



Erling Ravnanger Landet

Likestillingsparadokset og den økologiske feilslutning

En flernivåanalytisk tilnærming til å forklare horisontal kjønnsdeling i 49 arbeidsmarked i verden

Masteroppgave i Sosiologi

Trondheim, juni 2016

Erling Ravnanger Landet

Likestillingsparadokset og den økologiske feilslutning

En flernivåanalytisk tilnærming til å forklare
horisontal kjønnsdeling i 49 arbeidsmarked i verden

Masteroppgave i Sosiologi
Trondheim, juni 2016

Norges teknisk-naturvitenskapelige universitet

Abstract

In this thesis, I argue for a need for an individual-level segregation measure in the social sciences, and propose a solution. The measurement is used as a dependent variable in a multi-level analysis of 49 countries across the world to find country-level and individual-level factors that influence horizontal occupational gender segregation. The data used in the analysis is both individual-level demographics collected from 15 merged and harmonized rounds of the International Social Survey Programme, and country-level contextual data collected from various sources. Compositional effects with regards to working sector, work time and having children, only accounts for 2,8% of the variance between countries. Using rational-choice theory, I conclude that tolerance for parental leave is the primary contextual factor influencing both the female proportion in the workforce and horizontal occupational gender segregation. While the dependent variable appears useful, it does suffer from problems due to the hierarchical nature of the International Standard Classification of Occupations, which it is based on.

Forord

Ingen ting er som å regelmessig våkne opp svett av angst i et halvt år med dårlig samvittighet og mangel på mestringsfølelse, etterfulgt av å banke hodet i veggen på lesesalen, i håp om at noe fornuftig ramler ut på tastaturet.

Samtidig har ikke prosessen bare vært en øvelse i å balansere nevrotisisme med mindfulness. Enkelte mennesker har sørget for at tiden trolig har blitt et standardavvik bedre enn det ville vært.

Jeg vil takke Sindre Høgmo Johansen for å ha vært en bautastein i å motta sutring over modeller som ikke vil konvergere, og å nikke pent og le når jeg skryter av hvor smart jeg noen ganger tror syntaksen min er. Det burde vært et betalt heltidsverv når man er romkamerater. I tillegg har han vært en superhelt ved å lese korrektur.

Jeg vil takke Ingvild Firman Fjellså for å ha på finurlig vis lurt tankene mine over på andre ting enn tall. Hun har en evne til å koble av som jeg skamløst har snyltet på. Jeg vil takke henne for å ha vært tilstede når jeg har ønsket det, og av og til å ha holdt seg borte når jeg har bedt om det. Takk for støtten og korrekturlesing.

Pamper og ildsjeler på kontoret til Studentrådene HF og SVT har sørget for avbrekk med bedre diskusjoner enn vafler. Tusen takk for støtte både før og underveis i prosessen, og å ha gjort hele studieløpet mitt til noe mer enn å oppnå formelle kunnskaps- og ferdighetsmål.

Kull 2014 har alltid vært tilgjengelig i lunsjer og i gangene i bygg 10. Tusen takk for å være en inkluderende og støttende gjeng med sosiale avbrekk i hverdagen. Uansvarlig alkoholforbruk kombinert med etisk forkastelige diskusjoner har gjort masterkurset til en lek. En spesiell shoutout til Soheil for både gode diskusjoner om metodevalg og avslappende pauser på lesesalen.

Mine foreldre Rannveig og Einar Landet fortjener en spesiell takk for interessen de har vist ved å vie mye tid til korrekturlesing av både masteroppgaven og semesteroppgaver gjennom min tid ved NTNU. Selv om vi ikke blir enige om optimale formuleringer og korrekt bruk av faguttrykk, har oppgaven blitt målbart mer leservennlig som følge av deres innsats.

Til sist vil jeg takke Arild Blekesaune. Han har noe, som for meg er ubegripelige egenskaper til å være konkret, nytenkende og å forstå hvorfor jeg har gjort det jeg har gjort, før jeg har skjønt det selv. Å ha en dyktig veileder som jeg har dyp respekt for, og er på bølgelengde med, har gjort angstfylte og repetitive dager til optimistiske og produktive.

Erling Ravnanger Landet

Trondheim, 02.06.16

Innhold

1. Innledning.....	1
1.1 Aktualisering.....	1
1.2 Problemstilling.....	2
1.3 Strukturen i oppgaven.....	3
2. Teori og tidligere forskning.....	5
2.1 Kjønnssdeling og sosial endring.....	5
2.2 Teori om rasjonelle valg.....	6
2.3 Tidligere forskning.....	8
2.3.1 Ulike preferanser og økonomisk ulikhet.....	8
2.3.2 Likestillingspolitikkenes dimensjoner og konsekvenser for kjønnssdeling.....	8
2.3.3 Offentlig sektor og kvinners andel i arbeidsmarkedet.....	10
2.4 Hypoteseformulering.....	10
3. Data og metode.....	15
3.1 Datagrunnlag.....	15
3.2 Problematisering av datamateriale.....	16
3.3 Harmonisering og beskrivelse av uavhengige variabler på individnivå.....	16
3.3.1 Arbeidsrelaterte uavhengige variabler.....	19
3.3.2 Inntektsvariabler.....	21
3.4 Avhengig variabel.....	23
3.4.1 Tidligere forskning på kjønnssdeling og segregering.....	23
3.4.2 Konstruksjon av avhengig variabel.....	25
3.4.3 Validitets- og reliabilitetstesting av avhengig variabel.....	27
3.5 Sammenslåing av datasett og konstruksjon av variabler på andre nivå.....	28
3.5.1 Variabler på gruppenivå.....	30
3.6 Analyseteknikk.....	33
4. Analyse.....	37
4.1 Deskriptiv statistikk.....	37
4.2 Kontroll for datastruktur.....	38
4.3 Valg av modell i tabell 6.....	41
4.3.1 Hypotesetesting for hypoteser på individnivå.....	43
4.4 Kontroll for kvinneandelen i offentlig sektor og i arbeidsmarkedet.....	46
4.5 Kontroll for generell likestilling og økonomisk ulikhet.....	49
4.6 Kontroll for dimensjoner i likestillingspolitikk.....	52
4.7 Kontroll for relevante kontekstuelle variabler.....	54
5. Diskusjon og avslutning.....	59
5.1 Oppsummering av funn, nullfunn og forventede sammenhenger.....	59
5.2 Tolkning av resultater.....	61
5.2.1 Svakheter.....	65
5.3 Målet på kjønnssdeling.....	66
5.4 Generalisering og veien videre.....	68
6. Litteratur.....	69
7. Vedlegg.....	73

Tabelloversikt

Tabell 1: Oppsummering av hypoteser.....	13
Tabell 2: Korrelasjoner: kontekstuelle variabler og avhengig variabel på individ- og gruppenivå..	33
Tabell 3: Oversikt over kjønnsdeling i deltakerlandene.....	37
Tabell 4: Oversikt over datastrukturen.....	38
Tabell 5: Tre tomme modeller. Kontroll for datastruktur. N=286587.....	39
Tabell 6: Flernivåanalyser med kontroll for variabler på individnivå. N=286587.....	44
Tabell 7: Flernivåanalyser med kontroll for kvinneandelen i offentlig sektor og i arbeidsstyrken. N=286578.....	48
Tabell 8: Flernivåanalyser med kontroll for generell likestilling og inntektsulikhet. N=286587.....	51
Tabell 9: Flernivåanalyser med kontroll for dimensjoner i likestillingspolitikk. N=286587.....	53
Tabell 10: Flernivåanalyser med kontroll for relevante kontekstuelle variabler. N=286587.....	58

Figuroversikt

Figur 1: T.V.: Histogram over avhengig variabel på individnivå. T.H.: Histogram over avhengig variabel på utvalgsnivå.....	26
Figur 2: Samspill: Arbeidstid, sektor og kjønn.....	45
Figur 3: Spredningsplott over kvinneandelen av arbeidsstyrken og delt omsorg-dimensjonen.....	56
Figur 4: Sammenhengen mellom delt omsorgs-dimensjonen, horisontal og vertikal kjønnsdeling. .	60
Figur 5: Dimensjoner i toleransenivå for foreldrepermisjoner i sosiale institusjoner.....	62
Figur 6: Teoretisk tolkning av kjønnsdelingsprosessen.....	64

1. Innledning

Mange deltakere i spenningsfeltet mellom offentlig debatt og akademia har ytret seg om årsaken til likestillingsparadokset. Det vil si hvorfor Norge og andre land som oppfattes som likestilte, også har et horisontalt kjønnsdelt arbeidsmarked. Horisontal kjønnsdeling peker på graden menn og kvinner arbeider i ulike typer yrker med lik status på tvers i arbeidsmarkedet. Noen foreslår at fenomenet kan forklares med biologiske kjønnsforskjeller (Aakvaag, 2015; Bongard og Røskaft, 2010:154; Ihle og Eia, 2010; Pinker, 2008). Andre hevder at det er samfunnsmessige strukturer, sosialiseringsmønstre, stereotypisering, patriarkatet, selvforsterkende mekanismer, eller en kombinasjon av flere som må til for å forstå situasjonen (Blackburn, Browne, Brooks og Jarman, 2002; Reisel, 2014a; Reisel og Teigen, 2015). Debatten kan sies å ha mistet fokus på et underliggende premiss. Et premiss om at det er en sammenheng mellom likestilling og kjønnsdelte yrkesvalg, og det er utgangspunktet for mitt prosjekt.

1.1 Aktualisering

Det er hovedsaklig to måter en tidligere har arbeidet kvantitativt for å forske på hva som fører til horisontal kjønnsdeling. Det ene er å undersøke innad i enkeltland ved hjelp av enten deskriptiv statistikk eller regresjoner. Eksempelvis har en forsøkt å forstå det kjønnsdelte arbeidsmarkedet ved å forske på hva som forårsaker kjønnsdelte utdanningsvalg (Reisel, 2014b; Støren og Arnesen, 2003). Den andre måten er å bruke indekser der hele landet ett år deler verdi for å sammenligne på tvers av landegrenser (eks: Hakim, 1992; Melkas og Anker, 2001 Charles, 2003).

Alle forklaringene på horisontal kjønnsdeling har et empirisk problem. De er enten basert på observasjoner gjort på makronivå, innad i ett land, eller gjort ved bruk av deskriptiv statistikk. Alle kjente mål jeg har funnet som benyttes til å sammenligne likestillingen og kjønnsdelingen i arbeidsmarked i ulike land, er indekser der alle individene i landet deler en verdi. Verdiene indikerer hvor kjønnsdelt arbeidsmarkedet i landet er som helhet. Eksempelvis henviser verdien i *Duncans Dissimilarity Index* til hvor stor andel av arbeidstakerne som må endre yrke for at fordelingen skal bli jevn (Iceland, Weinberg og Steinmetz, 2002:119). Andre brukte teknikker er å se på deskriptiv statistikk på kjønnsfordelingen i ulike utdanningsretninger og enten sammenligne ulike land, eller se utvikling innad i land over tid (Støren og Arnesen, 2003). Målene kan være nyttige til å identifisere hvilke arbeidsmarked som er kjønnsdelte og ikke. De er dog ikke nyttige til å finne makrofaktorer som fører til kjønnsdelte yrkesvalg på individnivå.

Kvinner arbeider oftere i offentlig sektor enn menn, og har trolig som konsekvens generelt lavere lønninger (Statistisk sentralbyrå, 2016). Som følge av det kan horisontal kjønnsdeling også

føre til en kjønnsmessig ujevn vertikal deling i arbeidslivet. Videre kan ifølge NOU (2012:320) om Politikk for likestilling «forutintatte holdninger om hva som er passende kvinneyrker eller passende mannsyrker [...] påvirke utdanningsvalg» og «kjønnskjeve utdanningsvalg bidrar til kjønnskjeve yrkesvalg». Ut fra resonnementet kan det se ut som at kjønnskjevheten i seg selv øker sannsynligheten for at den kjønnskjeve fordelingen i yrkeslivet forblir. Det kalles gjerne stiafhengighet (Reisel, 2014a:39), og kan fungere som en barriere mot reell valgfrihet (NOU (2012:320)).

Godtas argumentet, og en anser reell valgfrihet som en viktig verdi i samfunnet, er det av interesse å ha data som sier noe om hva som fører til et kjønnsdelt arbeidsmarked. Det eksisterer mange undersøkelser om hva som påvirker utdannings- og karrierevalg innad i landet. Hvis en ønsker å se effekten av likestillingspolitikk, er det imidlertid ikke bra nok. Selv om en tar med tidsdimensjonen i deskriptiv statistikk, er det vanskelig å argumentere for en kausalsammenheng mellom politiske beslutninger og endringer i kjønnsdelingen, fordi en mangler sammenligningsgrunnlag. En kan ikke vite om det er politikken som har endret noe, eller om det er andre utviklingstrekk i samfunnet som har skyld i endringen. Det gjelder spesielt når det mangler en måte å måle kjønnsdeling på individnivå.

Premisset for forklaringene av horisontal kjønnsdeling i arbeidslivet lider derfor av en økologisk feilslutning. Feilslutningen er at en sammenheng på individnivå, kan ikke forklares når det empiriske grunnlaget ligger på makronivå (Robinson, 2009; Snijders og Bosker, 2012:15). I tilfellet av temaet for oppgaven, er feilslutningen at sammenhengen mellom likestilling og kjønnsdeling i arbeidslivet på makronivå, ikke sier noe om kjønnsdelingen i yrkene til arbeidstakerne på individnivå. I tilfellet av deskriptiv statistikk, har en ikke mulighet til å bruke kontrollvariabler for å finne ut hva kjønnsdelingen skyldes. Forklaringene er dog ikke nødvendigvis feil av den grunn. Poenget er at forklaringene på hva som fører til horisontal kjønnsdeling, og om at en økt grad av likestilling har en sammenheng med – og/eller fører til – horisontal kjønnsdeling, mangler empirisk støtte. Forklaringene mangler empirisk støtte fordi horisontal kjønnsdeling ikke er undersøkt med gode nok mål.

1.2 Problemstilling

I denne oppgaven har jeg derfor tatt for meg utfordringen med å finne ut *på hvilken måte et segregeringsmål på individnivå kan skapes, som er meningsfullt å tolke med kryssnasjonale surveydata?* For å svare på det, har jeg valgt å undersøke *hvilke kontekstuelle faktorer på landsnivå, og demografiske forhold på individnivå, som påvirker kjønnsdeling?* Det gjør jeg ved å benytte demografisk informasjon fra et sammensatt og harmonisert datasett fra 15 runder av International

Social Survey Programme (ISSP) sammen med kontekstuelle variabler på gruppenivå.

For å undersøke hvorvidt ulike land er kjønnsdelte, og hvordan det henger sammen med forskjellige kontekstuelle faktorer, er en kvantitativ tilnærming å foretrekke, fordi jeg ønsker å generalisere funnene til deltakerlandene. Det har medført en hypotetisk-deduktiv tilnærming der jeg skaper hypoteser basert på teori og tidligere forskning, og deretter tester hvorvidt de kan falsifiseres eller ei. Hypotesene blir testet med flernivåanalyser for å kontrollere for datastrukturen. Ingen andre enn undertegnede er ansvarlig for tolkninger og analyser gjort i denne oppgaven.

1.3 Strukturen i oppgaven

I neste kapittel presenterer jeg hvordan jeg har valgt å teoretisk forstå kjønnsdeling og relevant tidligere forskning på feltet. Det benytter jeg til å formulere hypotesene mot slutten av kapitlet. I oppgaven står metoden sentralt, og det gjenspeiles i omfanget av metodekapitlet, hvor jeg redegjør for alle sentrale metodevalg. Det innebærer harmonisering og valg av datamateriale på både individ- og gruppenivå, konstruksjonen av avhengig variabel og analyseteknikk. Det vies spesielt mye plass til beskrivelsen og diskusjonen av konstruksjonen av avhengig variabel og medfølgende reliabilitets- og validitetstester. I analysekapitlet viser jeg hvordan jeg har analysert datamaterialet, og tolkningen av modellene. Hypotesetestingen blir gjort underveis der. I siste kapittel besvarer og diskuterer jeg de to problemstillingene med basis i funn og nullfunn, og hvordan jeg har tolket det i lys av teori og tidligere forskning. Helt sist peker jeg på hva jeg har funnet og ikke funnet, og foreslår basert på det, videre steg i kunnskapsutviklingen.

2. Teori og tidligere forskning

I dette kapittelet kommer jeg til å kort introdusere den teoretiske logikken bak prosjektet, før jeg går inn i teorien om rasjonelle valg, som jeg skal bruke i diskusjonen. Deretter skal jeg trekke fram sentrale empiriske funn som jeg skal knytte til teorien. Til slutt formulerer jeg hypoteser basert på teori og tidligere forskning. De blir oppsummerer i tabell 1 på tampen av kapittelet.

Kjønnsdeling omtales gjerne i to former, horisontalt og vertikalt (NOU, 2012:144). Vertikal kjønnsdeling vil si hvorvidt menn og kvinner er fordelt ulikt i et stillingshierarki, fordelt på status og lønn. En kan eksempelvis snakke om at de som sitter i lederposisjoner oftere er menn enn kvinner, eller at statusen til yrker dominert av kvinner kan være preget av lavere lønninger (ibid.). I min oppgave er jeg interessert i den horisontale kjønnsdelingen, det vil si til hvilken grad menn og kvinner er fordelt i ulike typer yrker, sektorer eller næringer (ibid.). Jeg har derfor valgt bort et klasse- eller sosialt stratifiseringsperspektiv, og setter fokus på systematisk forskjellige yrkesvalg mellom menn og kvinner. Deling og segregering betyr i denne sammenhengen det samme, og vil bli brukt om hverandre. Når jeg bruker begrepet kjønnsdeling, mener jeg horisontal kjønnsdeling, med mindre jeg eksplisitt skriver noe annet.

2.1 Kjønnsdeling og sosial endring

Når man skal forklare hva som påvirker kjønnsdeling i arbeidsmarkedet, involverer det en tanke om at kjønnsdeling er noe som endres over tid og er forskjellig fra samfunn til samfunn. Giddens (2006:48) peker på tre hovedårsaker til sosial endring i moderne tid. *Kultur* påvirker hvordan vi tenker og hvilke ideer vi har. *Økonomiske rammer*, som den kapitalistiske strukturen påvirker, og blir påvirket av politiske og kulturelle faktorer. *Politikken* legger føringer både internt i nasjoner og eksternt i det internasjonale samfunnet. De spiller i dag en sentral rolle i å stimulere til økonomisk utvikling, og i alle industrielle samfunn er det en høy grad av statlig involvering i produksjon (ibid.). Disse tre hovedårsakene til sosial endring kan derfor sies å henge sammen og påvirke hverandre.

Et sett med kulturelt definerte ideer som likestillingsideer, kan det med dette grunnlaget argumenteres for at påvirker våre politiske ideer, som legger føringer for de tiltak som settes til verks i arbeidslivet. Et eksempel på en slik føring er likestillingsloven, vedtatt i 1978, og revidert en rekke ganger etter det. Formålet med loven «er å fremme likestilling uavhengig av kjønn» (Likestillingsloven, 2013, §1). Likestilling defineres til å innebære likeverd, like muligheter og rettigheter, tilgjengelighet og tilrettelegging (ibid.). Likestillingsloven kan derfor sies å fremme *formal- og sjanselighet*.

Formallikhet betyr at rettighetene og behandlingen av individer skal skje uavhengig av trekk ved individet (Schiefløe, 2011:274). Sjanselighet vil si at individer «gis samme muligheter, uavhengig av andre og utenforliggende forhold» (ibid.). Formålet med likestillingsloven er derfor å oppnå formallikhet og sjanselighet ved at den forbyr diskriminering basert på kjønn. Jeg forstår derfor likestilling som noe som gir både formallikhet og sjanselighet for menn og kvinner til å finne den jobben de vil ha. Graden av likestilling i et samfunn indikerer derfor hvorvidt individene i samfunnet har like muligheter, rettigheter og lik behandling, uavhengig av kjønn.

Å ha formallikhet og sjanselighet som mål, innebærer ikke nødvendigvis at menn og kvinner skal ha de samme type jobbene. Da er man over i en diskusjon om resultatlikhet (Schiefløe, 2011:275). Hvis man har resultatlikhet som mål i denne oppgavens sammenheng, skal menn og kvinner også være jevnt fordelt i alle yrker. Et mål på kjønnsdeling i arbeidslivet kan derfor ses som et mål på resultatet av formal- og sjanselikheten. Oppgavens hovedmål er dermed å måle formallikhetens og sjanselighetens påvirkning på resultatlikheten i kjønnsdelingen i arbeidslivet.

Den videre kursen for kapitlet er derfor å predikere hva som vil skje med kjønnsdelingen, dersom formal- og sjanselikheten økes eller reduseres. Kjønnsdeling i arbeidslivet er et komplekst og sammensatt fenomen. For å forstå hvorfor det oppstår, finnes det trolig ikke én riktig teoretisk innfallsvinkel. En kunne brukt et sosialiseringsteoretisk utgangspunkt der en ser for seg at opplevelser gjennom livet påvirker menn og kvinner på systematisk ulik måte, som gjør at utdannings- og karrierevalg forblir ulikt. Jonsson (1999) argumenterer imidlertid for at et sosialiseringsperspektiv ikke kan forklare hvorfor kvinner valgte å bryte sosialiseringsmønstrene ved å ta del i arbeidslivet på lik linje som menn. De ville ha blitt sosialisert til å forholde seg i familieinstitusjonen, og derfor holde seg unna arbeidsplassene (ibid.). Det kan derfor hende det er nødvendig med teori som tar for seg aktørenes strategiske valg når det er snakk om kjønnsdelte yrkesvalg.

2.2 Teori om rasjonelle valg

En av flere sentrale teorier når man skal forklare utdannings- og karrierevalg er teorien om rasjonelle valg (Reisel, 2014b:122). Teorien bestod tidligere av en «familie» spesifikke teorier, med utgangspunkt i de samme prinsippene. Friedman og Hechter (1988) lagde noe de beskrev som en skjelettmodell for teorien om rasjonelle valg, og hvordan den kan brukes i makrososiologisk forskning.

Ifølge Friedman og Hechter blir aktørene i teorien om rasjonelle valg sett som målrettet, og å ha intensjoner med handlingene sine (Friedman og Hechter, 1988:202). Teorien har dermed

metodologisk individualisme som utgangspunkt, der aktørene har rangerte preferanser, verdier og verktøy, men teorien tar ikke stilling til hvor disse kommer fra. Preferanser, verdier og verktøy blir alle omtalt under begrepet preferanser. Aktørene blir forstått som å handle med formål om å oppnå målene som blir definert av preferansehierarkiet. Handlingene kan ifølge teorien bli begrenset av to uavhengige kilder. Det ene er mangel på ressurser. Ulik fordeling av forskjellige typer ressurser gjør at noen mål er enkle å oppnå for noen individer, og ikke for andre. I tillegg «koster» det mer å oppnå noen mål enn andre. Aktørene må derfor holde følge med kostnadene ved deres nest-mest ønskede mål. En aktør kan velge å ikke forfølge det mest ønskede målet dersom ressursene er knappe, og hvis sjansene for å oppnå målet er små (ibid.). Mennesker fremstår dermed som kalkulerende og egoistiske i teorien om rasjonelle valg.

Den andre type begrensning av uavhengige kilder er sosiale institusjoner (Friedman og Hechter, 1988:202). Institusjoner legger føringer på handling ved å ha fastsatt regler. Håndhevbare regler kan blant annet være normer, lover og agendaer, og disse både fremmer og begrenser handlinger gjennom sanksjoner (ibid.). Jeg tolker poenget som at alle mennesker i samfunnet handler i kraft av sosiale institusjoner. Det kan være familie, utdanningssystem, arbeidsplass, lovverk og liknende. Mulighetskostnadene kan variere fra individ til individ eksempelvis på grunn av sosial bakgrunn eller kjønn. Friedman og Hechter (1988:202) skrev i en fotnote at de institusjonelle begrensningene ikke varierer fra individ til individ. I min oppgave har jeg imidlertid tatt meg friheten til å ta dette ett steg videre. Det er fordi hensikten med oppgaven er å se hvorvidt resultatligheten forandres, dersom formal- og sjanselikheten endres. Jeg velger derfor å se på de institusjonelle begrensningene som et uttrykk for formal- og sjanselighet, og likestilling (eller andre trekk ved samfunnet som helhet) som et uttrykk for institusjonelle begrensninger. Eksempelvis kan valg begrenses på ulik måte i land med høy grad av likestilling, kontra land med lavere grad av likestilling. De institusjonelle begrensningene tilknyttet lite likestilte land, er kanskje større for kvinner enn for menn. Jeg forstår derfor de institusjonelle begrensningene i et gitt samfunn en gitt tid som noe som forskjellsbehandler individer basert på gruppestereotypiske trekk ved individene.

Utdanningssystemet kan ses som en sosial institusjon. Det kan hende utdanningssystemet systematisk sanksjonerer eller belønner visse utdanningsvalg framfor andre basert på kjønn. Den institusjonelle begrensningen er i så måte forskjellig for menn og kvinner. Graden av forskjellsbehandling kan tenkes å ha en sammenheng med likestilling av kjønnene i samfunnet utdanningssystemet tilhører. Med dette kan en anta at graden av likestilling i et samfunn vil ha en negativ sammenheng med graden av kjønnsdeling.

2.3 Tidligere forskning

Tidligere arbeider har lagt vekt på å analysere hvorvidt enkeltland er kjønnssegregerte i arbeidslivet, og til hvilken grad kjønnsdelingen varierer mellom land (Nermo, 2000:296). Alt tyder på at graden av kjønnsdeling varierer i stor grad mellom land (f.eks: Chang, 2000, Charles og Grusky, 2004; Pinker, 2008). I tillegg viser det seg at land med høy grad av likestilling, egalitær kultur og liknende også er preget av høy grad av kjønnsdeling i arbeidslivet (Charles, 1992 Charles, 2003). Sammenhengen mellom likestilling på gruppenivå og horisontal kjønnsdeling på individnivå, der en kan kontrollere for individuelle variabler, har til min kunnskap ikke blitt undersøkt.

2.3.1 Ulike preferanser og økonomisk ulikhet

Noe forskning har antydnet at kvinner ser for seg å kombinere lønnet arbeid med barneomsorg (Mincer, 1974:123). Det kan derfor se ut til at kvinners mål er noe preget av et ønske om å balansere både karriere og familie. Mincer (1974:123) hevder også at dette kan forklare hvorfor kvinner velger annerledes enn menn. Det kan imidlertid ikke alene forklare hvorfor kjønnsdelingen endrer seg over tid, eller er ulik fra samfunn til samfunn, men kan gi aktørene i teorien om rasjonelle valg et mål som kan testes empirisk.

I land med lav grad av likestilling er det kanskje større avkastning ved å velge utradisjonelt, i at det er større økonomisk gevinst for kvinner å velge mannstypiske yrker. Selv om ikke inntekt ifølge Mincer (1974) er i toppen av preferansehierarkiet til kvinner, kan det hende at den relative økningen i gevinst likevel incentiverer kvinner til å velge mannstypiske yrker i land med høy grad av økonomisk ulikhet, enn i mer økonomisk jevnfordelte land. Det kan derfor hende det er bedre å se på ulikheter i inntekt på gruppenivå, enn likestilling for å forklare likestillingsparadokset.

2.3.2 Likestillingspolitikkenes dimensjoner og konsekvenser for kjønnsdeling

Reisel og Teigen (2014:11) påpeker at likestillingspolitikken «har fokusert for mye på kvinner som mødre, og for lite på kvinner som lønnsarbeidere». Det har blitt påstått at både den horisontale og vertikale kjønnsdelingen har en sammenheng med «familievennlig» politikk (Mandel og Semyonov, 2006; Reisel og Teigen, 2014:16). Argumentet er at tilbud som lange fødselspermisjoner, barnehager til en rimelig penge, deltidsarbeid og andre liknende ordninger, også gjør at kvinner systematisk incentiveres til å gå inn i noen få yrker, gjerne i offentlig sektor, der det er færre muligheter for å klatre i karrierestigen. Implikasjonen av den vertikale dimensjonen av dette har blitt tilbakevist av Korpi, Ferrarini og Englund (2013) med en flernivåanalyse på tvers av 18 land. Både Mandel og Semyonovs og Korpi m.fl. sine analyser tok for seg den vertikale dimensjonen, og ikke den horisontale. Den avhengige variabelen var hvorvidt man har en lederposisjon eller ikke.

