antall enheter er liten og av en spesifikk natur. Men selv om en har et stort antall enheter og tilfeldig effekt virker hensiktsmessig, kan det likevel være at faste effekt er best egnet dersom α_i er korrelert med forklaringsvariablene i x_{it} . Tilfeldig effekt er ikke robust med tanke på denne korrelasjonen, som fører til inkonsistente estimatorene. Men på en annen side, så tillater som sagt tilfeldig effektmodellen estimering av effekten av tidskonstante forklaringsvariabler.

For å teste dette, introduserte Hausman (1978) en test der nullhypotesen er at det ikke finnes korrelasjon mellom individspesifikk komponent og forklaringsvariablene⁴⁵. Ved å sammenligne de to estimatene (FE og RE), der FE som er konsistent under både null og alternativhypotesen, mens RE kun er konsistent under null hypotesen. Dersom det er signifikante forskjeller mellom estimatorene til FE og RE indikerer det at nullhypotesen sannsynlig ikke holder, altså at α_i og x_{it} er korrelert.

Under nullhypotesen kan det vises at:

$$V\{\hat{\beta}_{FE} - \hat{\beta}_{RE}\} = V\{\hat{\beta}_{FE}\} - V\{\hat{\beta}_{RE}\}$$

En beregner Hausmans test statistikk som følger:

$$W = \frac{\hat{\beta}_{FE} - \hat{\beta}_{RE}}{\hat{V}\{\hat{\beta}_{FE}\} - \hat{V}\{\hat{\beta}_{RE}\}} \sim \chi^2_{(k)} \quad (4.10)$$

Under nullhypotesen så har test statistikken en asymptotisk kji-kvadrat fordeling med K-frihetsgrader, hvor K er antall elementer i β .

Dersom (4.4) holder (og en beholder nullhypotesen) så er både tilfeldig effekt og faste effekt estimatoren konsistent. Da kan det vises at variansen til tilfeldig effekt estimatoren er mindre enn variansen til fast effekt estimatoren, dette gjør tilfeldig effekt mest effesient.

4.3 Binære valg modeller

Det kan ofte oppstå at det utfallet vi studere bare tar et begrenset antall verdier, gjerne to. En variabel som tar kun to verdier $y \in \{0,1\}$, kalles gjerne binær eller en dummy variabel. Som regel vil dette være variabler som reflekterer valg eller beslutninger som ulike aktører⁴⁶ tar. Om en ser på sykefravær som en binær variabel, kan en definere dette slik:

⁴⁵ $H_0: \text{Corr}(\alpha_i, x_{it}) = 0$ og alternativ hypotesen $H_A: \text{Corr}(\alpha_i, x_{it}) \neq 0$.

⁴⁶ Individ, bedrift, land etc.

$$y_i = \begin{cases} 1 & \text{hvis det i periode } t(\text{år}) \text{ har vært et eller flere sykefraværstilfeller} \\ 0 & \text{hvis det i periode } t(\text{år}) \text{ var ingen sykefraværstilfeller} \end{cases}$$

I slike modeller er ikke den underliggende adferd direkte observerbar, men vi observerer de valg som blir gjort. Her vil en observere valget om å sykemelde seg eller ikke, men en observerer gjerne ikke bakgrunnen for disse valgene. Siden denne atferden eller valget om å sykemelde seg er uobserverbar så blir det gjerne kalt en latent variabel. Bakgrunnen for individets valg kan for eksempel være et nyttemaksimeringsproblem der individene skal velge den beste allokeringen av fritid og arbeid (konsum). I henhold til dette kan en estimere en modell der den binære avhengige variabelen er en refleksjon av de valgene som blir tatt.

Tar en utgangspunkt i en latent underliggende modell, som er avgjørende for hvilke verdi y tar, kan dette skrives som:

$$y^* = x_i \beta + \varepsilon_i \quad (4.11)$$

$$y_i = \begin{cases} 1 & \text{hvis } y^* > 0 \\ 0 & \text{hvis } y^* \leq 0 \end{cases} \quad (4.12)$$

I et nyttemaksimerende rammeverk vil dette si at et individ velger å sykemelde seg $y_i = 1$ dersom nytten av å sykemelde seg er større enn null, om nytten er null eller mindre vil individet beslutte å ikke sykemelde seg $y_i = 0$. Setter en opp en lineær modell med binær avhengig variabel:

$$y_i = x_i \beta + \varepsilon_i$$

Med standard antagelsen har vi: $E\{\varepsilon_i|x_i\} = 0$ slik at $E\{y_i|x_i\} = x_i \beta$

$$\rightarrow E\{y_i|x_i\} = 1 \cdot P\{y_i = 1|x_i\} + 0 \cdot P\{y_i = 0|x_i\} = 1 \cdot P\{y_i = 1|x_i\} = x_i \beta$$

Det vil si at $x_i \beta$ er en sannsynlighet og bør derfor ligge innenfor intervallet $[0,1]$. Dette er bare mulig dersom x_i verdiene er bundet og β har visse restriksjoner. Dette er en av grunnene til at når en skal modellere på bakgrunn av en binær avhengig variabel, at det ikke lenger er passende med en lineær modell. En annen grunn er at en vil ha heteroskedastiske feilledd⁴⁷. Dette kan løses ved bruk av GMKM, men en vil fortsatt ha utfordringer med at de predikerte verdiene vil kunne ligge utenfor intervallet $[0,1]$ ⁴⁸. Når $x_i \beta$ kan ta verdier både over 1 og under 0, står en i

⁴⁷ Heteroskedastiske feilledd kommer som en følge av at den avhengige variabelen nå kun tar to verdier, som gjør at feilleddet også tar to verdier, det kan da vises at variansen er avhenger av forklaringsvariablene i X .

⁴⁸ Derfor er det lite passende å tolke resultatene som sannsynlighet.

fare for at feilledet får negativ varians, noe som er teoretisk umulig. På grunn av dette søker en etter andre løsninger, og to mulige løsninger kan være probit- eller logit-modellen.

For å sørge for at sannsynligheten holder seg innenfor 0 og 1, kan en ta utgangspunkt i en funksjon som holder seg innenfor det gitte intervallet. Enhver kumulativ sannsynlighetsfunksjon vil kunne oppfylle dette og en kan transformere modellen slik:

$$y = F(\alpha, \beta, x) = F(\alpha + x_i\beta)$$

$$F^{-1}(y) = \alpha + x_i\beta$$

Der F^{-1} er en invers (sannsynlighet)tetthetsfunksjon.

$F(\cdot)$ velges i forkant og α og β er som tidligere parametere en estimerer. $F(\cdot)$ bør være en kontinuerlig funksjon som er symmetrisk og konvergere asymptotisk mot 0 og 1. Transformasjonsfunksjonen $F(\cdot)$ blir ofte valgt som den kumulative standard normalfordelingsfunksjonen eller en logistisk fordelingsfunksjon, som begge har egenskapene som er nevnt ovenfor.

$$F(\cdot) = \Phi(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z e^{-\frac{1}{2}Z^2} dZ \quad \text{setter inn for } Z$$

$$: \Phi(\alpha + x_i\beta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha+x_i\beta} e^{-\frac{1}{2}(\alpha+x_i\beta)^2} d(\alpha + x_i\beta) \quad (4.13)$$

Som er standard normalfordeling og gir oss probit modellen, der (4.13) er «regresjonslinjen».

$$F(\cdot) = \Lambda(Z) = \frac{e^Z}{1 + e^Z} \quad \text{setter inn for } Z$$

$$: \Lambda(\alpha + x_i\beta) = \frac{e^{\alpha+x_i\beta}}{1 + e^{\alpha+x_i\beta}} \quad (4.14)$$

Som er den logistisk fordelingen og gir oss logit-modellen med (4.14) som «regresjonslinje». Disse to «regresjonslinjene» har begge en S-form og er svært like bortsett fra at den logistiske funksjonen har noe tykkere haler.

Som oftest blir parameterne i både probit- og logit modellen estimert med ML («maximum likelihood») metoden. Lar vi $\mathcal{L}(y_i|\beta)$ være sannsynligheten for at en observerer y_i gitt β , kan en beregne sannsynligheten for hvert individ, slik at sannsynligheten for hele utvalget blir:

$$\mathcal{L}(y_1) \times \mathcal{L}(y_2) \times \mathcal{L}(y_3) \times \dots \times \mathcal{L}(y_n) = \prod_{i=1}^n \mathcal{L}(y_i) \quad (4.15)^{49}$$

ML metoden finner derfor den β som maksimerer sannsynligheten for å observere det gitte utvalget, gitt at de individuelle y_i har en sannsynlighetsfordeling som er spesifisert lik $F(\cdot)$.

4.4 Paneldatamodeller med binær avhengig variabel

Vi så tidligere at det i denne oppgaven er lite hensiktsmessig å bruke fast effektmodellen⁵⁰, slik at en begrenses til tilfeldig effekt probit- og logit-modellen. Med tilfeldig effektmodellen så har den logistiske multivariate fordelingen en ulempe der korrelasjonene er bundet til $\frac{1}{2}$ (Maddala 1987, s:317). Det gjør modellen lite fleksibel selv om noe generalisering er mulig. Tar en utgangspunkt i tilfeldig effektmodellen, vil en derfor primært bruke tilfeldig effekt probit modellen.

En kan nå indeksere modellen med både tid og individ, og vi får følgende latente underliggende modell:

$$y_{it}^* = x_{it}\beta + \varepsilon_{it}$$

Som tidligere observerer vi kun y_{it} som tar følgende verdier:

$$y_{it} = \begin{cases} 1 & \text{hvis } y_{it}^* > 0 \\ 0 & \text{hvis } y_{it}^* \leq 0 \end{cases} \quad (4.16)$$

Her vil ε_{it} ha gjennomsnitt lik null og normalisert varians til 1, der en antar at ε_{it} er uavhengig av $(x_{i1}, x_{i2}, \dots, x_{iT}) = x_{it}$. Som ved valg modellene ved tverrsnitt, må en også her anta noe om den felles fordelingsfunksjonen til feilleddet $(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT}) = \varepsilon_{it}$ for å kunne estimere β koeffisientene med ML. Dersom vi har feilleddet som i (4.2): $\varepsilon_{it} = \alpha_i + u_{it}$, der u_{it} er uavhengig både tid og individ, der vi antar at $(\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{iT}) \sim N(0, 1)$ og $Cov\{\varepsilon_{it}, \varepsilon_{is}\} = 0, s \neq t$. Det vil si at α_i er $NID(0, \sigma_\alpha^2)$ og u_{it} $NID(0, 1 - \sigma_\alpha^2)$. Da vil den tilfeldige effekt probit modellens sannsynlighetsfunksjon være uttrykt som:

$$F(\cdot) = \Phi(Z) = \Phi\left(\frac{x_{it}\beta + \alpha_i}{\sqrt{1 - \sigma_\alpha^2}}\right) \quad \text{for } y_{it} = 1 \quad (4.17)^{51}$$

⁴⁹ En forenkler gjerne maksimerings prosedyren ved å ta $\log(\mathcal{L})$.

⁵⁰ De sentrale forklaringsvariablene er tidskonstante og vil integreres ut ved bruk av faste effektmodellen.

⁵¹ For $y_{it} = 0 \rightarrow F(\cdot) = 1 - \Phi\left(\frac{x_{it}\beta + \alpha_i}{\sqrt{1 - \sigma_\alpha^2}}\right)$

der Φ som tidligere er den kumulative tetthetsfunksjonen til standard normal fordelingen, og der (4.17) er «regresjonslinjen» som estimerer sannsynligheten for at $y_{it} = 1$.

4.5 utfordringer ved estimering

Endogenitet

I mange tilfeller vil det være urealistisk å anta feilledet ikke er korrelert eller uavhengig av forklaringsvariablene i modellen. I slike tilfeller vil ikke forklaringsvariablene i X lenger være eksogene, men endogene, og konsekvensen er estimat som er forventningsskjeve og inkonsistente. Problemet med endogene forklaringsvariabler kan oppstå av flere årsaker, men det er verdt å nevne tre: *utelatte variabler*, reversert kausalitet eller *to-veis kausalitet* og *målfeil*. I denne oppgaven er det først og fremst de to førstnevnte som er av relevans⁵².

Endogenitet på grunn av *utelatt variabel* oppstår dersom en relevant variabel, som er korrelert med en eller flere av forklaringsvariablene, er utelatt fra modellen. Men det kan også oppstå dersom det finner uobserverte faktorer som er korrelert med en eller flere variabler i X (Verbeek 2012). Denne type uobserverte komponenter refereres gjerne til som uobservert heterogenitet. I oppgavens kontekst så vil helse som er sterkt korrelert med sykefravær, men også andre variabler som utdanning og arbeidstid, være uobservert og utelatt. Denne forventningsskjevheten er av spesiell interesse om en ønsker å tolke resultatene som kausaleffekter. En kan oppleve *to-veis kausalitet* dersom x har en effekt på y samtidig som y har en effekt på x . For eksempel er det vist at inntektsnivået påvirker individets sykefravær, på samme tid er det slik at hvor mye sykefravær et individ har, kan påvirke lønnsnivået.

Seleksjon

En annen utfordring en ofte møter er ufullstendige panel og seleksjonsskjevhet, som også kan skape endogenitetsproblemer. Årsaken til dette kan være problemer med innsamling av data, frafall eller feilrapportering. Dersom ikke-registrering og frafall skjer på en ikke-tilfeldig måte, kan det få konsekvenser som inkonsistense, forventningsskjeve estimat og ugyldige tester (Baltagi 2008). Det gjør at seleksjon også kan gi en form for endogenitet. Det bør nevnes at frafall og feilrapportering antas å være heller små ved registerdata, men at disse årsakene til ikke-registrering er større for innvandringsbefolkningen, spesielt med kortere opphold, enn majoritetsbefolkningen. Noe som kan skape ytterligere problemer, er at det ofte er dårlig registrering av individ som utvandrer fra Norge. Det gjør at en kan observere individ etter at de

⁵² Målfeil er av mindre relevans i oppgaven, og jeg vil derfor ikke gå noe nærmere inn på dette.

har utvandret, som gjør at en får mye uregistrerte observasjoner på ikke-tidskonstante kjennetegn for disse individene.

5.0 Data og deskriptiv - statistikk og statistisk analyse

Formålet med dette kapittelet er å kunne gi innsikt i datagrunnlaget som blir brukt i oppgavens empiriske analyse. Kapittelet vil være todelt, første del tar for seg FD-trygd og hvor denne databasen kommer fra, før en går videre og ser på oppgavens datasett. Her ser en på sorteringer og begrensninger før det blir en gjennomgang av sentrale variabler. I andre del av kapittelet presenteres og diskuteres den deskriptive statistikken. Denne deskriptive delen vil i hovedsak ha to formål, for det første er det fokus på en deskriptiv analyse av data. I tillegg skal det tjene som en komplementær beskrivelse av dataen til den empiriske analysen, der innsikten skal gi godt grunnlag for diskusjon i det empiriske kapittelet.

5.1 Datasettet

FD-trygd, begrensninger og sortering av datasettet

Oppgaven har tatt utgangspunkt i forløpsdatabasen FD-trygd, som er en database bestående av registerbaserte opplysninger på individnivå⁵³ for hele befolkningen siden 1992. Databasen er bygget opp av informasjon fra NAV, skattedirektoratet og register i Statistisk sentralbyrå der de viktigste datatypene i denne databasen er tryggedata, demografidata, sysselsettingsdata, sosialhjelpdata og arbeidssøkerdata. Dette gjør databasen ekstra god til å studere endringer i arbeidsmarkedstatus slik som sykefravær (Hansen, Hetland og Sivertstøl 2013). I denne oppgaven har jeg hatt tilgang til et datasett for årene 1998-2008, men oppgaven vil primært fokusere på utviklingen fra 2004.

Ytterligere sorteringer er gjort ved å utelukke individ som enten utvandret eller døde i løpet av perioden⁵⁴. Oppgavens fokus er på forekomsten av sykefravær og tilhørende utbetaling av sykepengene, noe bare arbeidstakere som tjener over $\frac{1}{2}$ G har rett på. Derfor sorteres alle observasjoner under $\frac{1}{2}$ G bort. Videre er det sortert til å beholde individ kun mellom 16-67 år, da dette typisk definerer individ i arbeidsalder. Som en skal se senere, vil variabelen landbakgrunn være sentral. Av den grunn sorteres alle observasjoner der landbakgrunn ikke er registret bort. Etter dette står en igjen med et paneldatasett på 11 år, med 6,84 millioner observasjoner og 912 498 unike observasjoner (individ). Disse sorteringene er gjort både før den deskriptive delen og den empiriske analysen. I tillegg er det i den empiriske analysen sortert bort individ med kun et observasjonsår da en bruker paneldatamodeller. Etter den sistnevnte utsorteringen står en igjen med 6,74 millioner observasjoner, med 829 759, unike observasjoner.

⁵³ Individ er den statistiske enheten

⁵⁴ Denne sorteringen tilsvarer sortering gjort i studien av Hansen et al (2014).

Grunnen til at utsorteringen av individ med kun et observasjonsår kun er gjort i forbindelse med den økonometriske analysen, er for å få et mest mulig fullstendig bilde i den deskriptive analysen. Ulempen vil være at en ikke har sett på den deskriptive statistikken for nøyaktige samme utvalg som den empiriske analysen er basert på. I oppgavens tilfelle viser det seg imidlertid å være små forskjeller, endringene vil primært være i 2007/2008, og ellers holder de samme mønster i den deskriptive statistikken. Sentrale figurer der en har sortert bort individ med kun et observasjonsår er lagt i vedlegget. Uavhengig av den siste sortering så er datasettet ubalansert, som betyr at antall observerte individ ikke er likt for hvert år. Ubalansen kommer i første rekke av at antall observerte individ per år øker utover perioden⁵⁵. For hele utvalget, altså alle landgrupper, er det gjennomsnittlige antall observasjonsår 9,3 av 11 år⁵⁶.

5.2 Sentrale variabler

Målet med denne oppgaven er å finne ut om reduksjonen i sykefraværet fra 2004 og utover, delvis kan forklares av den massive arbeidsinnvandringen som kom i kjølvannet av EU-utvidelsen. I den sammenheng vil det være flere variabler som er sentrale og som denne oppgaven har tatt utgangspunkt i.

Sykefraværet er den avhengige variabelen, men det finnes flere måter å måle sykefraværet på. Når en ønsker å måle effekter på sykefraværet skiller en gjerne mellom sannsynligheten for sykefraværstilfelle(r), også kalt insidens, og varighet av eller lengden på sykefraværet. I denne oppgaven vil en først ta utgangspunkt i en binær variabel for sykefravær(tilfelle)⁵⁷. Det vil si en variabelen som tar verdien 1 dersom individet har hatt ett eller flere sykefraværstilfeller i referanseperioden, og tar verdien 0 dersom individet ikke har hatt noen sykefraværstilfeller. Ved bruk av en slik binær sykefraværstilfelle variabel, vil en da estimere sannsynlighet, insidens, for sykefravær.

Videre, for å estimere effekten på varigheten av sykefraværet, brukes variabelen «spdager» som avhengig variabel. Denne reflekterer antall sykefraværsdager et individ har hatt i løpet av året, i analysen anvendes $\log(\text{spdager})$.

⁵⁵ En positiv nettoinnvandring til Norge er forklaringen på dette.

⁵⁶ Ved utsortering av individ med ett observasjonsår er det gjennomsnittlig 9,5 antall observasjonsår.

⁵⁷ Variabelen en generert fra variabelen FD-Trygdvariabelen «spdager», som er antall sykefraværsdager for gitt år.

Når en skal se på effekten av arbeidsinnvandringen på sykefraværet, er det rimelig å dele inn individene etter land eller landgrupper. I denne oppgaven benyttes variabelen landbakgrunn⁵⁸, mens selve grupperingen tar utgangspunkt i den «nye» to-delingen av verden (Høydahl 2008). Høydahl forklarer hvordan den vestlig- og ikke-vestlig to-delingen som deler opp Europa, stadig blir mer meningsløs, da mange blant de ikke-vestlige innvanderne vil komme fra EU-land. Han deler således verden inn i to grupper: 1. EU/EØS-land, USA, Canada, Australia og New Zealand og 2. Asia, Afrika, Latin-Amerika, Oseania (ekskludert Australia og New Zealand) og Europa utenom EU/EØS. Fra disse to gruppene er det gjort to uttrekk, som danner sine egne grupper. Den første er individ med Norge som landbakgrunn; dette danner analysens referansegruppe. I tillegg så trekkes de 12 landene EU har blitt utvidet med siden 2004 ut, og det er denne gruppen som er av spesiell interesse i studien.