Hva konsekvensene av «familievennlig» politikk på den horisontale kjønnsdelingen i arbeidslivet på individnivå er, har etter min kunnskap, forblitt ubesvart.

For å sette det i lys av rasjonelle valg, incentiveres menn og kvinner ulikt på bakgrunn av kjønn. Sett i et slikt perspektiv, og forutsatt at likestillingspolitikk i ulike land følger omtrent samme logikk (en dimensjon), men varierer i mengde, kan en forvente at økt likestilling fører til økt kjønnsdeling. Ifølge de tidligere undersøkelsene vil det skje gjennom at kvinner bevisst velger en familievennlig jobb, gjennom å ta deltidsarbeid eller jobb i offentlig sektor.

Korpi m.fl. (2013) har imidlertid argumentert for at likestillingspolitikken følger tre dimensjoner og ikke en. De tre dimensjonene er *den tradisjonelle familiedimensjonen*, *delt forsørger-dimensjonen* og *delt omsorgs-dimensjonen*. Førstnevnte befatter seg med barnebidrag, offentlige barnehager og liknende. Delt forsørger-dimensjonen representerer graden staten klarer å overføre barneomsorgen fra mødre til offentlig sektor. Sistnevnte representerer hvorvidt politikken har klart å stimulere fedre til å ta en mer aktiv rolle i barneomsorgen. De tre dimensjonene peker på ulike typer likestillingspolitikk som kan ha konsekvenser for kjønnsdelingen. Den første dimensjonen (tradisjonell familie), blir forbundet med statistisk diskriminering i den forstand at kvinner får avbrutte karriereløp. De to andre har ikke det i samme grad, og samvarierer (ibid.). Samvariasjonen av toforsørgerdimensjonen og omsorgsdimensjonen kalles *forsørger-omsorgs-dimensjonen* (earner-carer), der de skandinaviske landene topper skalaen (ibid.). Dette forklarer imidlertid ikke hvorfor Norge også har et *horisontalt* kjønnsdelt arbeidsmarked, men det gir analytiske rammer å forholde seg til. En tilsvarende studie der det undersøkes hvorvidt dimensjonene har konsekvenser for horisontal kjønnsdeling på individnivå har jeg ikke funnet.

Slik jeg forstår det, kan den tradisjonelle familiedimensjonen ha en innvirkning på kjønnsdeling ved at økt ansvar for barneomsorg skjer i familieinstitusjonen. Kvinners muligheter, og derfor formal- og sjanselighet, til å delta i arbeidsmarkedet på lik linje med menn blir derfor redusert. Når det gjelder de to andre, toforsørgerdimensjonen og delt omsorgsdimensjonen, er svaret mer uklart. Rent teoretisk kan vi forvente ut i fra en teori om rasjonelle valg, at når staten får mer ansvar for barneomsorg, og fedre tar større del i barnas liv, reduseres de institusjonelle begrensningene for kvinner. Formal- og sjanselikheten kan med dette sies å jevnes ut. Jeg velger å se nærmere på toforsørgerdimensjonen og delt omsorgsdimensjonen, fordi det er nettopp i disse dimensjonene Norge og Norden ligger i teten.

2.3.3 Offentlig sektor og kvinners andel i arbeidsmarkedet

I tråd med dette, har Norges offentlige utredninger (2012:141) utpekt sektor som et sentralt skille i arbeidsmarkedet, der kvinner i høyere grad enn menn arbeider i offentlig sektor, og menn i høyere grad enn kvinner arbeider i privat sektor. Andre faktorer som tidligere arbeider har trukket fram er deltidsarbeid, der argumentet er at deltidsarbeid er utbredt blant kvinner, fordi det gjør det enklere å kombinere med familieliv (Blair-Loy og Wharton, 2004). Mange høyt utdannede kvinner velger bevisst bort jobber som krever høy tilstedeværelse og dedikasjon, fordi opplevelsen av å ikke strekke til både på arbeid og i familien blir for stor. Vi kan forstå dette som at det er rasjonelt for kvinner å velge bort mannsdominerte yrker, fordi barneomsorg også står høyt i preferansehierarkiet for kvinner. Å kontrollere for arbeidstid og sektor er derfor relevant.

Et annet moment er at kvinners inntog i arbeidslivet skjedde samtidig som velferdsstatens ekspansjon. Poenget er at bransjen tilpasset seg den arbeidsstyrken som var tilgjengelig på det tidspunktet den ekspanderte (Reisel, 2014a:37). Det ble tilrettelagt for deltidsarbeid, slik at mødre kunne kombinere familie med arbeid. Hvis andre land har en liknende utvikling, kan vi forvente at kjønnsdelingen vil øke i tråd med økning i kvinners andel av arbeidsstyrken.

2.4 Hypoteseformulering

Fordi tidligere forskning og teori spriker når det kommer til hvilken sammenheng likestilling og kjønnsdeling har, har jeg valgt å formulere den første hypotesen i begge retninger. Likestilling og kjønnsdeling har vist seg å ha en sammenheng på gruppenivå, men vi har ikke liknende data der et segregeringsmål på individnivå brukes. Hvis man følger teorien om rasjonelle valg, kan en si at økt likestilling trolig vil gi en jevnere fordelt sjanse- og formallikhet, som en kunne anta gir et mindre kjønnsdelt arbeidsmarked (H1a). Det er med mindre de to gruppene menn og kvinner generelt sett har ulike preferanser. Hvis det er tilfellet, vil likestilling, og derfor økt mulighet for kvinner til å være sammen med sine barn, gi et mer kjønnsdelt arbeidsmarked (H1b). Det henger kanskje sammen med det Reisel og Teigen (2014) peker på, at likestillingspolitikken har behandlet kvinner som mødre framfor lønnsarbeidere.

Det kan imidlertid også tenkes at inntektsulikhet mellom land er den utgjørende faktoren. Hvis det har seg slik at økonomisk ulikhet og likestilling samvarierer, kan det hende likestilling kun er en spuriøs faktor. En kan se for seg at kvinner i økonomisk ulike land, er i høyere grad incentivert til å ta et mannstypisk yrke, fordi disse generelt gir mer betalt enn kvinnetypiske yrker. I land med lavere grad av økonomisk ulikhet, vil det også være mer likegyldig for kvinner, om de velger et typisk kvinneyrke eller et typisk mannsyrke, fordi lønnen vil være tilstrekkelig uansett. Vi kan

derfor tenke oss at økonomisk ulikhet har en negativ sammenheng med kjønnsdeling for kvinner i land med høy grad av økonomisk ulikhet (H2a), og at sammenhengen mellom likestilling og kjønnsdeling ikke er signifikant etter kontroll for inntektsulikhet (H2b).

En av fordelene ved å måle kjønnsdeling på individnivå, er at det muliggjør å kontrollere for andre individuelle faktorer. Jeg ser det som sannsynlig at utdanning har en sammenheng med kjønnsdelingen. Med bakgrunn i den teoretiske argumentasjonen i begynnelsen av kapittelet, er det å forvente at utdanningssystemene systematisk diskriminerer i form av ulike incentiver for menn og kvinner. Det er mangel på empirisk grunnlag som gjør det vanskelig å formulere en hypotese med en retning. På en side kan man si ut ifra teorien om rasjonelle valg, at desto lenger tid brukt i den sosiale institusjonen som ulikt incentiverer, desto mer kjønnsdelt vil det bli (H3a), fordi lengre utdanning trolig også betyr en mer spesialisert utdanning. En kunne også forstått det som at de individene som har fortsatt utdanningen sin, og ikke sluttet på grunn av ulik behandling, vil arbeide i mindre kjønnsdelte yrker (H3b), fordi implikasjonen av dette, er at de som tok en kortere utdanning ble incentivert av systemet inn i yrker som er kjønnsdelte (H3c).

Med bakgrunn i Korpi m.fl (2013) sine dimensjoner forventer jeg å finne at delt omsorgsdimensjonen har en negativ sammenheng med kjønnsdeling, fordi dimensjonen innebærer å jevne ut sjanselikheten mellom menn og kvinner i arbeidsvalgene (H4a). Desto mer menn deltar i barneomsorgen, og kvinner derfor mindre, vil trolig menn og kvinner i høyere grad havne i samme yrker. Det samme gjelder delt forsørger dimensjonen (H4b). Det er fordi desto mer av ansvaret for barneomsorg staten har tatt på seg, betyr nødvendigvis også mindre tid brukt på barneomsorg for kvinnene. Mindre barnepass på kvinner, betyr trolig mer tid til arbeid. Det muliggjør derfor kanskje at kvinner kan ta på seg mer konkurransepregede yrker i privat sektor, som tradisjonelt har vært mannsdominerte.

Det tyder på at likestillingspolitikk fører til kjønnsdeling gjennom lange fødselspermisjoner, fleksibel arbeidstid og liknende tilbud. Dette blir mest synlig i offentlig sektor, som ikke er like konkurransepreget som privat sektor. Det er i tråd med funnene som sier at kvinner prøver å kombinere familielivet med arbeid i høyere grad enn menn. Likestillingspolitikken legger dermed opp til at kvinner havner i offentlig sektor, og vi kan forvente at andelen kvinner i offentlig sektor er en god prediktor på kjønnsdeling (H5a). Det skal ha skjedd som følge av at kvinner har i høyere grad tatt del i arbeidsmarkedet. Vi kan derfor forvente at kvinneandelen i arbeidsstyrken¹, også har en positiv sammenheng med kjønnsdeling (H5b), og at denne effekten forsvinner når man kontrollerer for andelen kvinner i offentlig sektor (H5c).

1 Andelen av arbeidsstyrken som er kvinner. Ikke andelen av kvinner totalt som arbeider.

Hvis det stemmer, vil vi trolig finne at deltidsarbeid er forbundet med kjønnsdeling for kvinner, mens heltidsarbeid er forbundet med kjønnsdeling for menn (H6a og H6b), fordi den «familievennlige» politikken har ført til at kvinner i høyere grad enn menn arbeider deltid. Det samme argumentet gjelder sektor på individnivå, der vi kan forvente å se at offentlig sektor har en sammenheng med økt kjønnsdeling for kvinner, mens privat sektor samvarierer med kjønnsdeling for menn (H6c og H6d).

En videreføring av logikken her kan testes ved å kontrollere for antall personer i husholdningen. Ved at det er flere personer i husholdningen, kan bety at respondenten har ett eller flere barn. Som jeg har forklart, viser den tidligere forskningen at deltidsarbeid er en strategi mange kvinner bruker for å kombinere familieliv og derfor barneomsorg med arbeidslivet. Vi kan anta at et økt antall personer i husholdningen har en sammenheng med økt kjønnsdeling for kvinner i samboerskap, men ikke for menn (H6e). Grunnen til at jeg skriver kvinner i samboerskap, er for å forsøke å utelukke respondenter som bor med sine foreldre, eller av andre grunner har en høy verdi på antall personer i husholdningen, men ikke har barn. Målet er ikke optimalt, men det er imidlertid det nærmeste målet som finnes i datamaterialet jeg bruker, og det er trolig en sammenheng mellom antall personer i husholdningen, samboerskap og antall barn man har.

Hvis det har seg slik, vil vi også kunne forvente at disse individuelle faktorene «stjeler» forklaringskraften til de kontekstuelle variablene, fordi de er mer nærliggende og et mer direkte mål på årsaken til kjønnsdeling enn variablene på gruppenivå. Jeg forventer derfor å finne at likestilling, andel kvinner i offentlig sektor, andel kvinner i arbeidsmarkedet, toforsørger- og delt omsorgsdimensjonen sine sammenhenger med kjønnsdeling forsvinner når man kontrollerer for arbeidstid, sektor, samboerskap og antall personer i husholdningen (H7a, H7b, H7c og H7d). Dersom de kontekstuelle variablene fortsatt er signifikante, er det et tegn på at det er noe mer enn forklaringsmodellene beskrevet ovenfor som påvirker hvorfor kjønnsdelingen oppstår.

Tabell 1: Oppsummering av hypoteser

- H1a Likestilling har en negativ sammenheng med kjønnsdeling.
- H1b Likestilling har en positiv sammenheng med kjønnsdeling.
- H2a Økonomisk ulikhet har en negativ sammenheng med kjønnsdeling for kvinner.
- H2b Sammenhengen mellom likestilling og kjønnsdeling er ikke signifikant etter kontroll for økonomisk ulikhet og samspill med kjønn.
- H3a Høy utdanning har en positiv sammenheng med kjønnsdeling, sammenlignet med lav utdanning.
- H3b Høy utdanning har en negativ sammenheng med kjønnsdeling, sammenlignet med lav utdanning.
- H3c Lav utdanning har en positiv sammenheng med kjønnsdeling, sammenlignet med høy utdanning.
- H4a Delt omsorgs-dimensjonen har en negativ sammenheng med kjønnsdeling.
- H4b Delt forsørger-dimensjonen har en negativ sammenheng med kjønnsdeling.
- H5a Andelen kvinner i offentlig sektor har en positiv sammenheng med kjønnsdeling.
- H5b Kvinneandelen i arbeidsstyrken har en positiv sammenheng med kjønnsdeling.
- H5c Sammenhengen mellom kvinneandelen i arbeidsstyrken er ikke signifikant etter kontroll for kvinneandelen i offentlig sektor.
- H6a Heltidsarbeid har en positiv sammenheng med kjønnsdeling for menn, sammenlignet med deltidsarbeid.
- H6b Deltidsarbeid har en positiv sammenheng med kjønnsdeling for kvinner, sammenlignet med fulltidsarbeid.
- H6c Privat sektor har en positiv sammenheng med kjønnsdeling for menn, sammenlignet med offentlig sektor.
- H6d Offentlig sektor har en positiv sammenheng med kjønnsdeling for kvinner, sammenlignet med privat sektor.
- H6e Antall personer i husholdningen har en positiv sammenheng med kjønnsdeling for kvinner i samboerskap.
- H7a Sammenhengen mellom likestilling og kjønnsdeling er ikke signifikant etter kontroll for arbeidstid, sektor, samboerskap og antall personer i husholdningen.
- H7b Sammenhengen mellom andel kvinner i offentlig sektor og kjønnsdeling er ikke signifikant etter kontroll for arbeidstid, sektor, samboerskap og antall personer i husholdningen.
- H7c Sammenhengen mellom andel kvinner i arbeidsmarkedet og kjønnsdeling er ikke signifikant etter kontroll for arbeidstid, sektor, samboerskap og antall personer i husholdningen.
- H7d Sammenhengen mellom delt forsørger-dimensjonen og kjønnsdeling er ikke signifikant etter kontroll for arbeidstid, sektor, samboerskap og antall personer i husholdningen.
- H7e Sammenhengen mellom delt omsorgs-dimensjonen og kjønnsdeling er ikke signifikant etter kontroll for arbeidstid, sektor, samboerskap og antall personer i husholdningen.

3. Data og metode

I dette kapitlet skal jeg gjøre rede for datagrunnlaget og den metodiske tilnærmingen for analysene. I avsnittet om datagrunnlag vil jeg ta leseren med på en utgreiing om hvilket datamateriale som ligger til grunn, og hvordan det har blitt behandlet. Deretter skal jeg vise hvordan konstruksjonen av segregeringsmålet som er avhengig variabel ble konstruert, der jeg også validitets- og reliabilitetstester den. Etter det viser jeg hvilke kontekstuelle variabler jeg har tatt med og hvordan de har blitt tilpasset mine analyser. Til slutt kommer jeg til å forklare og diskutere valg av analyseteknikk. Diskusjon om validitet og reliabilitet blir gjort underveis. Databehandlingen ble gjort i Stata IC 14.1, og analysene i Stata MP 13.1.

3.1 Datagrunnlag

Datamaterialet som er brukt er 15 årlige runder av International Social Survey Programme (ISSP Research Group, 2002-2016), fra runde 1999 til og med runde 2013. ISSP er et samarbeidsprosjekt der statistiske byråer fra ulike land i verden deltar ved å bruke tilnærmet like, men oversatte spørsmål som en del av sine nasjonale undersøkelser. Svarene fra deltakerlandene blir deretter satt sammen og stilt til disposisjon i anonymisert form av Leibniz Institute for the Social Sciences (GESIS).

Et av målene med masterprosjektet mitt er å finne ut om det finnes en internasjonal sammenheng mellom likestilling og kjønnsdeling. For å gjøre det trenger jeg representativt datamateriale fra hele verden. Det finnes ingen datasett som er representativt for verdens befolkning, men hensikten med valget av datamateriale var å legge seg så tett opptil det idealet som mulig.

De ulike rundene av ISSP har ulike temaer. Det vil si at i en runde av ISSP, for eksempel i runde 2009, har spørsmålene dreid seg om sosial ulikhet, mens i runde 2008 var temaet religion. De resterende variablene gir demografisk informasjon om respondentene. Fordi jeg har tatt utgangspunkt i flere datasett, er det kun den demografiske informasjonen som sammenfaller, og som jeg benytter i mine analyser. Det totale utvalget er respondenter som har eller har hatt arbeid, og som har besvart undersøkelsene fra ISSP runde 1999 til og med ISSP runde 2013 (ISSP Research Group, 2002-2016).

Årsaken til at det ble brukt flere datasett var for å øke utvalg på gruppenivå. Et utvalg på 20 land fra ett datasett er i minste laget når man bruker flernivåanalyser med random effects, fordi det kan gi unøyaktige standardfeil og signifikansverdier på gruppenivå (Bryan og Jenkins, 2015:4). I ett

enkelt datasett er det færre land enn hvis en kombinerer flere. Mange land har vært med i flere runder av ISSP. Det medførte nødvendigvis å skille utvalgene i samme land fra hverandre basert på tidspunktet undersøkelsen ble gjennomført. Tidsdimensjonen var dermed ikke tatt med for å se på utviklingstrekk i kjønnsdelingen over tid, men for å forbedre analysenes presisjon og derfor generaliserbarhet. En oversikt over landene som er med i undersøkelsen kan ses i tabell 3 på side 37.

3.2 Problematisering av datamateriale

I en flernivåanalyse kan man se på gruppene (landene) som et utvalg i seg selv (Maas og Hox, 2005; Snijders, 2005). Landene som er med i ISSP har selv valgt å være med, og de resterende landene som ikke er med, kan derfor ses på som ikke-svar. Selvseleksjonen kan ha ført til at det er spesielle trekk ved land som er med i undersøkelsen, som fører til systematiske skjevheter i utvalget land. For eksempel er det tenkelig at det er nødvendig for land å ha nok ressurser til å ha mulighet til å prioritere pengebruk på statistiske undersøkelser. Dette kan ha medført at landene som er med er mer velstående enn landene som ikke er med. Konsekvensen av selvseleksjonen er at jeg ikke kan generalisere til hele verden, men bare til de landene som er med i undersøkelsen.

Språket i spørsmålene respondentene ble stilt varierte, i og med at undersøkelsen er blitt gjort i mange land. ISSP er som sagt datamateriale samlet fra en rekke deltakernasjoner, der statistiske byråer i landene selv administrerer undersøkelsen. Det vil si at de oversetter spørsmål, henter data og videreformidler dette tilbake til GESIS, som er ansvarlig for ISSP. Når en jobber komparativt og bruker datasett fra mange land, er det et relevant spørsmål om begrepene som blir brukt er definert ulikt i ulike språk (Harzing, Reiche og Pudelko, 2012). Det er et poeng å være forsiktig med å attribuere funn om gruppeforskjeller, når forskjellen kan ha blitt forårsaket av språklige ulikheter. Derfor har jeg kun tatt utgangspunkt i demografiske variabler. Jeg mener det er rimelig å anta at folk er sikre nok i svarene på demografiske spørsmål, og at ulikheter i spørsmålsstilling ikke påvirker dette. Variabler som i større grad er i faresonen er spørsmålene om bosted, fagforening og utdanning, som vil bli diskutert underveis når jeg beskriver variablene. Spørsmålene i undersøkelsen blir kun presentert i sin basisform på engelsk. Spørsmålene til noen av bakgrunnsvariablene har også blitt endret i 2009 og 2010, og det opplyser jeg om der det er relevant underveis.

3.3 Harmonisering og beskrivelse av uavhengige variabler på individnivå

Selv om GESIS har satt sammen data fra deltakerlandene hvert år, er det en større utfordring når disse datasettene igjen skal slås sammen. Variabler som måler det samme er kodet ulikt, har

forskjellig navn, og i noen tilfeller har variabelen blitt fjernet, og andre lagt til. Harmonisering vil si prosessen der variabler som måler det samme fenomenet med forskjellige verdier blir sydd sammen til en felles standard. Nå skal jeg gå gjennom prosessen med harmonisering av datasettene, og de metodiske valgene gjort der, før jeg går videre med hvordan avhengig variabel ble skapt.

De 15 datasettene hadde i gjennomsnitt 48209 observasjoner hver, og 700321 totalt, men etter fjerning av doble utvalg var det tallet 575114. En diskusjon om fjerning av doble utvalg kommer senere i kapittelet under overskriften «Sammenslåing av datasett og konstruksjon av variabler på andrenivå». Det største datasettet var på 58897 observasjoner, og det minste hadde 30297 observasjoner. De demografiske variablene som ble beholdt fra alle datasettene var alder, kjønn, utdanningsgrad, personlig inntekt, husholdningsinntekt, samboerskap, jobbsektor, fagforeningsmedlemskap, bosted, om en er sjef over andre, arbeidstid, antall i husholdningen, land og yrkesgruppe som ble brukt til å konstruere avhengig variabel. Det utgjør alle variabler i det endelige datasettet på individnivå som jeg har tatt med i analysene.

Alder, kjønn og antall personer i husholdningen

Noen variabler som alder, kjønn og antall personer i husholdningen var uproblematiske å harmonisere. Alt som var nødvendig var å sette ubesvart til missing. Respondentene ble bedt om å oppgi fødselsår, og gjennomsnittsalderen for det totale utvalget er 46,25 år med et standardavvik på 17,23 år, og kun 0,37% missing. For kjønn var fordelingen litt flere kvinner enn menn (54,23% kvinner). Det skyldes trolig at det i enkelte utvalg var en overrepresentasjon av kvinner grunnet skjevheter i utvalgsdesignet, og en vektevariabel tar høyde for det i analysene. For antall personer i husholdningen var gjennomsnittet på 2,9 personer, et standardavvik på 1,6 og 7,3% missing.

Deltakerlandenes ansvarlige statistiske byrå stiller selv med en vektevariabel dersom utvalgsskjevheter har oppstått som følge av enten ujevne sannsynligheter for inklusjon, eller manglende svarprosent. Ujevne sannsynligheter for inklusjon kan innebære at ulike områder i et land har like mange respondenter, men at populasjonen i noen av områdene er større enn andre, og deres verdier veies tyngre enn andre. Hva det vektet for kommer an på utvalgsdesignet innenfor den enkelte deltakende nasjon. Eksempelvis er det blant annet i Tsjekkia i ISSP-runde 2000, 40,4 prosent menn med i utvalget, mens populasjonsdata viser at landet har 48 prosent menn som helhet. Vektevariabelen oppveier det, ved å gi menns verdier relativt til kvinner, litt mer innvirkning på resultatene i det utvalget. I Tsjekkia ble det også tatt høyde for aldersgrupper, regioner, arbeidsledighet og utdanningsnivå. Jeg skal ikke gjennomgå all vekting i alle utvalg, fordi det ville tatt alt for mye plass. For den interesserte finnes all informasjon om vekting i ISSPs variabelrapporter (GESIS, udatert). Validiteten for alder, kjønn og antall i husholdningen variablene

anser jeg som god. Som jeg skal vise senere skal variabelen bli brukt i et samspill med kjønn og samboerskap for å måle hvorvidt en har barn eller ikke. Da blir spørsmålet om validitet tatt opp igjen.

Utdanning

For andre variabler forekom det endringer i verdier mellom rundene av ISSP. Utdanningsvariabelen var kodet på tre ulike måter, og respondentene ble stilt spørsmålet: «What is the highest level of education that you have attained?». Med mange forskjellige skolesystemer, der det er ulikt antall år en studerer for å nå grader som kanskje ikke kan sammenlignes, kan variabelen bli problematisk. De fleste landene har heldigvis benyttet seg av den internasjonale klassifiseringen av utdanning (ISCED). De landene som ikke har gjort det, eksempelvis Slovenia i runde 2008, målte utdanning i ni kategorier og plasserte dem i ISSPs fem kategorier (Hoffmeyer-Zlotnik og Warner, 2013:242). Det kan derfor ha skjedd noen skjevheter på tvers av land, som jeg skal vise i neste avsnitt at jeg i kodingen min gjør et forsøk på å ta høyde for.

I ISSP-rundene fra 1999 til 2001 var variabelen delt opp i sju verdier. Fra 2002 til 2009, var den delt i seks verdier, mens fra 2010 til 2013 var den på sju verdier, dog med en annen sammensetning enn den første. Variabelen ble kodet ned til fire verdier, der 1=ingen eller laveste formelle utdanning, 2=over laveste formelle utdanning, 3=høyere og over sekundærutdanning 4=universitetsutdanning. Jeg valgte å slå sammen verdiene for ingen formell utdanning og laveste formelle utdanning fordi kun 7% av utvalget var uten utdanning. Teknikken kan beskrives som konservativ, da jeg ønsker minst 10% i hver kategori for å unngå at referansekategorien i dummysettet fører til misvisende resultater for resten av utvalget. Jeg valgte i tillegg å slå sammen høyere sekundærutdanning og over høyere sekundærutdanning på grunn av usikkerhet knyttet til de ulike landenes utdanningssystemer som beskrevet i avsnittet over. Det kan hende det fører til reduserte effekter og mindre signifikante resultater, men det er også en tryggere måte å sørge for at målingen ikke blir ujevn på tvers av landene. I det endelige utvalget var det henholdsvis i de fire kategoriene i økende inntektsrekkefølge: 18%, 22%, 40%, og 19%, og jeg anser validiteten for både variabelen og omkodingen som god.

Bosted

For variabelen bosted var det fra ISSP-runde 1999 til 2001 kun tre kategorier, mens fra 2002 til 2013 var det fem. Den enkleste varianten med tre verdier der 1=urbant, 2=suburbant og 3=ruralt ble derfor benyttet. Basisformen av spørsmålet lød: «Would you describe the place where you live as...» med avkrysning for de ulike bostedene. I Norge vil det være rimelig å anta at folk har en

annen oppfatning av hva det betyr å bo urbant enn andre i mer tettbebygde land. Innad i hvert land vil variabelen kanskje være mer meningsfull. En kan anta at respondentene innad i hvert land har omtrent den samme forståelsen av hva det vil si å bo ruralt og hva det vil si å bo urbant. En kunne innvendt at dette er problematisk fordi respondentene selv definerer bostedet. Jeg velger å ikke se på dette som et problem fordi en sannsynligvis vil se forskjeller mellom urbant og ruralt, og at utvalgsmetoden til de nasjonale undersøkelsene er gjort i flere steg gjennom regioner i hvert land. På den måten sikrer man en jevn fordeling, som er mellom 30 og 40 prosent i hver gruppe i det endelige utvalget.

Samboerskap

Variabelen for samboerskap har jeg basert på to variabler. Variabelen for samboerskap var med i alle rundene av ISSP. Fram til 2009 lød spørsmålet «Do you live together with a partner?» (ISSP Demographic Reseach Group, 2001), og etter 2009 «Do you have a spouse or a steady partner and, if yes, do you share the same household?» (ISSP Demographic Reseach Group, 2012). Respondentene som svarte på spørsmålet hadde tidligere svart at de *ikke* var gift og bor med ektefelle. Derfor benyttet jeg meg også av variabelen for sivilstatus. Dermed ble de som var gift og bor med ektefelle, sammen med de som hadde svart at de bor med samboer, satt til 1. Resten fikk verdien 0. Navnet jeg har gitt variabelen kan derfor være misvisende, fordi den også har med de som er gift. Variabelen har 2% missing, 67% er samboere i det endelige utvalget. Det tilsvarer det norske gjennomsnittet (Statistisk sentralbyrå, 2012), og jeg anser derfor validiteten som god.

3.3.1 Arbeidsrelaterte uavhengige variabler

Andre variabler var mer krevende, som arbeidstimer i uken, arbeidstype, fagforeningsmedlemskap og veiledning av andre på jobb. For å redusere antall missing, ble missing satt til en meningsfull verdi dersom en var satt som arbeidsledig i en annen variabel som måler arbeidsstatus. Dette innebærer hjemmeværende, uføre, studenter, pensjonerte og andre som ikke er deltakende i arbeidslivet. Et spørsmål leseren muligens har i denne sammenheng er hvorfor arbeidsledige blir gitt meningsfulle verdier, når målet med oppgaven er å måle kjønnsdeling i arbeidslivet. Svaret på det er at avhengig variabel som jeg skal vise, også har med informasjon om tidligere arbeidsforhold dersom respondenten er arbeidsledig for øyeblikket. Teknikken blir derfor en metode for å øke antall respondenter, og å unngå at det blir en overrepresentasjon av personer med en sikker jobb framfor personer som ikke har sikker jobb. Personer med en sikker jobb vil sannsynligvis oftere ha svart at de har en jobb på spørsmålet, enn personer som generelt har usikre jobber. Jeg mener derfor at teknikken er med på å øke representativiteten til undersøkelsen fordi den tar med hele spekteret

yrker, og ikke i høyere grad «vanlige» og «trygge» yrker enn det populasjonen har.

Arbeidstimer i uken

Arbeidstimer i uken var opprinnelig en kontinuerlig variabel, men ble omkodet til en kategorisk variabel der 0=0 timer i uken eller arbeidsledig, 1= mindre enn 37 timer i uken, og 2= 37 timer i uken eller mer. Med utgangspunkt i hypotesene mine ønsker jeg å fange opp forskjellen mellom deltidsarbeid og heltidsarbeid. Jeg har ikke et ønske om å se på hva som skjer med den gjennomsnittlige kjønnsdelingen dersom arbeidstimer i uken økes med x antall timer, og derfor er et dummysett å foretrekke. Respondentene ble stilt spørsmålet «How many hours, on average, do you usually work for pay in a normal week, including overtime?» (ISSP Demographic Research Group, 2001). Jeg har som med de andre arbeidsrelaterte variablene satt arbeidsledige til verdien 0, for å ikke miste dem i analysene. En mulig innvending er at antall arbeidstimer i uken som definerer en fulltidsjobb, kan variere fra land til land, og 37 timer er bare et skille vi i Norden opererer med. Dog måtte grensen settes en eller annen plass, og å ha en ulik grense fra land til land kunne medført skjevheter i seg selv. Det kan ha ført til at noen som arbeider i det enkelte land definerer som deltid, har havnet i heltidskategorien. Det er ingen vei utenom det, men i det store og det hele bør variabelen uansett fange opp de viktigste forskjellene mellom heltid og deltid. Jeg anser validiteten av behandlingen som god, da en person som ikke er i arbeid, sannsynligvis ikke jobber noen timer i uken, fordi jeg er interessert i forskjellen i effekten av deltidsarbeid og heltidsarbeid på kjønnsdeling.

Sektor

Sektor er en variabel som måler hvorvidt respondenten arbeider i offentlig eller privat sektor, og er basert på ulike variabler ut fra hvilket år i ISSP det er fra. I datasettene fra runde 1999 til 2009 målte variabelen hvorvidt en arbeidet for staten, offentlig eid firma, privat firma eller annet (f.eks frivillige eller ideelle organisasjoner). «Spørsmålet» i sin basisform på engelsk var «Working for private sector, public sector or self-employed» (ISSP Demographic Research Group, 2001). Senere ble variabelen delt i to, der en av dem måler om organisasjonen er ideell eller ei, og den andre måler hvorvidt organisasjonen er offentlig eller privat eid. Mest relevant for min oppgave var spørsmålet om organisasjonen er statlig eller privat eid, der basisversjonen av spørsmålet var «Do/ did you work for a public or a private employer?» (ISSP Demographic Research Group, 2012). Derfor ble den ble benyttet videre, og harmonisert til en felles standard for å kunne ses i samme variabel som de tidligere datasettene. Variabelen har derfor tre verdier der 0=arbeidsledig, 1=offentlig sektor og 2=privat sektor. Jeg har valgt å ha med de arbeidsledige i en egen kategori, fordi jeg ikke vil miste dem i analysene, og å plassere dem sammen med en av de andre kategoriene ville ført til skjevheter.

I det endelige utvalget var 4% arbeidsledig, 32% i offentlig sektor og 63% i privat sektor. Jeg har ingen bekymringer når det kommer til validiteten av variabelen, fordi jeg regner med sektor er et tydelig skille i arbeidslivet uavhengig av land og språk. Selv om variabelen er hentet fra to ulike variabler, er resultatet trolig likt, fordi spørsmålet ber om omtrent det samme.

Det er nødvendig å ha de arbeidsledige med som verdi, fordi å ta dem ut resulterer i skjevheter i de andre variablene dersom flere variabler er med i samme analyse. Derfor blir ikke de arbeidsledige brukt som referansekategori i analysene, men blir tatt med for å unngå at de faller helt bort.

Fagforeningsmedlemskap

Fagforeningsmedlemskap ble målt med ulik presisjon fra ISSP-runde til ISSP-runde. Det var derfor nødvendig å beholde variabelen på det enkleste nivået der 0=ikke medlem eller arbeidsledig og 1=medlem. Noen av datasettene målte også hvorvidt respondentene tidligere hadde vært medlem, men ikke alle. Derfor ble den enkleste varianten benyttet. I nordiske land er det svært vanlig å være medlem av fagforeninger, mens i land som USA er det mindre vanlig. I utvalgene fra Norge er det eksempelvis 50% som er medlem, og i USA er det bare 9%. I det totale endelige utvalget var 77% ikke medlem eller arbeidsledig, og 23% medlem av fagforening. Jeg har ingen grunn til å tro at det er noen problemer med validiteten i variabelen som sådan, men det er verdt å tenke på om den kan ta forklaringskraft fra variabler på andre nivå, fordi det er store forskjeller i fagforeningsmedlemskap mellom land.

Lederoppgaver

Variabelen for om har lederoppgaver, har jeg ikke gjort noe med, og var lik i alle rundene av undersøkelsen. Spørsmålet respondentene ble stilt var «Do/did you supervise other employees?» (ISSP Demographic Research Group, 2001). Variabelen er derfor dikotom med 0=Nei og 1=Ja. Respondentene jeg manglet verdier for ble satt til 0 dersom de ikke var i arbeid. I det endelige utvalget var 25,6% for øyeblikket eller tidligere sjef over andre i en arbeidssammenheng. 2,69% hadde ikke svart på spørsmålet. Jeg ser ingen problemer med variabelen, og anser den som valid.

3.3.2 Inntektsvariabler

Den samme logikken som i de arbeidsrelaterte variablene, men mer omfattende, ble benyttet i inntektsvariablene. Her ble missing i husholdningsinntekt satt til samme verdi som personlig inntekt dersom respondenten ikke hadde samboer og var arbeidsledig, og motsatt. Dersom en hadde samboer, ble missing i husholdningsinntekt satt til samme verdi som personlig inntekt dersom partneren var arbeidsledig. Jeg mener teknikken er valid og bedre enn å unnlate å gjøre det. Det er

fordi mange respondenter kun har svart på en spørsmålene for inntektsvariablene. Ved å utnytte verdier fra de to variablene økes det totale antallet meningsfulle verdier. Det var en sterk sammenheng mellom husholdningsinntekt og personlig inntekt ($\alpha=0,85$, $r=0,8$, mellom ubehandlet personlig- og husholdningsinntekt), noe jeg mener kan forsvare valget. Deretter ble missing i personlig inntekt satt til 0 dersom en ikke var i arbeid og ikke hadde partner. Til slutt ble missing i husholdningsinntekt satt til 0 dersom både respondenten og partneren var arbeidsledig. Det kan ha ført til at noen respondenter som har inntekt utenom lønn har fått en misvisende verdi. Jeg mener likevel at det er et bedre alternativ enn ikke å gjøre det, fordi det i flesteparten av tilfellene vil stemme i høyere grad enn hvis respondentene ikke er med i analysene, og skaper skjevheter på den måten. For å være på den sikre siden ble ubehandlede inntektsvariabler tatt med for å se om behandlingen hadde en innvirkning på resultatet. Missing i personlig inntekt ble redusert fra 21,86% til 7,89%. For husholdningsinntekt var det fra 19,64% til 11,63%. Metodevalget la også en føring for hvordan inntektsvariablene ble harmonisert.

Inntekten ble målt i landets valuta det året. Neste steg ble derfor å standardisere verdiene. Her er det to mulige fremgangsmåter. Enten regner man om til en felles valuta og tar høyde for inflasjon. Fordelene ved det er at en mister minst mulig informasjon, og at en kan beholde variabelen som kontinuerlig i analysene. Hvis ikke kan man benytte en prosentbasert tilnærming, og deler opp i innteksgrupper relativt innad i utvalget. Jeg ser tre fordeler ved dette. Det ene er at utførelsen er kjapp. I et omfattende harmoniseringsprosjekt som dette, er tidsbruk en relevant faktor. Et annet poeng er at respondentene der missingverdiene ble satt til 0, i høyere grad havner i korrekt innteksgruppe. Hvis en respondent er arbeidsledig, og ikke svarte på spørsmålet om inntekt, ser jeg det som sannsynlig at personen får inntekter gjennom for eksempel velferdsmidler. En slik type inntekt er sannsynligvis i laveste inntektskategori, og tilnærmingen har derfor tatt høyde for potensielle misvisende 0-verdier i inntekt. Den siste fordel er at for å plassere respondenten i et sosialt stratifiseringssystem, er det nok å måle den relative inntektsposisjonen for respondentene (Warner, 2009:27). Den prosentbaserte tilnærmingen ble dermed valgt, og syntaksen er vist i vedlegg på side 76.

Selv om en hadde brukt absolutte verdier i inntekt, ville det gi lite tolkningsmuligheter av en viktig grunn. Det er at kjøpekraften og eventuell symbolsk status en enhet penger gir, vil være ulikt fra land til land. Det viktigste for tolkningen sin del er å kunne se relative forskjeller i inntekt innad i utvalget. Det er fordi innad i hvert utvalg vil hver enhet penger gi omtrent den samme kjøpekraften og eventuell medfølgende symbolsk verdi. Innenfor hvert utvalg ble dermed respondentene delt inn i fire innteksgrupper der 1 = laveste kvartil og 4 = høyeste kvartil. Til dette ble *xtile* kommandoen i

Stata brukt. Variablene blir derfor brukt som dummiesett i modellene.

Det kan hende leseren har lurt på hvorvidt det er et problem å ha med begge inntektsvariablene når korrelasjonen mellom dem var såpass høy. Til det har jeg beregnet variansinflatningsfaktor (VIF) for en samlet OLS-modell. VIF er en måte å måle multikolaritet i OLS-modeller, og den sier ifra når man har med variabler som korrelerer så sterkt at de måler omtrent det samme. Hvis det er høy korrelasjon mellom uavhengige variabler, økes standardfeilen, og en får mer upresise estimater (Skog, 2012:286). Det er ikke mulig å beregne VIF for flernivåmodeller fordi utregningen tar utgangspunkt i R^2 (Acock, 2014:288). Derfor ble en samlet OLS-modell estimert, og VIF ut fra denne. Gjennomsnittlig VIF kom på 1,71, og betyr at kolaritet ikke var et problem i modellen. Den høyeste verdien var imidlertid for inntektsvariablenes høyeste kvartil på rundt 2,9, men det er fremdeles langt under 10 som er tommelfingerregelen for beregning av multikolaritet (ibid.). De to variablene samvarierer mindre etter den prosentbaserte kategoriseringen med $\alpha=0,71$ og $r=0,56$. Jeg anser validiteten som god, fordi jeg ikke er ønsket å se på absolutte verdier i inntekt sin sammenheng med kjønnsdeling.

3.4 Avhengig variabel

Sentralt for oppgaven står spørsmålet om hvordan en skal måle kjønnsdeling i arbeidslivet. Jeg skal derfor bruke vesentlig plass i oppgaven på både presentasjonen av og diskusjonen rundt avhengig variabel. Det første som skal besvares er hvorfor det er behov for et nytt mål på segregering. Herunder skal jeg i korte trekk definere kjønnsdeling, hvordan det har blitt målt tidligere, og hvorfor jeg mener disse ikke er tilstrekkelig. Deretter går jeg videre med å forklare hvordan jeg har løst utfordringene med operasjonaliseringen av kjønnsdeling i arbeidslivet. Den benyttede tilnærmingen som jeg skal vise i detalj, er et nytt bidrag til måling av kjønnsdeling og segregering generelt.

3.4.1 Tidligere forskning på kjønnsdeling og segregering

Den tradisjonelle måten å måle kjønnsdeling på er gjennom bruk av indekser. Det har vært en flere generasjoner lang akademisk diskusjon om hvordan en best måler kjønnsdeling og segregering. Ut fra denne diskusjonen har det kommet flere bidrag og innfallsvinkler til metodiske strategier. Det første spørsmålet en bør stille er nøyaktig hva det er vi vil måle.

I den forbindelse kan det hentes nyttige begreper fra forskningen som relaterer til kjønnsdeling og segregering generelt. Det finnes ulike typer segregering, og en kan si at mesteparten av forskningen er benyttet innenfor samfunnsgeografi (Iceland m.fl. 2002:119). Hyppig

brukte begreper for å måle segregering innenfor samfunnsgeografi er *eksponering*, *konsentrasjon*, *sentralisering*, *clustering*, og begrepet jeg skal ta for meg, *jevnhhet*. Begrepet jevnhet (*evenness*) innebærer i hvilken grad ulike grupper er jevnt fordelt forskjellige kategorier (ibid.). I mitt tilfelle skal jeg se på i hvilken grad menn og kvinner er jevnt eller ujevnt fordelt i ulike yrker.

I begrepet ligger det en implikasjon om at en må ta høyde for ulik størrelse på gruppene menn og kvinner. Dersom det for eksempel i et land er langt flere menn enn kvinner i arbeid, er ikke et yrke jevnt fordelt dersom det er en 50/50 fordeling av menn og kvinner. For at et land skal ha en jevn fordeling av menn og kvinner i landets yrker, må yrkene være fordelt med samme fordeling som den totale fordelingen av menn og kvinner i arbeid i det landet. Dersom en hadde tatt utgangspunkt i en 50/50 fordeling i alle yrker (og ikke landets totale fordeling menn og kvinner i arbeid), vil det bli en skjevfordeling i land som har ulik andel menn og kvinner i arbeid. I det tilfellet ville også målet innebære et element av kjønnskjev arbeidsdeltakelse, og målet ville ikke blitt like presist.

Det finnes flere mål på (u)jevnhhet. Eksempler er Gini-koeffisienten, entropi og Atkinson indeksen, men en kan si at den mest benyttede er Duncans Dissimilarity Index (Iceland, Weinberg og Steinmetz, 2002:119). Indeksene regnes ut på ulikt vis. Eksempelvis viser Duncans D (som går fra 0 til 1), andel av befolkningen som må endre jobb for at fordelingen av menn og kvinner skal være jevn i alle yrkesgrupper.

Felles for indeksene er at hele populasjonen deler en verdi, som indikerer hvor segregert landet eller gruppen som helhet er (ibid.). Det medfører problemer når en ønsker å benytte den som avhengig variabel i en regresjon, samtidig som man vil kontrollere for individuelle faktorer. Selv om deskriptiv statistikk ved bruk av indekser og sammenligninger kan gi nyttig informasjon, gir det ikke mulighet til samtidig å kontrollere for andre faktorer. Hvis en ønsker å finne ut hva som påvirker kjønnsdeling eller segregering, vil jeg argumentere for at det er en fordel å kunne kontrollere for både individuelle og kontekstuelle faktorer. En sammenheng mellom to variabler på gruppenivå gir kun innsikt i fenomener på det samme nivået. En korrelasjon mellom kjønnsdeling og likestilling på landsnivå for eksempel, gir ikke informasjon om en sammenheng på individnivå. Det ville vært en økologisk feilslutning (Robinson, 2009; Snijders og Bosker, 2012:15) Derfor er det behov for et segregeringsmål på individnivå, slik at målet kan brukes i regresjonsanalyser. I en regresjonsanalyse har man mulighet til å kontrollere for eventuelle individuelle faktorer som kan være med å påvirke kjønnsdelingen. Den informasjonen får en ikke ved å se på sammenhenger på makronivå.

3.4.2 Konstruksjon av avhengig variabel

Behovet ble oppfylt ved å se på fordelingen av menn og kvinner i de ulike yrkesgruppene, sammenlignet med fordelingen av menn og kvinner i arbeid totalt i utvalget. Prosentandelen kvinner i hver yrkeskategori ble trukket fra den totale andelen kvinner i utvalget, og tilsvarende for menn. Deretter ble negative verdier gjort til positive, og så ble variabelen delt på ti og avrundet til to desimaler. Dette ble gjort for alle utvalg og resultatet utgjør den avhengige variabelen. Fordi variabelen er sentral for oppgaven skal jeg gå gjennom hvert steg i detalj, før jeg til slutt viser reliabilitets- og validitetstester av den ferdige variabelen.

Variabelen tar utgangspunkt i et spørsmål alle respondenter har blitt stilt i alle rundene av ISSP. Dette spørsmålet er: «What is/was your occupation – i.e., what is/was the name or title of your main job?» (ISSP Demographic Methods Group, 2012:10). Svarene blir tolket og kategorisert i en lang liste med yrker kalt *International Standard Classification of Occupations* (ISCO88). Listen er standardisert, og er den samme for alle deltakende land. Den er delt opp i ti hovedkategorier (major groups), 28 «subhovedgrupper» (sub-major groups), 116 subgrupper, og 390 enhetsgrupper (unit groups). Til mitt formål er den inndelingen for presis, fordi det fører til at det er for få respondenter i hver gruppe, noe som ville resulterte i for mange ekstremverdier. Det var derfor nødvendig å samle liknende yrker i yrkesgrupper. Det ble gjort ved å til dels følge ISCO-indeksens egen inndeling. Målet var å beholde en presisjon der yrkene i hver gruppe har store likhetstrekk, samtidig som gruppene er store nok til at det ikke blir for mange ekstremverdier grunnet lavt antall individer per gruppe. Resultatet var en yrkesinndeling med 48 yrkesgrupper (se vedlegg 1 for syntaks).

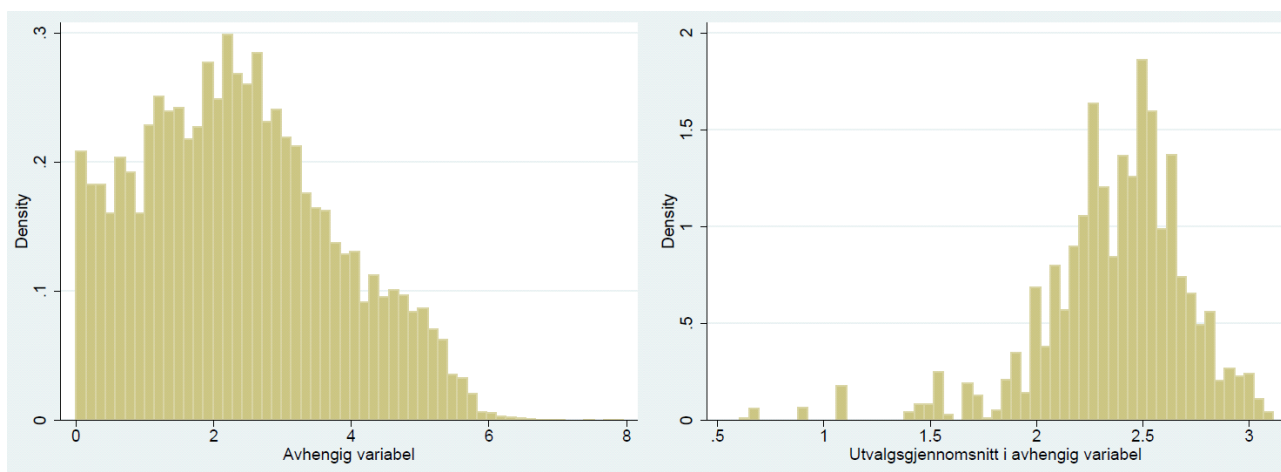
La oss ta et eksempel. I utvalget fra Norge i 2007 er det totalt 1143 antall respondenter. Av disse har 1045 svart på spørsmålet om yrkestype. Det utgjør mitt «subutvalg» for Norge i 2007. I dette «subutvalget» er 54,26% kvinner, og 45,74% menn. Dersom yrkesgruppene skal være jevnt fordelt, må yrkesgruppene ha den samme fordelingen som den totale andelen menn og kvinner i «subutvalget». For hvert prosentpoengs forskjell mellom fordelingen i yrkesgruppen og fordelingen i «subutvalget», økes verdien i avhengig variabel med én. Hvis vi går tilbake til eksempelet vårt, er det i yrkesgruppen salg og service totalt 51 personer hvorav 68,63% er kvinner, og 31,37% er menn. Utregningen blir da $68,63 - 54,26 = 14,37$ for kvinner og $31,37 - 45,74 = -14,37$ for menn. Negative verdier blir omgjort til positive, og delt på ti, dermed får både menn og kvinner i yrkesgruppen salg og service en verdi på 1,44. Tallet indikerer dermed hvor kjønnsdelt yrkesgruppen er, uavhengig av hvilket kjønn det er overvekt av.

Helt til slutt ble jeg nødt til å dele verdien på ti. Det var ingen annen grunn til dette enn at avhengig variabel ble for stor å bruke i flernivåanalyser med så mange observasjoner som jeg har i mitt datamateriale. Jeg klarte ikke å få modellene til å konvergere i Stata. Jeg delte derfor verdiene på 10 og avrundet til to desimaler. Formelen for avhengig variabel hos hvert individ er dermed:

$$y_i = \frac{\sqrt{(p_k - P_k)^2}}{10}$$

Der p_k = kjønns prosentandel i yrkesgruppen, og P_k = kjønns prosentandel totalt i «subutvalget». Kvadratrotten og kvadreringens rolle er å gjøre om negative verdier til positive. Hvis en ikke gjør dette, ville variabelen ikke blitt kjønnsnøytral. Den minste mulige verdien er null, og teoretisk maks er tilnærmet lik 10. Syntaksen for konstruksjonen av avhengig variabel er vist i vedlegg på side 73.

Tanken er at dersom et lands arbeidsmarked generelt har fordelinger i yrkesgruppene som ligner landets totale andel menn og kvinner i arbeid, vil landet ha et lavt gjennomsnitt i avhengig variabel. Land som har høy grad av kjønnsdeling vil dermed i gjennomsnitt få høyere verdier. Eksempelvis har Georgia i år 2013 et gjennomsnitt på 0,68 med et standardavvik på 0,73. Til sammenligning har Norge i år 2000 et gjennomsnitt på 2,71 og et standardavvik på 1,35. Totalt er gjennomsnittet 2,39 med standardavvik på 1,4, og et histogram over variabelen er vist til venstre i figur 1. Den avhengige variabelen har 419786 gyldige verdier; de som for øyeblikket er i, eller tidligere har hatt arbeid.



Figur 1: T.V.: Histogram over avhengig variabel på individnivå. T.H.: Histogram over avhengig variabel på utvalgsnivå.

3.4.3 Validitets- og reliabilitetstesting av avhengig variabel

For å kontrollere om målet på kjønnsdeling er valid, tok jeg utgangspunkt i det mest benyttede målet på kjønnsdeling, Duncans D. Duncans D ble utregnet med stata-utvidelsen *Duncan* for de samme yrkesgruppene som ble benyttet for å konstruere avhengig variabel (Jann, 2005). Som sagt er Duncans D en variabel på gruppenivå. For å kunne kontrollere hvorvidt mitt mål på segregering måler noe av det samme, måtte det i hvert utvalg lages gjennomsnittlig verdi for avhengig variabel. Det ble derfor laget en ny variabel med gjennomsnittet av y-variabelen i hvert utvalg delt på 10 (se figur 1 for fordeling), og gjort en reliabilitetstest og en korrelasjonstest mellom Duncans D og den nye variabelen². Korrelasjonen var $r=0,9735$, og cronbachs alpha kom på $\alpha=0,89$, noe jeg anser som sterkt, spesielt når kun to variabler er med i utregningen.

Neste punkt var å kontrollere test-retest reliabilitet for den avhengige variabelen. Det ble gjort ved å sjekke sammenhengen mellom verdien på avhengig variabel de ulike tidspunktene landene har blitt målt på. Hvis et land har omtrent samme gjennomsnittlig verdi ett år som et annet, vil jeg anta at variabelen er reliabel på gruppenivå (test-retest reliabilitet). Fordi ikke alle land har verdier alle årene, er det ikke mulig å beregne cronbach's alpha for alle årene samtidig. Derfor ble først en indeks regnet ut for første halvparten av årene, og en for siste halvdel. Deretter ble alpha regnet ut på bakgrunn av de to beregnede indeksene ($\alpha=0.7924$). Verdien er dermed over tommelfingerregelens kritiske verdi på 0,7, og jeg kan derfor konkludere med at avhengig variabel er reliabel på gruppenivå.

Disse to kontrollteknikkene måler imidlertid kun presisjonen på gruppenivå. Det var derfor relevant å se om reliabiliteten også forholdt seg stabil på individnivå. Med det mener jeg hvorvidt det er de samme yrkesgruppene innenfor hvert land som er kjønnsdelte fra år til år. For å gjøre det utnyttet jeg at Statas *generate*- og *replace*-kommandoer bruker gjeldende sorteringsrekkefølge (Newson, 2001). Datamaterialet ble sortert etter utvalg og yrkesgrupper for å lage en identifikasjonsvariabel som igjen ble brukt til å sortere materialet slik at annenhver enhet i materialet var fra ulike utvalg, men samme land og yrkesgruppe. På den måten var det mulig å gi verdier til neste respondent basert på forrige respondents verdi. Teknikken er ikke perfekt, fordi det er ulikt antall personer i hver yrkesgruppe i hvert utvalg, og at landene ikke har deltatt i ISSP like mange ganger. En kan derfor ikke være sikker på at alle har blitt gitt korrekte verdier. Kommandoen ble kun kjørt to ganger, for å være sikker på at det ikke ble for mange skjevheter. Deretter ble cronbach's alpha regnet ut på samme måte og med samme argument som i forrige avsnitt. En indeks ble regnet ut for første halvpart av årene og en for siste halvdel av årene. Reliabilitetskoeffisienten

2 Å igjen dele på 10 ble gjort for å tilpasse variabelen til Duncans variasjonsbredde som er mellom 0 og 1.

mellom disse to var $\alpha=0,8442$. Det betyr at avhengig variabel er reliabel også på individnivå, fordi de ulike yrkesgruppene i stor grad er kjønnsdelt på likt vis innenfor landene som har deltatt flere ganger.