Det er altså laget fire gjensidige utelukkende og uttømmende landgrupper. Disse gruppene består av Norge, EU-utvidede land⁵⁹ (gruppe 1), og to grupper i tillegg. Gruppe 2 består av land tilhørende Asia, Oseania (ekskludert Australia og New Zealand), Afrika, Sør- og Sentral-Amerika og land i Europa utenfor EU/EØS. Gruppe 3 består av de resterende landene som er Nord-Amerika, Australia og New Zealand og EU/EØS-land med flere⁶⁰.

En tilsvarende deling er gjort av blant annet av Tønnessen, Cappelen og Skjerpen (2014) der de gjør befolkningsframskrivninger for inn- og utvandring. Her argumenterer de for likheter innad i gruppene, landene tilhørende gruppe 1 er der innvandringen har økt mest de siste årene. Migrasjonspotensialet er stort for denne gruppen blant annet på grunn av de store inntektsforskjellene mellom Norge, samtidig som restriksjonen for innvandring i stor grad er avskaffet. Gruppe 2 beskrives som svært heterogen og er en forenkling, noe som passer i denne oppgaven. Til slutt argumenteres det for at gruppe 3 har relativ lik atferd slik den kan måles f.eks. i gjennomsnittlig fruktbarhet, utdanning, inntekt og utvandring. Siden oppgaven har et fokus på EU-utvidede land virker denne fire-gruppeinndelingen rimelig, og tilstrekkelig for oppgavens problemstilling.

⁵⁸ Landbakgrunn er, for personer født i utlandet, eget fødeland. For personer født i Norge er det foreldrenes fødeland, der mors fødeland blir valgt om foreldrene har ulikt fødeland (SSB).

⁵⁹ EU-utvidelsen består av landene: Polen, Litauen, Latvia, Estland, Tsjekkia, Slovakia, Ungarn, Slovenia, Kroatia, Kyrpros, Bulgaria og Romania

⁶⁰ For en detaljert landsoversikt i gruppene se Høydahl (2008). Siden en likevel trekker ut landene som er nye i EU, vil gruppe 2 og 3 kunne bli karakterisert som ikke-vestlige og vestlig, men jeg velger primært å kalle dem etter de mer nøytrale navnene gruppe 2 og gruppe 3.

Som nevnt er det totale utvalget på 6,84 millioner observasjoner (912 498 unike observasjoner), av disse er gruppen Norge den desidert største med 4,2 millioner observasjoner, mens gruppe 1 er på 178 000 observasjoner. Gruppe 2 og gruppe 3 har henholdsvis 1,1 og 1,3 millioner observasjoner. Disse andelene er viktige, spesielt når en ser på den deskriptive delen. Årsaken er at endring i de ulike gruppene vil ha forskjellig effekt på totalen, for eksempel vil Norge som den største gruppen ha større effekt på totalen, enn alle de andre gruppene, gitt en lik endring/utvikling

Andre forklaringsvariabler er kjønn, alder, utdanning, fulltid, reallønn, barn<11 år, oppholdstid, andre generasjonsinnvandrere og en step-dummy for år>2003. Kjønn er en binær variabel som tar verdien 1 hvis individet er mann og 0 hvis individet er kvinne. Aldersvariabelen er generert på bakgrunn av fødselsår og kalenderårsvariabelen. Utdanning er antall år med utdanning målt i høyeste fullført grad. Fulltid reflekterer om individet arbeider fulltid eller ikke, altså deltid. Variabelen er generert fra variabelen arbeidstid (arbtid) som primært tar 3 verdier, men er transformert til å bli kontinuerlig i intervallet 1-3⁶¹, der kode 1 betyr 4-19 timer, 2 er 20-29 timer og 3 betyr 30 timer eller mer. Fulltid tar verdien 1 dersom arbeidstid er større eller lik 2,5, det vil si at arbeideren har jobbet fulltid minst halve året eller mer, og tar verdien 0 om arbeidstid er lavere enn 2,5. Videre er det generert en reallønnsvariabel, med 1998 som basisår. Barn<11 år er en binær variabel for om individet har barn under 11 år eller ikke. Variabelen oppholdstid, som måler antall år med opphold i Norge, er generert på bakgrunn av variablene år og første oppholdsår. I tillegg har en binær variabel som tar verdien 1 dersom om individet er andre generasjonsinnvandrere og 0 ellers. Til slutt er det konstruert en step-dummy som tar verdien 1 dersom en befinner seg i årene 2004-2008 og 0 i årene 1998-2003. I tillegg til disse er det laget noen interaksjonsledd med blant annet: barn<11 år og kjønn, som gjør at en kan skille mellom effekten barn under 11 år har på henholdsvis fedre og mødre. Step-dummyen brukes videre til å generere tre interaksjonsledd mellom step-dummyen og de tre respektive landgruppene som sammenlignes med Norge.

Det finnes en rekke andre variabler som hadde vært rimelig og interessant og kontrollere for. Men begrensninger i datasettet, slik som tilgang til variabler og store andeler med ikke-registrerte observasjoner, gjør at disse er uaktuelle å inkludere i analysen. Når det gjelder ikke-registrerte observasjoner vil jeg spesielt nevne to variabler som er utelatt helt eller delvis på

⁶¹ Årsaken til denne transformasjonen er at arbeidstid blir registrert fortløpende (forløpsdata), så når en transformere til årsdata vil dette være gjennomsnittet for individet i løpet av året som kan ta enhver verdi mellom 1 og 3.

grunn av overnevnte begrensninger, dette gjelder næringer⁶² og utdanning. Utdanning er inkludert når en ser på gruppene hver for seg, med unntak av gruppe 1 på grunn av et stort antall uregistrerte observasjoner i årene 2007 og 2008. Dette gjelder både i den deskriptive og empiriske analysen.

5.3 Deskriptiv statistikk

Det er en rekke bakgrunnsfaktorer som vil kunne påvirke nivået på sykefravær for enkeltindivid og oppgavens landgrupper. Her skal en se på utviklingen av bakgrunnsvariablene alder, kjønn, sysselsetting, andel som jobber i fulltid og lønn for de ulike landgruppene. Disse kan si noe om nivået på sykefraværet og kanskje gi hypoteser om forskjeller i sykefraværet mellom de fire gruppene, spesielt Norge og gruppe 1. Videre ser en på utviklingen i sykefraværet over tid for det totale utvalget og for menn og kvinner isolert.

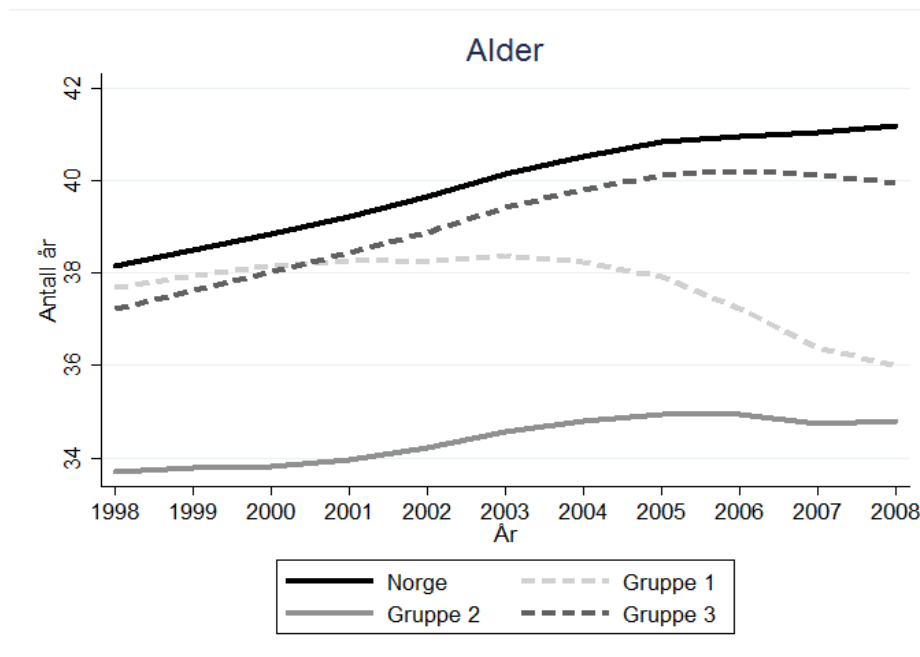
Demografi

Det er velkjent at sykefraværet varierer med alder, og en antar vanligvis at økende alder er forbundet med relativt dårligere helse, som gjerne forbindes med et økende sykefravær. Det er derfor av interesse å se på aldersutviklingen til de ulike landgruppene. Denne sammenhengen er likevel ikke entydig, gjennom en studie av effekten av AFP på tidligpensjonering finner Bratberg, Holmås og Thøgersen (2004) at sannsynligheten for tidligpensjonering øker med helseproblemer. Det vil si at eldre arbeidstakere, spesielt de med dårlig helse, går tidligere av med pensjon. Personer med et høyt sykefravær benytter seg altså i større grad av uførepensjon og tidligavgangsordninger, enn personer med mindre sykefravær (Rønning 2004). Derfor kan ikke en arbeidsstyrke bestående av flere eldre utelukkende forklare et økende sykefravær. Markussen, Røed, Røgeberg og Gaure (2009) finner at for mindre alvorlige sykdommer avtar tilbøyeligheten for sykefravær med alder, opptil 50 år. For mer alvorlig sykdommer øker tilbøyeligheten med alder.

Alderssammensetningen vil ofte variere over innvandringsgrunn: arbeid, flukt, familie, utdanning og andre. Siden oppgavens fire landgrupper har forskjellige hovedgrunn til innvandring, vil en naturlig finne forskjellig aldersfordeling mellom de fire gruppene. I figur 5.1 illustreres aldersutviklingen for de fire landgruppene. Merk at dette er aldersfordelingen etter utsorteringen av observasjoner utenfor alderen 16-67 år.

⁶² Det hadde vært interessant å skille mellom næringer, da mye av innvandringen fra gruppe 1 arbeider i bygg og anleggsbransjen.

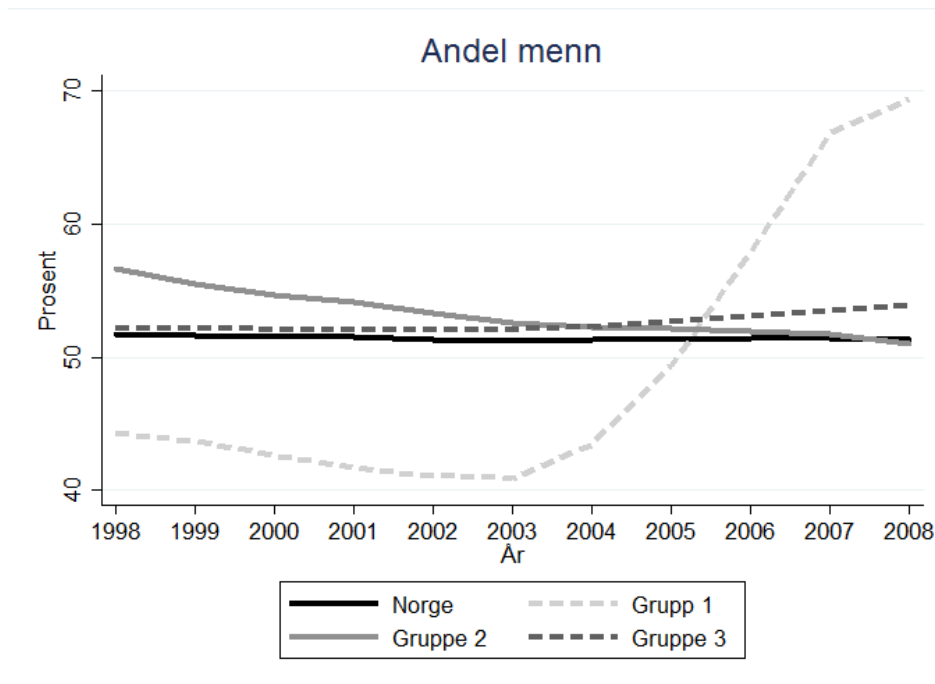
Figur 5.1 Aldersutvikling i perioden 1998-2008 for fire landgrupper, etter sortering ned til 16-67 år.



Norge har stort sett hele perioden den høyest gjennomsnittlig alder, og med tanke på en aldrende arbeidsstokk er økningen i gjennomsnittlig alder for Norge kanskje ingen overraskelse. Gruppe 2 skiller seg mest ut ved å ha den desidert laveste gjennomsnittlige alderen, denne gruppen forbindes gjerne med å ha en stor andel flyktninger, som typisk har lavere gjennomsnittlig alder enn de andre gruppene (Texmon og Brunborg 2013). Ser en på gruppe 1, viser figur 5.1 at den gjennomsnittlige alderen for denne gruppen begynner å falle i 2004, og reduksjonen tiltar utover i perioden. Dette kan og bør ses i sammenheng med den nye arbeidsinnvandringen fra nye EU land, da dette er innvandrere i relativt ung arbeidssalder (ibid).

Når det gjelder kjønnsfordelingen for de ulike landgruppene, er det spesielt gruppe 1 som skiller seg ut. Som en ser av figur 5.2 under går andelen menn for gruppe 1 fra 40-45 prosent til nærmere 70 prosent på bare 10 år, der skiftet skjer rundt 2004 og stiger kraftig utover perioden⁶³. De andre gruppene har en langt mer stabil andel av menn rundt 50-55 prosent, der gruppe 2 har hatt en fallende trend, gruppe 3 stiger noe utover perioden, mens for Norge ligger andel menn stort sett rundt 52 prosent.

⁶³ Tilsvarende utvikling er også rapportert av Texmon og Brunborg (2013).

Figur 5.2 Utviklingen i andel menn i perioden 1998-2008 for de fire landgrupper.

Det er velkjente funn at menn har et lavere sykefravær enn kvinner, dette gjelder både tilbøyeligheten for sykefraværstilfelle og lengden på sykefravær (se blant annet Markussen et al 2009 og Smeby, Bruusgaard og Claussen 2009). Dette gjelder også uansett hvilke aldersgruppe en ser på (Hauge og Kann 2007). Årsaken til nivåforskjellen er flere og sammensatte; dobbeltbyrde, kjønnsforskjeller i normer og kjønnsforskjeller i helseplager er noen årsaker som kan være med å forklare denne forskjellen (Mykletun og Vaage 2012).

Faktorene som bestemmer sykefraværet er som sagt flere og sammensatte, i tillegg til kjønn og alder vil blant annet lønnsnivået og sysselsettingsnivået spille inn.

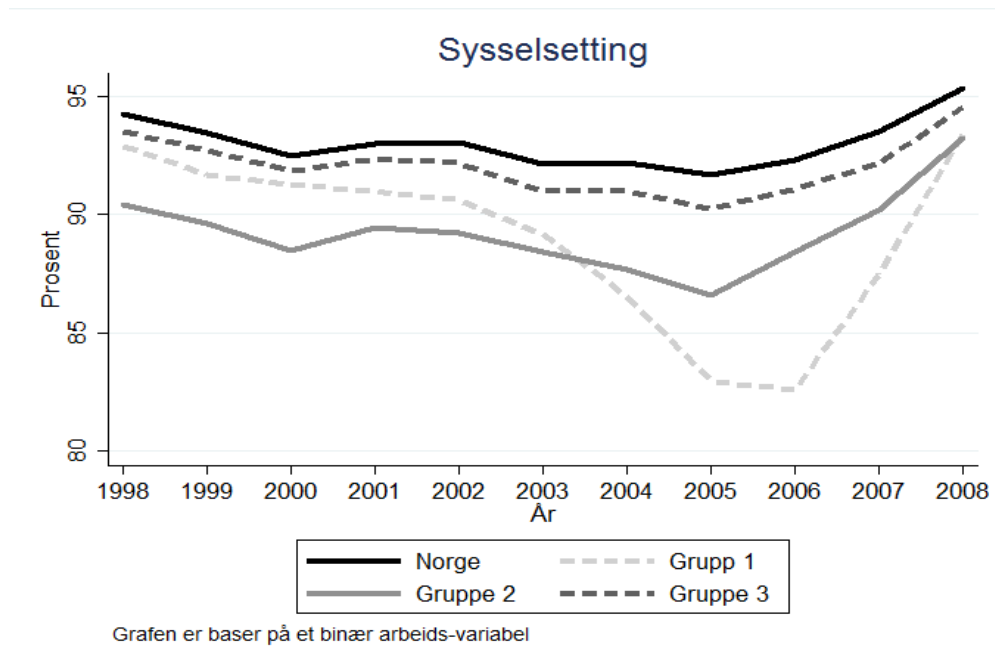
Sysselsetting og lønn

Når en nå skal se på sysselsettingen er det greit å merke seg at en kan se på sammenhengen mellom sykefravær og sysselsetting på to måter. Sysselsettingen kan for det første være en indikasjon på økonomiens konjunktursituasjon, og som en så i teorikapittelet varierer sykefraværet med konjunkturer. Et annet aspekt er at sysselsettingen er nevneren i sykefraværsprosenten, dersom sysselsettingen øker vil sykefraværet, i prosent, reduseres, alt annet likt. Disse to måtene vil så klart være knyttet til hverandre, men fokuset her vil primært være på sysselsetting og konjunktursituasjonen.

I figur 5.3 vil sysselsettingsprosenten være definert som antall sysselsatt i forhold til totalen i de ulike gruppene, og figuren viser utviklingen i sysselsettingen for de fire landgruppene.

Sysselsettingsraten i utvalget ligger svært høyt for alle gruppene, noe som ikke er unaturlig da utvalget er sortert ned til å være de som har en inntekt over ½ G. Utsorteringen gjør at store deler av de som hele eller deler av observasjonsåret er uten arbeid sorteres vekk. En vil også fjerne en rekke individ som arbeider relativt kort deltid, men dette er mer hensiktsmessig da de heller ikke har rett på sykepengene.

Figur 5.3 Utviklingen i sysselsetting (prosent) i perioden 1998-2008 for de fire landgruppene.



En ser videre at majoritetsbefolkningen, Norge, har den høyeste graden av sysselsetting gjennom hele perioden, mens gruppe 3 ligger like under dette nivået. Sysselsettingen varierer noe for disse gruppene i løpet av perioden, men de største variasjonene i sysselsettingen finner en for gruppe 1 og gruppe 2. En årsak til dette kan være at de har en relativt svakere tilknytning til arbeidslivet, Horgen (2014) finner at av innvandrere med svak tilknytning til arbeidsmarkedet⁶⁴ stod innvandrere fra Asia, Afrika og EU-land fra Øst-europa for over 70 prosent av disse, og de aller fleste har vært i Norge mindre enn 5 år. I tillegg så arbeider typisk arbeidsinnvandrere fra lavinntektsland i konjunkturutsatte bransjer, som gjør dem mer sårbare for konjunktursvingninger (Bratsberg et al 2014a). Teoriene om sykefravær og konjunkturer, i kapittel 3, viser til at variasjonen mellom sykefraværet og konjunkturer er prosyklisk. Det vil si at en relativt høy sysselsetting forbindes med høyere sykefravær, enn ved lav sysselsetting. Om

⁶⁴ Defineret som: Individ som ikke er sysselsett, under utdanning, arbeidsledig eller på andre helserelaterte trygdeytelser.