Fra et mer teoretisk standpunkt kan også avhengig variabel problematiseres. ISCO-indeksen er oppbygd ved at sjefer og direktører er plassert i egne yrkesgrupper, og ikke innenfor sektoren eller yrkestype de jobber innenfor. Det kan være et problem ved at man også måler vertikal kjønnsdeling. Problemet er imidlertid vanskelig å jobbe rundt, fordi en ikke har tilgang på informasjon om hvilke yrkesgrupper sjefer og direktører «egentlig» tilhører. En løsning kunne vært å fjerne sjefer og direktører, men da har man ett nytt problem ved at man skaper en skjevhet i datamaterialet. En kan se for seg at man fra en ideell fordeling (der direktører og sjefer er plassert i «riktig» sektor), har plukket noen individer (toppsjefer o.l.) ut av utvalget og skapt en systematisk skjevhet. Når man først er i gang med dette, kan en stille seg spørsmålet om hvor en skal sette grensen. Hvis en fjerner «eliten» fra utvalget, kunne man kanskje jevnet det ut igjen med å ta vekk de som er lavest på rangstigen. I det tilfellet har jeg ikke lenger et mål på hele spekteret av yrker, noe jeg tidligere har argumentert for at er metodens styrke. Jeg har derfor vurdert at det er bedre å ikke ta dem ut. Dette er imidlertid et problem alle mål på horisontal kjønnsdeling i arbeidslivet som baserer seg på ISCO-indeksen har. Det er derfor et problem med ISCO-indeksen, og ikke segregeringsmålet i seg selv. Av 48 yrkesgrupper er det tre av gruppene som består av sjefer og direktører. Selv om det kan påvirke resultatet, er majoriteten av respondentene fordelt i de andre kategoriene. Det vil si at jeg fortsatt har et validt nok mål på horisontal kjønnsdeling, men med et element av vertikal kjønnsdeling.

I analysene har jeg forsøkt å ta høyde for dette ved å kontrollere for de som er i yrkesgruppen med toppsjefer og direktører med en egen dikotom variabel. Dette bør utjevne skjevheten som skapes av de tre yrkesgruppene. Den variabelen er lik 1 for de individene i de tre involverte yrkeskategoriene. Resten har verdien 0. Variabelen har 8,76% i kategorien for den vertikale kjønnsdelingskontrollen.

På gruppenivå er variabelen greit normalfordelt, og sammen med reliabilitets- og validitetskontrollene, peker det på at det ikke er tilfeldig hvilke verdier et land i gjennomsnitt har i variabelen. I analysen viser jeg hvordan variabelen fungerer i praksis.

3.5 Sammenslåing av datasett og konstruksjon av variabler på andre nivå

Etter at alle uavhengige variabler på individnivå var kodet til samme standard, var det mulig å slå sammen de 15 datasettene til ett stort samlet med i overkant av 700000 observasjoner. I denne delen

av metodekapittelet skal jeg gå gjennom sletting av doble observasjoner, for små utvalg og konstruksjon av variabler på andre nivå.

I ISSP er det vanlig at deltakerland deltar i to runder av ISSP med det samme utvalget. Det vil si at de nasjonale undersøkelsene har med spørsmål fra flere runder av ISSP i en og samme undersøkelse. Det har medført at noen utvalg har kommet med i det kombinerte datasettet flere ganger. For å løse dette måtte disse utvalgene slettes. Følgende variabler var involvert: Årstallet undersøkelsen ble gjort, med en variasjonsbredde fra 1998 – 2015. Issprunden datamaterialet er hentet fra med en variasjonsbredde fra 1999-2013. Og variabelen som viser til landet respondentene er fra. Basert på disse ble det laget to gruppevariabler, landår og landrunde. De grupperer respondentene fra samme utvalg i samme verdi. Basert på disse utvalgsvariablene ble det laget to nye variabler som tilsvarer antall respondenter i gruppen. Hver observasjon hadde derfor to slike identifikasjonsvariabler. Syntaxen brukte følgende logikk: dersom landårets gruppe var større enn landrundens, slett observasjonene hvis verdien for issprunden ikke er det samme som årstallet undersøkelsen ble utført. På den måten kunne de doble observasjonene enkelt lukes ut og slettes. Det fører til noen falske positive, men ingen falske negative, ved at noen deltakerland kan ha utført undersøkelsen, og vært med i flere runder av ISSP med det samme utvalget, uten at årstallet er det samme som issprunden i noen av tilfellene. Det ble gjennomgått manuelt, og sørget for at ingen som ikke var doble ble tatt ut. Antall observasjoner gikk dermed ned fra 700321 til 575114. Syntaksen er vist i vedlegg på side 75.

I enkelte utvalg var det veldig små gjenstående utvalg i analysene. Det skjedde av to grunner. Den ene grunnen var at med mange variabler ble det mange missingverdier. Enkelte utvalg ble til slutt så små som tre respondenter. Tre respondenter kan ikke være representativt for et helt land. Derfor ble de små utvalgene med under 100 respondenter tatt ut. Den andre grunnen var at enkelte utvalg kun hadde kategorisert respondentene i få yrkeskategorier. Eksempelvis var det i Sør-Afrika i 2003 kun respondenter som ble kategorisert enten som ekspeditører eller maskinoperatører. Det er sannsynligvis ikke representativt. Løsningen var å slette de utvalgene som hadde respondenter kategorisert i færre enn 20 yrkesgrupper. Antall observasjoner gikk ned til 286587 fordelt på 330 utvalg i 49 land.

Et utvalg på 49 land kan høres mye ut, men statistisk sett er ikke det helt tilfredsstillende. I flernivåanalyser må landene må ses som et utvalg i seg selv, fra populasjonen verdens land. Det kan føre til misvisende standardfeil og upresise estimater av variablene på andrenivå (Bryan og Jenkins, 2015). De fleste undersøkelser av hvor mange enheter som er nødvendig på øverste nivå kommer fram til en plass mellom 20 og 50 (Bryan og Jenkins, 2015; Maas og Hox, 2005). Dette er grunnen

til at jeg har gått gjennom arbeidet med å slå sammen 15 datasett av ISSP. Selv om antall enheter på nivå tre ikke er perfekt, er datamaterialet jeg har satt sammen så nært optimalt som mulig med kryssnasjonale surveydata.

For å kunne utnytte flernivåanalysenes styrke til å kontrollere for kontekstuelle variabler, har jeg tatt i bruk variabler på gruppenivå. Det vil si variabler der respondentene i samme utvalg deler verdi. Indeksene jeg har valgt å ta i bruk er: gini index for økonomisk ulikhet, en indeks for likestilling, en indeks for andelen av personer i arbeidsmarkedet som er kvinner, en tilsvarende for offentlig sektor og indekser for delt omsorgs- og delt forsørger-dimensjonen i likestillingspolitikk.

3.5.1 Variabler på gruppenivå

Gini-indeksen måler graden inntektsfordelingen blant individer eller husholdninger i en økonomi avviker fra en perfekt lik fordeling, og er en indeks basert på data hentet fra The World Bank (1999-2014) og The World Factbook utarbeidet av CIA (Central Intelligence Agency, 2013). En verdi på 0 indikerer perfekt likhet og 100 indikerer perfekt ulikhet. Indeksen fra The World Bank hadde ikke verdier for New Zealand, Taiwan og Sør Korea. Tallene fra The World Factbook er imidlertid begrenset av at verdiene for gini-indeksen ikke varierer hvert år. Her hadde jeg dermed et metodevalg som jeg nå skal beskrive.

Fordi verdensbankens verdier på gini indeksen manglet verdier for tre land, ville konsekvensen av å kun bruke denne medført at antall enheter på nivå tre i analysene går ned. Med et utvalg på 50 land, er tre land et relativt stort innhugg i antall missing i analysene. For å unngå det og de konsekvensene et lite antall enheter har på i flernivåanalyser, har jeg forsøkt å supplere med verdier fra The World Factbook. Fordi indeksene fra de to kildene ikke er regnet ut på identisk måte, var det imidlertid ikke ønskelig å kun trekke ut verdiene for de tre landene som manglet verdier i verdensbankens indeks, og gi de verdiene til landene med missingverdier. Det var heller ikke ønskelig å kun bruke The World Factbook sin indeks, fordi den ikke varierer fra år til år. Den potensielle forklaringskraften på andre nivå i modellene (årstall innenfor hvert land), ville derfor forsvunnet. For å utnytte styrkene ved begge kildene og begrense negative konsekvenser av hvert av alternativene, har jeg laget en ny skala basert på de to indeksene. De to gini-indeksene hadde en korrelasjon på $r=0,9557$, og jeg mener derfor at metoden er forsvarlig. Verdiene for New Zealand, Taiwan og Sør Korea blir ikke *helt* riktig utregnet relativt til de andre landenes verdier, men det er slik jeg ser det, det alternativet med færrest ulemper. Indeksen for økonomisk ulikhet hadde før sentreringen et gjennomsnitt på 40,5 og et standardavvik på 11,19. Hensikten med å ta med Gini-indeksen var å kontrollere om ikke kjønnsdeling egentlig forklares av inntektsulikhet framfor

likestilling.

Indeksen jeg bruker for å måle generell likestilling heter Global Gender Gap index, og er publisert av World Economic Forum, og harmonisert av Humanitarian Data Exchange (HDX, 2014). Indeksen kan ses på som et stillbilde av hvordan likestillingssituasjonen er i landene på et gitt tidspunkt (World Economic Forum, udatert). Det vil si at målet i grunnen er et mål på resultatligheten i likestilling. Som mål på likestilling tar den utgangspunkt i fire subindekser. Økonomisk deltakelse og muligheter, som tar for seg forskjellen mellom menn og kvinners deltakelse i arbeidsmarkedet, inntektsforskjeller mellom kjønnene, inntektslikhet for likt arbeid og forskjellen mellom kjønnenes representasjon i topplederstillinger. Den andre subindeksen er basert på utdanningsoppnåelse, og er basert på data om forskjeller i menn og kvinners tilgang til utdanning på grunnskole, videregående og høyere nivå, samt forskjellen mellom menn og kvinners lesekompetanse. Den tredje subindeksen handler om kjønnsforskjeller i helse og overlevelse, og er basert på to indikatorer. Den ene er kjønnsforskjellen i fødte barn, og forsøker å fange opp «forsvunnende» døtre i land med en sterk preferanse for sønner. Den andre er gapet mellom kjønnenes forventede levealder. Den siste subindeksen går på politisk deltakelse, og baserer seg på kjønnsforskjellen i ministerposisjoner og i parlamentposisjoner, i tillegg til kjønnsforskjellen i antall år i statsministerposisjonen har vært beholdt av det enkelte kjønn de siste femti årene (ibid.). Poenget med å ta med likestillingsindeksen er å få et generelt mål på likestilling, for å kunne se om likestillingsparadokset er et internasjonalt fenomen. Indeksen strekker seg fra 0 (ikke likestilt) til 1 (likestilt), og hadde et gjennomsnitt på 0,71 før sentreringen, og et standardavvik på 0,05.

Indeksen jeg har brukt for å måle kvinneandelen i offentlig sektor har jeg konstruert basert på andelen kvinner i variabelen for sektor beskrevet tidligere i kapittelet. Innenfor hvert utvalg har jeg satt gjennomsnittet i kjønnsvariabelen ganget med 100, sentrert rundt gjennomsnittet som var 55,3. Variabelen har et standardavvik på 10,5.

Indeksen for kvinneandelen i arbeidsmarkedet har jeg hentet fra The World Bank (udatert), og er basert på tall fra International Labour Organisation. Variabelen indikerer prosentandelen av det aktive arbeidsmarkedet som er kvinner innenfor hvert land hvert år. Den har et gjennomsnitt på 44, med standardavviket 4.

Indeksene for delt omsorgs- og delt forsørger-dimensjonen i likestillingspolitikk har jeg selv konstruert ut fra data hentet fra International Labour Organization (udatert) supplert med data fra UNESCO (United Nations Educational, Scientific and Cultural Organization), eurostat og andre nasjonale kilder, der det manglet data (Andrén og Schmidt 2005; Charles og Williams, 2006; Kamerman, 2006; Ministry of Education, Republic of China, 2012/2013; Myers, 2000; Rostgaard,

2004; Taratukhina, Polyakova, Berezina, Notkina, Sheraizina, Borovkov, 2006; UNESCO, 2010, 2015; UNESCO International Bureau of Education, 2006a, 2006b, 2006c, 2006d). Ideen ble inspirert av Korpi m.fl (2013), og jeg har forsøkt å lage indeksene så nært deres som mulig.

Delt omsorgs-dimensjonen i likestillingspolitikk er en indeks basert på antall uker med foreldrepermisjon som kan fritt fordeles mellom mor og far og antall uker foreldrepermisjon som må benyttes av faren. Delt forsørger-dimensjonen er en indeks basert på tre indikatorer. Den første er barnehagedekningen for små barn (0-2 år) i form av andelen registrerte kontra totalt antall barn i den relevante aldersgruppen. Den andre er andelen registrerte kontra totalt antall barn i barnehage for barn fra tre år fram til skolealder. Den tredje indikatoren er en inntektsrelatert foreldreforsikring, som igjen er en multiplikativ variabel basert på to indikatorer. Det er prosentandelen av opprinnelig inntekt ganget med tiden den blir utbetalt.

Variablene har blitt standardisert, slik at de har et gjennomsnitt på null og varians lik 1. Standardavviket for delt omsorgs-dimensjonen er 0,76 og for delt forsørger-dimensjonen er det 0,7. Norge og Sverige toppe delt omsorgs-dimensjonen, etterfulgt av Tyskland og Tsjekkia. I delt forsørger-dimensjonen er Norge Sverige og Danmark på topp, med Island og Sør Korea hakk i hæl. Målet med indeksen var å fange opp til hvilken grad politikken åpner for en overføring av ansvar for barneomsorg fra familien til offentlig sektor, som muliggjør at mødre kan ha en mer sammenhengende arbeidsforpliktelse. Jeg har ingen måte å måle validiteten på, men det største problemet ligger i at den delte omsorgs-dimensjonen ikke endrer verdier hvert år. Det kan ha medført et reliabilitetsproblem i at informasjonen om foreldrepermisjonene enten kan være utdatert eller for ny. Informasjonen fra International Labour Organization er oppdatert, men har mangler i at dataene ikke viser endringene over tid i foreldrepermisjoner. I delt omsorgs-dimensjonen er det også mange land som deler minsteverdi. Det er fordi land som har 0 uker foreldrepermisjon som kan fordeles fritt og ingen fedrekvote, nødvendigvis vil dele verdi. På tross av potensielle problemer med dimensjonene, ser vi i tabell 2 under, at målenes korrelasjoner med kjønnsdeling på gruppenivå er mer eller mindre på samme styrkenivå som resten av de kontekstuelle variablene.

De to indeksene har en korrelasjon med hverandre på $r=0,57$. En fullstendig oversikt over korrelasjoner med de kontekstuelle variablene og avhengig variabel både på individ- og gruppenivå er vist i tabell 2. Korrelasjonene er vektet for å ta høyde for at noen land har deltatt flere ganger enn andre.

Tabell 2: Korrelasjoner: kontekstuelle variabler og avhengig variabel på individ- og gruppenivå.

Mål	1	2	3	4	5	6	7	8
1. Kvinneandel								
Arbeidsstyrken	1,00							
2. Kvinneandel								
Offentlig sektor	0,71	1,00						
3. GINI indeks								
inntektsulikhet	-0,36	-0,40	1,00					
4. Global								
Gender Gap	0,44	0,39	-0,32	1,00				
5. Delt omsorg								
Dimensjon	0,36	0,44	-0,46	0,23	1,00			
6. Delt forsørger								
Dimensjon	0,55	0,59	-0,70	0,36	0,57	1,00		
7. Kjønnssdeling								
Gruppernitt	0,37	0,44	-0,33	0,32	0,32	0,43	1,00	
8. Kjønnssdeling								
Individnivå	0,12	0,15	-0,11	0,11	0,10	0,14	0,33	1,00

Indeksene har ikke alltid verdier for alle utvalgene i datasettet jeg har satt sammen fra ISSP. Det medfører at de utvalgene som ikke har verdier, heller ikke kommer med i analysene. Derfor har jeg for alle variabler på gruppenivå brukt gjennomsnitt for tidligere og senere år for å få med flest mulig utvalg i analysene. I tilfeller der jeg ikke har data for tidligere eller senere år, har det blitt brukt den mest nærliggende verdien (nærliggende i form av tid) for det landet. Behandlingen ble gjort i excel, og deretter satt inn i Stata. Alle gruppenivåvariablene var kontinuerlige, og har blitt sentrert rundt gjennomsnittet for å gjøre konstanten mest mulig tolkbar i analysene.

3.6 Analyseteknikk

Jeg har valgt å benytte meg av flernivåanalyse for å undersøke sammenhengen mellom kontekstuelle variabler og kjønnssdeling. Grunnen til at flernivåanalyse er gunstig er av flere årsaker. Det første er at utvalget er gjort i flere steg. Å behandle alle respondentene som en felles gruppe i en ordinær OLS-regresjon, bryter med forutsetningen om at restleddvariasjonene er uavhengige av hverandre (Skog, 2004:236). Det betyr at et individs sjanse for å bli med, ikke skal være påvirket av en annen sin sjanse. Når en respondent fra et land i ISSP blir trukket, økes sannsynligheten for at en annen fra samme land også blir trukket, sammenlignet med respondenter fra land som ikke er med. Med en flernivåanalyse kontrollerer man for strukturen i dataene, ved at hvert utvalg har et eget gjennomsnitt og feilledd. For hvert nivå i flernivåanalysen legges det til et feilledd som tar høyde for den nøstede strukturen i datamaterialet (Snijders, 2005; Snijders og Bosker, 2012:18;28). I mitt tilfelle blir landene den overordnede gruppen, hvor tanken er at en del uforklart variasjon vil ligge. Med det mener jeg at noe av variasjonen i avhengig variabel kan forklares av hvilket land en er fra.

Det andre nivået består av utvalgene innenfor hvert land. Land som har vært med i ISSP flere ganger vil derfor bli separert, og ikke behandlet som det samme utvalget. Den uforklarte

variasjonen på andre nivå vil derfor bestå av variasjon over tid innad i landene. Rent teknisk sett kommer kontekstvariablene til å bli plassert her, fordi landenes verdier på de kontekstuelle variablene varierer over tid. Innenfor hvert utvalg finner vi respondentene, som utgjør første nivå av flernivåanalysen. Her skal jeg legge inn variabler for individuelle faktorer som kan spille inn på kjønnsdelingen. Målet er at jeg skal kunne se om de kontekstuelle variablene har en innvirkning på kjønnsdelingen på individnivå selv etter kontroll for individuelle bakgrunnsvariabler. Med det unngår jeg å begå en økologisk feilslutning.

Det finnes flere typer flernivåanalyser, og jeg har valgt å bruke en flernivåanalyse med random effects, også kalt «mixed» modeller. Andre varianter er fixed effects og between effects. Med fixed effects er regresjonslinjen like «bratt» i hver gruppe, og en kan ikke ha med kontekstuelle variabler der effekten varierer fra gruppe til gruppe (Snijders og Bosker, 2012:43). Man ser altså på variasjon innad i gruppene. I between effects, behandles variablene kun som gjennomsnitt, og man kan derfor kun se på variasjoner mellom grupper. Random effects har elementer av begge varianter, der man kan både se på variasjoner innad i grupper, gruppeforskjeller og hvordan effekten varierer fra gruppe til gruppe. Jeg ønsker å se på hvordan kjønnsdelingen på individnivå blir påvirket av både kontekstuelle og individuelle variabler. Derfor har jeg valgt å benytte random effects i mine analyser. Det finnes flere analyseteknikker som går under random effects kategorien, og mer spesifikt skal jeg benytte meg av en analyseteknikk som heter maximum likelihood (ML). Andre varianter som restricted maximum likelihood (REML) er nærliggende, men er generelt å foretrekke dersom utvalget er lite (Swallow og Monahan, 1984:50). I mitt tilfelle er utvalget svært stort, og skjevheten knyttet til varianskomponentene ved bruk av ML, blir mindre. I tillegg muliggjør ML og ikke REML å bruke en deviancetest³ for å vurdere og sammenligne både delen av modellen som er random, og fixed. Jeg har derfor valgt å bruke ML-metoden.

I en av tabellene med modeller i analysen har jeg også lagt inn en *random slope* i modellen. Det vil si at jeg tillater effekten av en variabel å variere fra utvalg til utvalg (Snijders og Bosker, 2012:77). I mitt tilfelle har jeg tillatt effekten av kjønn å variere. Rent teknisk betyr der at man får et ekstra feilledd som beskriver variansen i effekten av kjønnsvariabelen på gruppenivå. Kjønnsvariabelens koeffisient indikerer gjennomsnittlig kjønnsforskjell i kjønnsdeling, og varianskomponenten i den nederste delen av modellen viser til hvor stor variansen er i effekten av kjønn på utvalgsnivå. Det medfører at modellen blir mye mer kompleks. Derfor er det viktig å ha teoretiske argumenter til grunn når man velger å gjøre det. Jeg kommer tilbake til det i analysen.

For å vurdere og sammenligne modeller har jeg valgt å bruke *varienskomponentene*,

3 Også kalt likelihood-ratio test (lrtest).

Deviance, og *BIC*⁴. *Deviance* er et mål som indikerer modellens mangel på tilpasning til dataene (Snijders og Bosker, 2012:97). Den utregnes ved å gange den naturlige logaritmen til likelihood-funksjonen for modellen med -2. *Deviance* kan ikke tolkes alene, og må ses i sammenheng med *deviance*verdier for flere modeller tilpasset det samme datamaterialet. *Deviance* forbedres (reduseres) desto fler variabler som legges til. For hver variabel som legges til⁵, øker antall parametre også med en, og modellen blir mer kompleks. En *deviancetest* er å se om endringen i *deviance* er statistisk signifikant i en *kji-kvadratfordeling* med frihetsgrader lik antall parametre i forskjell mellom de to modellene som testes. Når man har en rekke modeller i samme tabell, er det nyttig å ha *deviance*verdiene for hver modell, slik at alle modellene kan ses i sammenheng, når utvalget er holdt konstant. Det gjør modellvalget ryddigere.

BIC måler forholdet mellom hvor godt modellen er tilpasset dataene, og kompleksitet i modellen. Det straffer kompleksitet i modellene i form av parametre, på en annen måte enn *deviancetesten*, som kun ser på signifikante endringer i *kjikvadratfordelingen*. Desto mindre verdi i *BIC*, desto bedre er forholdet mellom datatilpasning og kompleksitet (Snijders og Bosker, 2012:202). En *flernivåanalyse* er en kompleks estimeringsteknikk, og ofte er enkle modeller bedre enn kompliserte modeller når det kommer til tolkningsmuligheter. Jeg ønsker derfor et mål som er mest mulig streng når det kommer til kompleksitet i modellene⁶. Det fungerer derfor godt som en motpol til *Deviance*, som har en tendens til å bli bedre og bedre med flere forklaringsvariabler.

Den siste måten jeg skal sammenligne hvor gode modellene i analysene er, som skiller seg fra ordinære regresjoner, er å se på *varianskomponentene* på de ulike nivåene. En styrke i *flernivåanalyser* er å kunne se hvor mye av den uforklarte variansen som kan tilskrives de ulike nivåene. Fordi jeg bruker tre nivåer i mine modeller, kan man se hvor mye uforklart *restleddvariasjon* som ligger mellom det totale gjennomsnittet for hele utvalget og gjennomsnittene innad i landene. På andre nivå finner man derfor hvor mye *restleddvariasjon* som befinner seg mellom landets gjennomsnitt og de flere utvalgene (landår) innenfor hvert land. På nivå 1 er den individuelle *variasjonen* som er *restleddvariasjonen* innad i hvert utvalg. Ved å regne på *varianskomponentene* skal jeg finne ut nøyaktig hvor mye av *variasjonen* mellom land som skyldes individuelle faktorer og hvor mye som fanges opp av de kontekstuelle variablene.

4 Bayesian Information Criterion.

5 Det gjelder kontinuerlige variabler. For kategoriske variabler som legges inn som *dummysett*, er antall parametre lik antall kategorier minus 1.

6 Det er også grunnen til at jeg har valgt *BIC* over andre kjente informasjonskriterier som *AIC*. *AIC* har en tendens til å overtilpasse modeller, mens *BIC* har en tendens til å undertilpasse modeller. Fordi jeg har så stort utvalg, er sannsynligheten for at jeg undertilpasser modellene mindre enn at jeg overtilpasser dem. Derfor har jeg valgt å benytte meg av *BIC*. For en diskusjon om informasjonskriterier og modelltilpasning se Dziak, Coffman, Lanza og Li (2012). Artikkelen er dog ikke fagfellevurdert.

4. Analyse

Jeg skal begynne analysen med deskriptiv statistikk over deltakerlandenes gjennomsnittlige verdier i avhengig variabel. Deretter skal jeg begi meg ut i analysene. Analysene kommer først til å se på hvor mye av variasjonen i avhengig variabel som kan forklares av strukturen i datamaterialet. Det vil si hvilket land og hvilket år respondentene har svart. Etter at det er gjort, skal jeg presentere flere sett med flernivåanalyser med et konstant utvalg der hypotesene skal settes på prøve.

4.1 Deskriptiv statistikk

Kjønnsdelingsvariabelen i tabell 3 viser til gjennomsnittlig decilpoengs avstand fra yrkesgruppens fordeling av kjønnene i arbeidsmarkedet til utvalgets totale fordeling menn og kvinner i arbeid. Jeg har sortert de nordiske landene først.

Tabell 3: Oversikt over kjønnsdeling i deltakerlandene.

Land	Gj. snitt kjønnsdeling	Land	Gj. snitt kjønnsdeling
Denmark	2,40	Israel	2,32
Finland	2,71	Italy	2,09
Iceland	2,42	Japan	2,40
Norway	2,58	Latvia	2,91
Sweden	2,50	Lithuania	2,69
Total Norden	2,56	Mexico	2,05
Argentina	2,53	Netherlands	2,37
Australia	2,36	New Zealand	2,49
Austria	2,53	Philippines	2,40
Belgium	2,53	Poland	2,38
Brazil	2,13	Portugal	2,33
Bulgaria	2,49	Russia	2,50
Canada	2,09	Slovak Republic	2,62
Chile	2,48	Slovenia	2,48
China	1,47	South Africa	2,15
Croatia	2,27	South Korea	2,08
Cyprus	2,49	Spain	2,27
Czech Republic	2,76	Switzerland	2,47
Dominican Republic	2,36	Taiwan	2,04
Estonia	2,25	Turkey	1,47
France	2,27	Ukraine	2,40
Georgia	0,69	United Kingdom	2,48
Germany	2,65	United States	2,10
Hungary	2,52	Uruguay	2,49
India	0,77	Venezuela	1,41
Ireland	2,62	Total uten Norden	2,34

Norges gjennomsnittlige verdi på 2,58 betyr at den gjennomsnittlige avstanden mellom yrkesgruppene i Norge og den totale andelen menn og kvinner i arbeid er i underkant av 26

prosentpoeng. Det medfører at Norges kjønnsdeling plasserer seg midt på treet, sett i forhold til resten av Norden, som har en gjennomsnittlig verdi på 2,56. Sammenligningsvis er resten av verdens gjennomsnittlige verdi på 2,34. Det ser derfor ut som at Norge og Norden har en større grad av kjønnsdeling enn resten av de deltagende landene.

4.2 Kontroll for datastruktur

Jeg skal fortsette analysen med å vise datastrukturen, og gjøre en kontroll for den. Tanken er at hvis strukturen i dataene kan forklare noe av variasjonen i kjønnsdelingen, er det et tegn på at det finnes nasjonale og/eller tidsmessige forskjeller i kjønnsdeling, og en flernivåanalyse er nødvendig for å kontrollere for dette. Datastrukturen kan ses i tabell 4. Den viser at det i modellene er respondenter fra 49 land med. Mange land har deltatt i ISSP flere ganger, og antall utvalg er dermed 330. Det minste utvalget er på 104, og det største er på 3607. Som nevnt i metoddelen har jeg ikke tatt med utvalg på færre enn 100 respondenter. Jeg har også tatt utvalg som har færre enn 20 yrkesgrupper ut av analysene. Jeg har holdt utvalget konstant i alle modellene i analysen, slik at de kan sammenlignes på tvers av tabeller. Selv om det har medført noen flere missingverdier, mener jeg at utvalget er stort nok likevel, som vi ser er 286587 individer totalt. Jeg hadde opprinnelig data fra 50 land, men Taiwan har ikke verdier i likestillingsindeksen, og har derfor falt bort fra alle analyser. Det er ikke ønskelig å miste enheter på nivå tre når jeg bare har 50, men gevinsten med å kunne sammenligne på tvers har jeg ansett som større enn risikoen ved å miste en enhet.

Tabell 4: Oversikt over datastrukturen.

Gruppe	Antall	Obs. per gruppe		
		Min.	Gj. snitt	Maks.
Land	49	363	5848,7	16822
Landår	330	104	868,4	3607
Individer	286587			

I tabell 5 har jeg estimert tre tomme modeller for å se til hvilken grad variasjonen i kjønnsdeling kan forklares av hvor og når respondentene har svart. Modell en viser en samlet ML-regresjon⁷, og modell to viser en flernivåanalyse der tredje nivå er land, og det andre nivå er landår⁸. Modell tre har i tillegg med en designvekt, og poenget med den er å bruke senere i analysen for å kunne lage en pseudo R^2 . Det er fordi resten av modellene i analysekapittelet har vektet for utvalgsskjevheter.

7 En samlet ML regresjon er en regresjon som bruker maximum-likelihood metoden. Fordi flernivåanalyser benytter seg av ML, gjør jeg det samme i modellen med kun ett nivå, for å kunne sammenligne hvor godt modellene er tilpasset dataene. En samlet regresjon kalles ofte på engelsk for «pooled», og betyr at det ikke er en flernivåanalyse.

8 Landår (land + år) er et begrep jeg bruker for å betegne utvalget. Det vil si gruppen respondenter som har besvart samme nasjonale undersøkelse på omtrent samme tidspunkt.

Ved å se på VPC⁹ for nivå to og tre i modell to, finner vi at rundt 11 prosent av variasjonen i kjønnsdeling kan forklares av datastrukturen. Overvekten av den uforklarte variasjonen *over* første nivå, kan forklares av landnivået (9,5%), mens bare litt forklares av landåret (1,6%). Det er verdt å merke seg at disse tallene kan skyldes både komposisjonelle effekter og mangel av vektning. Å ikke vekte modellene vil si at hvert individ innenfor hvert utvalg har like stor innvirkning på utslaget. Det er fordi det ikke mulig å legge inn designvekter i Stata uten å legge inn flere nivåer. Derfor er modell tre en tilsvarende modell, med vektning på individnivå. Komposisjonelle effekter vil si til hvilken grad variasjonen i avhengig variabel som her fanges opp som variasjon mellom land, egentlig skyldes sammensetningen av individene. For eksempel hvis noen utvalg har en unormalt stor andel personer med en viss karakteristikk som å være over en viss alder, eller overvekt av ett kjønn. Uten å kontrollere for individuelle variabler, kan man ikke vite hvorvidt variasjonen som nå er tilskrevet de ulike nivåene stemmer.

Uansett er den uforklarte variasjonen mellom land og mellom utvalg over 5 prosent, som ofte brukes som en tommelfingerregel på når det er nyttig å bruke flernivåanalyse. Hvis under 5 prosent av den totale variasjonen i avhengig variabel kunne forklares av datastrukturen, ville det ikke vært av like forsvarlig å benytte flernivåanalyse, fordi gruppene ikke statistisk kan skilles fra hverandre, og individene kunne like gjerne vært en del av samme gruppe. Her ser vi derimot at variasjonen mellom land er utpreget. I modell tre har VPC på de ulike nivåene endret seg noe. Ved å vekte har enten mer uforklart variasjon forplantet seg opp i datastrukturen, eller så er det de robuste standardfeilene som gjør at det er en forskjell mellom modell to og modell tre. Det skal jeg dog ikke gå mer inn på i denne oppgaven. Poenget er som sagt bare å ha en nullmodell jeg kan sammenligne restleddvariasjon med.

Tabell 5: Tre tomme modeller. Kontroll for datastruktur. N=286587

	Modell 1		Modell 2		Modell 3	
Fixed del	Koeffisient	S.E.	Koeffisient	S.E.	Koeffisient	Robust S.E.
Lederskjevhet	-0,769 *	0,009	-0,770 *	0,009	-0,801 *	0,118
Konstant	2,457	0,003	2,356	0,064	2,377	0,068
Random del	Estimat	S.E.	Estimat	S.E.	Estimat	Robust S.E.
Nivå 3 varians:			0,191	0,043	0,198	0,081
VPC:			0,095		0,096	
Nivå 2 varians:			0,032	0,003	0,032	0,004
VPC:			0,016		0,016	
Nivå 1 varians:	1,895	0,005	1,781	0,005	1,830	0,064
VPC:	1,000		0,889		0,888	
BIC, Deviance	996471	996433,4	979890,1	979827,2	980133,2	980070,36

$P \leq 0,05 = *$

9 VPC står for variation partition coefficient. Også kalt ICC (intraklassekorrelasjon), og indikerer andelen av den uforklarte variasjonen i avhengig variabel som kan tilskrives de ulike nivåene.

Videre ser vi at modell to har mye lavere verdier for BIC og deviance. Ikke bare forteller devianceverdien oss at modellen er mye bedre tilpasset dataene, men også at selv om modellen har to ekstra parametre (gruppenivåene), viser differansen i BIC at modell to er mye bedre egnet til å forklare variasjonen i avhengig variabel. Jeg kan dermed konkludere med at en flernivåanalyse er nødvendig, fordi VPC viser at over fem prosent av variasjonen skyldes datastrukturen, og BIC og deviance viser at flernivåmodellen var bedre tilpasset dataene enn den uten kontroll for datastruktur.

En kunne argumentert for at det ikke er nødvendig å kontrollere for årstall, når den uforklarte variasjonen av årstall er såpass lav ($VPC=0,016$). Å se bort fra årstall medfører dog at jeg ikke har mulighet til å se på variasjon mellom år innad i landene. De fleste kontekstuelle variabler jeg har brukt, har ulike verdier fra år til år. Det er dermed av interesse å se om noe av kjønnsdelingen kan forklares innad i landene, og ikke bare mellom land. Når jeg går videre i analysen blir derfor modellene spesifisert med landår som andre nivå og land som det tredje. Det er imidlertid et godt tegn for reliabiliteten til avhengig variabel sin del at variasjonen på tvers av år innad i landene er såpass stabil.

Kapittelet kommer til å bli sortert som følger. Først skal jeg se på og vise hvordan jeg har arbeidet meg fram til hvilke individuelle variabler jeg skal kontrollere for i resten av modellene. Deretter skal jeg ta for meg to og to variabler på andre nivå i hver sine tabeller med modeller. Det innebærer at jeg i den første tabellen med uavhengige variabler skal se på hvilke individuelle faktorer som er med å forklare kjønnsdeling. Hypotesene H3a, H3b, H3c, H6a, H6b, H6c, H6d, H6e, H6f og H6g skal her testes. Deretter skal jeg sjekke hypotesene 5a, 5b, 5c, 7b og 7c, ved å kontrollere for andelen kvinner i hele arbeidsstyrken og i offentlig sektor. Etter det skal jeg ta for meg generell likestilling målt med global gender gap index og økonomisk ulikhet målt med gini-indeksen. Det betyr at den tredje tabellen med multiple flernivåanalyser tar for seg hypotesene H1a, H1b H2a, H2b og H7a. Deretter skal jeg vise sammenhengen mellom kjønnsdeling og de to dimensjonene i likestillingspolitikk jeg har mål på (delt forsørger-dimensjonen og delt omsorgs-dimensjonen), og skal se på hypotesene H4a, H4b, H7d og H7e. Underveis skal jeg greie ut om, og begrunne hvorfor modellene er satt opp slik de er gjort, og om resonnementene for valg av modell. Til slutt skal jeg ta trekke sammen variabler som består testene om å legge til meningsfull forklaringskraft, og plassere dem i samme modell for å sjekke om de kontekstuelle variablene som har vist seg å gi forklaringskraft, egentlig måler det samme.

4.3 Valg av modell i tabell 6

Tabell 6 viser fem estimerte modeller med robuste standardfeil der modell en inneholder alle individuelle variabler som har blitt tatt med fra de 15 datasettene. Som nevnt tidligere, var det noen utvalgskjevheter som har ført til at enkelte grupper mennesker innenfor hvert utvalg har hatt større sjanse for å bli med i undersøkelsen enn andre. Derfor har jeg lagt inn en vekt på individnivå, slik at noen grupperinger ikke spiller større inn på resultatene enn andre. Vekting av modellene har medført at jeg må bruke robuste standardfeil.

Den uforklarte variansen i modell en, sammenlignet med modell tre i den tomme modellen (tabell 5), kan brukes for å forstå hvor mye alle de individuelle variablene forklarer. Den første modellen i tabell 6 forklarer 4,4 prosent av variasjonen i kjønnsdeling. Vi regner ut vårt Pseudo R²-mål slik: $((0,198+0,032+1,83)-(0,208+0,032+1,729))/(0,198+0,032+1,83)=0,44$.

I den første modellen var det noen variabler som ikke fikk et signifikant resultat. Av disse var det noen variabler jeg ikke hadde hypoteser på, som jeg har tatt ut i de neste modellene. Å ha lederoppgaver så ikke ut til å ha en sammenheng med kjønnsdeling. Det kan muligens forklares av at variabelen jeg har brukt for å kontrollere for ISCO-indeksens hierarkiske struktur har såpass lavt signifikansnivå og høy koeffisient. Interessant nok har respondentene i topplederstillinger en lavere verdi i kjønnsdeling enn resten av yrkesgruppene. Sagt på en annen måte er deltakerlandene mer horisontalt kjønnsdelt enn vertikalt. Verken alder eller bosted hadde en signifikant sammenheng med kjønnsdeling.

Jeg ønsket å gjøre modellene så lite komplekse som mulig, samtidig som jeg ville ha med alle variabler som er med på å forklare kjønnsdelingen. Jeg har hypoteser på antall i husholdningen, sektor og arbeidstid, og derfor har jeg tatt med de variablene videre, selv om de i første modell ikke var signifikante. Tanken var at dersom effekten ikke kom fram i eksempelvis sektorvariabelen, kanskje fordi menn og kvinner generelt har ulike verdier, vil det å ikke kontrollere for et samspill i beste fall føre til heteroskedastisitet, og i verste fall signifikansverdier over 0,05. Det er derfor nødvendig å ta med interaksjonene, så jeg ikke feilaktig forkaster en hypotese.

I den andre modellen i tabell 6 har jeg tatt ut alle variablene som ikke var signifikante. Det fungerer derfor som en grunnmodell jeg har jobbet ut ifra for å se om tillegg av nye variabler gir en bedre modell. Devianceverdien og BIC økes, og indikerer en dårligere modell. Jeg tolker det som at noen av variablene er relevante, men modell en var ikke spesifisert riktig, når det med utgangspunkt i hypotesene trolig er kjønnsforskjeller i effekten av sektor, arbeidstid, samboerskap og antall i husholdningen. Derfor har jeg i modell tre innført samspill mellom sektor og kjønn, arbeidstid og

kjønn, og et treveis samspill mellom antall i husholdningen, samboerskap og kjønn. Alle samspill var signifikante i en Wald-test (*testparm* kommando i Stata). BIC ble redusert med 1385,4 etter at interaksjonene ble lagt til, og indikerer en mye bedre tilpasning til dataene, selv med den økte kompleksiteten det medfører i modellen.

Dummysettet for personlig inntekt er ikke lenger signifikant i modell tre. Jeg tolker det som at samspillene har tatt forklaringskraften fra variabelen for personlig inntekt fanget opp i modell en og to. I modell fire er det derfor tatt ut. Der ser vi at BIC igjen reduseres, mens Deviance øker. BIC straffer som sagt kompleksitet i modellene, og dummysettet for personlig inntekt gir modellen tre ekstra parametre. Deviance øker, og det er ikke rart, fordi deviance bestandig reduseres når flere variabler legges til. Modell fire er derfor å foretrekke over modell tre, fordi å kontrollere for personlig inntekt ikke gjør modellen bedre. Variabelen for antall i husholdningen er ikke signifikant i verken modell tre eller fire. Vi ser dog at koeffisienten har økt betraktelig fra modell en (-0,002) til modell tre og fire (0,012). Det tyder på at noe har skjedd bak kulissene med hvordan denne variabelen har vært med å påvirke resultatene. I modell fem har jeg forsøkt å ta ut variabelen for antall i husholdningen og samspillet mellom det, samboer og kjønn. Her ser vi at både BIC og deviance peker på en tydelig forverring av modellen, med en økning i BIC på 75, og 125,2 deviance.

Samtidig ser vi at samspillet ikke gir mye tolkningsverdi til min oppgave, fordi variabelen antall i husholdningen ikke er signifikant. Det er bare for personer uten samboer at samspillet viser en endring i kjønnsdeling. Jeg har valgt å beholde variabelen og samspillet, selv om det ikke er signifikant av to grunner. Det ene er at selv om jeg ikke ønsker å stole blindt på BIC og deviance som mål på modelltilpasning, forteller verdiene at noe forklares, som jeg ikke kan se direkte i tabellen. Det andre er at med utgangspunkt i hypotesene 7a til og med 7e, ønsker jeg å sjekke hvorvidt de kontekstuelle variablene er signifikante før og etter kontroll for arbeidstid, sektor, samboerskap og antall personer i husholdningen. Hvis jeg ikke tar med samspillet, har jeg ikke lenger mulighet til å teste de hypotesene. Derfor er modell fire i tabell 6 den endelige modellen, som jeg skal bruke for å teste hypotesene som går på individnivå. Modellen skal også brukes som utgangspunkt for å teste hypotesene som går på kontekstuelle faktorer i resten av analysen. Med det mener jeg at modell fire er best egnet til å kontrollere for individuelle variabler, og skal derfor brukes for å kunne kontrollere for individuelle variabler samtidig som de kontekstuelle variablene legges til.

4.3.1 Hypotesetesting for hypoteser på individnivå

I denne seksjonen skal jeg tolke den endelige modellen i tabell 6, og gjøre resultatene forståelige for leseren. Underveis skal jeg forkaste og beholde hypoteser. Det er en kompleks modell selv nå før tillegg av kontekstuelle variabler, og jeg mener det er viktig å ta med leseren på hva resultatene innebærer. Jeg skal gå gjennom verdiene for alle variablene, og uttrykke dem skriftlig.

Jeg begynner øverst. Fordi det er mange samspill i modellen, der kjønn brukes i samtlige, kan ikke variabelen tolkes uavhengig av resten av samspillsleddene. Kjønn har en statistisk signifikant sammenheng med kjønnsdeling. Koeffisienten indikerer en 0,657 decilpoengs gjennomsnittlig større kjønnsdeling for menn enn kvinner, når menn har verdien 0 i alle samspill. Uansett kan vi si at menn jobber jevnt over i mer kjønnsdelte yrker enn kvinner. Variabelen for samboerskap er også med i et samspill. Det medfører at variabelens koeffisient alene indikerer kvinner sin økning i kjønnsdeling dersom man har samboer. Kvinner som har samboer arbeider dermed i yrker som er litt mer kjønnsdelt enn kvinner som ikke har samboer. Til en dikotom variabel å være, er ikke effekten spesielt stor med 0,5 prosentpoeng større avstand fra utvalgets totale kjønnsfordeling i arbeidsstyrken.

Samspillet mellom kjønn, samboerskap og antall i husholdningen er bare til dels tolkbart, selv om tillegget av samspillet som sådan er signifikant i en Wald-test (*testparm* kommando i Stata). Menn som ikke har samboer, har i snitt rundt 0,041 decilpoengs større avstand fra landets totale fordeling av kjønnene i arbeidslivet, for hvert ekstra husholdningsmedlem. Det er vanskelig å si hva en kan tolke ut av dette, annet enn at alenefedre eller menn som bor med sine foreldre og søsken, men ikke har samboer arbeider i mindre kjønnsdelte yrkesgrupper desto flere som bor i samme husholdning. For verken kvinner eller de som er samboere, er ikke resultatene signifikante. Det betyr at det ikke er en statistisk sammenheng mellom personer som har samboere sin kjønnsdeling og antall i husholdningen. Det medfører at hypotese 6e ikke får støtte i mitt datamateriale.

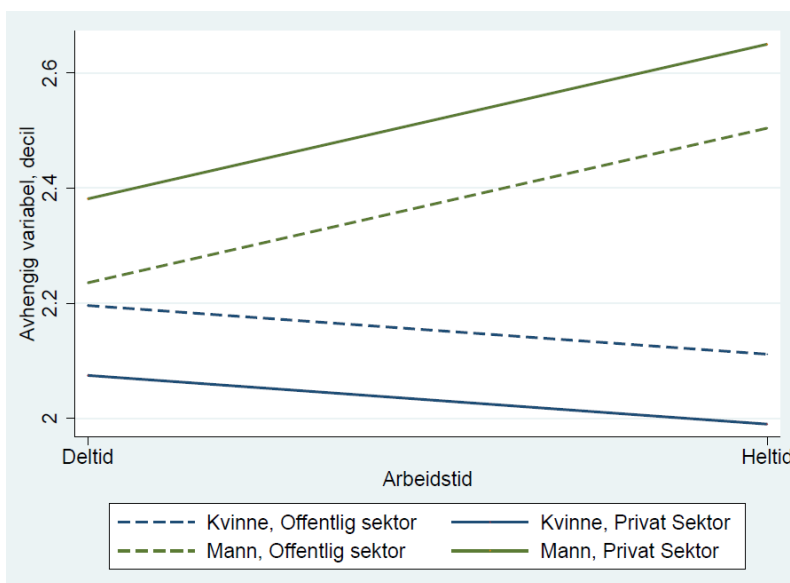
Sektor fikk en signifikant sammenheng med kjønnsdeling etter at samspillet med kjønn ble lagt til, og må tolkes i lys av det. Kvinner i offentlig sektor arbeider i gjennomsnitt i yrkesgrupper med 0,121 decilpoeng større avstand fra utvalgets totale fordeling av kjønnene i arbeidsstyrken, sammenlignet med privat sektor. For menn i offentlig sektor, blir kjønnsdelingen derimot redusert, og det relativt mye, med en reduksjon på 0,267 decilpoeng, sammenlignet med privat sektor. Hypotese 6c og 6d får derfor støtte i modellen, og beholdes.

Tabell 6: Flernivåanalyser med kontroll for variabler på individnivå. N=286587

Fixed del	Modell 1		Modell 2		Modell 3		Modell 4		Modell 5	
	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.
Mann	0,550 *	0,038	0,551 *	0,038	0,714 *	0,044	0,709 *	0,044	0,657 *	0,042
Har samboer (dikotom)	0,090 *	0,011	0,081 *	0,012	0,045 *	0,015	0,046 *	0,015	0,075 *	0,012
Antall i husholdningen	-0,002	0,003			0,012	0,008	0,012	0,008		
Sektor (ref. Privat sektor)										
Arbeidsledig/0 timer i uken	-0,021	0,029			0,046	0,040	0,045	0,039	0,044	0,039
Offentlig sektor	0,002	0,028			0,120 *	0,041	0,121 *	0,040	0,118 *	0,040
Arbeidstid (ref. Heltid <37 timer)										
Arbeidsledig/0 timer i uken	-0,027	0,029			-0,057 *	0,027	-0,062 *	0,026	-0,064 *	0,025
Deltid (≥37 timer)	0,018	0,025			0,084 *	0,026	0,085 *	0,027	0,084 *	0,027
Samspill: Samboerskap, kjønn og antall i husholdningen (ref. Har ikke samboer, kvinne)										
Har ikke samboer*Mann*Antall i husholdningen					-0,041 *	0,016	-0,041 *	0,016		
Har samboer*Kvinne*Antall i husholdningen					-0,005	0,005	-0,005	0,005		
Har samboer*Mann*Antall i husholdningen					-0,013	0,017	-0,013	0,016		
Samspill: Kjønn og sektor (ref. Kvinne, Privat sektor)										
Mann*Arbeidsledig/0 timer i uken					-0,145 *	0,057	-0,145 *	0,057	-0,143 *	0,057
Mann*Offentlig sektor					-0,266 *	0,051	-0,267 *	0,051	-0,259 *	0,052
Samspill: Kjønn og arbeidstid (ref. Kvinne, Heltid)										
Mann*Arbeidsledig/0 timer i uken					0,078	0,047	0,082	0,047	0,083	0,047
Mann*Deltid					-0,355 *	0,045	-0,353 *	0,045	-0,358 *	0,046
Utdanningsgrad (ref. Ingen/laveste formelle utdanning)										
Over laveste formelle	0,135 *	0,043	0,146 *	0,041	0,143 *	0,041	0,144 *	0,041	0,142 *	0,041
Sekundær/ begynt høyere	-0,005	0,047	0,010	0,044	0,008	0,044	0,006	0,045	0,005	0,045
Universitetsud +	-0,278 *	0,065	-0,265 *	0,067	-0,268 *	0,064	-0,274 *	0,065	-0,273 *	0,066
Personlig inntekt (ref. Første kvartil)										
Andre kvartil	0,033 *	0,015	0,034 *	0,015	0,015	0,014				
Tredje kvartil	0,040	0,025	0,048 *	0,022	0,020	0,025				
Fjerde kvartil	0,009	0,032	0,016	0,030	-0,018	0,034				
Husholdningsinntekt (ref. Første kvartil)										
Andre kvartil	0,048 *	0,017	0,051 *	0,017	0,056 *	0,017	0,060 *	0,019	0,057 *	0,019
Tredje kvartil	0,036	0,021	0,042 *	0,019	0,050 *	0,021	0,048 *	0,024	0,043	0,023
Fjerde kvartil	-0,028	0,026	-0,022	0,023	-0,009	0,027	-0,020	0,033	-0,028	0,031
Lederoppgaver (dikotom)	-0,029	0,027								
Medlem av fagforening (dikotom)	0,077 *	0,022	0,078 *	0,025	0,074 *	0,022	0,074 *	0,022	0,076 *	0,022
Bosted (ref. Urbant)										
Suburbant	0,015	0,013								
Ruralt	-0,023	0,023								
Alder (sentrent)	-0,001	0,001								
Lederskjevhhet	-0,813 *	0,112	-0,828 *	0,113	-0,835 *	0,113	-0,838 *	0,112	-0,834 *	0,112
Konstant	2,028	0,080	2,002	0,085	1,974	0,094	1,984	0,093	1,997	0,092
Random del	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.
Nivå 3 varians:	0,208	0,082	0,209	0,083	0,202	0,080	0,202	0,080	0,205	0,081
Nivå 2 varians:	0,032	0,004	0,032	0,004	0,032	0,004	0,032	0,004	0,032	0,004
Nivå 1 varians:	1,729	0,056	1,730	0,056	1,720	0,056	1,720	0,056	1,721	0,056
BIC, Deviance:	964262,8	963936,2	964283,2	964069,6	962897,8	962533,42	962886,9	962560,2	962961,9	962685,4

$P \leq 0,05 = *$

Arbeidstid fikk i likhet med sektor en tolkningsmessig verdi etter at interaksjonen med kjønn ble lagt til. Deltidsarbeidende kvinner arbeider i gjennomsnitt i yrkesgrupper med 0,085 desilpoengs større avstand fra utvalgets totale fordeling av kjønnene i arbeidsstyrken enn de som arbeider heltid. For menn er situasjonen motsatt, og vi finner en klar negativ sammenheng mellom kjønnsdeling i deltidsarbeid, sammenlignet med heltidsarbeidere ($b = -0,353$). Det støtter opp under hypotesene 6a og 6b. Heltidsarbeidende menn arbeider i yrkesgrupper som er mer kjønnsdelt enn deltidsarbeidende menn, og deltidsarbeidende kvinner arbeider i yrkesgrupper som er mer kjønnsdelt enn heltidsarbeidende kvinner. Sammenhengen mellom kjønn, sektor og arbeidstid kan ses i figur 2 under. For å finne den komposisjonelle effekten av samspillene, tar man den uforklarte variasjonen på gruppenivå for modell to, og trekker fra den uforklarte variasjonen på gruppenivå for modell tre, og deler differansen på førstnevnte gruppenivåvariasjon vist med følgende formel: $((0,209+0,032)-(0,202+0,032))/(0,209+0,032)=0,028$. Effekten tilknyttet sektor, arbeidstid, antall i husholdningen, samboerskap og interaksjonene med kjønn forklarer 2,8 prosent av gruppenivåvariasjonen. Bare litt av gruppevariasjonen kan derfor bli tilskrevet den forventede sammensetningen av individene i gruppene.



Figur 2: Samspill: Arbeidstid, sektor og kjønn.

Videre finner vi at de med over laveste formelle utdanningsgrad, arbeider i yrkesgrupper med større grad av kjønnsdeling enn de som har lavest eller ingen formell utdanning ($b=0,144$). Den største effekten i dummysettet for utdanning finnes imidlertid hos de med en fullført utdanning på universitets- og høyskolenivå. De arbeider i yrkesgrupper med 0,274 desilpoengs mindre avstand fra landets totale fordeling av menn og kvinner i arbeid. Jeg finner derfor støtte for hypotese 3b, og

ikke 3a. Høy utdanning har en negativ sammenheng med kjønnsdeling, sammenlignet med lav utdanning. I forbindelse med det, har også hypotese 3c fått støtte, fordi sammenlignet med høy utdanning, er både den laveste og nest laveste formelle utdanningens verdi høyere.

Husholdningsinntekt ser ut til å ha en sammenheng med kjønnsdeling, der andre og tredje kvartil innenfor hvert landår hadde litt høyere verdier for kjønnsdeling enn det laveste kvartilet. Det samme kan sies om medlemmer av fagforeninger, som også ser ut til å arbeide i yrkesgrupper med noe høyere grad av kjønnsdeling. Den største effekten finnes i skjevhetsskontrollen for den hierarkiske strukturen i ISCO-indeksen med 0,838 decilpoeng mindre avstand mellom yrkesgruppen og landårets totale fordeling av menn og kvinner i arbeid.

Vi finner pseudo R^2 ved å sammenligne total restleddvarians med modell tre i tabell 5: $((0,198+0,032+1,83)-(0,202+0,032+1,72))/(0,198+0,032+1,83)=0,051$, som indikerer at modellen har forklart 5,1% av den totale uforklarte variasjonen i avhengig variabel.

4.4 Kontroll for kvinneandelen i offentlig sektor og i arbeidsmarkedet

I neste tabell med modeller skal jeg ta for meg hvorvidt kvinneandelen i offentlig sektor og kvinneandelen i arbeidslivet påvirker kjønnsdelingen. Dette er det første settet med kontekstuelle variabler jeg skal teste effekten av. Tabell 7 viser fire modeller med robuste standardfeil grunnet vektning, der utvalget har blitt holdt konstant, og der første modell har med de kontekstuelle variablene. Modellen har i tillegg med kontroll for kjønn, utdanningsgrad, husholdningsinntekt og fagforeningsmedlemskap, sammen med skjevhetsskontrollen for avhengig variabel. Modell to, tre og fire har i tillegg med kontroll for samboerskap, antall i husholdningen, sektor og de samme samspillene som modell fire i forrige tabell med modeller. Planen med modellene i denne tabellen er å vise hvorvidt de kontekstuelle variablene har en sammenheng med kjønnsdeling, og samtidig se om samvariasjonen endres etter kontroll for de individuelle variablene. I tabellen skal derfor hypotesene 5a, 5b, 5c, 7b og 7c settes på prøve.

Modell en viser at kvinneandelen i offentlig sektor ikke har en signifikant sammenheng med kjønnsdeling, og at kvinneandelen i arbeidsstyrken har en positiv sammenheng. De individuelle variablene viser mye de samme resultatene som i det første settet med modeller, og jeg går derfor ikke gjennom det. Modell to har i tillegg kontroll for de individuelle variablene og samspill fra modell fire i forrige tabell. Den kontekstuelle variabelen for kvinneandelen i arbeidsstyrken er fortsatt signifikant, mens kvinneandelen i offentlig sektor fortsatt ikke er det. Det medfører at hypotese 5a forkastes, fordi kvinneandelen i offentlig sektor ikke gav mer forklaringskraft til modell to enn fire, og at variabelen ikke var signifikant. Teknisk sett beholder jeg også hypotese 7b, fordi

jeg ikke kan forkaste den, men det er lite relevant, fordi kvinneandelen i offentlig sektor ikke var signifikant i noen av modellene. Hypotese 5b beholdes, og jeg forkaster nullhypotesen.