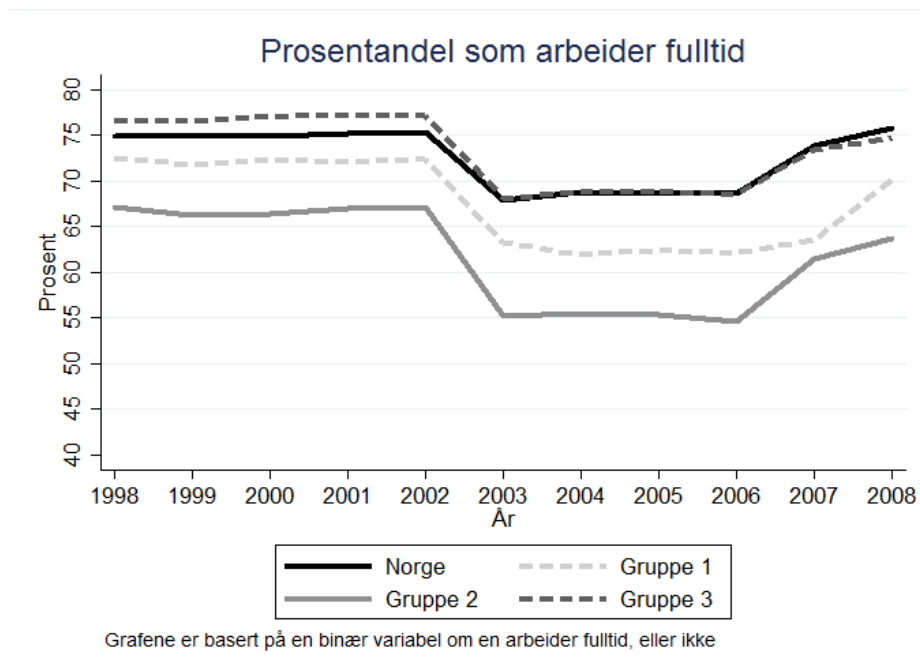
gruppe 1 og 2 er mest utsatt for denne type svingninger vil det være mulig at sykefraværet for disse varierer mer med konjunktorene (sysselsettingen) enn for Norge og gruppe 3.

Det som er felles for alle gruppene, er en mindre nedgang i sysselsettingen rundt år 2000 og en mer betydelig reduksjon i sysselsettingen fra 2002 og til rundt 2005, spesielt for gruppe 1. En nedgang i sysselsetting denne perioden er i samsvar med rapporterte tall fra SSB⁶⁵. Rundt 2005/2006 stiger sysselsettingen igjen for alle grupper, og gruppe 1 får en voldsom vekst med nesten 20 prosentpoeng denne perioden. Dette bør ses i sammenheng med økningen i innvandrere til Norge fra denne gruppen, som en så i bakgrunnskapitlet. En så også i bakgrunnskapitlet at sysselsettingsveksten fra 2002 i stor grad har blitt drevet av sysselsettingsøkning blant innvandrere, et inntrykk en også får av figur 5.3. Om en ser på denne utviklingen som en indikasjon på konjunkturstatusen, så vil teorien predikere at sykefraværet reduseres frem til rundt 2005 for så å stige fra 2005/2006. Der disse utviklingene er størst for gruppe 1 og 2.

En så tidligere at sysselsettingen, herunder sammensetningen av de yrkesaktive, kunne være med å bestemme sykefraværet. I tillegg til sammensetningen er det vist at sykefraværet varierer med sysselsettingsgraden, altså om en er ansatt fulltid eller deltid. En antar vanligvis at deltidsarbeidere har et høyere sykefravær enn de som arbeider fulltid. Dersom det å arbeide deltid gir mindre mestingsfølelse og tilhørighet til arbeidsplassen, kan dette være belastende for den enkelte arbeider, og bidra til et høyere sykefravær (Ingstad og Kvande 2011). I tillegg er det en klart større andel kvinner som arbeider deltid enn menn, og en vet at kvinner har høyere sykefravær enn menn. Derfor ser en gjerne på sammenhengen mellom sysselsettingsgrad og sykefravær i relasjon med kjønn og sykefravær. På en annen side kan deltidsarbeid være selvvalgt og en optimal allokering av fritid og konsum, enten som følge av dårlig helse eller familieforpliktelser, denne seleksjonen kan føre til et lavere sykefravær blant deltidsansatte. Moland (2007) har gjort flere studier på sammenhengen mellom stillingsbrøk og sykefravær, og finner at heltidsansatte i helse- og omsorgsektoren i Oslo hadde høyere sykefravær enn deltidsansatte. Siden studien er gjort på en spesifikk sektor er det vanskelig å avgjøre om disse resultatene er overførbare til det generelle arbeidsmarkedet eller andre spesifikke sektorer. I sin studie av sykefraværet til innvandrere i Norge, finner også Dahl et al (2010) at de som jobber fulltid har et høyere sykefravær enn de som ikke jobber fulltid.

⁶⁵ Se for eksempel tabell 03923 på ssb.no.

Figur 5.4 Utviklingen i andelen som jobber fulltid i perioden 1998-2008 for de fire landgruppene.



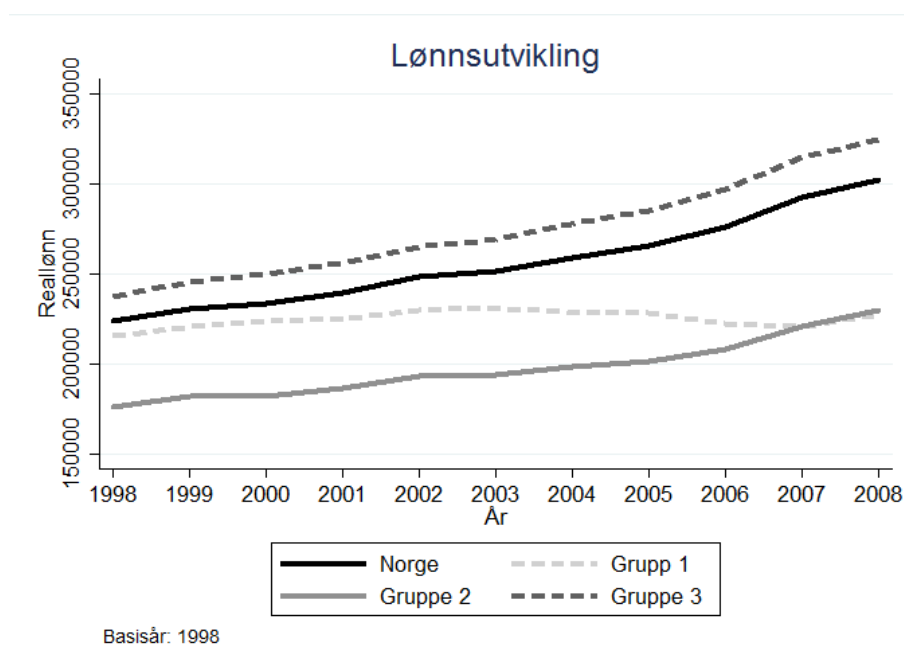
Når en ser på andelen som arbeider fulltid for de ulike landgruppene⁶⁶, er verdt å legge merke til nivåforskjellene. Norge og gruppe 3 ligger svært likt med en fulltidsandel på rundt 75 prosent, gruppe 1 ligger på rundt 73-74 prosent i begynnelsen av perioden, men avstanden til både Norge og gruppe 3 øker noe utover i perioden. En legger merke til fallet i andel i fulltid i 2003 og økningen i 2007, disse endringene skjer samtidig for alle landgruppene og kan være en refleksjon av tilbakefall i sysselsettingen i gitt periode som vi også så i figur 5.3 ovenfor. Ut ifra dette vil en forvente at både gruppe 1 og 2 har et lavere sykefravær enn etnisk norske og gruppe 3, gitt alt annet. Fordelingen kan også reflektere at både gruppe 1 og 2 har en svakere tilknytning til arbeidslivet og mindre sikkerhet rundt sin arbeidsplass, som igjen kan forbindes med et høyere sykefravær. På en annen side vil mindre sikkerhet kunne føre til at sykefraværet går ned, slik det følger av disiplineringseffekten en så i teoridelen.

Lønnsutviklingen ser vi i figur 5.5, alle landgruppene har en stigende lønnsvekst stort sett hele perioden, bortsett fra gruppe 1 som skiller seg ut ved at veksten avtar og nesten stagnerer rundt 2003. En kan forvente at den kraftige økningen i arbeidsinnvandringen fra EU-utvidede land vil kunne påvirke lønns- og sysselsettingsutviklingen, både for tidligere ankommet innvandrere og nordmenn. Ser en på den klassiske arbeidsmarkedsmodellen, så vil et generelt økt tilbud av arbeidskraft redusere lønningene generelt, slik at markedet kan absorbere den nye

⁶⁶ En må huske at dette er etter utsortering av de som ikke har rett på sykepengene, altså de som tjener mindre enn en ½ G, det gjelder også for lønnsutviklingen.

arbeidskraften. I et litt mer nyansert bilde vil denne lønnseffekten avhenge av om arbeiderne kan karakteriseres som substituerende eller komplementær arbeidskraft, der ufaglært arbeidskraft som mye av både tidligere og den nye arbeidsinnvandringen har bestått av, vil som substitutt ifølge teorien oppleve en nedgang i lønningene⁶⁷, alt annet likt. Men effekten vil også avhenge av hvor godt arbeidsmarkedet kan absorbere ny arbeidskraft, det vil si at effekten også avhenger av konjunktursituasjonen. Bratsberg og Raaum (2013) viser at lønnseffekten har større effekt på innvandrerne selv enn på nordmenn. En kan derfor tenke seg at den avtagende veksten i reallønnen for gruppe 1 er en konsekvens av den nye arbeidsinnvandringen.

Figur 5.5 Reallønnsutviklingen i perioden 1998-2008 for de fire landgruppene.

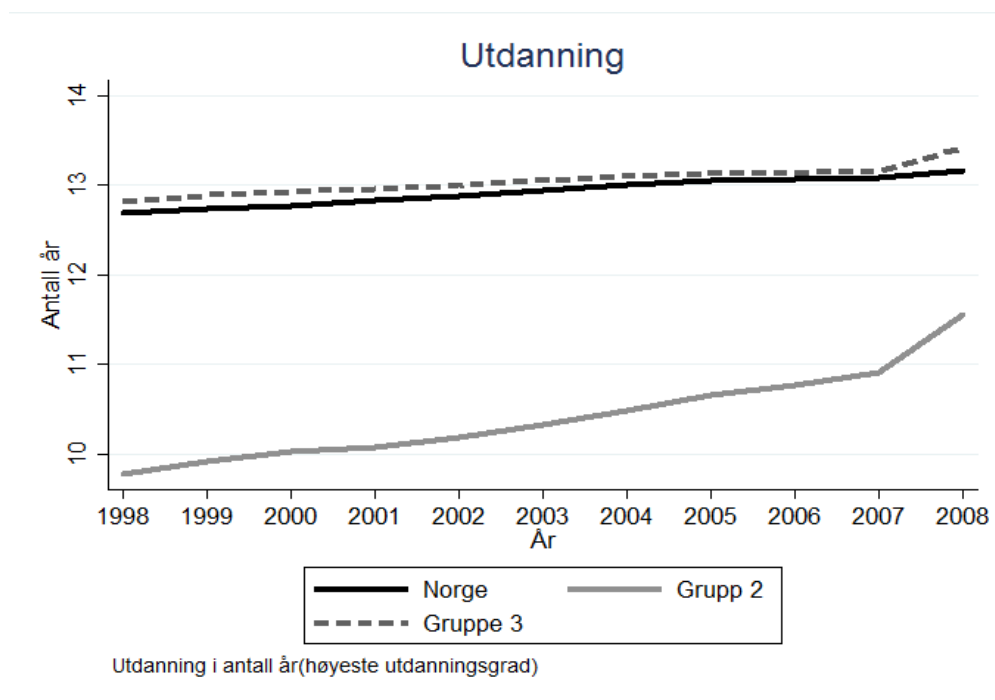


I kapittel 3 så en at arbeidstilbudsmodellen predikerer at høyere lønn vil gi lavere sykefravær, og da også at lavere lønn tilsier et høyere sykefravær. Men selv om en gjerne ser en negativ korrelasjon mellom lønn og sykefravær, er det i tilfelle med gruppe 1 flere aspekt som gjør at denne tolknigen blir mer komplisert. Den relative lønnen til gruppe 1 mot Norge synker gjennom hele perioden. Om en antar en positiv sammenheng mellom lavtlønnet arbeid og sykefravær, vil en forvente, alt annet likt, at individ fra gruppe 1 har et høyere sykefravær enn individ tilhørende Norge. Spørsmålet blir da om det er lønningene i Norge eller i hjemlandet individene fra gruppe 1 primært forholder seg til. Svaret vil antageligvis variere fra arbeider til arbeider, men det kan tenkes at det avhenger av lengden på planlagt opphold, altså hvor lenge en har tenkt å bli å arbeide i Norge.

⁶⁷ Komplementær arbeidskraft vil da, ifølge teorien, få en lønnsøkning siden de nå blir et relativt knapp resurs.

Lønn kan delvis reflektere hvor høy utdanning et individ har, men høyere utdanning er gjerne forbundet med mer fleksibel og selvkontrollert arbeidsforhold(tid), ikke bare en høyere lønn. En ser gjerne en negativ sammenheng mellom utdanning og sykefravær (Piha et al 2009 og Aagestad et al 2014) og det er flere årsaker til dette. Flexibilitet og mer selvbestemmelse av arbeidsdagen er noen nevnte årsaker, men det kan også være slik at personer som velger å ta mer utdanning har spesifikke karakteristikk, for eksempel bedre helse og er mer resurssterke enn de som ikke tar høyere utdanning. Dette er forhold en vil komme nærmere inn på i neste kapittel, da en i den empiriske analysen bruker modeller som tar hensyn til uobservert individuell heterogenitet. Under ser en figur 5.6 som viser utviklingen i gjennomsnittlig antall utdanningsår for de tre landgrupper. Utviklingen for gruppe 1 er ekskludert på grunn av mange uregistrerte observasjoner, som nevnt tidligere under delkapittel 5.3

Figur 5.6 Gjennomsnittlig utdanningsnivå, målt i antall år (høyeste grad) i perioden 1998-2008, for tre av landgruppene.



Gruppe 2 skiller seg ut ved å ha lavest antall år med utdanning, dette er forenelig med andre datakilder. Gruppe 2 består av mange land som har en stor andel av individ med flyktningebakgrunn. Ut i fra figur 5.6 og en negativ korrelasjon mellom utdanning og sykefravær vil det ikke være en betydelig forskjell i sykefraværet til gruppe 3 og Norge, men Gruppe 2 vil ha et høyere sykefravær enn de resterende landgruppene, alt annet likt. Et annet aspekt en legger merke til er at både gruppe 2 og 3 får en liten knekk opp fra 2007 til 2008, mens Norge ligger relativt

stabil dette året. Et økende utdanningsnivået generelt og blant innvandrere finner en også i annen utdanning statistikk.

Som nevnt vil sykefraværet bli bestemt av en rekke sammensatte faktorer, flere enn de overnevnte. En faktor i denne oppgaven som er verdt å trekke fram, er *innvandringsårsak*. Innvandringsårsak sier noe om grunnlaget for opphold, som kan reflektere både helsetilstand eventuelt incentiv og mulighet til å bidra i arbeidsmarkedet. Flere studier har vist at sykefraværet variere med om innvandreren har flyktningebakgrunn eller er arbeidsinnvandrere, se blant annet Dahl et al (2010) og Bengtsson og Scott (2006).

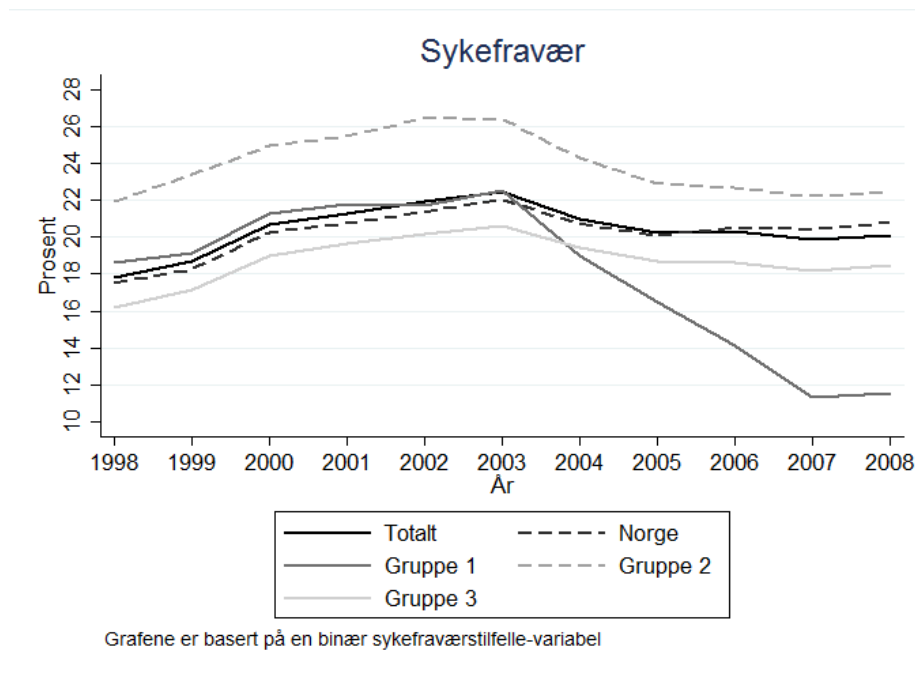
Fra figurene vi har sett i dette kapitlet, får en inntrykk av at den nye innvandringen fra gruppe 1 primært består av menn i relativt ung arbeidsfør alder. Dette er en gruppe arbeidere som har et lavere sykefravær enn andre grupper, slik som kvinner og eldre. Men en har også sett at gruppe 1 i større grad er ansatt i deltidsstillinger og sysselsettingen er mer volatil enn for gruppene Norge og gruppe 3, der disse også generelt sett har høyere lønninger enn både gruppe 1 og 2. Ut i fra utviklingen i de demografiske- og arbeidsmarkedsfaktorene vil en kanskje vente at gruppe 3 har sykefraværsliv og -utvikling som ligner mest på majoritetsbefolkningen. Mens det for gruppe 2 peker mer i retning av et høyere sykefravær enn de to overnevnte gruppene. Ut i fra utviklingen i de demografiske faktorene, vil gruppe 1 etter 2004 ha en relativ reduksjon i sykefravær. Antar en også at de primært forholder seg til kostnadsnivået i sitt hjemland, vil dette peke i samme retning. En skal nå se hvordan det gjennomsnittlige sykefravær har utviklet seg for de fire gruppene. Om noen av forventningene slår til, kan det være en indikasjon på at disse er viktig å ha med som forklaringsvariabler i den empiriske analysen.

5.4 Sykefravær

For å se på utviklingen i sykefraværet i datasettet er det tatt utgangspunkt i den binære variabelen for sykefraværet. Det vil i første rekke være interessant å se på sykefraværstilfeller, spesielt for gruppe 1 relativt til Norge. Når det gjelder lengden av sykefravær som til dels avhenger av hvor mange sykefraværstilfeller det er i hver gruppe, vil en komme tilbake til i den empiriske analysen. Den gjennomsnittlige prosentandelen beregnes på bakgrunn av antall individer som har hatt et eller flere sykefraværstilfeller i løpet av året. For hver gruppe vil sykefraværprosenten være andelen av sykmeldte av gruppens totale utvalg. Figur 5.7 illustrerer denne utviklingen for de respektive gruppene: Norge, gruppe 1, gruppe 2 og gruppe 3.

Når en skal kunne si noe om bidraget fra de ulike gruppene til det totale sykefraværet, så må en sammenligne de ulike gruppene med variasjonen i totalt sykefravær. Som en så tidligere, er gruppene av ulike størrelser og en bør derfor ha i minne hvor store gruppene er i forhold til totalen, for å si noe om en eventuell påvirkning er rimelig å anta.

Figur 5.7 Utviklingen i prosenandelen med ett eller flere sykefraværstilfeller, for perioden 1998-2008.