BIC og deviance-verdiene reduseres med henholdsvis 1577,1 og 1740 fra modell en til to, og indikerer at modell to ikke overraskende er bedre tilpasset dataene enn modell en. I modell tre har jeg fjernet variabelen for andelen av arbeidsstyrken som er kvinner, og i modell fire har jeg lagt den inn igjen, og tatt ut variabelen for andelen av offentlig sektor som er kvinner. Det gjorde jeg for å se om variabelen for kvinneandelen i arbeidsmarkedet har stjålet forklaringskraften til kvinneandelen i offentlig sektor. Utenom dette, er modellen identisk til modell fire i tabell 6. Det betyr at det er mulig å sammenligne restleddvariansen på gruppenivåene, for å se hvor mye av variasjonen i avhengig variabel mellom gruppene som forklares av de kontekstuelle variablene, som jeg gjør i slutten av delkapittelet.

Ved å se på koeffisienten for kvinneandelen i offentlig sektor i modell tre, sammenlignet med i modell to, ser vi at den endres noe, men at variabelen ikke er signifikant selv uten å kontrollere for kvinneandelen i offentlig sektor. Det tolker jeg som at kvinneandelen i offentlig sektor ikke har noen innvirkning for den horisontale kjønnsdelingen i arbeidsmarkedet.

BIC indikerer at modell tre er litt dårligere tilpasset dataene enn modell to, med en differanse på 2,4. Modell fire er ifølge BIC den beste modellen med en reduksjon på 10,2, sammenlignet med modell tre. På grunn av lavest BIC og devianceverdier, sammen med at jeg ikke har med en irrelevant kontekstuell variabel, er modell fire den endelige modellen i tabell 7.

Koeffisienten for kvinneandelen i arbeidsstyrken indikerer at for hver prosent økning av arbeidsstyrken som er kvinner, øker kjønnsdelingen i landet med 0,034 desilpoeng lenger unna landets totale fordeling menn og kvinner i arbeid. Variabelen forklarer relativt mye av den uforklarte restleddvariasjonen på gruppenivå. Ved å sammenligne uforklart restleddvariasjon på gruppenivåene i modell fire i tabell 6 og modell fire i tabell 7, (slik: $((0,202+0,032)-(0,129+0,033))/(0,202+0,032)=0,308$), finner en at den kontekstuelle variabelen, forklarer nesten 31 prosent av den uforklarte variasjonen i gruppenivåene, sammenlignet med når man ikke kontrollerer for kontekstuelle variabler.

Tabell 7: Flernivåanalyser med kontroll for kvinneandelen i offentlig sektor og i arbeidsstyrken. N=286578

Fixed del	Modell 1		Modell 2		Modell 3		Modell 4	
	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.
Offentlig sektor, kvinneandel	0,000	0,003	0,000	0,003	0,002	0,003		
Arbeidsstyrken, kvinneandel	0,033 *	0,014	0,034 *	0,014			0,034 *	0,012
Mann	0,556 *	0,040	0,709 *	0,044	0,709 *	0,044	0,709 *	0,044
Har samboer (dikotom)			0,046 *	0,015	0,046 *	0,015	0,046 *	0,015
Antall i husholdningen			0,012	0,008	0,012	0,008	0,012	0,008
Sektor (ref. Privat sektor)								
Arbeidsledig/0 timer i uken			0,046	0,040	0,046	0,039	0,046	0,039
Offentlig sektor			0,121 *	0,040	0,121 *	0,040	0,121 *	0,040
Arbeidstid (ref. Heltid <37 timer)								
Arbeidsledig/0 timer i uken			-0,062 *	0,026	-0,062 *	0,026	-0,062 *	0,026
Deltid (≥37 timer)			0,085 *	0,027	0,085 *	0,027	0,085 *	0,027
Samspill: Samboerskap, kjønn og antall i husholdningen (ref. Har ikke samboer, kvinne)								
Har ikke samboer*Mann*Antall i husholdningen			-0,041 *	0,016	-0,041 *	0,016	-0,041 *	0,016
Har samboer*Kvinne*Antall i husholdningen			-0,005	0,005	-0,005	0,005	-0,005	0,005
Har samboer*Mann*Antall i husholdningen			-0,013	0,016	-0,013	0,016	-0,013	0,016
Samspill: Kjønn og sektor (ref. Kvinne, Privat sektor)								
Mann*Arbeidsledig/0 timer i uken			-0,144 *	0,057	-0,145 *	0,057	-0,144 *	0,057
Mann*Offentlig sektor			-0,267 *	0,051	-0,267 *	0,051	-0,267 *	0,051
Samspill: Kjønn og arbeidstid (ref. Kvinne, Heltid)								
Mann*Arbeidsledig/0 timer i uken			0,082	0,047	0,082	0,047	0,082	0,047
Mann*Deltid			-0,353 *	0,045	-0,355 *	0,045	-0,353 *	0,045
Utdanningsgrad (ref. Ingen/laveste formelle utdanning)								
Over laveste formelle	0,143 *	0,041	0,143 *	0,041	0,143 *	0,041	0,143 *	0,041
Sekundær/ begynt høyere	0,002	0,045	0,006	0,045	0,006	0,045	0,006	0,045
Universitetsutd +	-0,277 *	0,068	-0,274 *	0,065	-0,274 *	0,065	-0,274 *	0,065
Husholdningsinntekt (ref. Første kvartil)								
Andre kvartil	0,081 *	0,018	0,060 *	0,019	0,060 *	0,019	0,060 *	0,019
Tredje kvartil	0,083 *	0,021	0,048 *	0,024	0,048 *	0,024	0,048 *	0,024
Fjerde kvartil	0,016	0,028	-0,021	0,033	-0,020	0,033	-0,021	0,033
Medlem av fagforening (dikotom)	0,085 *	0,024	0,074 *	0,022	0,074 *	0,022	0,074 *	0,022
Lederskjevhet	-0,826 *	0,112	-0,838 *	0,112	-0,838 *	0,112	-0,838 *	0,112
Konstant	2,061	0,074	1,992	0,082	1,983	0,092	1,992	0,082
Random del	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.
Nivå 3 varians:	0,134	0,070	0,129	0,065	0,190	0,083	0,129	0,065
Nivå 2 varians:	0,033	0,004	0,033	0,004	0,032	0,004	0,033	0,004
Nivå 1 varians:	1,731	0,056	1,720	0,056	1,720	0,056	1,720	0,056
BIC, Deviance:	964478,1	964289,6	962901	962549,2	962898,6	962559,4	962888,4	962549,2

$P \leq 0,05 = *$

Variabelen for kvinneandelen i arbeidsstyrken var signifikant både før og etter kontroll for kvinneandelen i offentlig sektor, og jeg forkaster derfor hypotese 5c. Sammenhengen mellom kvinneandelen i arbeidsstyrken og kjønnsdeling var i tillegg signifikant etter kontroll for de individuelle variablene arbeidstid, sektor, samboerskap og antall personer i husholdningen, og samspillene derimellom. Jeg forkaster derfor også hypotese 7c og beholder nullhypotesen.

4.5 Kontroll for generell likestilling og økonomisk ulikhet

I den neste tabellen med modeller tar jeg for meg den generelle likestillingsindeksen global gender gap, og inntektsulikhetsindeksen GINI. På samme måte som tidligere skal jeg ved å kontrollere for de individuelle variablene som viste seg brukbare til å forklare kjønnsdeling, undersøke hvorvidt likestilling og økonomisk ulikhet har en sammenheng med kjønnsdeling på individnivå.

Tabell 8 viser fem modeller. I den første modellen har jeg kontrollert for to kontekstuelle variabler, der kjønn og variabelen for inntektsulikhet har et samspillsledd. Den samme logikken - der jeg har en modell før og etter samspill og involverte variabler i hypotesene 7a-e - er benyttet. I tillegg har jeg lagt på en random slope for kjønn i modell en og to. Jeg ønsker med det å vise hvorvidt inntektsulikhet og likestilling har en sammenheng med kjønnsdeling, og til hvilken grad sammenhengen blir påvirket av kontroll for samboerskap, antall i husholdningen, sektor, arbeidstid og samspillene med kjønn. Det er nødvendig for å kunne teste hypotesene 1a, 1b, 2a, 2b og 7a. I modell en og to har jeg latt effekten av kjønn variere fra utvalg til utvalg (random slope). I modell tre, fire og fem har jeg tatt vekk både likestillingsindeksen og random slope, for å se hvordan den økte kompleksiteten med random slope i modell en og to påvirker resultatene.

Modell en viser at inntektsulikhetsindeksen har en svak negativ sammenheng med kjønnsdeling for kvinner, men ikke for menn, ved at samspillet mellom kjønn og GINI-indeksen ikke er statistisk signifikant i både modell en og to. Likestillingsindeksen global gender gap er ikke signifikant verken før eller etter kontroll for individuelle variabler, og jeg forkaster derfor hypotesene 1a og 1b. Hypotese 2b og 7a beholder jeg teknisk sett, men det har ingen betydning, fordi likestillingsindeksen ikke hadde en sammenheng med kjønnsdeling i utgangspunktet.

I modell tre har jeg laget en modell identisk til modell to, bare at random slope er tatt ut. I modell fire har jeg i tillegg tatt ut likestillingsindeksen. BIC-verdien for modell fire er lavere enn i modell tre, noe som betyr at likestillingsindeksen ikke bidrar med en forbedring av modell tre. Devianceverdiene i modell tre og fire er identiske, og støtter opp under indikasjonen fra BIC-verdiene. I modell fem har jeg tatt ut samspillet mellom GINI-indeksen og kjønn. Verdiene for BIC og deviance indikerer at modell fire er bedre enn modell fem.

Verdiene for uforklart restleddvariasjon i modell en og to antyder at mye av den uforklarte variasjonen i avhengig variabel blir forklart med at modellen har en random slope. Mesteparten av den variasjonen som blir forklart av modell en, og ikke med modell tre-fem, er situert i andre og tredje nivå. Det er grunnen til at verdiene for BIC og deviance er mye lavere i modell en og to, enn i tre, fire og fem.

Hvis en utelukkende ser på verdiene for BIC og deviance, er modell to den modellen som er best tilpasset dataene, med BIC-verdi ~ 3200 lavere enn modell tre, fire og fem. Dog, gir verken samspillet eller tillegget av random slope noe mer tolkningsmuligheter, fordi samspillet ikke var signifikant, og det gjør modellen mer kompleks. Jeg ser ingen gevinst i å beholde en modell som er mer kompleks enn den trenger å være, dersom den ikke gir noen økt tolkningsverdi. Jeg mener derfor at den beste modellen i dette settet modeller er den femte. Det medfører at jeg forkaster hypotese 2a, fordi samspillet mellom GINI-indeksen og kjønn ikke var signifikant, selv når effekten av kjønn varierer fra utvalg til utvalg. Det kan imidlertid hende at GINI-indeksen har noe å tilføre av forklaringskraft, da den var signifikant uten samspill med kjønn. Jeg tar med resultatene til videre utforskning senere i analysen og med meg i drøftingen. Modell fem er derfor den endelige modellen i tabell 8.

For hver økning med 1 i GINI-indeksen, som har en teoretisk variasjonsbredde på 100, men i praksis 43, og et standardavvik på 8,8, sentrert rundt gjennomsnittet 38,144, reduseres kjønnsdelingen i landet med 0,012 desilpoeng lenger unna landets totale fordeling av menn og kvinner i arbeidsstyrken. En høy verdi i GINI-indeksen indikerer at det er store økonomiske forskjeller mellom innbyggerne i landet. En standardisert koeffisient finner vi ved å gange koeffisienten med standardavviket for den uavhengige variabelen delt på standardavviket for den avhengige variabelen: $-0,012 * (8,797463 / 1,393518) = -0,076$. Det betyr at en økning med ett standardavvik i GINI-indeksen, samvarierer med en reduksjon på 0,076 standardavvik i den avhengige variabelen. Det tolker jeg som at effekten av inntektsulikhet er lav.

Tabell 8: Flernivåanalyser med kontroll for generell likestilling og inntektsulikhet. N=286587

Fixed del	Modell 1		Modell 2		Modell 3		Modell 4		Modell 5	
	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.
GINI indeks	-0,013 *	0,005	-0,012 *	0,004	-0,011 *	0,005	-0,011 *	0,005	-0,012 *	0,005
Global Gender Gap indeks	0,163	0,674	0,184	0,677	0,108	0,747				
Mann	0,546 *	0,046	0,658 *	0,045	0,692 *	0,046	0,692 *	0,046	0,709 *	0,044
Sampill: GINI*mann	-0,003	0,005	-0,004	0,005	-0,003	0,004	-0,003	0,004		
Har samboer (dikotom)			0,043 *	0,015	0,045 *	0,015	0,045 *	0,015	0,046 *	0,015
Antall i husholdningen			0,006	0,005	0,009	0,007	0,009	0,007	0,012	0,008
Sektor (ref. Privat sektor)										
Arbeidsledig/0 timer i uken			0,034	0,033	0,046	0,039	0,046	0,039	0,045	0,039
Offentlig sektor			0,139 *	0,037	0,124 *	0,040	0,124 *	0,040	0,121 *	0,040
Arbeidstid (ref. Heltid <37 timer)										
Arbeidsledig/0 timer i uken			-0,048	0,026	-0,063 *	0,025	-0,063 *	0,025	-0,062 *	0,026
Deltid (≥37 timer)			0,070 *	0,025	0,085 *	0,026	0,085 *	0,026	0,085 *	0,027
Sampill: Samboerskap, kjønn og antall i husholdningen (ref. Har ikke samboer, kvinne)										
Har ikke samboer*Mann*Antall i husholdningen			-0,027 *	0,008	-0,035 *	0,012	-0,035 *	0,012	-0,041 *	0,016
Har samboer*Kvinne*Antall i husholdningen			-0,007	0,005	-0,004	0,005	-0,004	0,005	-0,005	0,005
Har samboer*Mann*Antall i husholdningen			0,004	0,008	-0,009	0,013	-0,009	0,013	-0,013	0,016
Sampill: Kjønn og sektor (ref. Kvinne, Privat sektor)										
Mann*Arbeidsledig/0 timer i uken			-0,125 *	0,037	-0,147 *	0,057	-0,147 *	0,057	-0,145 *	0,057
Mann*Offentlig sektor			-0,304 *	0,042	-0,271 *	0,051	-0,271 *	0,051	-0,267 *	0,051
Sampill: Kjønn og arbeidstid (ref. Kvinne, Heltid)										
Mann*Arbeidsledig/0 timer i uken			0,057	0,046	0,085	0,046	0,085	0,046	0,082	0,047
Mann*Deltid			-0,319 *	0,041	-0,351 *	0,045	-0,351 *	0,045	-0,353 *	0,045
Utdanningsgrad (ref. Ingen/laveste formelle utdanning)										
Over laveste formelle	0,131 *	0,039	0,134 *	0,040	0,143 *	0,040	0,143 *	0,040	0,143 *	0,041
Sekundær/ begynt høyere	-0,003	0,044	0,004	0,044	0,006	0,045	0,006	0,044	0,006	0,045
Universitetsutd +	-0,285 *	0,065	-0,278 *	0,063	-0,275 *	0,065	-0,275 *	0,065	-0,274 *	0,065
Husholdningsinntekt (ref. Første kvartil)										
Andre kvartil	0,078 *	0,018	0,059 *	0,019	0,059 *	0,019	0,059 *	0,019	0,060 *	0,019
Tredje kvartil	0,079 *	0,021	0,047	0,024	0,048 *	0,024	0,048 *	0,024	0,048 *	0,024
Fjerde kvartil	0,007	0,028	-0,029	0,033	-0,021	0,033	-0,021	0,033	-0,020	0,033
Medlem av fagforening (dikotom)	0,081 *	0,024	0,072 *	0,023	0,074 *	0,022	0,074 *	0,022	0,074 *	0,022
Lederskjevhhet	-0,824 *	0,110	-0,837 *	0,111	-0,840 *	0,112	-0,840 *	0,112	-0,838 *	0,112
Konstant	2,082	0,073	2,026	0,082	1,988	0,087	1,987	0,087	1,977	0,092
Random del	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.
Nivå 3 varians:	0,104	0,064	0,101	0,063	0,176	0,080	0,177	0,077	0,177	0,077
Nivå 2 varians:	0,040	0,005	0,039	0,005	0,032	0,004	0,032	0,004	0,032	0,004
Varians i effekten av mann:	0,111	0,016	0,110	0,018						
Nivå 1 varians:	1,705	0,056	1,695	0,056	1,720	0,056	1,720	0,056	1,720	0,056
BIC, Deviance:	961199,4	960985,8	959673,6	959296,6	962900,2	962535,8	962887,7	962535,8	962895,7	962556,4

$P \leq 0,05 = *$

4.6 Kontroll for dimensjoner i likestillingspolitikk

I tabell 9 har jeg bygd fire modeller der jeg har tatt med de to kontekstuelle variablene som måler dimensjoner i likestillingspolitikk. Målet med modellsettet er å vise to poeng, i tillegg til å teste hypotesene 4a, 4b, 7d og 7e. Det ene er at delt omsorgs-dimensjonen er signifikant før og etter kontroll for samboerskap, antall i husholdningen, sektor, arbeidstimer og interaksjonene jeg tidligere har greid ut om. Det andre er at delt omsorgs-dimensjonen legger til forklaringskraft, i tillegg til de variablene på individnivå gjør. Forskjellen mellom modell en og to er kontrollvariablene med samspill på individnivå. Som det kommer frem gjennom BIC og deviance, er modell to bedre enn modell en. Delt omsorgs-dimensjonen er signifikant selv etter kontroll for individuelle variabler, og jeg forkaster derfor hypotese 7e. Hypotese 7d, beholdes teknisk sett, selv om det ikke betyr noe for tolkningen sin del, igjen fordi delt forsørger-dimensjonen aldri var signifikant. Det medfører i tillegg at hypotese 4b forkastes.

BIC indikerer at den beste modellen i tabellen er modell fire, der differansen mellom den nest beste (modell tre) er 5,9. Devianceverdien for modell to og fire er identiske. Med ett ekstra parameter i modell to enn i modell fire, er derfor ikke kontrollen for delt forsørger-dimensjonen en signifikant forbedring i en deviancetest. Modell fire er derfor den endelige modellen i dette modellsettet. Ved å sammenligne restleddvariansen på gruppenivåene med modell fire i tabell 6, finner en at delt omsorgs-dimensjonen i likestillingspolitikk har forklart 12,7 prosent av den totale uforklarte variasjonen mellom land og mellom landår. Sammenligningsvis og selv om ikke delt forsørger-dimensjonen var signifikant, forteller nivå 2 og 3 variansen i modell tre oss at den har forklart 7,3 av variasjonen i avhengig variabel mellom land og mellom landår.

Effekten av delt omsorgs-dimensjonen ser svært stor ut, men det skyldes at variabelen er standardisert, og har en variasjonsbredde på 2,5. Hypotese 4a er imidlertid i motsatt retning av det resultatet viser. Delt omsorgs-dimensjonen har en positiv og ikke negativ sammenheng med kjønnsdeling. Jeg forkaster derfor hypotese 4a, men tar med meg resultatet videre i diskusjonen av resultatene.

Tabell 9: Flernivåanalyser med kontroll for dimensjoner i likestillingspolitikk. N=286587

Fixed del	Modell 1		Modell 2		Modell 3		Modell 4	
	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.
Delt omsorg-dimensjon	0,244 *	0,066	0,236 *	0,066			0,233 *	0,063
Delt forsørger-dimensjon	-0,008	0,106	-0,004	0,106	0,052	0,102		
Mann	0,556 *	0,040	0,709 *	0,044	0,709 *	0,044	0,709 *	0,044
Har samboer (dikotom)			0,046 *	0,015	0,046 *	0,015	0,046 *	0,015
Antall i husholdningen			0,012	0,008	0,012	0,008	0,012	0,008
Sektor (ref. Privat sektor)								
Arbeidsledig/0 timer i uken			0,045	0,039	0,045	0,039	0,045	0,039
Offentlig sektor			0,121 *	0,040	0,121 *	0,040	0,121 *	0,040
Arbeidstid (ref. Heltid <37 timer)								
Arbeidsledig/0 timer i uken			-0,062 *	0,026	-0,062 *	0,026	-0,062 *	0,026
Deltid (≥37 timer)			0,085 *	0,027	0,085 *	0,027	0,085 *	0,027
Samspill: Samboerskap, kjønn og antall i husholdningen (ref. Har ikke samboer, kvinne)								
Har ikke samboer*Mann*Antall i husholdningen			-0,041 *	0,016	-0,041 *	0,016	-0,041 *	0,016
Har samboer*Kvinne*Antall i husholdningen			-0,005	0,005	-0,005	0,005	-0,005	0,005
Har samboer*Mann*Antall i husholdningen			-0,013	0,016	-0,013	0,016	-0,013	0,016
Samspill: Kjønn og sektor (ref. Kvinne, Privat sektor)								
Mann*Arbeidsledig/0 timer i uken			-0,145 *	0,057	-0,145 *	0,057	-0,145 *	0,057
Mann*Offentlig sektor			-0,267 *	0,051	-0,267 *	0,051	-0,267 *	0,051
Samspill: Kjønn og arbeidstid (ref. Kvinne, Heltid)								
Mann*Arbeidsledig/0 timer i uken			0,082	0,047	0,082	0,047	0,082	0,047
Mann*Deltid			-0,353 *	0,045	-0,353 *	0,045	-0,353 *	0,045
Utdanningsgrad (ref. Ingen/laveste formelle utdanning)								
Over laveste formelle	0,143 *	0,041	0,143 *	0,041	0,143 *	0,041	0,143 *	0,041
Sekundær/ begynt høyere	0,003	0,045	0,006	0,045	0,006	0,045	0,006	0,045
Universitetsutd +	-0,276 *	0,068	-0,274 *	0,065	-0,274 *	0,065	-0,274 *	0,065
Husholdningsinntekt (ref. Første kvartil)								
Andre kvartil	0,081 *	0,018	0,060 *	0,019	0,060 *	0,019	0,060 *	0,019
Tredje kvartil	0,083 *	0,021	0,048 *	0,024	0,048 *	0,024	0,048 *	0,024
Fjerde kvartil	0,015	0,028	-0,020	0,033	-0,020	0,033	-0,020	0,033
Medlem av fagforening (dikotom)	0,084 *	0,024	0,074 *	0,022	0,074 *	0,022	0,074 *	0,022
Lederskjevhet	-0,826 *	0,112	-0,838 *	0,112	-0,838 *	0,112	-0,838 *	0,112
Konstant	2,066	0,084	1,997	0,091	1,985	0,091	1,997	0,091
Random del	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.
Nivå 3 varians:	0,182	0,083	0,176	0,079	0,186	0,092	0,175	0,069
Nivå 2 varians:	0,032	0,004	0,032	0,004	0,032	0,004	0,032	0,004
Nivå 1 varians:	1,731	0,056	1,720	0,056	1,720	0,056	1,720	0,056
BIC, Deviance:	964481,8	964293,2	962905,5	962553,6	962898,9	962559,6	962893	962553,6

$P \leq 0,05 = *$

4.7 Kontroll for relevante kontekstuelle variabler

Slik som resultatene i analysen har tredd fram, skal jeg nå vise en siste tabell med modeller der jeg setter de variablene som har vist seg relevante i samme tabell. Poenget med tabellen er å kontrollere for om de kontekstuelle variablene i praksis måler det samme. De signifikante kontekstuelle variablene som har tredd fram i analysen er andelen av arbeidsmarkedet som er kvinner, GINI-indeksen for økonomisk ulikhet og delt omsorgs-dimensjonen i likestillingspolitikk. Det er også av interesse å se hvor mye av gruppevariasjonen som kan fanges opp av de kontekstuelle variablene sammen, for å sammenligne med hvor mye de fanger opp alene. Med det mener jeg hvor mye av den forklarte gruppevariasjonen som overlapper mellom de ulike kontekstuelle variablene.

Tabell 10 viser siste tabell med flernivåanalyser med robuste standardfeil, igjen grunnet vekting på første nivå. I modell en har jeg som vanlig ikke kontrollert for samspillene og de involverte variablene, mens i modell to har jeg det. Både delt omsorgs-dimensjonen og kvinneandelen i arbeidsstyrken er signifikante forklaringsvariabler for kjønnsdeling i arbeidslivet, etter kontroll for alle relevante individuelle og hverandre. GINI-indeksen var ikke signifikant i noen av tilfellene, så jeg har tatt den ut i modell tre og fire. I modell fire har jeg også satt inn et samspill mellom kvinneandelen i arbeidsmarkedet og delt omsorgs-dimensjonen i likestillingspolitikk.

GINI-indeksen, som hadde en signifikant sammenheng med kjønnsdeling i tabell 8, er ikke lenger signifikant i modell en og to. Det er ikke vist i tabellen, men ved siden av har jeg kjørt modeller der jeg kontrollerer for GINI-indeksen, og bare en av de andre kontekstuelle variablene av gangen. Både kvinneandelen av arbeidsstyrken og delt omsorgs-dimensjonen tar forklaringskraften fra GINI-indeksen. Det kan tolkes på flere måter. På en side kan en anta at inntektsulikhet har en sammenheng med kjønnsdeling, men at kvinneandelen i arbeidsstyrken og delt omsorgs-dimensjonen i likestillingspolitikk er mer nærliggende mål. Sett på et slikt vis, blir kvinneandelen i arbeidsstyrken og delt omsorgs-dimensjonen mellomliggende faktorer, som tar vekk forklaringskraften til inntektsulikhet. En slik tolkning innebærer at en har tillit til at validiteten til indeksene er gode. En annen måte å tolke det på kan være at indeksen for inntektsulikhet, delt omsorgs-dimensjonen og kvinneandelen i arbeidsstyrken i praksis måler det samme. En slik tolkning vil innebære at en ikke har tillit til at de kontekstuelle variablene måler det de skal, eller at antall enheter på nivå tre ikke er stort nok. Det kommer jeg tilbake til i diskusjonsdelen.

BIC indikerer at modell tre er den beste modellen. Det er et poeng å påpeke at selv om kvinneandelen i arbeidsstyrken har tilsynelatende minst effektstørrelse, må man huske på to ting. Det ene er at hvis man skal kunne sammenligne effektstørrelsen av disse to variablene, må man

gjøre om koeffisienten til en standardisert koeffisient (beta), fordi variasjonsbredden i variablene er forskjellige.

Betaverdien får vi ved å gange koeffisienten til kvinneandelsvariabelen med standardavviket sitt delt på standardavviket for avhengig variabel: $0,030*(3,491277/1,393518)=0,075$. Betaverdien for delt omsorgs-dimensjonen er $0,159*(0,8260605/1,393518)=0,094$. Slik kan man se at forskjellen i effektstørrelsene for kvinneandelen i arbeidsstyrken og delt omsorgs-dimensjonen ikke er veldig stor. Vi kan også finne ut hvor mye av gruppevariasjonen som blir plukket opp av de to variablene i modell tre ved å dele differansen mellom den uforklarte variansen på gruppenivåene i modell tre og modell fire i tabell 10 med den uforklarte variansen på gruppenivåene i modell fire i tabell 6.

Da finner man at delt omsorgs-dimensjonen og kvinneandelen i arbeidsstyrken har forklart $((0,205+0,032)-(0,122+0,033))/(0,205+0,032)*100=34,8$ prosent av den totale uforklarte gruppevariasjonen når man kun har kontrollert for individuelle variabler. Det kan vi sammenligne med tallene vi fant i de tidligere modellene, der de to kontekstuelle variablene ble målt i to ulike modeller. Modell fire i tabell 9 viste at delt omsorgs-dimensjonen alene forklarte 12,7% av den totale gruppevariasjonen, og modell fire i tabell 7 viste at kvinneandelen i arbeidsstyrken alene forklarte 30,8%. $30,8+12,7$ er ikke lik 34,8. Det betyr at 8,7 prosent av forklaringskraften på gruppenivå har overlappet mellom de to kontekstuelle variablene.