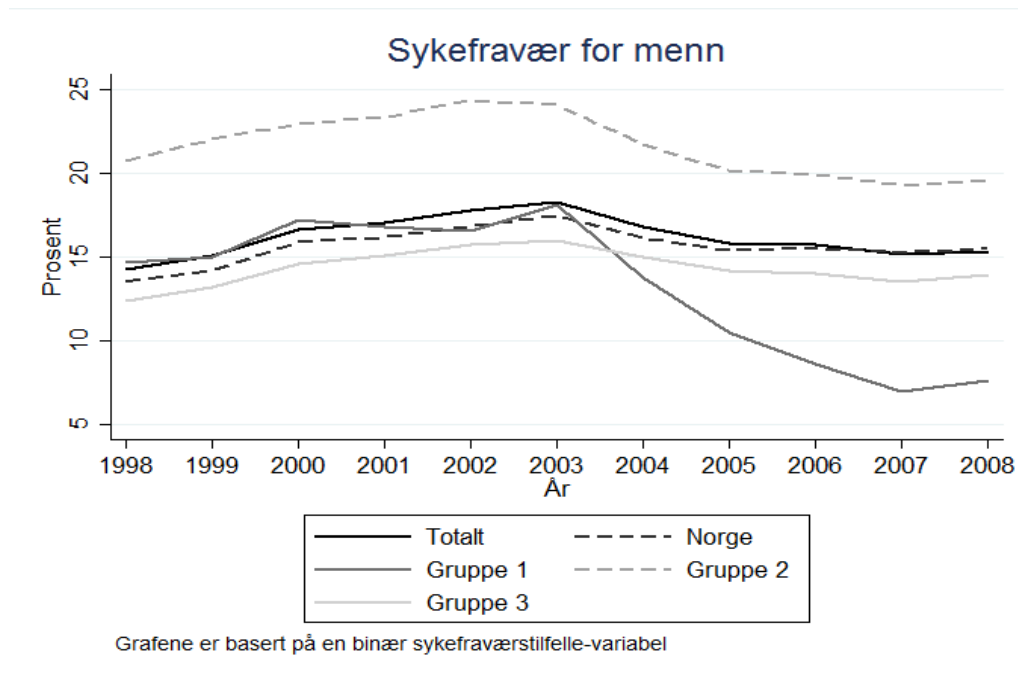


En ser at for samtlige grupper begynner sykefraværet å synke i 2003 og utover i 2004. Dette gjelder spesielt gruppe 1, men også gruppe 2 har relativt stort fall i sykefraværet i denne perioden. Reduksjonen flater mer ut for både Norge og gruppe 2 rundt 2005, mens sykefraværet tydelig fortsetter å falle frem til 2007 for gruppe 1. Dette er interessant, da en på samme tid ser at det totale sykefravær, som inkluderer alle grupper, reduseres til å ligge under sykefraværnivået til Norge fra 2005. Rundt samme tidspunkt tiltok utviklingen i den deskriptiv statistikken for alder og kjønnsfordelingen, for gruppe 1, med en reduksjon i alder og en økt andel menn.

Selv om gruppe 1 er en relativt liten gruppe, er reduksjonen så stor at det kan tenkes at dette har bidratt til at det totale sykefraværet utover år 2005 ligger under sykefraværet til majoritetsbefolkning alene, sammen med en mindre reduksjon både i gruppe 2 og gruppe 3. Dette er noe som vil være interessant å se nærmere på i den empiriske analyse. Finner vi forskjeller og er de eventuelt signifikante, også når vi kontrollerer for kjønn, alder og andre bakgrunnsfaktorer?

En så tidligere at andel menn i gruppe 1 var på rundt 70 prosent i slutten av perioden, av den grunn er det interessant å se sykefraværet for menn og kvinner isolert. Under har en tilsvarende graf som figur 5.7, men bare for menn.

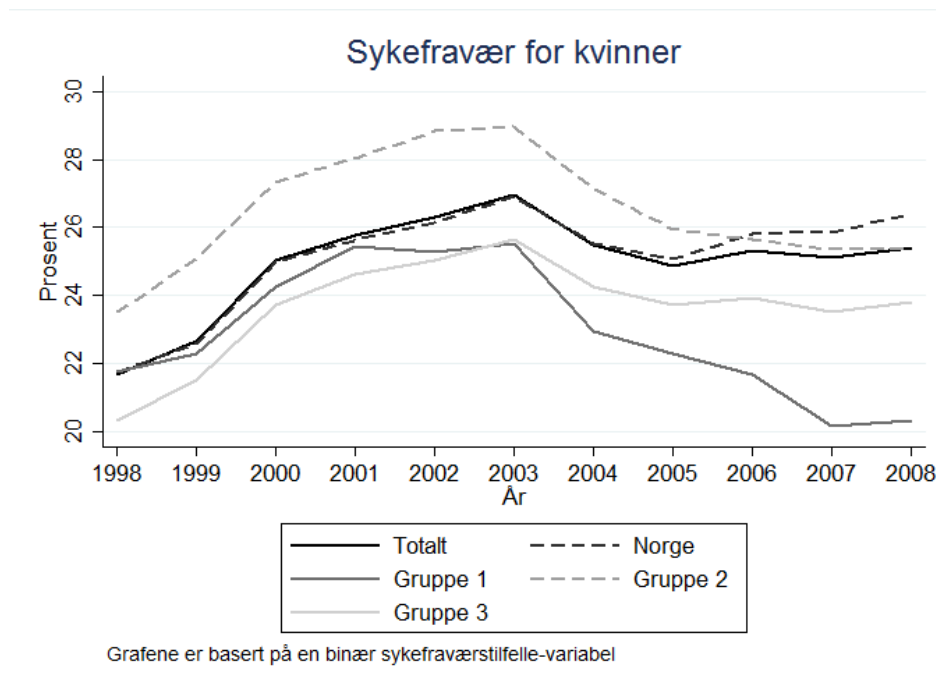
Figur 5.8 Utviklingen i prosenandelen med ett eller flere sykefraværstilfeller i peroiden 1998-2008, for menn.



Utviklingen for landegruppene er svært like for menn som for det totale bilde i figur 5.7. Menn fra gruppe 1 har hatt det desidert største fallet i sykefraværet fra 2004 og frem til 2007. En legger også merke til at selv om utviklingen er ganske lik, er nivået lavere, noe som understreker at menn generelt sett har et lavere sykefravær enn kvinner. Som i figur 5.7 ser en at det også for menn er slik at det totale sykefravær blir lavere enn sykefraværet for majoritetsbefolkningen, altså at innvanderne bidrar negativt til det samlede fraværet, men i noe mindre grad en når en så på både menn og kvinner samlet.

Når det gjelder sykefraværet for kvinner isolert, ser en først og fremst at nivået er høyere enn for menn. Økningen frem til 2003 er også brattere for kvinner enn menn, gruppe 1 skiller seg også ut isolert for kvinner, med en kraftig reduksjon og et relativt lavt nivå på sykefraværet i slutten av perioden. Et annet poeng her er at sykefraværet til Norge øker tydelig over det totale sykefraværet fra 2005, samtidig som gruppe 2 og 3 faller marginalt. Dette skjer mens gruppe 1 fortsetter å falle et par prosentpoeng. Et interessant spørsmål er hvorfor sykefraværet for kvinner med Norge som landbakgrunn øker, mens det for kvinner fra gruppe 1 fortsetter å synke?

Figur 5.9 Utviklingen i prosenandelen med et eller flere sykefraværstilfeller for perioden 1998-2008, kvinner.



5.5 Oppsummerende kommentarer

Hovedformålet med dette kapitlet har vært å gi innsikt i oppgavens datasett og en deskriptiv fremstilling og diskusjon som kan være et utgangspunkt for den empiriske analysen. I utvalget som består av 6,84 millioner observasjoner er det generert fire landgrupper; Norge, gruppe 1 bestående av EU-utvidede land siden 2004, Gruppe 2 med land tilhørende Asia, Oseania Afrika, Sør- og Sentral-Amerika og land i Europa utenfor EU/EØS og gruppe 3 med resterende land tilhørende Nord-Amerika, Australia og New Zealand og EU/EØS-land. I den deskriptive delen så en for gruppe 1 at andel menn økte og den gjennomsnittlige alderen sank rundt tidspunktet da EU ble utvidet, samtidig økte også sysselsettingen for denne gruppen. Både i dette kapitlet og bakgrunnskapitlet har en sett at gruppe 1 har hatt en økning i sysselsettingen og er nå en større andel av den totale sysselsettingen. Samtidig har de hatt et betydeligere lavere sykefravær enn alle de andre landgruppene etter 2004, både for menn og kvinner, men reduksjonen har vært størst for menn. Dette er i tråd med oppgavens hypotese. Det er derfor interessant å se om de resultatene fra den økonometriske analysen kan underbygge tendensene som er beskrevet i dette kapitlet, og hva som eventuelt kan forklare dette.

Gruppe 1 har i tillegg vært den gruppen med desidert størst nedgang i sykefraværet etter 2004, både for kvinner og menn. Samtidig ser det også ut til at nedgangen har vart i en lengre periode for denne gruppen. Gruppe 3 er den gruppen som har vist seg å være mest lik

majoritetsbefolkningen, Norge, både når det gjelder arbeidsmarkedsfaktorer og demografiske kjennetegn. De har likevel litt lavere sykefravær, både for kvinner og menn. Gruppe 2 har derimot hatt det høyeste sykefravær, for begge kjønn. De skiller seg ut ved å ha relativt lav utdanning og lønn samtidig som de har en lavere andel i heltidsarbeid.

6.0 Empirisk analyse

6.1 Introduksjon

I dette kapitlet presenteres og diskuteres de empiriske resultatene. Den første delen ser på hvordan landbakgrunn påvirker sannsynligheten for å påbegynne et sykefravær (insidensraten). Her er det utført både lineære regresjoner og binære sannsynlighetsmodeller på det totale utvalget, med og uten bakgrunnsvariabler og tids-kontroller. Videre ser jeg på hver landgruppe isolert, med et spesielt fokus på individets oppholdstid og om en er andregenerasjonsinnvandrere. Den andre delen av kapitlet går ut på tilsvarende opplegg, men fokuset rettes nå på varigheten av sykefravær. En bør merke seg at koeffisientene ikke kan tolkes som rene kausaleffekter; årsaker og utfordringer knyttet til dette blir gjennomgått i en kritisk vurdering av analysen.

Under vises en tabell med gjennomsnitt for sykefraværet (prosent) og forklaringsvariablene. Selvsagt er også gjennomsnittet på sykefravær for Norge av interesse, da det er denne størrelsen en tolker størrelsene på koeffisientene opp imot, siden Norge er referansegruppe i regresjonstabellene 6.2 og 6.4. Gjennomsnittene under vil både være over tid og individ, og tallene viser til samme tendenser som en så i den deskriptive delen⁶⁸. For eksempel har gruppe 1 laveste sykefraværet både i prosent og i antall dager, mens gruppe 2 har størst for begge størrelser. Når det gjelder oppholdstid, målt i antall år, er den lavest for gruppe 1 og høyest for gruppe 3, men gruppe 3 har samtidig den laveste andelen med andregenerasjonsinnvandrere.

Tabell 6.1 Gjennomsnitt for avhengige- og forklaringsvariabler for de fire landgruppene og det totale utvalget.

	Norge	Gruppe 1	Gruppe 2	Gruppe 3	Total
Antall sykedager ⁶⁹	96,3	91,8	98,9	95,1	96,5
Sykefravær (binær)	20,3	16,6	24,1	18,9	20,6
Andel menn	0,51	0,54	0,53	0,53	0,52
Alder	40,0	37,5	34,6	39,3	38,9
Oppholdstid, år	-	14,0	14,8	27,4	21,2
Utdanning, år	12,9	12,7	10,5	13,1	12,6
Barn<11 år	0,30	0,62	0,56	0,48	0,38
Reallønn	25,83	23,23	20,41	27,94	25,31
2.gen.innv	-	0,033	0,050	0,015	0,031
Fulltid	0,69	0,60	0,56	0,69	0,67

⁶⁸ Her kommenteres noen av størrelsene kort, da en har hatt en fyldigere gjennomgang av deskriptiv statistikk i forrige kapittel. Merk tallene i tabell 6.1 gjelder med utsortering av individ med kun et observasjonsår.

⁶⁹ Dette er gjennomsnittlig antall sykedager betinget på at en har hatt et eller flere sykefravær i løpet av året.

6.2 Landbakgrunn og sannsynligheten for sykefravær

For å undersøke hvilke effekt landbakgrunn har på sannsynligheten for sykefravær er det valgt å kjøre både lineære regresjonsmodeller og ikke-lineære sannsynlighetsmodeller (MKM og Probit, henholdsvis). I både det lineære og ikke-lineære tilfellet er det tatt hensyn til uobserverte individuelle karakteristikk (RE og RE-Probit, henholdsvis).

Kolonner to til fem i tabell 6.2 viser de ubetingede effektene av landbakgrunn, målt med de tre landgruppedummyene gruppe 1, gruppe 2 og gruppe 3. Dette er en annen måte å framstille sannsynligheten for sykefravær i forhold til tabell 6.1, nemlig i form av forskjeller i stedet for nivå.⁷⁰ I tillegg forteller signifikansnivået (markert med én til tre stjerner) om forskjellene er signifikante eller ei. I kolonne 6-8 inkluderes en step-dummy som tar verdien 1 for år > 2003 og 0 for år ≤ 2003, samt interaksjonsledd mellom denne step-dummen og de respektive gruppene for å se på sannsynligheten for sykefravær, relativt til Norge, før og etter 2004.⁷¹ Videre, i kolonne ni og ti, tas det også med en rekke forklaringsvariabler⁷² for å se om disse kan forklare noe av forskjellene i sykefraværet mellom de ulike gruppene og Norge. Som nevnt er Norge referansegruppen og koeffisientene til gruppe 1, 2 og 3 tolkes derfor som sannsynlighetene for sykefravær relativt til majoritetsbefolkningen. I tabell 6.2 er alle koeffisientene statistisk signifikante, atpåtill på 0,1 prosents signifikansnivå. Med dette utgangspunktet vil kommenteringen på tolkningene primært bestå av retning og styrke på resultatene.

Kolonner to til fem i tabell 6.2 viser at før en har kontrollert for bakgrunnsfaktorer, så har gruppe 1 (EU-utvidede land) en relativt lavere sannsynlighet for sykefravær enn norskfødte. Gruppe 1 har den laveste sannsynlighet for å starte opp et sykefraværstilfelle av alle gruppene, og dette gjelder uavhengig hvilke av modellene (fra kolonne to-fem) en ser på. Når en tar hensyn til uobserverte kjennetegn med en tilfeldig effekt-modell (RE og RE-Probit i hhv kolonne fire og fem), øker differansen mellom Norge og gruppe 1. Gruppe 1 har mellom 3,7 – 7,2 prosentpoeng lavere tilbøyelighet for å påbegynne et sykefravær enn majoritetsbefolkningen, avhengig av om en tar hensyn til uobserverte karakteristikk eller ikke.

⁷⁰ For eksempel betyr -0.037 at gruppe 1 har 3.7 % lavere sykefravær enn etniske nordmenn, som er det samme som det vi kan lese ut av tabell 6.1 (20,3 – 16,6 = 3,7).

⁷¹ En mer fleksibel fremgangsmåte ville være å representere hvert år med en separat dummy. Dette er gjort, uten at det bød på vesentlig andre resultater. Vi havnet derfor på modellen med step-dummy for å forenkle framstillingen i de mange tabellene som følger.

⁷² Se vedlegg for variabelforklaringer.

Tabell 6.2 Modeller på sykefravær (binær variabel), for det totale utvalget (alle grupper)

	MKM	Probit (dy/dx)	RE	RE- Probit (dy/dx)	MKM	Probit (dy/dx)	RE	RE- Probit (dy/dx)	MKM	RE
Gruppe 1(G1)	-0.037***	-0.040***	-0.064***	-0.072***	0.0092***	0.0092***	-0.043***	-0.052***	0.0056**	-0.011***
Gruppe 2(G2)	0.038***	0.037***	0.030***	0.029***	0.049***	0.047***	0.029***	0.028***	0.068***	0.075***
Gruppe 3(G3)	-0.014***	-0.014***	-0.017***	-0.018***	-0.012***	-0.012***	-0.016***	-0.018***	-0.0075***	-0.0044***
D04					0.00604***	0.0061***	0.024***	0.025***	0.0092***	0.012***
D04*G1					-0.070***	-0.077***	-0.034***	-0.036***	-0.053***	-0.018***
D04*G2					-0.021***	-0.020***	-0.0026*	-0.0029***	-0.026***	-0.0049***
D04*G3					-0.0048***	-0.0048***	-0.0043***	-0.0037***	-0.0064***	-0.0030***
Mann									-0.072***	-0.085***
Alder									0.014***	0.012***
Alder ²									-0.00014***	-0.00012***
Barn <11 år									0.028***	0.0081***
Mann*Barn<11 år									-0.027***	-0.012***
Oppholdstid, år									0.00030***	0.00096***
2. gen. innv									-0.018***	-0.030***
Reallønn (10 000 kr)									0.0013***	-0.00027***
Fulldid									0.025***	0.033***
Konstant	0.20***	-0.83***	0.21***	-1.09***	0.20***	-0.84***	0.19***	-1.15***	-0.093***	-0.020***
Observasjoner	6745535	6745535	6745535	6745535	6745535	6745535	6745535	6745535	6744212	6744212
Adjusted R ²	0.0018				0.0020					

Kilde: SSB, norsk registerdata
* p < 0.05, ** p < 0.01, *** p < 0.001

Dette betyr at relativt til Norge, har gruppe 1 rundt 18,2-35,5 prosent mindre sannsynlighet for å påbegynne et sykefraværstilfelle, noe som må sies å være en betydelig forskjell.

Disse landforskjellene, på linje med nivåene som rapporteres i tabell 6.1, er et interessant utgangspunkt for analysen, da det støtter opp under oppgavens hypotese om at den nye arbeidsinnvandringen kan ha bidratt til en reduksjon i sykefraværet. Tidligere studier har funnet at sannsynlighet for å påbegynne et sykefraværstilfelle blant østeuropeiske innvandrere er høyere enn for norskfødte⁷³ (se for eksempel Dahl et al (2010)). De finner imidlertid også at disse forskjellene «forsvinner» dersom en kontrollerer for ulike bakgrunnsfaktorer. Siden studien er gjort med data før 2004, kan de ulike resultatene fra før EU-utvidelsen om større sannsynlighet og etter EU-utvidelsen med mindre sannsynlighet også være med på å gi en styrket tro på oppgavens hypotese.

Men en så i den deskriptive delen i kapittel 5 at dette ikke er tilfelle i hele observasjonsperioden. Med et fokus på den nye arbeidsinnvandringen fra 2004, så en i kapittel 5 at det først er etter 2004 at gruppe 1 har hatt et relativt lavt sykefravær. Når en i kolonne 6-8 inkluderer step-dummyen og interaksjonsleddene, gjør dette at en kan skille de relative tilbøyelighetene til landgruppene før og etter 2004.

Gruppekoeffisientene er nå den relative sannsynligheten for sykefravær før 2004. Resultatene viser da at gruppe 1 har rundt 1 prosentpoeng høyere sannsynlighet for sykefravær før 2004. Interaksjonsleddene viser at sannsynligheten for sykefravær er redusert etter 2004, innad i alle gruppene. Dette stemmer godt overens med utviklingen en så i kapittel 5, der sykefraværet gikk ned for alle grupper fra rundt 2005 og utover. Ser en på gruppe 1 fra 2004, har de mellom 30,1-33,5 prosent mindre sannsynlighet for å påbegynne et sykefravær enn majoritetsbefolkningen⁷⁴. Dette illustrerer skiftet i sykefraværet for EU-utvidede land etter 2004, og gir ytterligere støtte til oppgavens hypotese. Det går ikke fram av resultatene så langt hva som har drevet dette skiftet, for gruppe 1 fra knappe ett prosentpoeng høyere sannsynlighet for sykefravær før 2004 til 7 prosentpoeng lavere sannsynlighet fra og med 2004. Et sentralt poeng er innføring av reformen rundt 2004, der en strengere portvaktrolle stod sentralt, som er vist å ha hatt en relativt stor effekt på sykefraværet (Markussen 2009b). Men som argumentert for tidligere, vil jeg i

⁷³ Det vil være noen begrensninger ved å sammenligne disse resultatene da Dahl et al (2010) har gruppen Øst-Europa, i denne oppgaven vil disse havne i to grupper: Gruppe 1 for EU-utvidede land og Gruppe 2 for andre Østeuropeiske land.

⁷⁴ For MKM: ca 6,1 prosentpoeng, Probit: ca 6,8 som tilsvarer 30,1 og 33,5 prosent av det norske gjennomsnittet fra tabell 6.1

mindre grad tro dette vil gjelde i stor utstrekning for EU-utvidede land (gruppe 1), fordi de fleste i den gruppen har ankommet Norge på eller etter reform tidspunktet. Det er kun en mindre andel av gruppe 1 som innvandret før 2004 og som har vært igjennom denne reformendringen. Dette vil i så tilfelle gi ytterligere indikasjon på at oppgavens hypotese har noe for seg.

Når det gjelder de andre gruppene, så har gruppe 2 høyere, mens gruppe 3 har lavere sannsynlighet for å begynne et sykefraværstilfelle enn majoritetsbefolkningen (kolonne 2-8), som er overensstemmende med tidligere resultatet (se Hansen et al 2014 og Dahl et al 2010). Mulige årsaker til disse sammenhengene for gruppe 2 og 3 kan som tidligere nevnt være ulik oppholdsgrunn, som også kan forbindes med ulik helsestatus. I sin studie på innvanderens sykefravær og selvrapportert helse, finner Brekke og Shøne (2013) at ikke-vestlige innvandrere rapporterer en generell dårligere helse enn andre grupper. Når de så har kontrollert for forskjellige helsestatus og andre bakgrunnsvariabler finner de ingen signifikant forskjell i sykefraværet mellom ikke-vestlige innvandrere og majoritetsbefolkningen. Dette kan tyde på at ulik helsestatus sammen med andre bakgrunnsfaktorer kan forklare forskjellen i sykefraværet. Et interessant spørsmål er om selvrapporterte helsestatus er et godt mål på individets reelle helse. Forfatterne bruker også et objektive mål på helse, og finner lignende, men svakere effekter.

Som nevnt reduseres sannsynligheten for å påbegynne et sykefraværstilfelle etter 2004, også for gruppe 2 og 3. Men mønsteret er fortsatt som før, altså at gruppe 2 har høyere, mens gruppe 3 har lavere sannsynlighet for sykefravær enn norskfødte, men gruppe 2 er nå nærmere Norge, mens gruppe 3 har en økt differanse med enda lavere sannsynlighet.

Faktorer som forklarer forskjellene?

Så langt har en altså sett at det er forskjell i sykefraværene mellom landgruppene og majoritetsbefolkningen, og mellom landgruppene. Så hva kan forklare disse forskjellene? I den deskriptive delen så en at utviklingen i demografiske og arbeidsmarkedsfaktorer var svært forskjellig blant gruppene, spesielt kjønn- og aldersutviklingen for gruppe 1 skilte seg ut. Kan disse bakgrunnsfaktorene forklare noe av den observerte forskjellen i sykefraværet en identifiserte i kapittel 5 og i resultatene i tabell 6.2? I kolonne ni og ti er en rekke bakgrunnsfaktorer inkludert som kontrollvariabler i regresjonen⁷⁵. Vi begynner med å se nærmere på om det har endret noe for effekten av landbakgrunn. Deretter går vi gjennom

⁷⁵ Her er probit-modellen ekskludert, da en ikke fant noe maksimumspunkt ved estimeringen i Stata. I tillegg har en sett fra de andre modellene at resultatene for probit-modellene i stor grad samsvarer med resultatene for RE og MKM.

effekten på sykefraværet av hver enkelt bakgrunnsvariabel.

Som ventet blir betydning av landbakgrunn for gruppe 1 med tanke på sannsynligheten for sykefravær noe svakere, det vil si at sannsynlighetsdifferansen mellom Norge og EU-utvidede land er redusert, både før og etter 2004. Den relative sannsynligheten går ned rundt 1,4 prosentpoeng (fra -6,1 til -4,7) når en kontrollerer for bakgrunnsfaktorer, dette tilsvarer en endring på omtrent 7 prosent. Det vil si at demografiske kjennetegn, arbeidsmarkedsstørrelser med flere kan forklare noe av forskjellen i sykefraværet for gruppe 1 og Norge, også det faktum at gruppe 1 har hatt en større reduksjon etter 2004. Fra før vet vi for eksempel at unge menn er overrepresentert i den økte arbeidsinnvandringen etter 2004; derfor er det som forventet at disse variablene fanger opp en del av effekten fra land-gruppedummyene i kolonne to til åtte. I kolonne ti tas det hensyn til uobserverte individuelle kjennetegn, vi ser at betydningen av landbakgrunn da blir svakere for gruppe 1, særlig etter 2004. Siden vi snakker om uobserverte kjennetegn, går det selvsagt ikke an å si sikkert hva som fanges opp. Men siden arbeidsføre og – sannsynligvis – arbeidsvillige menn i relativt ung alder var overrepresentert i den store kontingenten av innvandrere etter 2004, er det rimelig å trekke fram god helse, høy motivasjon, og lignende.; forhold som er assosiert med lavt sykefravær.

Også for gruppe 3 forklarer disse bakgrunnsvariablene noe av forskjellene i sykefraværet til majoritetsbefolkningen, men det gjenstår også her en reell forskjell. I motsetning til gruppe 1 og 3 så øker forskjellene for gruppe 2, dette gjelder både før 2004 og fra 2004. Dette kan bety at disse bakgrunnsfaktorene ikke kan forklarer forskjellene i sykefraværet mellom majoritetsbefolkningen og gruppe 2. Men det foreligger fortsatt en statistisk signifikant forskjell mellom innvandringsgruppene og Norge.

Demografi og familieliv

Ser en på effekten av bakgrunnsvariablene på sykefraværet, vil menn ha lavere sannsynlighet for å begynne et sykefraværstilfelle enn kvinner. Dette er et velkjent funn i empiriske studier på sykefravær (Smeby et al 2009). Av bakgrunnsfaktorene har denne størst effekt på sykefraværet. Det skiller rundt 7,2 prosentpoeng i sykefraværet mellom menn og kvinner, det vil si at menn har rundt 35 prosent mindre sannsynlighet for å begynne et sykefraværstilfelle for det totale utvalget. Dette er en betydelig og stor forskjell og spesielt interessant med tanke på at gruppe 1, EU-utvidede land, primært har bestått av menn etter 2004. Her vil det derfor være av interesse å se på sammenhengen mellom kjønn og sykefravær for hver gruppe isolert for å se om en kan identifisere forskjeller mellom landgruppene. Dette kommer vi tilbake til

under den komparative analysen.

Individets alder har en positiv, men avtakende korrelasjon med sannsynligheten for sykefravær. Korrelasjonen mellom alder og sykefravær snur ved 51-53 årsalderen⁷⁶, som betyr at sannsynligheten for sykefravær øker med alder opp til 51-53 årsalderen for så å avta. Ofte kan dette være motsatt av hva en forventer, da en gjerne antar en negativ korrelasjon mellom helse og alder. Lignende resultat er også funnet tidligere se blant annet Markussen et al (2009). Bratberg et al (2010) peker blant annet på at en årsak kan være selektering ut av arbeidslivet av individ med dårligere helse. Dette kan medfører at de med relativt god helse står igjen i arbeidslivet og bidrar til å redusere sannsynligheten for sykefravær.

Det å ha barn under 11 år er forbundet med et høyere sykefravær, særlig for kvinner. Faktisk viser resultatene at menn med barn under 11 år kun har marginalt høyere sykefravær enn menn uten barn (0,1 prosentpoeng), mens kvinner har 2,8 prosentpoeng høyere sannsynlighet for sykefravær enn menn med barn under 11 år. For kvinner kan dette reflektere at det fortsatt er slik at kvinner forbindes med omsorgsrollen når en stifter familie, mens menn har oppgaven med å forsørge. For eksempel viser det seg at menn blir mer stabile i arbeidsmarkedet når de får barn, noe som kan tyde på menn inntar den mer tradisjonelle forsørgelsesrollen ved familiestiftelse.

Arbeidsmarkedsforhold, ansettelsesgrad og lønn

Tidligere så en at det var variasjon mellom gruppene på hvor store andeler som arbeidet fulltid. Tabell 6.2 over viser at det å arbeide fulltid er positiv korrelert med sykefravær. Resultatet viser at en som arbeider fulltid har rundt 2,5 prosentpoeng høyere tilbøyelighet for sykefraværstilfeller enn de som ikke arbeider fulltid. Dette tilsvarer rundt 12 prosent. Dahl et al (2010) finner tilsvarende resultat, det å jobbe fulltid øker sannsynligheten for sykefravær. En hypotese kan være at det er en seleksjon som fører til at deltidsansatte har lavere sykefravær enn heltidsjobbene. Deltidsansatte bruker redusert stilling som en mestringsstrategi for å balansere arbeid og familieliv (dette vil kun gjelde dersom valget om å jobbe deltid er frivillig). Tar en hensyn til uobserverte karakteristikk, kan det virke som resultatet taler mot denne hypotesen, da korrelasjonen mellom sykefravær og det å jobbe fulltid styrkes.

⁷⁶Ved å derivere mhp alder finner enn maksimumspunktene ved: $0,0145 - 2(0,000240) * \text{alder} = 0 \rightarrow \text{alder} = 0,0145 / 0,00028 = 51,79$ for MKM og $0,0128 - 2(0,0012) * \text{alder} \rightarrow \text{alder} = 0,0128 / 0,00024 = 53,33$ for RE.

Sannsynligheten for sykefravær er negativ korrelert med lønn. Dette tilsvare den teoretiske prediksjonen en så ved arbeidstilbudsmodellen med fravær. Dersom substitusjonseffekten dominerer vil en økning i lønnen, alt annet likt, øke individets arbeidstilbud og redusere sitt fravær. Tilsvarende sammenheng mellom sykefravær og lønn har støtte fra empirien (Dahl et al 2010 og Bengtsson og Scott 2006). Et problem med denne tolkningen er at en ikke vet hvilke vei denne effekten går. Det er flere studier som viser at kausaliteten også går andre vei, altså vil lønnen påvirkes av sykefraværet. Et funn er at sykefraværet har en kausal negativ påvirkning på individets lønnsutvikling, se for eksempel Markussen (2012) og Ichino og Moretti (2009). Som nevnt i metodekapittelet, så er to-veis-kausaltet en kilde til endogenitet. Dette har implikasjoner for den estimerte lønnskoeffisienten, men også de andre koeffisientene kan være påvirket som følge av endogenitet i lønnsvariabelen.

Oppholdstid og andregenerasjonsinnvandrere

En har allerede sett at det er forskjeller i sykefraværet mellom innvandrergруппene og majoritetsbefolkningen. Et interessant moment her er om denne tilpasningen av sykefravær for innvandrergруппene gjelder etter som individene bli boende og integrert i det norske samfunnet? Og hvordan er det med andregenerasjonsinnvandrere, presterer de bedre eller tilsvarende som sine foreldre? Analysen viser at oppholdstid, målt som antall år siden ankomst, har en positiv samvariasjon med sykefraværstilfelle. Det vil si at sannsynligheten for et sykefraværstilfelle øker med antall år en har oppholdt seg i Norge. For innvandringskohorter som har vært i Norge lenger enn 10-15 år finner Bratsberg et al (2014b) at arbeidsmarkedetsdeltakelsen er synkende, noe som igjen har medført et økt forbruk av uføreforsikring. Dette vil i oppgavens kontekst i stor grad være individ tilhørende gruppe 2. Spørsmålet relatert til oppholdstid er også svært interessant å se videre på for hver enkelt gruppe, spesielt EU-utvidede land. Men for denne gruppen vil det være begrensninger i og med at datagrunnlaget kun streker seg 4 år etter EU-utvidelsen, det er derfor vanskelig å si noe om en langsiktig tilpasning.

Ser en på tilpasningen til andregenerasjonsinnvandrere viser studien til Bratsberg et al (2014b) at innvandrere barn, født i Norge, gjør det bedre enn sine foreldre. Tilsvarende resultat er rapportert i tabell 6.2, som viser at andregenerasjonsinnvandrere har rundt 1,8 prosentpoeng mindre sannsynlighet for sykefravær enn førstegenerasjonsinnvandrere generelt for alle gruppene. Ser en dette opp mot gruppenes gjennomsnittssykefravær tilsvarer dette 10,8, 7,5 og

9,5 prosent for gruppe 1, 2 og 3 henholdsvis. En annen studie viser at det ikke er forskjell mellom andre generasjonsinnvandrere og majoritetsbefolkningen (Dahl et al (2010)).

Komparativ analyse av landgruppene

Etter å ha sett på bakgrunnsvariablenes generelle effekt, er det som nevnt interessant å se om de ulike faktorene har forskjellige påvirkning, både i størrelse og retning, fra gruppe til gruppe. Under presenteres regresjoner med tilsvarende bakgrunnsfaktorer som i tabell 6.2 for hver av de fire gruppene (inkludert referanse gruppen Norge) isolert. Siden ikke de binære valgmodellene skilte seg fra MKM og RE, og tilsvarende likhet finnes når en kjører regresjonen for hver gruppe er det her valgt kun å rapportere resultatene fra de lineære regresjonene MKM og RE. Utdanningsvariabelen legges til som bakgrunnsvariabel for Norge gruppe 2 og gruppe 3. Det å inkludere utdanningsvariabelen utgjorde ikke en stor forskjell på de andre variablene, og det vil derfor fortsatt være mulig å sammenligne resultatene mellom gruppene. Fokuset vil være på sammenhengen mellom oppholdtid og andregenerasjonsinnvandrere med sykefravær for de 3 innvandringsgruppene. I tillegg vil det bli en komparativ del, der en ser hvilke gruppe som ligner mest og minst på majoritetsbefolkningen.

Resultatene i tabell 6.3 viser at korrelasjonene har samme retning for alle gruppene for alle faktorer, med unntak av step-dummysen D04. Menn, utdanning, reallønn, mann med barn og andregenerasjonsinnvandrere har en negativ korrelasjon med sykefravær. Mens det å ha barn under 11 år, om en arbeider fulltid, og økende oppholdtid og alder er forbundet med et høyere sykefravær.

Det er tre ting verdt å legge merke til når en ser på gruppe 1. De skiller seg ut med en relativ sterk sammenheng mellom sykefravær og kjønn og relativ svak med tanke på sykefravær og alder. Siden korrelasjonen med kjønn er størst for gruppe 1, samtidig som de har den største andel av menn etter 2004, vil det være nærliggende å tro at dette kan være noe av årsaken til at gruppe 1 hadde en kraftig reduksjon i sykefraværet som en så i figur 5.7 i kapittel 5. Når det gjelder alder er korrelasjonen positiv, men avtakende for alle landgruppene, men gruppe 1 skiller seg ut ved en relativ liten effekt. Dette kommer til uttrykk når en ser på toppunktene, det vil si der sannsynligheten snur fra positiv til negativ. For individ tilhørende gruppe 1 er det på en alder rundt 74 år, mens det for Norge, gruppe 2 og gruppe 3 er på henholdsvis 53, 47 og 51 årsalderen. Det betyr at individ fra gruppe 1 er hele 74 år før sannsynligheten synker med alder, siden dette er noen år forbi normal pensjonsalder virker det rimelig å si den generelle korrelasjonen mellom sykefravær og alder, for gruppe 1 kun er positiv.

Tabell 6.3 Modeller på sykefravær (binær variabel), for de fire landgruppene

	Norge		Gruppe 1 (EU-utvidede land)		Gruppe 2 (Ikke-vestlige innvandrere)		Gruppe 3 (Vestlige innvandrere)	
	MKM	RE	MKM	RE	MKM	RE	MKM	RE
Mann	-0.0766***	-0.0937***	-0.0969***	-0.102***	-0.0579***	-0.0599***	-0.0749***	-0.0885***
Alder	0.0160***	0.0109***	0.00663***	0.00606***	0.0226***	0.0225***	0.0147***	0.0127***
Alder ²	-0.000152***	-0.000105***	-0.000045***	-0.000040***	-0.000241***	-0.000231***	-0.000145***	-0.000117***
Barn <11 år	0.0391***	0.0117***	0.00561	0.00236	0.0357***	0.0160***	0.0267***	0.0132***
Mann*Barn <11 år	-0.0419***	-0.0176***	-0.0272***	-0.0214***	-0.0220***	-0.0150***	-0.0328***	-0.0190***
Oppholdstid, år	-	-	0.00215***	0.00248***	-0.000256***	0.000216***	0.000468***	0.000750***
2.gen.innv	-	-	-0.0241***	-0.0287**	0.0125***	0.00894**	-0.00994***	-0.00918*
Utdanning, år	-0.00983***	-0.00882***	-	-	-0.00615***	-0.00580**	-0.00576***	-0.00544***
Reallønn (10 000kr)	-0.00102***	-0.000128***	-0.000551***	-0.000129	-0.000135***	0.00118***	-0.00124***	-0.000361***
Fultid	0.0180***	0.0281***	0.0425***	0.0419***	0.0417***	0.0400***	0.0240***	0.0315***
D04	0.0130***	0.0138***	-0.0123***	-0.00408	-0.0113***	0.00118	0.00537***	0.0109***
Konstant	0.0314***	0.0382***	0.0116	0.00673	-0.154***	-0.192***	-0.0317***	-0.0277***
Observasjoner	4170361	4170361	159886	159886	855552	855552	1164953	1164953
Adjusted R ²	0.0318		0.0536		0.0399		0.0319	

Kilde: SSB, norsk registerdata

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

En annen forklaring kan være at snittalderen for gruppe 1 er relativt lav, da spesielt etter 2004, som innebærer mindre variasjon i aldersvariablene og relativt få observasjoner av eldre i denne gruppen. For gruppe 2 kommer dette snu-punktet relativt tidlig, som kan tyde på at seleksjonen ut av arbeidslivet kanskje er størst for denne gruppen. I den deskriptive statistikken så en at dette var gruppene med lavest gjennomsnittsalder, spesielt på slutten av perioden. Likevel reflekterer gruppe 1 og 2 ytterpunkter når en ser på korrelasjonen mellom sykefravær og alder.

Gruppe 3 ligner mest på majoritetsbefolkningen når det gjelder alderssammensetningen og påvirkningen av den. Men gruppe 3 er også gruppen som totalt sett har flest likhetstrekk med Norge når en ser på resultatene presentert i tabell 6.3. Dette stemmer overens med inntrykket en fikk fra den deskriptive statistikken der gruppe 3 har svært lik utvikling og nivå på både arbeidsmarkedsfaktorene og de demografiske kjennetegnene. Dette er heller ikke ulikt den generelle oppfatningen av at gruppen tilpasser seg på lik linje som majoritetsbefolkningen (Doksheim 2010). På den andre siden presterer individ fra gruppe 2, ikke-vestlige innvandrere, generelt sett dårligere i arbeidsmarkedet enn majoritetsbefolkningen (ibid). En har sett at gruppe 2 har høyeste sannsynligheten for gå i ut sykefravær, hvilke faktorer som kan forklare disse forskjellene er sikkert sammensatte. Brekke og Shøne (2013) finner at egenrapportert helse er et viktig faktor for å forklare sykefraværsforskjellene. Likevel, om en tar hensyn til uobserverte kjennetegn endres ikke korrelasjon mellom sykefravær og alder nevneverdig for gruppe 2. I tillegg er det denne gruppen sammen med gruppe 1 som har størst positiv korrelasjon med å arbeide fulltid eller ikke.

Et interessant funn for gruppe 2 er at oppholdstid er negativt korrelert med sykefravær (før kontroll for uobservert heterogenitet) og det å være andregenerasjonsinnvandrere har en positiv samvariasjon med sykefraværet. Resultatene motsier tidligere empiri (Dahl et al 2010) og er motsatt av korrelasjonsretningene for de andre gruppene. Kontrollerer en for uobserverte individuelle karakteristikk (RE), så er også økt oppholdstid for gruppe 2 forbundet med et økt sykefravær. Et poeng her er at hvis store deler av denne gruppen i tidlig alder går fra å være yrkesaktive og over til trygdeytelser vil de med lengst oppholdstid være selektert ut fra utvalget som påvirker effekten av oppholdstid. Både for gruppe 1 og 3 har andregenerasjonsinnvandrere en lavere sannsynlighet enn de som er født utenlands, som er i tråd med funn om at norskfødte innvandrere presterer bedre enn sine foreldre i arbeidsmarkedet (Bratsberg et al 2014b). Her må det imidlertid understrekes at gruppe 1 består i stor del av førstegenerasjonsinnvandrere, spesielt etter 2004. Men dersom flere av denne gruppene blir boende, vil det være interessant å

se hvordan disse, men også hvordan deres barn tilpasser seg i det norske arbeidsmarkedet på lengre sikt.

6.3 Landbakgrunn og varigheten på sykefravær

Vi har sett at sannsynligheten for sykefravær og hva som forklarer det varierer fra gruppe til gruppe. EU-utvidede land har hatt lavere sannsynlighet enn Norge etter 2004, mens gruppe 2 har høyere sannsynlighet både før og etter 2004.

Gruppe 3 er som forventet den gruppen som ligner mest på majoritetsbefolkningen når det gjelder hva som kan forklare sykefraværet, og differansen i sannsynlighet for sykefravær er heller ikke relativ stor mellom disse. Men hvordan er disse forholdene når en ser på varigheten til sykefravær? Er det slik at de individene eller gruppene med høyest sannsynlighet for sykefravær også er de som har de lengste sykefraværsperiodene?