I modell fire har jeg lagt inn et samspill mellom kvinneandelen av arbeidsstyrken og delt omsorgs-dimensjonen. I modellen er ikke basisvariabelen for kvinneandelen i arbeidsstyrken signifikant lenger, og samspillet med delt omsorgs-dimensjonen er signifikant. Det er tolkbart fordi variabelen for kvinneandelen av arbeidsstyrken er sentrert rundt gjennomsnittet som er 46,7. Det betyr at når kvinneandelen av arbeidsstyrken er gjennomsnittlig, indikerer koeffisienten for delt omsorgs-dimensjonen en økning på $0,317*(0,8260605/1,393518)=0,188$ standardavvik i avhengig variabel når man øker den uavhengige med ett standardavvik. Sammenlignet med modell tre, blir effekten av den delte omsorgs-dimensjonen doblet.

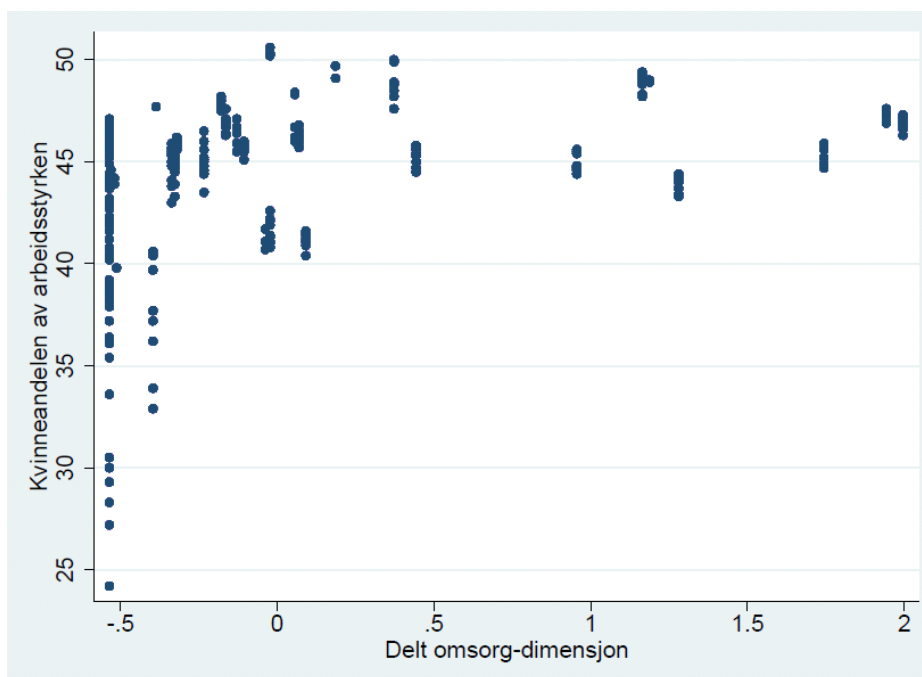
I modell fire har igjen den restleddvariasjonen naturlig nok blitt redusert når det er ett ekstra parameter i modellen, sammenlignet med modell tre. Med samspillet forklarer modell fire to prosent mer av gruppevariasjonen enn modell tre. Ved å konstruere en pseudo R^2 , finner vi at modell tre forklarer:

$((0,198+0,096+1,830)-(0,122+0,033+1,720))/(0,198+0,096+1,830)=0,090$ 9 prosent av kjønnsdelingen totalt, og modell fire forklarer:

$((0,198+0,096+1,830)-(0,118+0,032+1,720))/(0,198+0,096+1,830)=0,093$ 9,3 prosent av kjønnsdelingen totalt. Verdiene er tatt fra restleddvariasjonene på alle nivåene i modell tre og fire, sammenlignet med de tilsvarende verdiene for modell tre i tabell 5, der jeg ikke kontrollerte for noen individuelle variabler bortsett fra skjevhetsskontrollen av avhengig variabel. Alt i alt betyr det at for land med en gjennomsnittlig fordeling menn og kvinner i arbeid, har delt omsorgsdimensjonen en større effekt enn for land som ikke har en gjennomsnittlig fordeling menn og kvinner i arbeid.

For å undersøke grunnen til at kvinneandelen av arbeidsstyrken ikke var signifikant etter tillegget av et samspill har jeg kjørt et spredningsplott for de to variablene vist i figur 3 under. Det viser at det for landene som skårer høyt på den delte omsorgsdimensjonen, ikke finnes noen land som har lav andel kvinner i arbeidsmarkedet. Det er derfor ikke en sammenheng når kvinneandelen er høy, fordi det mangler sammenligningsgrunnlag.

Jeg har ikke behov for å velge en endelig modell i dette settet med modeller, fordi jeg ikke har noen hypoteser som skal testes her. For interessens skyld ser vi likevel at BIC indikerer at modell tre er best tilpasset dataene, og deviance peker i motsatt retning.



Figur 3: Spredningsplott over kvinneandelen av arbeidsstyrken og delt omsorg-dimensjonen.

Fordi variabelen for kvinneandelen i arbeidsstyrken ikke var signifikant i samspillet har jeg undersøkt det samme samspillet som i modell fire, bare at de kontekstuelle variablene har andre nullverdier. Det har jeg ikke vist i analysen, fordi det ville tatt 16 nye modeller for å vise alle kombinasjonene. I korte trekk kan resultatene beskrives som at når kvinneandelen i

arbeidsmarkedet er lav, er effekten av delt omsorgs-dimensjonen større. Når kvinneandelen øker og nærmer seg 50 prosent, forsvinner effekten av den delte omsorgs-dimensjonen.

I neste kapittel argumenterer jeg for en teoretisk forståelse av funnene. For å finne ut om det stemmer har jeg undersøkt om husholdningsinntekten har en ulik påvirkning på kjønnsdelingen for kvinner og menn i land med ulike verdier i delt omsorgs-dimensjonen. Derfor har jeg kjørt modeller innenfor landene med og uten fedre- og frie foreldrepermisjoner hver for seg med et samspill mellom husholdningsinntekt og kjønn. Modellene er heller ikke vist her, fordi analysen er omfattende nok som den er. Resultatene viste at kvinner med lav husholdningsinntekt arbeider i yrker med mindre kjønnsdeling enn kvinner med høy husholdningsinntekt i land med lavest verdi i delt omsorgs-dimensjonen. I land med mer enn ingen fedrepermisjoner og permisjoner som kan fordeles fritt mellom foreldrene, er det ingen sammenheng. Fremgangsmåten kan problematiseres ved at en reduserer antall enheter på nivå tre betraktelig. Det kan likevel ses på som et av få akseptable strategier, fordi alternativet er modeller med fireveis-interaksjoner og kryssnivåinteraksjon. Det er mye mer komplisert, og å kjøre egne modeller innenfor land med mye versus lite fedre og/eller foreldrepermisjoner, anså jeg som det mest hensiktsmessige alternativet. Jeg har også kjørt de disse modellene med de ubehandlede inntektsvariablene, og på tross av et redusert utvalg, var resultatene stabile.

Avslutningsvis vil jeg nevne at jeg også har kjørt alle modellene uten kontroll for fagforeningsmedlemskap. I og med at Norge og andre skandinaviske land kjennetegnes av å ha høy grad av fagorganisering, var mistanken at variabelen i land der mange innbyggere er medlem av fagforening tok forklaringskraft på gruppenivåene vekk fra de kontekstuelle variablene. Det hadde ingen betydelige innvirkninger på resultatene, og jeg har derfor ikke tatt det med i analysen.

Tabell 10: Flernivåanalyser med kontroll for relevante kontekstuelle variabler. N=286587

Fixed del	Modell 1		Modell 2		Modell 3		Modell 4	
	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.	Koeffisient	Robust S.F.
Arbeidsstyrken, kvinneandel	0,028 *	0,014	0,029 *	0,014	0,030 *	0,013	0,003	0,012
Delt omsorg-dimensjon	0,135 *	0,044	0,129 *	0,044	0,159 *	0,050	0,317 *	0,074
Samspill: Arbeidsstyrken, kvinneandel*delt omsorg-dimensjon							-0,068 *	0,023
GINI indeks	-0,006	0,006	-0,006	0,006				
Mann	0,556 *	0,040	0,709 *	0,044	0,709 *	0,044	0,709 *	0,044
Har samboer (dikotom)			0,046 *	0,015	0,046 *	0,015	0,046 *	0,015
Antall i husholdningen			0,012	0,008	0,012	0,008	0,012	0,008
Sektor (ref. Privat sektor)								
Arbeidsledig/0 timer i uken			0,045	0,039	0,046	0,039	0,046	0,039
Offentlig sektor			0,121 *	0,040	0,121 *	0,040	0,121 *	0,040
Arbeidstid (ref. Heltid <37 timer)								
Arbeidsledig/0 timer i uken			-0,062 *	0,026	-0,062 *	0,026	-0,062 *	0,026
Deltid (≥37 timer)			0,085 *	0,027	0,085 *	0,027	0,085 *	0,027
Samspill: Samboerskap, kjønn og antall i husholdningen (ref. Har ikke samboer, kvinne)								
Har ikke samboer*Mann*Antall i husholdningen			-0,041 *	0,016	-0,041 *	0,016	-0,041 *	0,016
Har samboer*Kvinne*Antall i husholdningen			-0,005	0,005	-0,005	0,005	-0,005	0,005
Har samboer*Mann*Antall i husholdningen			-0,013	0,016	-0,013	0,016	-0,013	0,016
Samspill: Kjønn og sektor (ref. Kvinne, Privat sektor)								
Mann*Arbeidsledig/0 timer i uken			-0,144 *	0,057	-0,144 *	0,057	-0,144 *	0,057
Mann*Offentlig sektor			-0,267 *	0,051	-0,267 *	0,051	-0,267	0,051
Samspill: Kjønn og arbeidstid (ref. Kvinne, Heltid)								
Mann*Arbeidsledig/0 timer i uken			0,082	0,047	0,082	0,047	0,082	0,047
Mann*Deltid			-0,353 *	0,045	-0,353 *	0,045	-0,353 *	0,045
Utdanningsgrad (ref. Ingen/laveste formelle utdanning)								
Over laveste formelle	0,142 *	0,041	0,143 *	0,041	0,143 *	0,041	0,143 *	0,041
Sekundær/ begynt høyere	0,002	0,045	0,006	0,045	0,006	0,045	0,006	0,045
Universitetsutd +	-0,277 *	0,068	-0,274 *	0,065	-0,274 *	0,065	-0,274 *	0,065
Husholdningsinntekt (ref. Første kvartil)								
Andre kvartil	0,081 *	0,018	0,060 *	0,019	0,060 *	0,019	0,060 *	0,019
Tredje kvartil	0,083 *	0,021	0,048 *	0,024	0,048 *	0,024	0,048 *	0,024
Fjerde kvartil	0,016	0,028	-0,021	0,033	-0,021	0,033	-0,021	0,033
Medlem av fagforening (dikotom)	0,084 *	0,024	0,074 *	0,022	0,074 *	0,022	0,074 *	0,022
Lederskjevhets	-0,826 *	0,112	-0,838 *	0,112	-0,838 *	0,112	-0,838 *	0,112
Konstant	2,063	0,076	1,994	0,083	2,000	0,083	2,083	0,084
Random del	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.	Estimat	Robust S.F.
Nivå 3 varians:	0,033	0,004	0,118	0,056	0,122	0,059	0,118	0,055
Nivå 2 varians:	0,033	0,004	0,033	0,004	0,033	0,004	0,032	0,004
Nivå 1 varians:	1,731	0,056	1,720	0,056	1,720	0,056	1,720	0,056
BIC, Deviance:	964485,2	964284,2	962908,4	962544	962896,8	962545	962901,4	962537

$P \leq 0,05 = *$

5. Diskusjon og avslutning

I den siste delen av oppgaven skal jeg diskutere resultatene i lys av det teoretiske rammeverket, hvordan det skiller seg fra den tidligere forskningen på feltet, og med det besvare problemstillingene. Utgangspunktet for prosjektet var å fylle to kunnskapshull reflektert i problemstillingene. Det ene var å undersøke *på hvilken måte et segregeringsmål på individnivå kan skapes, som er meningsfullt å tolke med kryssnasjonale surveydata*. For å svare på det, har jeg undersøkt *hvilke kontekstuelle faktorer på landsnivå, og demografiske forhold på individnivå, som påvirker kjønnsdeling*. Strukturen i diskusjonen vil bli lagt opp slik at jeg først besvarer sistnevnte problemstilling, før jeg besvarer den første. Underveis trekker jeg fram styrker og svakheter ved tolkningene og den metodiske fremgangsmåten. Til slutt vil jeg greie ut om begrensninger ved generalisering og foreslå hva som kan gjøres videre på feltet.

5.1 Oppsummering av funn, nullfunn og forventede sammenhenger

Først skal jeg greie ut om nullfunnene jeg fant i analysen. Deretter skal jeg beskrive de bekreftede funnene, og hvordan en kan forstå dem i lys av tidligere forskning og teori. Likestillingsparadokset går ut på en forventning om at likestilling har en positiv sammenheng med kjønnsdeling, fordi det ikke intuitivt gir mening hvorfor Norge, som gjerne oppfattes som et likestilt land, har høy grad av horisontal kjønnsdeling. Riktignok fantes det en korrelasjon mellom den generelle likestillingsindeksen global gender gap og kjønnsdeling på gruppenivå ($r=0,323$). Jeg fant derimot ikke en sammenheng mellom den generelle likestillingsindeksen og kjønnsdeling i flernivåanalysene presentert i analysedelen. Jeg fant heller ikke en sammenheng mellom delt forsørger-dimensjonen i likestillingspolitikk og kjønnsdeling i analysen. Der var korrelasjonen med kjønnsdeling på gruppenivå enda sterkere ($r=0,427$). Det samme gjelder kvinneandelen i offentlig sektor, som hadde en korrelasjon med kjønnsdeling på gruppenivå på $r=0,445$, men hadde ikke en sammenheng i flernivåanalysene.

Jeg antok at kvinner med barn skulle arbeide i yrkesgrupper med større grad av kjønnsdeling, og forventet derfor at antall personer i husholdningen skulle ha en sammenheng med kjønnsdeling for kvinner med fast samboer. Validiteten til målet på hvorvidt en har barn eller ikke kan diskuteres, men hvis en skal tro resultatene, stemte heller ikke det.

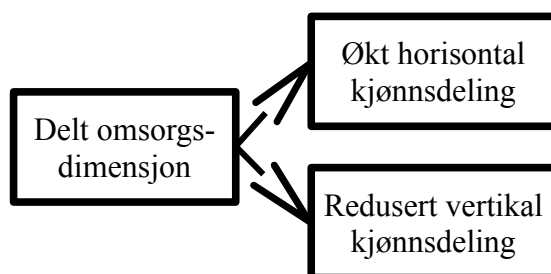
Ut fra hypotesene 4a og 7e forventet jeg at sammenhengen mellom delt omsorgs-dimensjonen og kjønnsdeling skulle være negativ, men forsvinne etter kontroll for arbeidstid, sektor, samboerskap og antall personer i husholdningen. Bak hypotesene lå det en argumentasjonsrekke. Ifølge den mest sammenlignbare tidligere forskningen er en viktig grunn til at

den vertikale kjønnsdelingen oppstår, at kvinner i land som har begrensede muligheter til å dele ansvaret for barneomsorg med fedre, trolig gjør at de ikke har like muligheter til å klatre i karrierestigen som menn. Mødrene i slike land ville trolig i høyere grad enn andre land ha prioritert å arbeide deltid, fordi mennene ikke i like stor grad er tilgjengelige for å hjelpe til med barneomsorgen.

Ut fra den posisjonen antok jeg at økte muligheter for fedre til å ta foreldrepermisjon ville gi redusert kjønnsdeling (H4a). Det viste seg ikke å stemme. Tvert imot var sammenhengen positiv. Jeg antok også at sammenhengen ville forsvinne etter at man kontrollerte for sektor, arbeidstid, antall i husholdningen og samboerskap (H7e). Dette stemte heller ikke. Etter at man kontrollerer for alle sentrale faktorer som den tidligere forskningen har pekt på som angivelig fører til kjønnsdeling, fortsetter den delte omsorgs-dimensjonen i likestillingspolitikk å forklare nasjonale forskjeller i horisontal kjønnsdeling. Som vi ser i den siste tabellen i analysen, er effekten av delt omsorgs-dimensjonen større for land med en gjennomsnittlig andel kvinner i arbeidsmarkedet, enn når man ser de kontekstuelle variablene som uavhengige av hverandre.

I samme tabell fant jeg at variabelen for inntektsulikhetsindeksen ikke var signifikant etter at den ble kontrollert for samtidig som delt omsorgs-dimensjonen i likestillingspolitikk og kvinneandelen av arbeidsmarkedet. Det kan bety at GINI-indeksen fungerer som en bakenforliggende forklaringsvariabel.

Et gjennomgående funn som jeg ikke har stilt hypoteser på, er at den vertikale kjønnsdelingen målt med skjevhetsskoeffisienten i avhengig variabel har en negativ sammenheng med horisontal kjønnsdeling. Jeg fant at den delte omsorgs-dimensjonen hadde en positiv sammenheng med kjønnsdeling. Sett på bakgrunn av at tidligere forskning indikerte at den delte omsorgs-dimensjonen hadde en negativ sammenheng med vertikal kjønnsdeling, er funnet kanskje ikke overraskende. Hvis en forstår den delte omsorgs-dimensjonen som en form for sjanse- og/eller formallikhet, blir kausalpilen som i figur 4 under.



Figur 4: Sammenhengen mellom delt omsorgs-dimensjonen, horisontal og vertikal kjønnsdeling

På bakgrunn av argumentasjonen min basert på tidligere forskning i kapittel 2, forventet jeg at de kontekstuelle variablene ville miste forklaringskraft etter kontroll for de mer nærliggende målene på individnivå. For å si det på en annen måte forventet jeg at forskjellene mellom gruppene skyldtes sammensetningen av individene i gruppene, også kalt komposisjonelle effekter. Det viste seg å ikke stemme. Det er med andre ord internasjonale forskjeller i kjønnsdeling som ikke kan forklares av de demografiske variablene på individnivå jeg har kontrollert for i mine analyser.

Det medfører at, selv om det var kjønnsforskjeller i kjønnsdeling mellom privat og offentlig sektor og deltids- og heltidsarbeidere, er det ikke det som forklarer mesteparten av gruppeforskjellene i kjønnsdeling. De individuelle variablene sektor, arbeidstid, antall i husholdningen, samboerskap og alle interaksjonene med kjønn, forklarte sammenlagt 2,8% av landsforskjellene i kjønnsdeling.

Sammenligningsvis forklarte kvinneandelen av arbeidsstyrken og den delte omsorgsdimensjonen i likestillingspolitikk til sammen 34,8 prosent av gruppeforskjellene i kjønnsdeling. Hvis man skal forstå det store bildet om hvorfor land har ulik grad av horisontal kjønnsdeling, hjelper det derfor ikke stort å henvise til sektorskiller, deltidsarbeid eller effekten av å ha barn. Det gjenspeiles i at kvinneandelen i offentlig sektor heller ikke hadde en signifikant sammenheng med kjønnsdeling. Funnet strider med etablerte oppfatninger innenfor samfunnsforskningen om hva som fører til kjønnsdelte yrkesvalg. Oppfatningene er at den økte kjønnsdelingen oppstår fordi kvinner går inn i offentlig sektor eller arbeider deltid. Det hevdes at det er en konsekvens av opplevelsen av å ikke beherske kombinasjonen av arbeids- og familieliv, og/eller at de i høyere grad enn menn ønsker samvær med barna.

Det skal sies at mangel på forklaring av gruppevariasjonen med de individuelle variablene jeg har brukt kan skyldes to ting. Det ene er at avhengig variabel er basert på yrkesinndelingene i ISCO-indeksen. Yrkesinndelingene i ISCO-indeksen er i store trekk sortert hierarkisk etter det påkrevde ferdighetsnivået i arbeidet. Det andre er at det kan finnes andre variabler på individnivå jeg ikke har hatt mulighet til å kontrollere for i min studie. Ett eksempel er skillet mellom manuelt og ikke-manuelt arbeid. Med kontroll for manuelt og ikke-manuelt arbeid, vil kanskje en større del av variasjonen mellom landene bli forklart som komposisjonelle effekter.

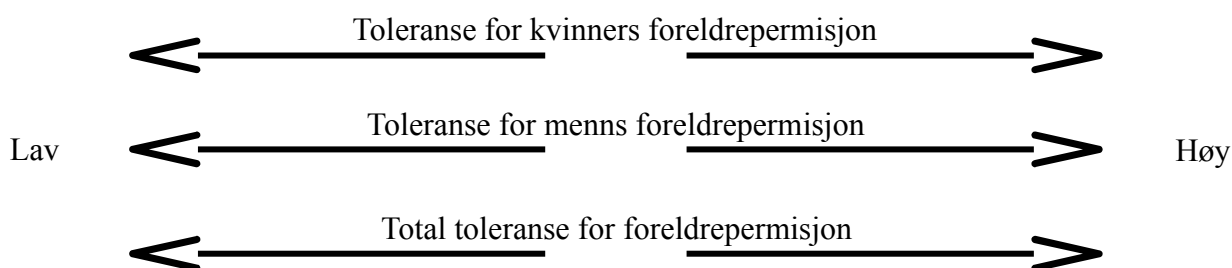
5.2 Tolkning av resultater

Et tilsynelatende paradoks er det at kvinner som har barn ikke påvirker utfallet av kjønnsdelingen, samtidig som fordelingen av foreldrepermisjoner har en sterk sammenheng. Det finnes noen mulige forklaringer, der de første er metodiske. Det kan være at målet på om en har barn eller ei ikke var godt nok, og at en treveis-interaksjon ble benyttet for å ha et tilnærmet mål på

fenomenet. Det kan også hende at variabelen for delt omsorgs-dimensjonen fanger opp nasjonale forskjeller som også kunne vært forklart av noe annet. Her trengs det ytterligere forskning med mer presise mål på om en har barn, og hvor gamle barna er, samt informasjon om foreldrepermisjon på individnivå. Det vil trolig bidra med å se om effekten av det å ha barn varierer fra land til land, og om det er en kryssnivåinteraksjon mellom det å ha barn og fedrepermisjoner på makronivå.

En teoretisk måte å forklare det på er med teorien om rasjonelle valg, og her antar jeg at målene er gode nok. Jevnere fordeling av barneomsorg blir nå sett som en form for institusjonell begrensning for menn. Familien som en sosial institusjon begynner å kreve mer tid av mannen enn uten den institusjonelle begrensningen. Og hvis vi går med på at menn og kvinners preferansehierarki har familieliv og arbeid sortert i ulik rekkefølge, kan beslutningen når en får valget mellom ulike arbeidsplasser være systematisk ulikt for kjønnene. Da blir den logiske forutsetningen at det er egenskaper ved arbeidsplassene som avgjør hvilke yrkesgrupper aktørene velger å arbeide i.

Som følge av måten jeg har valgt å forstå teorien om rasjonelle valg, kan de institusjonelle begrensningene også være forskjellige fra en sosial institusjon til en annen, og for ulike grupper mennesker innenfor samme sosiale institusjon. Det innebærer for det første at arbeidsplasser kan variere i normer om hva som er akseptabel foreldrepermisjon. For det andre kan hver arbeidsplass også ha ulike spilleregler for hva som er tolerert foreldrepermisjon for menn og kvinner. Jeg velger å forstå det som totalt tre dimensjoner som illustrert i figur 5.



Figur 5: Dimensjoner i toleransenivå for foreldrepermisjoner i sosiale institusjoner.

Den totale toleransen forstår jeg som toleransen for foreldrepermisjoner generelt. Den totale toleransen blir påvirket av toleransen for både menn og kvinners foreldrepermisjoner. I tillegg blir toleransene for hvert av kjønnenes foreldrepermisjoner påvirket av den totale toleransen.

Den delte omsorgs-dimensjonen blir da mål på toleranse for både menns foreldrepermisjon og total toleranse. I tillegg måler den toleransen for kvinners foreldrepermisjoner indirekte. Det er fordi variabelen var konstruert basert på lengden på fedrekvoten og lengden på foreldrepermisjon som kan fordeles fritt mellom mor og far. Ved at den totale toleransen også påvirker hvert av

kjønnet, blir toleransen for kvinnenes foreldrepermisjoner påvirket av den delte omsorgsdimensjonen.

Med et slikt apparat med utgangspunkt i teorien om rasjonelle valg kan man begynne å plassere ulike typer arbeidsplasser innenfor toleransedimensjonene. Først vil jeg innta posisjonen til den som ansetter. For en sjef, vil det i høyere grad lønne seg å ha arbeidere som kan arbeide mer, enn de som kan arbeide mindre. Det vil lønne seg desto mer i privat sektor, der sjefen har større profitt av å ansette arbeidere som arbeider mye, enn i offentlig sektor der sjefen har høyere grad av statlig regulerte lønninger. Det fant jeg i mine resultater (figur 2, side 45).

Hvis man inntar en kvinnes posisjon, vil det å ta en jobb ved en arbeidsplass som har lavere toleranse for kvinners foreldrepermisjon, ikke være like attraktivt som å arbeide der det er større aksept for det. Det henger sammen med kvinners mål, som ifølge Mincer (1974:123) skiller seg fra menn sine mål. Hvis kvinner ønsker å prioritere familie i høyere grad enn menn, vil kvinner trolig velge arbeidsplasser med større aksept for deres foreldrepermisjon. Menn ønsker også familieliv, men rangerer arbeidsliv framfor familieliv i sitt preferansehierarki. Derfor kan mennene akseptere å arbeide på arbeidsplasser med lavere toleranse for foreldrepermisjoner.

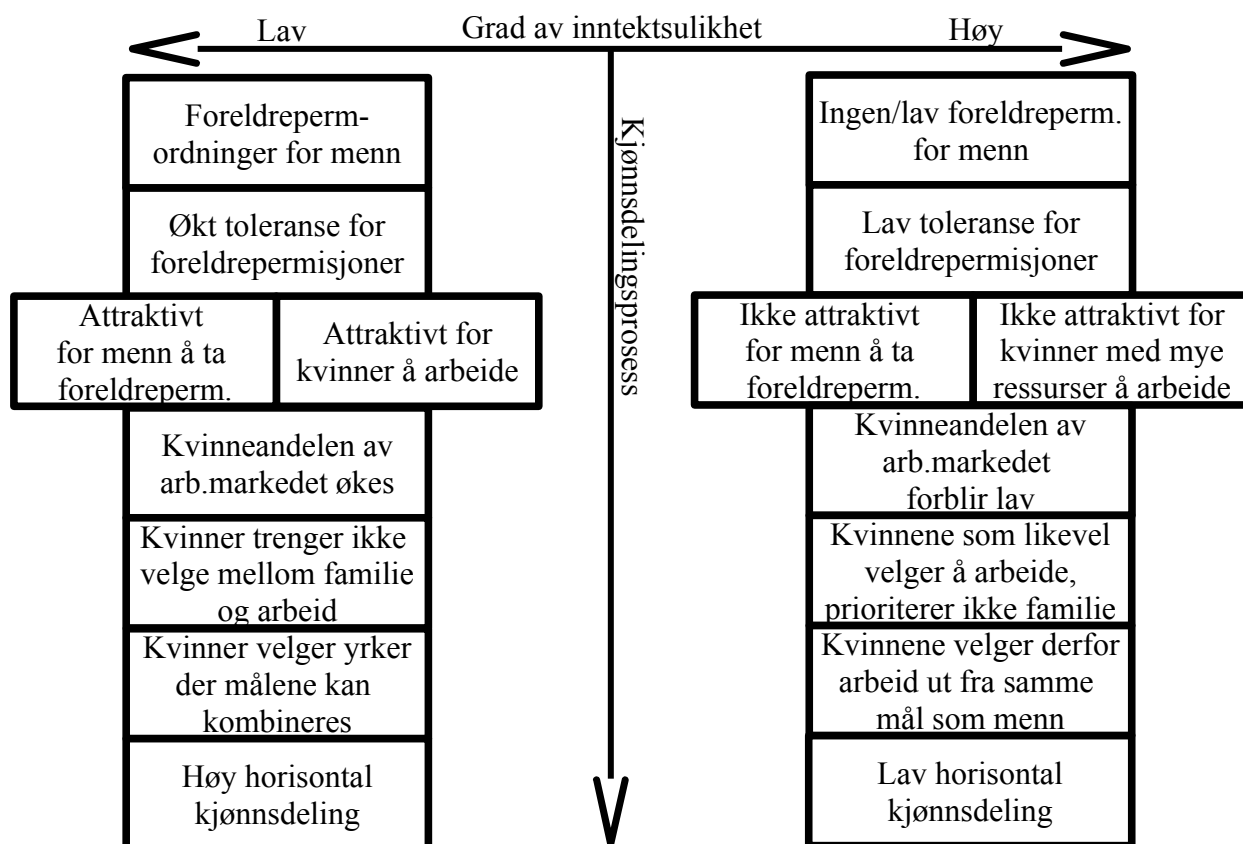
Vi går tilbake til kvinnene, og nå skal jeg gjøre leseren bevisst på at valget ikke nødvendigvis står mellom en arbeidsplass eller en annen. Valget står også mellom å ha en jobb eller ikke. Med en høyere total toleranse for foreldrepermisjon på arbeidsplassene i et land, blir alternativet om å arbeide mer attraktivt. Det henger sammen med at den delte omsorgsdimensjonen har en større effekt for landene med en lavere kvinneandel i arbeidsstyrken. Det tolker jeg til at når den totale toleransen for foreldrepermisjon øker, velger flere kvinner å arbeide, fordi det i høyere grad blir mulig å kombinere familieliv og arbeid.