For å se på dette, blir det utført regresjoner på varigheten av sykefravær, gitt at individet har hatt et sykefraværstilfelle i løpet av observasjonsåret. Tilsvarende betingelse er også gjort av Askildsen, Bratberg og Nilsen (2005) i deres studie om varighet av sykefravær og arbeidsledighet. En ulempe vil da være at en mister mange observasjoner. Det er her kjørt regresjoner på (logaritmen av) sykefraværsdager med både MKM og Tilfeldig effekt-modeller (RE). Jeg velger å bruke logaritmen fordi det er et ønske å tolke resultatene som marginaleffekter i prosent, dette vil blant annet gjøre det lettere å trekke paralleller til resultatene en så i de to foregående tabellene⁷⁷. Resultatet er presentert i tabellen 6.4 under. Som i tabell 6.2 er de første kolonene i tabell 6.4 regresjoner kun med de tre landgruppene som forklaringsvariabler, før en inkluderer dummy for år>2003 og interaksjonsledd med gruppene og denne dummyen. Tilslutt er det inkludert en rekke forklaringsvariabler for å se om dette kan forklare noe av forskjellen som måtte finne sted i varigheten på sykefravær.

Før en kontrollerer for andre variabler enn landbakgrunn viser resultatene fra tabell 6.4 at forholdene mellom landgruppene og Norge og de andre landgruppene er tilsvarende som når en så på sannsynligheten for sykefravær, der alle differansene er signifikant selv på 0,1 prosentnivå. Slik at for hele perioden (1998-2008) så har gruppe 1 og gruppe 3 lavere varighet i sykefraværstilfellene enn majoritetsbefolkningen, der differansen er størst for gruppe 1.

⁷⁷ Av den grunn så anvendes ikke Tobit-modellen da logaritmen til null ikke er definert. En Tobit-modell ville vært passende dersom en ikke betinget på at en hadde sykefraværstilfelle (da sorterer en bort alle med ingen (0) sykefraværsdager).

Tabell 6.4 Modeller på sykefravær (ln(antall sykefraværsdager)), for det totale utvalget (alle grupper)

	MKM	RE	MKM	RE	MKM	RE
Gruppe 1 (G1)	-0.0441***	-0.0452***	-0.0168	-0.0236*	0.00598	0.00630
Gruppe 2 (G2)	0.0192***	0.0238***	0.0313***	0.0333***	0.0993***	0.107***
Gruppe 3 (G3)	-0.0178***	-0.0180***	-0.0156***	-0.0154***	0.0113***	0.0128***
D04			0.0165***	0.0345***	0.0133***	0.0214***
D04*G1			-0.0493***	-0.0431**	-0.0102	-0.00599
D04*G2			-0.0244***	-0.0218***	-0.0150**	-0.0117*
D04*G3			-0.00509	-0.00707	0.00618	0.00367
Mann					-0.00950***	-0.0201***
Alder					0.00507***	0.00384***
Alder ²					0.0000325***	0.0000521***
Barn <11 år					0.00823***	0.0101***
Mann*Barn<11 år					-0.0922***	-0.0876***
Oppholdstid, år					0.00137***	0.00145***
2. gen.innv					-0.0135	-0.0215*
Reallønn (10 00kr)					-0.00493***	-0.00427***
Fulltid					-0.0253***	-0.0212***
Konstant	4.132***	4.117***	4.124***	4.101***	3.945***	3.926***
Observasjoner	1386349	1386349	1386349	1386349	1386022	1386022
Adjusted R ²	0.0002		0.0002		0.0140	

Kilde: SSB, norsk registerdata

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

I tillegg til å ha den høyeste sannsynligheten for sykefravær så har også individ fra gruppe 2 lengst varighet på sykefraværstilfellene. Som nevnt tidligere kan disse forskjellene ha med helsestatus å gjøre. En hypotese rundt dette kan være at hvis individ som innvandrer til Norge har gått ubehandlet, frem til de kommer til Norge, så vil helsetilstanden deres, gitt samme diagnose/sykdom, være verre for disse enn for personer som er født i Norge. Dette fordi norskfødte kan å fått behandling på et tidligere stadium. Derfor kan det tenkes at ikke bare hyppigheten av sykdom/skade er større, men også helsestatusen for samme diagnose kan være dårligere. Men dette er ikke mulig å verifisere i datamaterialet i denne oppgaven. Tilsvarende, med hensyn til uobservert heterogenitet, er det bare koeffisienten til gruppe 2 som endres nevneverdig, med en liten økning.

Om en skiller mellom tiden før og etter 2004, viser tabell 6.4 at før 2004 så er forskjellene mellom majoritetsbefolkningen og gruppe 1 ikke signifikant, med mindre en tar hensyn til uobserverte kjennetegn. Med hensyn til uobserverte kjennetegn, har gruppe 1 signifikant lavere antall dager med sykefravær enn Norge før 2004. Både gruppe 1 og 2 har signifikant lavere antall sykefraværsdager etter 2004 enn før, mens for gruppe 3 er ikke denne reduksjonen statistisk signifikant, uavhengig om en kontrollerer for uobserverte individuelle kjennetegn eller ikke. Reduksjonen for gruppe 3 er også betydelig mindre enn for de andre gruppene. Det ser derfor ut til at antall sykedager relativt til majoritetsbefolkningen har sunket for gruppe 1, som øker differansen mellom gruppene. Mens differansen mellom gruppe 2 og majoritetsbefolkningen er redusert etter 2004, altså gruppe 2 er nå kommet nærmere Norge. Det relative forholdet mellom Norge og gruppe 3 kan se ut til å være uendret etter 2004.

Korrelasjoner med forklaringsfaktorer

En ser altså forskjeller i antall sykedager før 2004 for gruppe 2 og 3 og etter 2004 for gruppe 1 og 2. For å prøve å finne ut hvilke faktorer som kan forklare disse forskjellene er det i neste steg inkludert en rekke andre forklaringsvariabler, tilsvarende prosedyren gjort ved modellene på sannsynligheten for sykefravær. En ser nå at verken før eller etter 2004 er antall sykedager for gruppe 1 signifikant forskjellig fra majoritetsbefolkningen. Ved å gjøre en stegvis inkludering av forklaringsvariabler, viser det seg at det er oppholdstid, målt i antall år, som slår ut. En mulig årsak til dette kan være at en stor andel arbeidsinnvandrere som har kommet etter 2004 vil ha lav oppholdstiden, dersom det er disse individene som også har hatt et lav antall sykefraværstilfellene, vil dette kunne slå ut i koeffisienten til oppholdstid. Det er også oppholdstid som slår ut når antall sykedager for gruppe 3 blir positiv relativt til Norge. Gruppe

2 har fortsatt en reduksjon i antall sykedager fra 2004, som er statistisk signifikant på et 1 prosentnivå, og utover perioden som ikke kan forklares av de gitte forklaringsvariablene. Når en tar hensyn til uobserverte individuelle karakteristikk er det fortsatt er statistisk signifikant reduksjon etter 2004 på 5 prosentnivå for gruppe 2. Dette tyder på at gruppe 2 har et høyere antall sykedager enn Norge, men at denne differansen reduseres fra 2004.

Ser vi på effekten av kontrollvariablene på sykefraværet så har menn, som da vi estimerte sannsynligheten for sykefravær, generelt sett et signifikant lavere antall sykedager enn kvinner. I kolonne seks og syv ser en at forskjellen er på rundt 1-2 prosent, som tilsvarer 1-2 færre dager med sykefravær for menn enn kvinner⁷⁸. Det kan tyde på at kvinner både har flest sykefraværstilfeller og flere antall sykedager enn menn i løpet av året. Men mens forskjeller er heller liten i varigheten mellom kjønn var den relativt stor når en så på tilbøyeligheten for sykefravær. Denne korrelasjonen styrkes når en kontrollerer for uobserverte individspesifikke egenskaper. Antall sykedager er positivt korrelert med alder, og er ikke avtakende slik som en så ved sannsynligheten på sykefravær. En årsak til dette kan være en negativ korrelasjon mellom helse og alder, og sannsynligheten for å få en mer alvorlig sykdom/diagnose øker mer alderen som krever lengre sykefraværperiode og eventuelt flere sykefraværstilfeller. Dette er så klart vanskelig å vite uten mer detaljert data på for eksempel diagnoser, men Markussen et al (2009) finner at for mer alvorlige sykdommer så øker tilbøyeligheten med alder. Avhengig om en tar hensyn til uobserverte individuelle kjennetegn vil en aldersforskjell på 10 år tilsvare en 3,8-5 prosentsøkning i varigheten på sykefravær, noe som tilsvarer rundt 3,5- 5 dager med fravær⁷⁹. Effekten av alder avtar når en tar hensyn til uobservert heterogenitet, slik som for eksempel helse.

Resultatet viser at å ha barn under 11år har er positivt assosiert med antall sykedager, igjen gjelder dette spesielt kvinner. Faktisk er det slik at menn med barn under 11 år er forbundet med et lavere antall fraværsdager enn menn uten, og forskjellen er betydelig på rundt 8 prosent ved MKM. Forskjellen er også statistisk signifikant på et 0,1 prosentnivå. Antall sykedager er økende med oppholdstid og korrelasjonen er statistisk signifikant på 1% nivå. Mens andregenerasjonsinnvandrere har et lavere antall sykefraværsdager enn innvandrere født utenfor Norge. Estimaten er ikke statistisk signifikant for MKM, men signifikant på et 5 prosentnivå med en tilfeldig effektmodell. En andregenerasjonsinnvandrere vil ifølge resultatene ha omtrent 2 prosent mindre sykefraværsdager enn individ født i utlandet. Reallønn har som en så tidligere

⁷⁸ Beregnet ut ifra gjennomsnittet av antall sykedager til det totale utvalget (alle grupper) i tabell 6.2.

⁷⁹ Der en ser bort ifra den kvadrerte aldersvariabelen.

(tabell 6.2) en negativ sammenheng med sykefravær, dette gjelder også antall sykedager. Når det gjelder fulltid er dette forbundet med lavere antall sykefravær, men høyere sannsynlighet. Dette kan tolkes i retning av at deler av de som er deltidsansatte har valgt denne tilpasningen (selektert) på grunn av helse eller som en fordeling av sine ressurser. Dette kan gjøre at de sjeldnere er sykemeldte, men at når de først sykemeldes trenger de lengre tid til å komme tilbake i arbeid.

Komparativ analyse av landgruppene

En så i tidligere i delkapittel «Landbakgrunn og sannsynlighet for sykefravær» at bakgrunnsfaktorene kunne variere både i styrke og fortegn mellom landgruppen, det er derfor interessant å se om dette også gjelder for antall sykedager. Som sagt vil en miste en del observasjoner når en betinger på at individene har hatt minst ett sykefraværstilfelle. Siden gruppe 1 er den minste gruppen med lavest sykefraværstilfeller, særlig etter 2004, vil gruppen være særlig påvirket av denne betingelsen. En står faktisk i fare for å utelate de fleste individene som kom etter 2004. Noe av dette kan tenkes å bli reflektert i resultatene i tabell 6.5, da en ser at få av variablene er statistisk signifikante for gruppe 1, med unntak av reallønn og oppholdstid⁸⁰.

Oppholdstid er positivt korrelert med antall sykefraværsdager, dette gjelder samtlige grupper, men korrelasjonen ser ut til å være sterkest for gruppe 1. Dette er interessant og et spørsmål er da om dette gjelder tidligere innvandrere eller den nye innvandringen som har kommet siden 2004? For gruppe 1 vil en oppholdstid på 10 år øke antall sykedager rundt 3 prosent, men dette ligger rundt 2- 2,5 prosent for de andre gruppene. For hele utvalget så en at andregenerasjonsinnvandrere tenderte til å ha et lavere antall sykedager enn utenlandsfødte, ser en på dette gruppevis finner en tilsvarende, men ikke signifikant sammenheng for gruppe 2. For gruppe 1 og 3 så er forskjellen positiv, men forskjellen er ikke statistisk signifikant for noen av gruppene. Dette kan bety at det innad i gruppene ikke er forskjell på antall sykedager når det gjelder om individet er andregenerasjonsinnvandrer eller ikke.

Tilsvarende funn for sannsynligheten for sykefravær er motsatt for gruppe 1 og 3, mens gruppe 2 hadde en positiv effekt på sykefraværet ved å være andregenerasjonsinnvandrere, der alle korrelasjonene er statistisk signifikant på et 0,1-1 prosentnivå. Av gruppene er det igjen gruppe 3 som ligner mest på Norge.

⁸⁰ En ser også at resultatene for MKM og RE er identiske (dette er ikke gjort en feil ved å ta MKM to ganger).

Tabell 6.5 Modeller på sykefravær (ln(*antall sykefravær*)), for de fire landgruppene

	Norge											
	Gruppe 1 (EU-uvide land)				Gruppe 2 (Ikke-vestlige innvandrere)				Gruppe 3 (Vestlige innvandrere)			
	MKM	RE	MKM	RE	MKM	RE	MKM	RE	MKM	RE	MKM	RE
Mann	-0.0197***	-0.0307***	-0.00456	-0.00456	-0.0190**	-0.0356***	-0.0114	-0.0221***				
Alder	0.00717***	0.00447***	0.00317	0.00317	0.0173***	0.0174***	0.00513***	0.00394*				
Alder ²	0.0000156	0.0000467***	0.0000482	0.0000482	-0.000118***	-0.000108***	0.0000284	0.0000459*				
Barn <11år	0.0104**	0.0122***	-0.0260	-0.0260	0.0394***	0.0389***	0.0150**	0.0176**				
Mann*Barn<11år	-0.133***	-0.130***	-0.0623*	-0.0623*	-0.0625***	-0.0557***	-0.0970***	-0.0912***				
Utdanning,år	-0.0118***	-0.0131***	-	-	-0.00577***	-0.00674***	-0.00704***	-0.00775***				
Oppholdstid,år	-	-	0.00301***	0.00301***	0.00238***	0.00243***	0.00207***	0.00218***				
2. gen.innv	-	-	0.0107	0.0107	-0.00700	-0.0101	0.0129	0.0122				
Reallønn (10 000kr)	-0.00389***	-0.00330***	-0.00523***	-0.00523***	-0.00559***	-0.00388***	-0.00465***	-0.00415***				
Fulltid	-0.0292***	-0.0251***	-0.0141	-0.0141	-0.0141**	-0.0121*	-0.0189***	-0.0142*				
D04	0.0191***	0.0258***	0.0196	0.0196	-0.00925	0.00475	0.0190***	0.0240***				
Konstant	4.090***	4.106***	3.969***	3.969***	3.840***	3.781***	4.015***	4.012***				
Observasjoner	846078	846078	26499	26499	204243	204243	224764	224764				
Adjusted R ²	0.0168	0.0143	0.0143	0.0143	0.0111	0.0111	0.0148	0.0148				

Kilde: SSB, norsk registerdata
* p < 0.05, ** p < 0.01, *** p < 0.001

Norge skiller seg spesielt ut på to ting, korrelasjonen mellom det å være fulltidsansatt og sykefravær og i hvilke grad det å ha barn under 11 år påvirker sykefraværet til menn. Men det er for gruppe 2 at antall sykedager øker mest for kvinner med barn under 11, rundt 6 prosent. En hypotese her kan være at omsorgsrollen til kvinner kanskje står enda sterkere for denne gruppen relativt til andre grupper.

Norge har også den sterkeste negative korrelasjonen mellom det å arbeide fulltid med antall sykedager, en har tidligere sett at det er Norge og gruppe 3 som har den høyest andelen som arbeider fulltid som kan forsterke denne korrelasjonen. Men for alle gruppene gjelder det at det å arbeide fulltid forbindes med et lavere antall sykedager enn om en arbeider deltid. Dette er i samsvar med resultatet for det samlede utvalget i tabell 6.4 Når det gjelder skille mellom før og etter 2004, er det kun gruppe 2 som har redusert antall sykedager fra 2004, men sammenhengen er ikke statistisk signifikant.

6.4 Kritisk vurdering

Resultatene som er rapportert i kapittel 6.2 og 6.3 vil bli påvirket av modellspesifikasjonen, altså hvilke modell en har valgt når en estimerer. Det er primært brukt lineære regresjonsmodeller, herunder tilfeldig effektmodellen og MKM. Som en så i kapittel 3 om økonometrisk metoder, er det en rekke antagelser som skal være oppfylt for at estimatene skal være forventningsrette og/eller konsistente. Den kanskje mest sentrale antagelsen er at det ikke skal være korrelasjon mellom feilleddet og noen av forklaringsvariablene. Dette vil i så fall forårsake endogenitet. Når det gjelder endogenitet, er det i denne oppgaven hovedsakelig to årsaker til dette: *utelatte variabler* og *to-veis kausalitet*. Helse er en opplagt *utelatt variabel* i denne sammenheng, men en kan også tenke seg andre variabler som ikke er inkludert som er av relevans som for eksempel insentiv eller motivasjon til å arbeide. Dette er spesielt sentralt i tilfeldig effektmodellene, da en plasserer den individspesifikke komponenten i feilleddet, som blant annet vil kunne inneholde individets helse. Dersom problemet utelukkende ligger i den individspesifikke komponenten i feilleddet, er fast effektmodellen en mulig løsning. Som en så i metodekapittelet er det mulig å utføre en test, Hausman-test, for å se om feilleddet er korrelert med forklaringsvariablene, som gjør fast effektmodellen mer passende enn en tilfeldig effektmodell⁸¹.

⁸¹ Resultatene fra Hausman testene, kjørt på regresjonene for hver enkelt gruppe, viser at nullhypotesen forkastet. Det er derfor sannsynlig at en eller flere av forklaringsvariablene er korrelert med feilleddet, og en bør bruke fast effektmodellen.

Ulempen med fast effektmodellen er at en ikke kan estimere tidskonstante variabler, noe som har vært helt sentralt i denne oppgaven. En eventuell utvei er å se etter alternativer som kombinerer muligheten til å ta høyde for den individspesifikke komponenten samtidig som en kan estimere effekten av tidskonstante variabler.

Resultatene vil også avhenge av oppgavens sorteringer og generering av sentrale variabler, som for eksempel hvilke land som tilhører hvilke landgruppe og hvor mange grupper som er konstruert. I oppgaven er det valgt å gjøre en relativt grov innledning etter land, med fire landgrupper der EU-utvidede land siden 2004 (gruppe 1) har vært av størst interesse. Det er ikke bare innvandringen fra disse landene som har økt etter 2004, Norge har også hatt en høy arbeidsinnvandring fra Sverige både før og etter 2004. Om gruppe 1 også hadde inkludert Sverige, som nå vil tilhøre gruppe 3, ville dette muligens kunne endret resultatene og det bildet oppgaven har dannet av både gruppe 1 og 3. Et annet poeng er at gruppe 2 og 3, som tilsvarer en ikke-vestlig og vestlig innvandringsdeling, er to svært store og heterogene grupper. Det å anta at så store grupper som inkluderer en rekke forskjellige land skal ha en underliggende likhet, kan i noen grad virke urealistisk. På en annen side vil det alltid være variasjon i gruppene uansett hvor fin-inndelt landgruppene er. I tillegg har også gruppe 2 og 3 hatt et mindre fokus i oppgaven. Den relativt grove inndelingen har derfor virket rimelig med tanke oppgavens formål.

En annen årsak til endogenitet vi har sett på tidligere er *to-veis-kausaltitet*, og i oppgavens analyse er det spesielt to variabler, nemlig reallønn og det å arbeide fulltid eller ikke, som dette omfatter. Vi har sett at det foreligger empiriske belegg for at sykefraværet påvirker lønnsutviklingen til individet, se for eksempel Markussen (2012) og ikke bare motsatt. Tilsvarende gjensidig kausalitet kan tenkes å gjelde fulltid og sykefravær også. Det vil si at variablene fulltid og reallønn kan ses på som et utfall, altså en avhengig variabel på tilsvarende måte som sykefravær. Som kontrollvariabler, vil de kunne skape endogenitet og bidrar ikke til å komme nærmere enn kausaltolkning. Dette blir i litteraturen karakterisert som «bad controls» (Angrist og Pischke 2009). Det vil si at det ikke utelukkende er slik at det å inkludere flere forklaringsvariabler gjør modellen bedre, altså at en kommer nærmere kausaleffekter. Ved å utføre en stegvis ekskludering av fulltid og reallønn fra den opprinnelige regresjonen som inkludere begge, finner en få og relativt små endringer. De endringene som er, er først og fremst resultatene på varigheten av sykefravær. Det ser derfor ikke ut til å være et stort problem, i denne oppgaven, å inkludere disse to kontrollvariablene.

I kapittel 3 så en at også *seleksjon* eller *ufullstendig panel* kan føre til endogenitet. Dette kommer som følge av ikke-registrering og frafall, særlig der dette skjer på en ikke-tilfeldig måte. Selv om problemer med frafall er heller små ved registerdata for hele populasjonen, kan det være spesielle grupper, slik som innvandrere, som er spesielt utsatt for ikke-registrering. En vet at ikke-registrering er hyppigere for innvandringspopulasjonen enn majoritetsbefolkningen. Dette vil spesielt gjelde gruppe 1 i oppgaven da en stor andel av dem er ny arbeidsinnvandring, i tillegg er denne gruppen mer preget av kortere oppholdstid i Norge enn andre innvandringsgrupper. Denne ikke-registreringen har vært tydelig i oppgavens datasett, da flere av mulige forklaringsvariabler har vært sterkt preget av ikke-registrerte observasjoner for gruppe 1. Herunder er det spesielt to variabler, utdanning og næringstilhørighet, som jeg har hatt intensjon om å ha med i analysen. Når det er nødvendig å *utelate relevante variabler* vil dette kunne føre til endogenitet på grunn av at en har utelatt en viktig variabel som kan være korrelert med de andre forklaringsvariablene.

Problemet med utelatte variabler vil også gjelde mangel på andre sosioøkonomiske kontrollvariabler som hadde vært relevant å kontrollere for. Som med utdanning og næring, vil dette kunne føre til endogenitet dersom disse er korrelert med en eller flere av forklaringsvariablene.

Vi kan ikke utelukke at endogenitet enten som følge av utelatt variabel eller to-veis-kausaltitet er noe som er av betydning i denne oppgaven. Dette vil i så fall føre til at estimeringsresultatene er forventningskjevne og/eller inkonsistente. Størrelse og retning på en slik skjevhet vil i en enkel regresjon avhenge av forholdet mellom avhengig variabel og uavhengig variabel (beta koeffisienten) og korrelasjonen mellom forklaringsvariabel og den utelatte variabelen. Men ved multippel regresjon vil korrelasjonen mellom forklaringsvariablene og korrelasjonen med flere enn en utelatt variabel, for eksempel både individets helse og incentiv eller motivasjon til arbeide også spille inn. Det er derfor vanskelig og predikere noe retning og størrelse på disse skjevhetene som vil finnes i regresjonsresultatene. Av den grunn kan en ikke tolke resultatene som kausal effekter, og resultatene må tolkes i lys av dette.

7.0 Oppsummering og avsluttende kommentarer

Formålet med denne oppgaven har vært å undersøke hvordan økningen i arbeidsinnvandringen etter EU-utvidelsen i 2004 påvirket sykefraværet. Det har vært av særlig interesse å se om dette kan ha bidratt til den reduksjonen en så i sykefraværet fra 2004. På grunn av databegrensninger er analysen basert på sykefraværet utover 16 dager, altså legemeldt sykefravær som er finansiert av NAV. Funnene sammenlignes med de relativt få studiene av sykefravær til innvandrere i Norge der det er rimelig og interessant å sammenligne.

Den empiriske strategien har vært å analysere både sykefravær i form av sykefraværstilfeller og varigheten av sykefravær, ved primært å bruke lineære regresjonsmodeller. To sentrale steg har vært å generere fire landgrupper, med én gruppe som inneholder alle land EU har blitt utvidet med siden 2004. Neste steg var å lage en indikatorvariabel som gjorde det mulig å skille mellom sykefravær før og etter 2004, som er tidspunktet da EU ble utvidet. I tillegg er det kjørt separate regresjoner for hver gruppe for å kunne si noe om hvilke bakgrunnsfaktorer som kan forklare gruppenes sykefravær.

Den deskriptive analysen gav grunn til å tro at den nye arbeidsinnvandringen kan ha bidratt til den reduksjon i sykefraværet en så etter 2004. Blant annet viste utviklingen i sykefraværet at gruppen med arbeidsinnvandrere har hatt den største reduksjonen i sykefraværet. Denne observasjonen synes å være en svært interessant og er sett nærmere på i den empiriske analysen. Både sannsynligheten for og varigheten av sykefravær er analysert. Resultatene viste at tilbøyeligheten for sykefraværet er redusert for alle grupper etter 2004, og dette gjaldt spesielt gruppen som innholdt nye EU land. Etter 2004 har denne gruppen betydelig lavere sannsynlighet for å påbegynne et sykefravær enn majoritetsbefolkningen, noe som har gitt styrket tro på oppgavens hypotese. For de to andre gruppene har reduksjonen i tilbøyeligheten etter 2004 gjort at gruppen med ikke-vestlige innvandrere fortsatt har høyere sykefravær enn majoritetsbefolkningen, men differansen er redusert. Mens gruppe 3 har en enda lavere sannsynlighet for sykefravær i forhold til referansegruppen Norge enn tidligere. Som ventet ble forskjellen mellom majoritetsbefolkningen og innvandrergruppene mindre når det ble kontrollert for en rekke bakgrunnsfaktorer (bortsett fra gruppen av ikke-vestlige). Lignende resultat fant en også for varigheten av sykefravær.

I analysekapittelet så en også på begrensninger ved resultatene som ble presentert, spesielt viktig var endogenitetsproblem på grunn av utelatte variabler og to-veis-kausaltet. Sammenhengene som er dokumentert er derfor ikke å oppfatte som kausale, noe som må taes

hensyn til under tolkningen. En annen utfordring har vært manglende observasjoner, særlig blant gruppen med de nye EU-landene. Videre er analyseperioden 1998-2008, som kun inkluderer 4 år etter EU-utvidelsen, slik at resultatene i liten grad kan si noe om en langsiktig tilpasning.

For å kunne belyse en langsiktig tilpasning av sykefraværet for de nye arbeidsinnvandrerne, trengs det data som dekker en lengre periode. Herunder ville det vært interessant å se hvordan sykefraværet utviklet seg blant EU-utvidede land i løpet av og etter finanskrisen, spesielt med tanke på at de i større grad arbeider i konjunkturutsatte bransjer, for eksempel bygg og anlegg. Dessverre har det ikke vært mulig å kontrollere for næringstilhørighet, på grunn av for dårlig kvalitet (manglende observasjoner) for denne variabelen.

Som nevnt i den kritiske vurderingen i kapittel 6, vil resultatene også avhenge av hvilke landgruppeinndeling som er gjort. Her nevnte jeg spesielt Sverige, som kunne inkluderes i gruppen med arbeidsinnvandrere, da Sverige og Polen er de to største avsenderlandene av arbeidsinnvandrere til Norge. En annen inndeling som hadde vært interessant, og som er gjort av Bratsberg et al (2014b), tar utgangspunkt i sammensetningen av ulike innvandringskohorter, som for eksempel arbeidsinnvandringen på 1970 tallet, den etterfulgte familieinnvandringen og arbeidsinnvandringen fra nye EU-land etter 2004. Dette krever imidlertid data på innvandringsgrunn, som ikke har vært tilgjengelig for denne oppgaven

Et sentralt forskningsspørsmål videre vil være hvordan den langsiktige tilpasning av sykefravær vil bli blant disse nye arbeidsinnvandrerne. Vil de fortsette i liten grad benytte seg av sykelønnsordningen, eller vil de tilpasse seg på lik linje med tidligere arbeidsinnvandrere, hvor tendensen har vært et økende trygdeforbruk over tid? Svaret på dette spørsmålet vil sannsynligvis ha stor betydning for oppslutningen om og utformingen av den framtidige innvandringspolitikken.

8.0 Litteraturliste

- Aagestad, C., Johannessen, H.A., Tynes, T., Gravseth, H.M. og Sterud, T. (2014). Work-related psychosocial risk factors for long-term sick leave: a prospective study of the general working population in Norway. *Journal of Occupational and Environmental Medicine*, Vol. 56 (8), 787–793.
- Alexanderson, K., Allebeck, P., Hansson, T., Hensing, G., Jensen, I., Mastekaasa, A., Norlund, A., Perk, J., Syversson, A., Wahlström, R. and Vingård, E. (2003) *Sjukskrivning – orsaker, konsekvenser och praxis. En systematisk litteraturöversikt*. Stockholm: Statens beredning för medicinsk utvärdering. SBU.
- Allen, S.G. (1981). An empirical model of work attendance. *The Review of Economics and Statistics*, Vol. 63 (1), 77-87.
- Angrist, J.D. og Pischke, J.S. (2009). *Mostly harmless econometrics. An empiricist's companion*. Princeton, Princeton University press.
- Arbeids- og inkluderingsdepartementet (2008). Arbeidsinnvandring. St.meld.nr. 18 (2007-2008). Oslo, Arbeids- og inkluderingsdepartementet.
- Arbeids- og sosialdepartementet (8. oktober 2014). *Folketrygdens utgifter* [internett], Arbeids- og sosialdepartementet. Tilgjengelig fra: <https://www.regjeringen.no/nb/aktuelt/Folketrygdens-utgifter/id2005566/> [Lest 10.03.2015].
- Askildsen J.E., Bratberg, E. og Nilsen, Ø. A. (2005). Unemployment, labor force composition and sickness absence: a panel data study. *Health Economics*, Vol. 14 (11), 1087-1101.
- Bakken, F., Bråthen, M., Handal, J. og Aarehun, T. (2009). Hvordan vil en nedgangskonjunktur påvirke arbeidsinnvandrere i Norge? NAV – rapport 01/2009.
- Baltagi, B.H. (2008). *Econometric Analysis of Panel Data*. 4.utg. UK Chichester, Wiley.
- Bengtsson, T. og Scott, K. (2006). Immigrant consumption of sickness benefits in Sweden, 1982-1991. *The Journal of Socio-Economics*, vol. 35 (3), 440-457.
- Berge, C. (2008). *Data for analyse av sykefravær*. Oslo, Statistisk sentralbyrå, notat.
- Berge, C. Johannessen, J.H. og Næsheim, H. (2012). *Internasjonal sammenligning av sykefravær. Er Arbeidskraftsundersøkelsen egnet som datakilde?* Oslo, Statistisk sentralbyrå, rapporter 6/2012.
- Bjørnstad, R. (2006). Er det økte sykefraværet tegn på et mer inkluderende eller ekskluderende Arbeidsliv? *Økonomiske analyser*, 6/2006.
- Bockmann, G. og Hagelund, A. (2010). *Velferdens grenser: innvandringspolitikk og velferdsstat i Skandinavia 1945-2010*. Oslo, Universitetsforlaget.

- Borjas, G.J. (1989). Economic theory and international migration. *International Migration Review*, Vol. 23, No. 3, Special Silver Anniversary Issue: International Migration an Assessment for the 90's (Autumn, 1989), 457-485.
- Bratberg, E. og Risa, A.E. (2000). *Insentivvirkninger i helserelaterte stønadsordninger*. Rapport skrevet på oppdrag av Sosial- og Helsedepartementet. Institutt for økonomi, Universitetet i Bergen.
- Bratberg, E. og Vaage, K. (2015). Norge er fortsatt på trygdetoppen. Aftenposten, 18. mars 2015 (internett). Tilgjengelig fra: <http://www.aftenposten.no/viten/Norge-er-fortsatt-pa-trygdetoppen-7948352.html> [Lest 08.05.15].
- Bratberg, E., Holmås, T.H. og Thøgersen, Ø. (2004). Assessing the effects of early retirement programs. *Journal of Population Economics*, Vol. 17 (3), 387-408.
- Bratsberg, B. og Raaum, O. (2013). Migrasjonsstrømmenes påvirkning på lønns- og arbeidsvilkår. *Samfunnsøkonomen*, nr.3/2013.
- Bratsberg, B., Raaum, O. og Røed, K. (2014a). Arbeidsinnvandring – Varig gevinst? *Søkelys på arbeidslivet*, årgang 31 Nr. 4 2014.
- Bratsberg, B., Raaum, O. og Røed, K. (2014b). Immigrants, Labour Market Performance and Social Insurance. *The Economic Journal*, Vol. 124 (580).
- Brekke, I. og Schøne P. (2013). Kan forskjeller i sykefravær mellom innvandrerbefolkningen og majoritetsbefolkningen forklares med forskjeller i selvrapportert helse. *Søkelys på arbeidslivet*, årgang 30 Nr. 1-2 2013.
- Bø, T.P. og Håland, I. (2002). *Dokumentasjon av arbeidskraftundersøkelsen (AKU)*. Oslo, Statistisk sentralbyrå, Notater 2002/24.
- Carlsen, B. og Nyberg, K. (2009). The Gate is Open: Primary Care Physicians as Social Security Gatekeepers. *Memorandum*, 07/2009. Universitetet i Oslo, Økonomisk institutt.
- Dahl, S., Hansen, H., og Olsen, K. M. (2010). Sickness Absence among Immigrants in Norway, 1992-2003. *Acta Sociologica*, Vol. 53 (1), 35-52.
- Doksheim, M. (2010). *Innvandrerne i arbeidsmarkedet*. Civita- notat nr. 15/2010.
- Ekspertgrupperapport (2010). *Tiltak for reduksjon i sykefravær: Aktivisering- og nærværsreform*. Levert til Arbeidsdepartementet 01.02.10.
- Folketrygdløven (1997). *Lov om folketrygd av 28. februar 1997 nr. 19*.
- Friberg, J.H., Dølvik, J.E. og Eldring, L. (2013). *Arbeidsmigrasjon til Norge fra Øst- og Sentral Europa*. Temanotat. Program Velferd, arbeidsliv og migrasjon – VAM. Oslo, Norges forskningsråd.

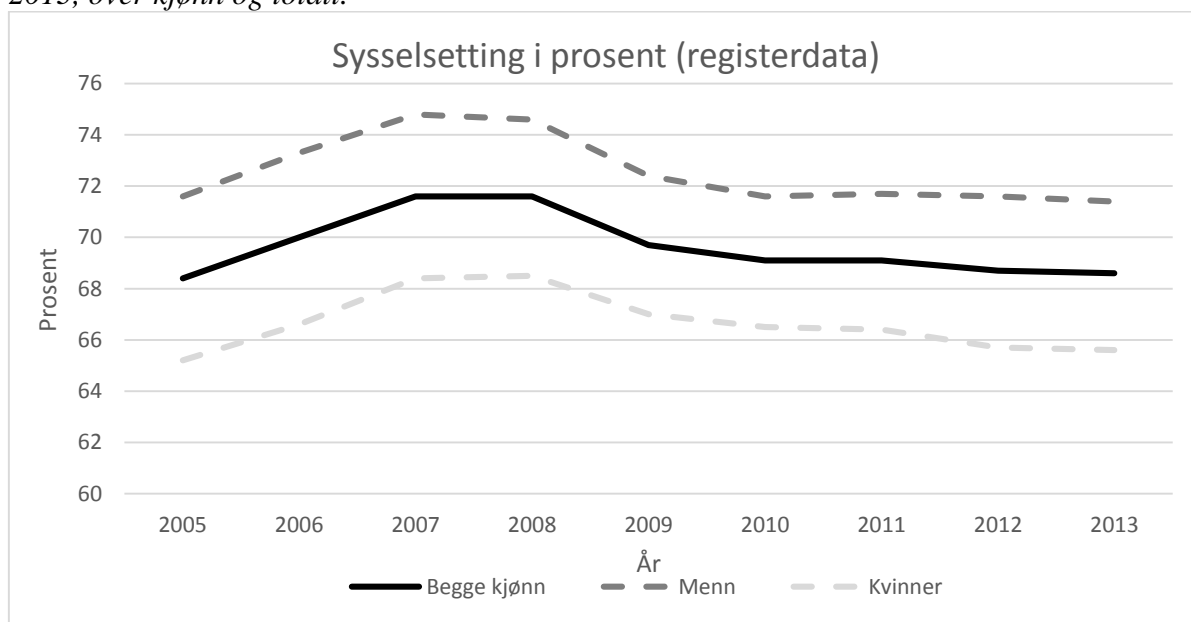
- Friberg, J.H., Tronstad, K.R. og Dølvik J.E. (2012). Central and Eastern European labour migration to Norway, trend, conditions and challenges 2012. I OECD (2012). *Free Movement of Workers and Labour Market Adjustment: Recent Experiences from OECD Countries and the European Union*, OECD Publishing.
- Gleinsvik, A. Klingenberg, S. og Mastekaasa, A. (2014). *Internasjonal sammenligning av sykefravær*. Hovedrapport, 2014-05 PROBA samfunnsanalyse.
- Hansen, H., Holmås, T.H., Islam, M.K. og Naz, G. (2014). Sickness Absence Among Immigrants in Norway: Does Occupational Disparity Matter? *European Sociological review*, Vol. 30 (1), 1-12.
- Hansen, O. E., Hetland, A. og Sivertstøl, Ø. (2013). *FD-Trygd: Dokumentasjonsrapport*. Oslo, Statistisk sentralbyrå, Notater 2013/44.
- Hauge, L. og Kann, I.C. (2007). Hva betyr alders- og kjønns sammensetningen for sykefraværet? *Arbeid og velferd*, 1/2007, NAV.
- Hesselius, P., Nilsson, J.P. og Johansson, P. (2009). Sick of you colleagues` absence? *Journal of the European Economic Association*, Vol. 7 (2-3), 583-594.
- Horgen, E.H. (2014). *Innvandrere med svak tilknytning til arbeidsmarkedet – hvem er de?* Oslo, Statistisk sentralbyrå, rapporter 22/2014.
- Hsiao, C. (2003). *Analysis of Panel Data*, 2nd edition. UK Cambridge, Cambridge University Press.
- Høydahl, E. (2008). Vestlig og ikke-vestlig- ord som ble for store og gitt ut på data. *Samfunnsspeilet* 4/2008. Oslo, Statistisk sentralbyrå.
- Ichino, A. og Moretti, E. (2009). Biological Gender Differences, Absenteeism, and the Earnings Gap. *American Economic Journal: Applied Economics*, Vol. 1(1), 183-218.
- Ingstad, K. og Kvande, E. (2011). Arbeid i sykehjem – for belastende for heltid? *Søkelys på arbeidslivet*, årgang 28 Nr. 1-2 2011.
- Kann, I.C., Bragstad, T. og Thune, O. (2013). Stadig flere på trygd. *Arbeid og velferd*, 3/2013, NAV.
- Maddala, G.S (1987). Limited Dependent Variable Models Using Panel Data. *The Journal of Human Resources*, 22 (3), 307-338.
- Markussen, S. (2009a). Closing the gates? – Evidence from a natural experiment on physicians` sickness certification. *Memorandum*, 19/2009. Universitetet i Oslo, Økonomisk institutt.
- Markussen, S. (2009b). How physicians can reduce sick leave – evidence from a natural experiment. Ragnar Frisch senter for økonomisk forskning, 2009.
- Markussen, S. (2010). 2004: Da sykefraværet falt som en stein. *Samfunnsøkonomen*, nr.3/2010.

- Markussen, S. (2012). The individual cost of sick leave. *Journal of Population Economics*, Vol. 25 (4), 1287-1306.
- Markussen, S., Røed, K., Røgeberg, O.J. og Gaure, S. (2009). *The anatomy of Absenteeism*. The Institute for the Study of Labor (IZA) Discussion Paper No. 4240, June 2009.
- Moland, L.E. (2007). *Deltidsarbeid og sykefravær i Oslo kommune*. Fafo- notat 2007:01.
- Mykletun, A. og Vaage, K. (2012). *Rapport etter ekspertmøte om årsaker til kjønnsforskjeller i sykefravær*. Rapport overlevert 29. mars 2012 til Arbeidsdepartementet.
- Nossen, J. P. (2009). Sykefravær og konjunkturer – en oversikt. *Norsk Epidemiologi*, Vol. 19 (2), 115-125.
- OECD (2014). *OECD Economic Surveys: Norway 2014*. OECD Publishing.
- OECD (2015). Employment rate (indicator). [Lest: 05.01.2015].
- Olsen, B. (1997). *Prøveundersøkelse om 1-3 dagers sykefravær i sentral sykefraværstatistikk*. Oslo, Statistisk sentralbyrå, Notater 97/66.
- Olsen, B. (2013). Innvandringen økte sysselsettingen. *Samfunnsspeilet* 5/2013. Oslo, Statistisk sentralbyrå.
- Piha, K., Laaksonen, M., Martikainen, P., Rahkonen, O. og Lahelma, E. (2010). Interrelationships between education, occupational class, income and sickness absence. *The European Journal of Public Health*, Vol. 20 (3), 276–280.
- Rønningen, D. (2005). *Sysselsetting og tidligavgang for eldre arbeidstakere: En deskriptiv analyse for perioden 1992-1999*. Oslo, Statistisk sentralbyrå.
- Shapiro, C. og Stiglitz, J.E. (1984). Equilibrium Unemployment as a Worker Discipline Device. *The American Economic Review*, Vol. 74 (3), 433-444.
- Smeby, L., Bruusgaard, D. og Claussen, B. (2009). Sickness absence: Could gender divide be explained by occupation, income, mental distress and health? *Scandinavian Journal of Public Health*, Vol. 37 (7), 674-681.
- Sosialdepartementet (1977). Om lov om endring i lov av 17.juni 1966 nr.12 om folketrygden. Ot.prp. nr. 68 (1967-1977). Oslo, Sosialdepartementet.
- Sosialdepartementet (2004a). *Arbeidsskadeforsikring*. NOU 2004:3. Oslo, Statens forvaltningstjeneste.
- Sosialdepartementet (2004b). *Om lov om endringer i folketrygloven*. Ot.prp. nr. 48 (2003-2004). Oslo, Sosialdepartementet.
- Statistisk sentralbyrå. *Om Forløpsdatabasen trygd (FD-Trygd)* [internett], Statistisk sentralbyrå. Tilgjengelig fra: <http://www.ssb.no/a/mikrodata/datasamlinger/fdtrygd.html> [Lest 3.10.2014].

- Terum, L.I. og Hatland, A. (2014). Sysselsetting og trygd under arbeidslinja. *Søkelys på arbeidslivet*, årgang 31 Nr.1-2 2014.
- Texmon, I. og Brunborg, H. (2013). *Vekst og sammensetning av befolkningen ved ulike forutsetninger om innvandring*. Oslo, Statistisk sentralbyrå, rapporter 23/2013.
- Tønnessen, M., Cappelen, Å. og Skjerpen, T. (2014). Befolkningsframskrivingene 2014-2100: Inn- og utvandring. *Økonomiske analyser*, 4/2014.
- Utenriksdepartementet (2012). *Utenfor og innenfor: Norges avtaler med EU*. NOU 2012:2. Oslo, Departementenes servicesenter.
- Vassenden, K. (2012). Hvor stor er egentlig innvandringen til Norge – nå, før og internasjonalt? *Samfunnsspeilet* 3/2012. Oslo, Statistisk sentralbyrå.
- Verbeek, M. (2012). *A Guide to Modern Econometrics*. 4.utg. UK Chichester, Wiley.
- Villund, O. (2012). *Arbeidsmarkedsstatistikk for innvandrere, basert på arbeidskraftundersøkelsen (AKU): utvikling av et nytt beregningsopplegg*. Oslo, Statistisk sentralbyrå.

9.0 Vedlegg

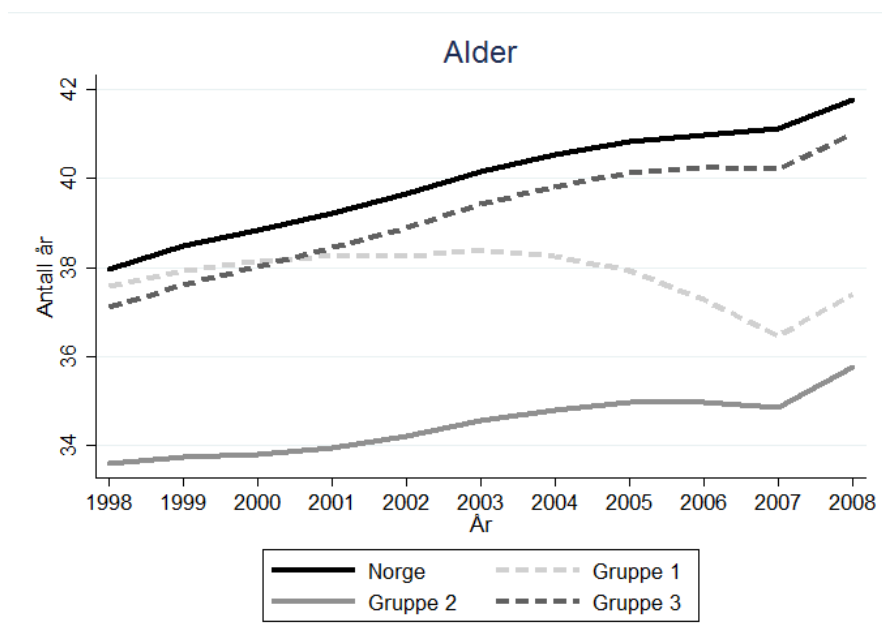
Figur A1: Sysselsetting i prosent for befolkningen i alder 15-74 år for perioden 2005 til 2013, over kjønn og totalt.

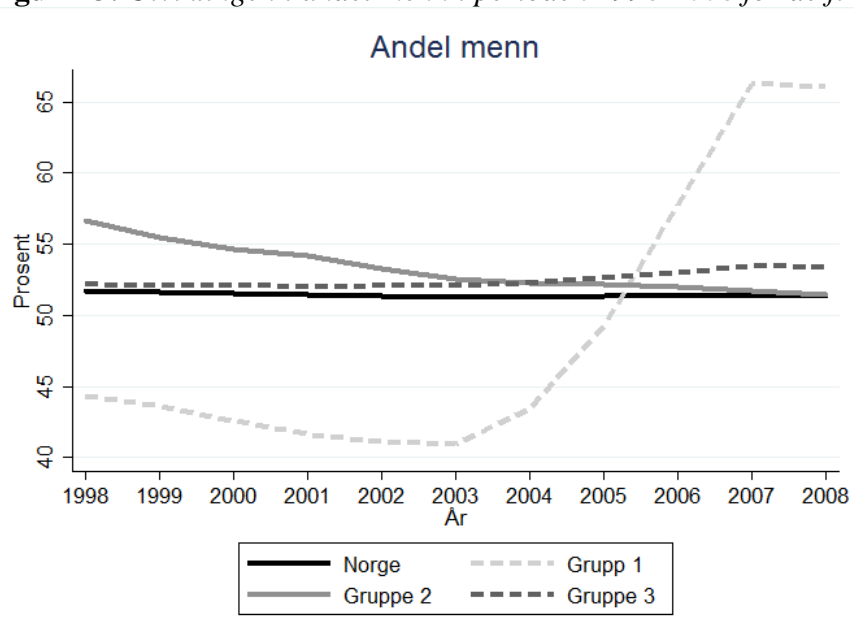
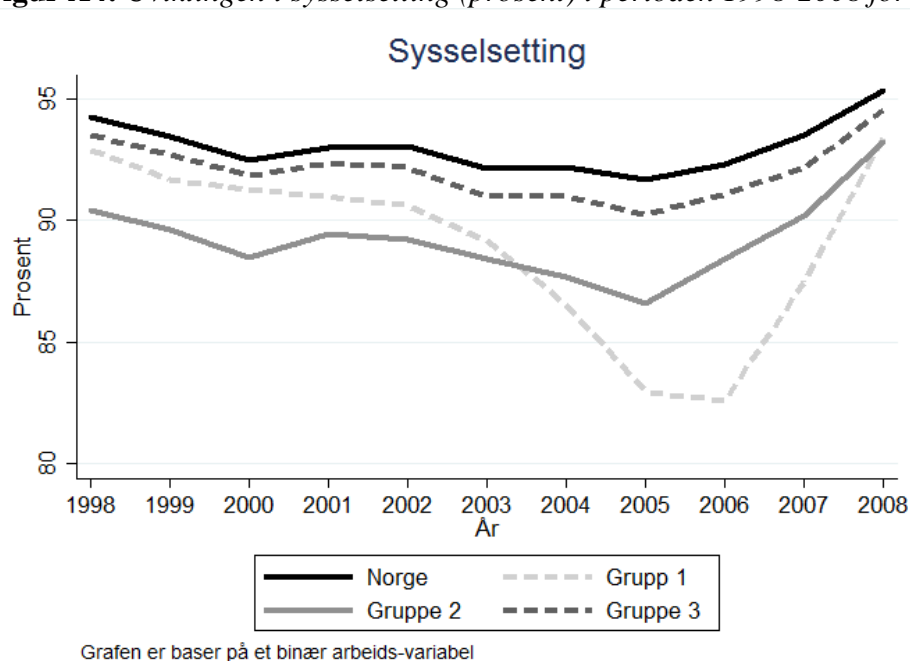


Kilde: SSB

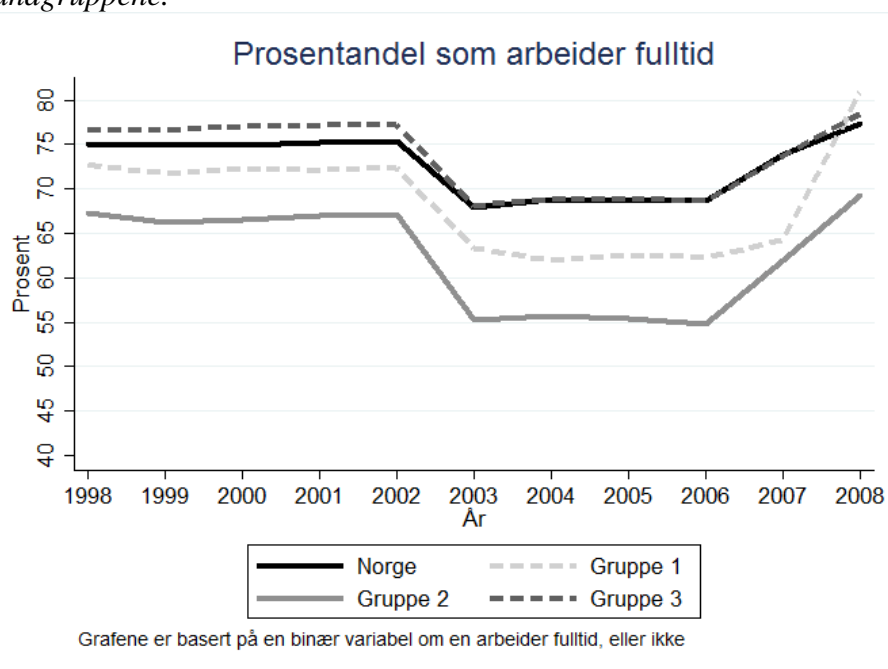
Etter utsortering av individ med kun ett observasjonsår:

Figur A2: Aldersutvikling i perioden 1998-2008 for fire landgrupper, etter sortering ned til 16-67 år.

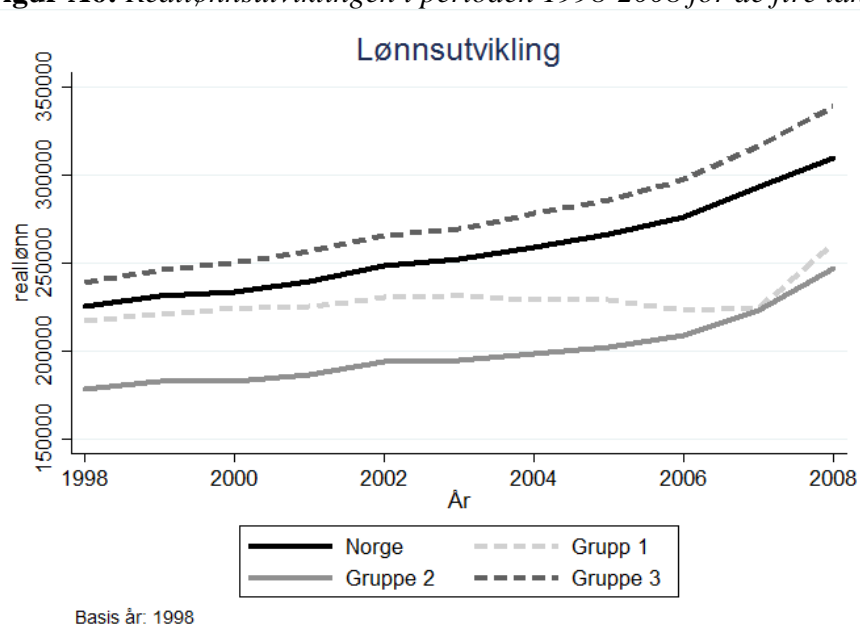


Figur A3: Utviklingen i andel menn i perioden 1998-2008 for de fire landgrupper.**Figur A4:** Utviklingen i sysselsetting (prosent) i perioden 1998-2008 for de fire landgruppene.

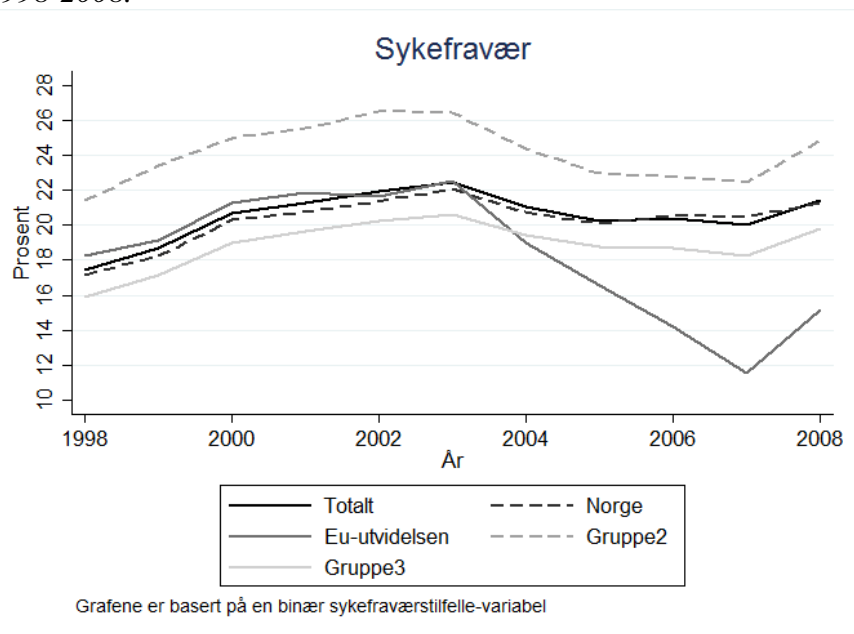
Figur A5: Utviklingen i andelen som jobber fulltid i perioden 1998-2008 for de fire landgruppene.



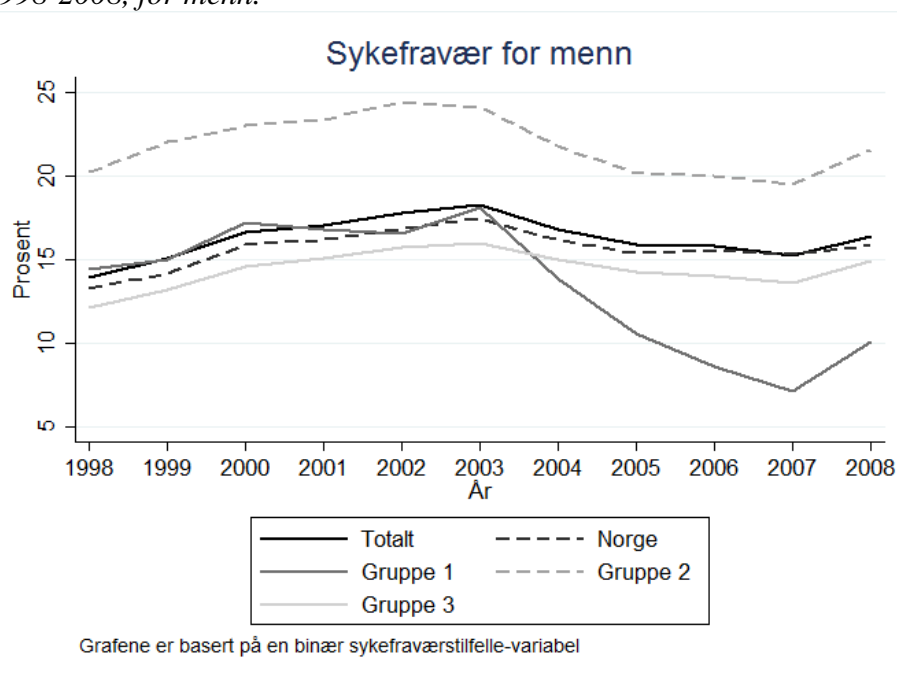
Figur A6: Reallønnsutviklingen i perioden 1998-2008 for de fire landgruppene



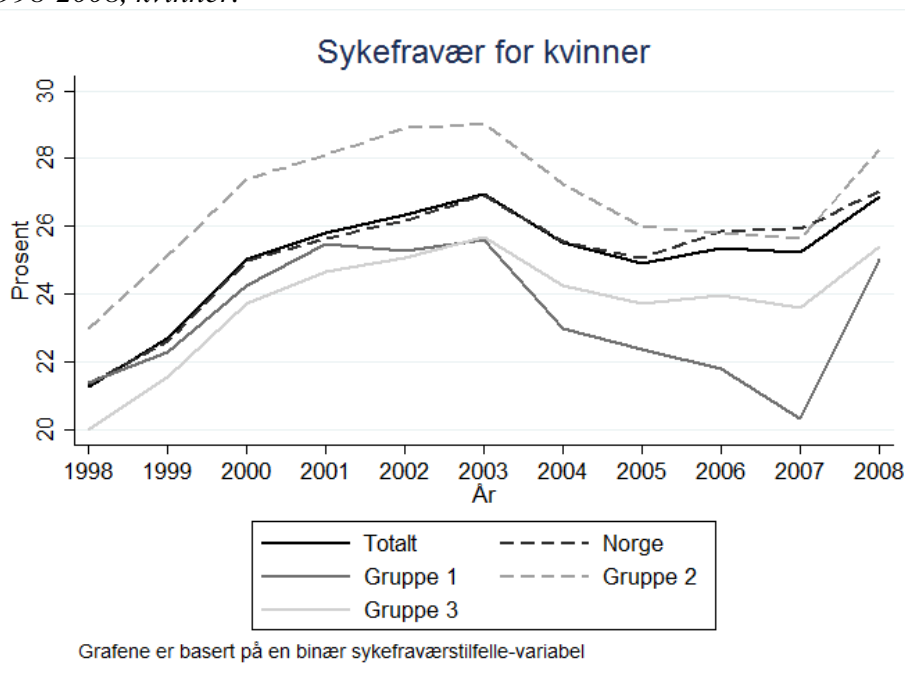
Figur A7: Utviklingen i prosenandelen med ett eller flere sykefraværstilfeller, for perioden 1998-2008.



Figur A8: Utviklingen i prosenandelen med ett eller flere sykefraværstilfeller i peroiden 1998-2008, for menn.



Figur A9: Utviklingen i prosenandelen med et eller flere sykefraværstilfeller for perioden 1998-2008, kvinner.



Variabelforklaringer:

Gruppe 1 (G1): EU-utvidede land fra 2004 og 2007.

Gruppe 2 (G2): Asia + Tyrkia, Afrika, Sør- og Sentral-Amerika, Oseania og Europeiskeland utenfor EU.

Gruppe 3 (G3): Nord-Amerika, Australia, New-Zealand og land i EU.

Mann binærvariabel for om en er mann eller kvinne, **Alder** er individet alder, **Alder²** er den kvadrerte alderen, **Mann*Alder** interaksjon mellom Mann og Alder, **Barn<11 år** er en binær variabel for om individet har barn som er under 11 år gammel, **Mann*Barn<11år** er interaksjonsvariabel mellom Mann og Barn<11år, **Utdanning** forteller hvor mange år med utdanning individet har som er målt i høyeste grad, **Oppholdstid** er hvor mange år individet har oppholdt seg i Norge, **2.gen.innv** er en binærvariabel for om en er andre generasjonsinnvandrere eller ikke, **Reallønn** er reallønn per 10 000 kr og **Fulltid** som er en binærvariabel som reflekter om en arbeider fulltid eller deltid. **D04** er en step-dummy som tar verdien 1 for år>2003 og 0 for år ≤2003. **D04*G1**, **D04*G2**, **D04*G3** er interaksjonsvariabler mellom step-dummyen og de tre gruppene gruppe 1, gruppe 2 og gruppe 3 henholdsvis.