Når halvparten av arbeidstakerne er kvinner, forsvinner nettopp effekten av den delte omsorgsdimensjonen. Det er her vi til nå har trodd at det er en sammenheng som vi ikke forstår, kalt likestillingsparadokset. Dersom et land har høy grad av toleranse for foreldrepermisjoner, er landet også preget av høy sysselsetting blant kvinner. Det er fordi det er mer attraktivt enn ikke for kvinner å ha en jobb når det kan kombineres med barneomsorg, og at mennene kan bidra i barneomsorgen. Når kvinner velger arbeid, er valgene fortsatt preget av at sorteringen av at målene deres forblir den samme. Det fører til at disse kvinnene velger yrker der de kan kombinere familieliv med arbeidsliv.

For å forstå hvorfor sammenhengen er sterkere i land med lavere kvinneandel av arbeidsstyrken, må vi bryte gruppen kvinner ned i to subgrupper innenfor landene med lavere kvinnedeltakelse i arbeidsmarkedet. Den ene gruppen har god tilgang på ressurser, og den andre ikke. For gruppen med mindre tilgang på ressurser, er det mer rasjonelt å arbeide. Den gruppen som

velger å arbeide har færre ressurser, og har derfor vurdert at deres preferanse for familieliv var uopnåelig. Ifølge teorien vil de derfor forsøke å oppnå deres nest mest ønskede mål. Når det er arbeid som blir målet, og ikke familieliv, vil kvinnene velge som menn, og derfor uavhengig av arbeidsplassens toleranse for foreldrepermisjoner.

For gruppen med god tilgang på ressurser, er det ikke like nødvendig å ha arbeid. Det er fordi deres hovedpreferanse for familieliv er oppnåelig, på grunn av at ressurstilgangen er tilfredstilt. De kvinnene som har god ressurstilgang og likevel velger å ta seg arbeid, vil beholde familielivet som sin hovedprioritet. De vil derfor arbeide i yrker med høyere toleranse for foreldrepermisjoner. Den gruppen er dog mindre i land med høyere inntektsulikhet og lavere toleranse for foreldrepermisjoner enn i land med lavere inntektsulikhet og høyere toleranse for foreldrepermisjoner. Derfor er arbeidsmarkedet i landene med lavere grad av kvinner i arbeid også mindre kjønnsdelte. Begrunnelsen henger for det første sammen med at inntektsulikhetsindeksen mistet sin sammenheng etter at det ble kontrollert for kvinneandelen av arbeidsstyrken og den delte omsorgs-dimensjonen. For det andre var sammenhengen mellom den delte omsorgs-dimensjonen større da kvinneandelen av arbeidsstyrken var mindre.



Figur 6: Teoretisk tolkning av kjønnsdelingsprosessen.

Fordi det er mange posisjoner å innta i den teoretiske tolkningen, har jeg satt opp en ny figur som beskriver kausalsammenhengene jeg har argumentert for at jeg har funnet. Den er vist i figur 6 over. I figuren har jeg satt opp det jeg har forstått som den kausale sammenhengen som fører til enten høy eller lav horisontal kjønnsdeling. I venstre kolonne har jeg satt opp det jeg har forstått som den teoretiske sammenhengen som fører til økt horisontal kjønnsdeling. I kolonnen til høyre har jeg satt opp sammenhengen jeg har kommet fram til at fører til lav kjønnsdeling i land med høyere grad av inntektsulikhet. Land som kan plasseres lenger mot venstre kolonne enn høyre er Norge og de andre skandinaviske landene, Tyskland og Tsjekkia. I kolonnen til høyre finner vi i høyere grad land som India, Tyrkia og Venezuela.

5.2.1 Svakheter

Den teoretiske forklaringen har noen svakheter. Det er at jeg argumenterer for at den delte omsorgs-dimensjonen indirekte indikerer en total toleranse for foreldrepermisjoner i landet. I praksis er det kun brukt informasjon om lengden på foreldrepermisjon som kan deles fritt mellom foreldrene, og hvor mye tid av total foreldrepermisjon som må benyttes av faren. Det mangler derfor et mål på kvinnes foreldrepermisjoner.

Det andre er forklart variasjon. Slik vi så i analysen forklarte kvinneandelen av arbeidsmarkedet og den delte omsorgs-dimensjonen rundt 34 prosent av variasjonen mellom landene. Det betyr at godt over halvparten av landsforskjellene er uforklart. Hvis den teoretiske forklaringen min stemmer, vil trolig en del av det kunne forklares av kvinners foreldrepermisjoner i de ulike landene. Det kan være interessant å kontrollere for i en annen anledning.

Det tredje er at jeg ikke fant store komposisjonelle effekter knyttet til sektor, deltidsarbeid eller samspillet på om en har barn. Det kan enten tyde på at teorien er dårlig egnet til å forstå kjønnsdeling, eller at det er store forskjeller i toleranse for foreldrepermisjoner også innad i sektorene. Et eksempel kan være skillet mellom manuelt og ikke-manuelt arbeid. Her trengs det kanskje mer presise mål, for å se om landsforskjellene kan forklares av komposisjonelle effekter tilknyttet skiller internt i sektorene.

Det fjerde er at jeg ikke har kontrollert for uteliggere eller tunge enheter i materialet. Cooks D eller andre mål som $dfbeta$ er ikke bare enkelt å benytte i flernivåanalyser med tre nivåer. Det finnes Stata-utvidelser som kan finne innflytelsesrike enheter på andre nivå, men jeg har ikke funnet noen som gjør det tilsvarende i modeller med tre nivå. Det betyr at resultatene kan ha blitt påvirket av at ett land har ekstreme verdier i kjønnsdeling eller kontekstuelle variabler.

5.3 Målet på kjønnsdeling

Her tar jeg for meg den første problemstillingen som handler om vurderingen av segregeringsmålet jeg har skapt, og hvorvidt det ga meningsfull informasjon.

Under avsnittet om konstruksjon av variabler på gruppenivå i metodekapittelet viste jeg i tabell 2 på side 33 hvordan avhengig variabel på både individ- og gruppenivå samvarierer med de kontekstuelle variablene jeg har brukt i denne oppgaven. Jeg tok med korrelasjonstabellen for å peke på tre poeng.

Det ene er hvordan de kontekstuelle variablene korrelerer med kjønnsdeling på individnivå før man har kontrollert for datastruktur og individuelle faktorer. I løpet av analysen kommer det fram at det ikke nødvendigvis er de samme kontekstuelle variablene som korrelerer med kjønnsdeling før og etter kontroll for datastruktur og individuelle faktorer. Eksempelvis var den sterkeste korrelasjonen med kjønnsdeling på individnivå, variabelen for kvinneandelen i offentlig sektor, med en korrelasjon på $r=0,15$. Etter det ble satt inn i en flernivåanalyse, var det ikke lenger en signifikant sammenheng. Samtidig viste den delte omsorgs-dimensjonen den svakeste korrelasjonen med avhengig variabel med 0,1, som i flernivåanalysene viste seg å ha en sammenheng. Det betyr at det ikke er en sammenheng mellom de kontekstuelle faktorenes korrelasjoner med kjønnsdeling på individnivå, og de kontekstuelle faktorenes forklaringskraft når man har kontrollert for datastruktur og individuelle faktorer.

Det andre poenget er liknende, men på gruppenivå. Her hadde også andelen av offentlig sektor som er kvinner den sterkeste korrelasjonen med kjønnsdeling på gruppenivå ($r=0,44$). Den svakeste korrelasjonen var igjen delt omsorgs-dimensjonen ($r=0,32$). Skal man ta utgangspunkt i disse sammenhengene for å forstå hva som fører til et kjønnsdelt arbeidsmarked, ville en endt opp med helt andre (og mer misvisende) svar enn hvis man benytter seg av et segregeringsmål på individnivå med kontroll for datastruktur, som jeg i denne oppgaven har gjort. For å samle poenget mitt med et begrep, er dette et eksempel på hvorfor den økologiske feilslutning er viktig å ha forståelse for når man skal forstå kjønnsdeling og andre former for segregering. Sammenhengen mellom kjønnsdeling på gruppenivå og diverse kontekstuelle variabler sier fint lite om hva som fører til kjønnsdelte yrkesvalg på individnivå.

Det tredje poenget er at likestillingsindeksen global gender gap ikke hadde – relativt sett – en sterk korrelasjon med kjønnsdeling. Hvis vi godtar at indeksen måler den gjengse oppfatning av hva likestilling betyr, finnes det for det første, flere andre kontekstuelle variabler som korrelerer mer eller mindre like godt med kjønnsdeling. Deriblant var det ikke en stor forskjell mellom likestillingsindeksens og GINI-indeksens korrelasjon med kjønnsdeling. For det andre oppfatter jeg

det som underlig at nettopp likestilling har blitt trukket ut som den mest sentrale kontekstuelle faktoren som samvarierer med kjønnsdeling, og ikke andre trekk ved kjønnsdelte land. Bare fordi noen likestilte land er kjønnsdelt, betyr det verken at kjønnsdelte land er likestilte, at alle likestilte land er kjønnsdelte, eller at det er et kausalforhold mellom dem.

Segregeringsmålet brukt i oppgaven har vist seg å gi nyttig informasjon om kjønnsdeling på individnivå. Gjennom reliabilitetstester på både gruppe- og individnivå, samt validitetstest på gruppenivå, har jeg vist at variabelen måler stabilt og hensiktsmessig. Gjennom analysene har jeg vist at kjønnsdelingsmålet gir tolkbar informasjon om kjønnsdeling på individnivå, og at målet fungerer godt når det blir benyttet i flernivåanalyser der det også kan fange opp gruppevariasjon. Det er tolkbart fordi koeffisientene kan tolkes direkte, og fordi resultatene ga teoretisk mening.

Det største problemet med avhengig variabel er ikke segregering som sådan. Problemet ligger i at ISCO-indeksens oppbygning ikke er optimalt oppbygd for formålet. Det er flere tilnærmet like yrkestyper som er plassert i forskjellige kategorier i ISCO-indeksen, fordi ferdighetsnivået som kreves er ulik (International Labor Organisation, 2004). Det er med andre ord konstruert etter en hierarkisk struktur. Det har jeg kontrollert for i mine analyser, og det gav tolkbar informasjon. Likevel kan jeg ikke være sikker på at skjevhetkontrollen fanger opp alle nødvendige elementer av vertikal kjønnsdeling. Det er ikke bare ledere som er plassert i «feil» yrkesgruppe. Fordi hele oppbygningen er sentrert rundt nødvendige ferdigheter for å utføre arbeidet, vil noen personer som arbeider på den samme arbeidsplassen, havne i ulike yrkesgrupper i inndelingen. Det er kanskje grunnen til at de forventede komposisjonelle effektene tilknyttet sektor ikke var større.

Jeg har likevel vist at det er mulig å måle segregering på individnivå for å undersøke hva som fører til kjønnsdeling på både gruppe- og individnivå. Målet gir informasjon om hvilken grad to gjensidig utelukkende grupper er jevnt fordelt i to eller flere gjensidig utelukkende kategorier. Antall mulige kategorier har ingen øvre begrensning, men tilnærmingen forutsetter at antall individer totalt er langt høyere enn antall kategorier. Dersom det per kategori er nok enheter til å ikke skape for mange ekstremverdier, vil målet være brukbart.

Et bruksområde kan eksempelvis være gruppene sosial minoritet/majoritet fordelt på kategoriene bydeler eller nabolag. Variabelen kan dermed eksempelvis brukes både som avhengig og uavhengig variabel i regresjonsanalyser med en eller flere nivåer. Et annet bruksområde kan være å bruke den aggregerte versjonen med gruppegjennomsnitt av variabelen som et alternativ til de etablerte segregeringindeksene.

5.4 Generalisering og veien videre

Funnene i oppgaven kan generaliseres til de 49 landene som var med i undersøkelsene. Det betyr at jeg ikke kan si noe om hva som fører til kjønnsdeling på verdensbasis. Det harmoniseringsarbeidet jeg har gjort er imidlertid i den større enden av skalaen av hva som er mulig med kryssnasjonale spørreundersøkelser. Sammen med at segregeringsmålet jeg har konstruert fungerte, bør resultatene veie tungt, fordi sammenlignbare undersøkelser har enten undersøkt få land, eller brukt kjønnsdelingsmål på gruppenivå.

Det jeg har gjort er å se på nasjonale forskjeller for å se på hva som fører til kjønnsdeling. Med det antar jeg at landenes kjønnsdeling utvikler seg etter omtrent samme mønster. Det er dog ikke sikkert at de gjør det. For å kontrollere bort det, er man nødt til å se på endringer over tid. Det er ingen enkel oppgave å finne datamateriale som oppfyller de nødvendige kriteriene for å kunne oppnå den kontrollen. Det må være data på individnivå, og de samme individene må ha tatt undersøkelsen over mange nok år til at en kan se en endring. Det må gjerne være informasjon om utdanningsretning, om en har barn og hvor gamle de er, foreldrepermisjon på individnivå i tillegg til de demografiske variablene jeg fant en sammenheng med i mine analyser.

Det kan finnes andre måter å løse dette på. En variant er å undersøke kjønnsdelingen med pseudo-panel. Innenfor hver yrkesgruppe i hvert utvalg, kan en gi verdien i kjønnsdeling for ett års utvalg, til individene i de samme yrkesgruppene i neste års utvalg. På den måten kan man kontrollere for autokorrelasjonen. Den variasjonen som gjenstår, er den man vil undersøke hva som påvirker. Jeg har ikke gått inn på stivhengighet i min oppgave annet enn innledningsvis, fordi det ikke har vært relevant å kontrollere for det. Stivhengighet kan beskrives enkelt som at kjønnsdelte arbeidsmarkeder er kjønnsdelte fordi de har vært kjønnsdelte tidligere. Å kontrollere for autokorrelasjon vil effektivt også kontrollere for stivhengigheten. Det kunne vært interessant å undersøke i neste omgang.

6. Litteratur

- Aakvaag, G. (2015). *Likestillingsparadoks og biologi*. Hentet den 02.06.16 fra https://morgenbladet.no/debatt/2015/likestillingsparadoks_og_biologi
- Acock, A. C. (2014). *A gentle introduction to Stata*. (Fourth ed.). College Station, Texas, USA: Stata press.
- Andrén, B., og Schmidt, P. (2005). Statistics in focus. Populations and social conditions. *Education in Europe Key statistics 2002-2003*. ISSN 1024-4352.
- Blackburn, R. M., Browne, J., Brooks, B., og Jarman, J. (2002). Explaining gender segregation. *The British journal of sociology*, 53(4), 513-536.
- Blair-Loy, M., og Wharton, A. S. (2004). Organizational commitment and constraints on work-family policy use: Corporate flexibility policies in a global firm. *Sociological Perspectives*, 47(3), 243-267.
- Bongard, T., og Røskoft, E. (2010). Det biologiske mennesket - individer og samfunn i lys av evolusjon. Trondheim: Tapir Akademisk Forlag.
- Bryan, M. L., og Jenkins, S. P. (2015). Multilevel modelling of country effects: a cautionary tale. *European Sociological Review*. Vol 32(1). 3-22
- Central Intelligence Agency. (2013) *The World Factbook. 2013-14*. Washington, DC. Hentet 10.05.16 <https://www.cia.gov/library/publications/the-world-factbook/index.html>
- Chang, M. L. (2000). The Evolution of Sex Segregation Regimes 1. *American Journal of Sociology*, 105(6), 1658-1701.
- Charles, M. (1992). Cross-national variation in occupational sex segregation. *American Sociological Review*, 483-502.
- Charles, M. (2003). Deciphering Sex Segregation. Vertical and Horizontal Inequalities in Ten National Labor Markets. *Acta Sociologica*. Vol. 46(4). 267-287.
- Charles, M., og Grusky, D. B., (2004). *Occupational Ghettos. The Worldwide Segregation of Women and Men*. Stanford, CA: Stanford University Press.
- Charles, L. D., og Williams, S. (2006). *Early childhood care and education in the Caribbean. (CARICOM states)*. Background paper prepared for the Education for All Global Monitoring Report 2007. Strong Foundations: early childhood care and education.
- Dziak, J. J., Coffman, D. L., Lanza, S. T., og Li, R. (2012). Sensitivity and specificity of information criteria. *The Methodology Center, Technical Report Series*. 1-20
- Friedman, D., og Hechter, M. (1988). The contribution of rational choice theory to macrosociological research. *Sociological Theory*. Vol. 6(2). 201-218.
- GESIS. (udatert). *Content of Module Pages*. Hentet den 31.05.16 fra <http://www.gesis.org/issp/modules/>
- Giddens, A. (2006). *Sociology*. (Fifth edition). Cambridge, UK: Polity Press
- Hakim, C. (1992). Explaining Trends in Occupational Segregation: The Measurement, Causes, and Consequences of Sexual Division of Labour. *European Sociological Review*. Vol. 8(2).
- Harzing, A. W., Reiche, B. S., og Pudelko, M. (2013). Challenges in international survey research: a review with illustrations and suggested solutions for best practice. *European Journal of International Management*, Vol. 7(1), 112-134.
- Hoffmeyer-Zlotnik, J. H., og Warner, U. (2013). *Harmonising demographic and socio-economic variables for cross-national comparative survey research*. Springer Science & Business Media.
- HDX. (2014). *The Global Gender Gap Index*. Hentet 22.09.15 fra <https://data.hdx.rwllabs.org/dataset/global-gender-gap-index-world-economic-forum>
- Hoffmeyer-Zlotnik, J. H., og Warner, U. (2013). *Harmonising demographic and socio-economic variables for cross-national comparative survey research*. Springer Science & Business Media.

- Iceland, J., Weinberg, D. H., og Steinmetz, E. (2002). Racial and ethnic residential segregation in the United States 1980-2000. Washington DC: U.S. Government Printing Office. Bureau of Census.
- Ihle, O-M., og Eia, H. (2010). *Likestillingsparadokset*. Hentet 02.06.16 fra <http://www.aftenposten.no/meninger/kronikker/Likestillingsparadokset-6264501.html>
- International Labour Organization. (2004). *ISCO-88*. Hentet 28.05.16 fra <http://www.ilo.org/public/english/bureau/stat/isco/isco88/>
- International Labor Organisation. (udatert). *Databases and Subjects*. Hentet 19.04.16 fra http://www.ilo.org/ilostat/faces/help_home/data_by_subject?_afzLoop=365394351783787#%40%3F_afzLoop%3D365394351783787%26_adf.ctrl-state%3Dxjhhna9xy_251
- ISSP Demographic Methods Group. (2001). *The ISSP Standard Background Variables (June 2001)*. Hentet 09.05.16 fra http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/issp/members/codinginfo/bv2001_20060425.pdf
- ISSP Demographic Methods Group. (2012). *ISSP Background Variables Questionnaire (BVQ)*. Hentet den 16.03.16 fra http://www.gesis.org/fileadmin/upload/dienstleistung/daten/umfragedaten/issp/members/codinginfo/BV_questionnaire_for_issp2013.pdf
- ISSP Research Group (2002). *International Social Survey Programme: Social Inequality III - ISSP 1999*. GESIS Data Archive, Cologne. ZA3430 Data file Version 1.0.0, doi:10.4232/1.3430
- ISSP Research Group (2003). *International Social Survey Programme: Environment II - ISSP 2000*. GESIS Data Archive, Cologne. ZA3440 Data file Version 1.0.0, doi:10.4232/1.3440
- ISSP Research Group (2003). *International Social Survey Programme: Social Relations and Support Systems / Social Networks II - ISSP 2001*. GESIS Data Archive, Cologne. ZA3680 Data file Version 1.0.0, doi:10.4232/1.3680
- ISSP Research Group (2008). *International Social Survey Programme: Role of Government IV - ISSP 2006*. GESIS Data Archive, Cologne. ZA4700 Data file Version 1.0.0, doi:10.4232/1.4700
- ISSP Research Group (2009). *International Social Survey Programme: Leisure Time and Sports - ISSP 2007*. GESIS Data Archive, Cologne. ZA4850 Data file Version 2.0.0, doi:10.4232/1.10079
- ISSP Research Group (2012). *International Social Survey Programme: National Identity II - ISSP 2003*. GESIS Data Archive, Cologne. ZA3910 Data file Version 2.1.0, doi:10.4232/1.11449
- ISSP Research Group (2012). *International Social Survey Programme: Citizenship - ISSP 2004*. GESIS Data Archive, Cologne. ZA3950 Data file Version 1.3.0, doi:10.4232/1.11372
- ISSP Research Group (2012). *International Social Survey Programme: Religion III - ISSP 2008*. GESIS Data Archive, Cologne. ZA4950 Data file Version 2.2.0, doi:10.4232/1.11334
- ISSP Research Group (2012). *International Social Survey Programme: Social Inequality IV - ISSP 2009*. GESIS Data Archive, Cologne. ZA5400 Data file Version 3.0.0, doi:10.4232/1.11506
- ISSP Research Group (2012). *International Social Survey Programme: Environment III - ISSP 2010*. GESIS Data Archive, Cologne. ZA5500 Data file Version 2.0.0, doi:10.4232/1.11418
- ISSP Research Group (2013). *International Social Survey Programme: Family and Changing Gender Roles III - ISSP 2002*. GESIS Data Archive, Cologne. ZA3880 Data file Version 1.1.0, doi:10.4232/1.11564
- ISSP Research Group (2013). *International Social Survey Programme: Work Orientation III - ISSP 2005*. GESIS Data Archive, Cologne. ZA4350 Data file Version 2.0.0, doi:10.4232/1.11648
- ISSP Research Group (2015). *International Social Survey Programme: Health and Health Care - ISSP 2011*. GESIS Data Archive, Cologne. ZA5800 Data file Version 3.0.0, doi:10.4232/1.12252
- ISSP Research Group (2015). *International Social Survey Programme: National Identity III - ISSP 2013*. GESIS Data Archive, Cologne. ZA5950 Data file Version 2.0.0, doi:10.4232/1.12312

- ISSP Research Group (2016). *International Social Survey Programme: Family and Changing Gender Roles IV - ISSP 2012*. GESIS Data Archive, Cologne. ZA5900 Data file Version 3.0.0, doi:10.4232/1.12339
- Jann, B. (2005). *duncan2: Stata module to calculate dissimilarity index. S447202*. Boston College Department of Economics. <https://ideas.repec.org/c/boc/bocode/s447202.html>
- Jonsson, J. O. (1999). Explaining Sex Differences in Educational Choice. An Empirical Assessment of a Rational Choice Model. *European Sociological Review*. Vol. 15(4). 391-404.
- Kammerman, S. (2006). A global history of early childhood education and care. *Background paper for the Education for All Global Monitoring Report 2007*. Strong Foundations: early childhood care and education.
- Korpi, W., Ferrarini, T., og Englund, S. (2013). Women's opportunities under different family policy constellations: Gender, class, and inequality tradeoffs in western countries re-examined. *Social Politics: International Studies in Gender, State & Society*, Vol. 20(1), 1-40.
- Likestillingsloven. Lov 21. juni 2013 nr. 59 om likestilling mellom kjønnene.
- Maas, C. J., og Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, Vol. 1(3), 86-92.
- Mandel, H., og Semyonov, M. (2006). A Welfare State Paradox: State Interventions and Women's Employment Opportunities in 22 Countries. *American journal of sociology*, Vo. 111(6), 1910-1949.
- Melkas, H., og Anker, R. (2001). Occupational Segregation by Sex in Nordic Countries: An Empirical Investigation. I M. F. Loutfi (Red.). *What is Equality and how do we get there? Women, Gender, and Work*. (Første utgave) Geneve: International Labour Office.
- Mincer, J. (1974). *Schooling, Experience and Earnings*. Cambridge: National Bureau of Economic Research, Inc.
- Ministry of Education, Republic of China. (2012/2013). *Education in Taiwan*.
- Myers, R. (2000). Early Childhood Care and Development. *Dakar, Senegal: World Economic Forum: Education for All 2000 Assessment*.
- Nermo, M. (2000). Models of cross-national variation in occupational sex segregation. *European societies*, Vol. 2(3), 295-333.
- Newson, R. (2001). Stata tip 13: generate and replace use the current sort order. *The Stata Journal*. Vol. (1). 1-2.
- NOU 2012:15 (2012). *Politikk for likestilling*. Oslo: Departementenes servicesenter, Informasjonsforvaltning.
- Pinker, S. (2008). *The sexual paradox: Men, women and the real gender gap*. New York: Simon and Schuster.
- Reisel, L. (2014a). Kjønnssdeling på tvers. I L. Reisel og M. Teigen (Red.), *Kjønnssdeling og etniske skiller på arbeidsmarkedet*. Oslo: Gyldendal Akademisk Forlag.
- Reisel, L. (2014b). Kjønnssdelte utdanningsvalg. I L. Reisel og M. Teigen (Red.), *Kjønnssdeling og etniske skiller på arbeidsmarkedet*. Oslo: Gyldendal Akademisk Forlag.
- Reisel, L., og Teigen, M. (2014). Det kjønnssdelte arbeidsmarkedet. I L. Reisel og M. Teigen (Red.), *Kjønnssdeling og etniske skiller på arbeidsmarkedet*. Oslo: Gyldendal Akademisk Forlag.
- Reisel, L., og Teigen, M. (2015). *Fakta om likestillingsparadokset*. Hentet 27.05.16 fra https://morgenbladet.no/debatt/2015/fakta_om_likestillingsparadokset
- Robinson, W. S. (2009). Ecological correlations and the behavior of individuals. *International journal of epidemiology*, Vol. 38(2), 337-341.
- Rostgaard, T. (2004). *Early Childhood and Family Policy Series: Family Support Policy in Central and Eastern Europe – A Decade and a Half of Transition*.
- Schiefloe, P. M. (2011). *Mennesker og samfunn. Innføring i sosiologisk forståelse (Andre utgave)*. Bergen: Fagbokforlaget.

- Skog, O.-J. (2004). *Å forklare sosiale fenomener. En regresjonsbasert tilnærming. Revidert og utvidet utgave.* (Andre utgave). Oslo: Gyldendal Norsk Forlag AS.
- Snijders, T. A. B. (2005). Power and sample size in multilevel linear models. I B. S. Everitt og D. C. Howell (Red.). *Encyclopedia of statistics in behavioral science: 1570-1573*. Chichester, U.K.: Wiley.
- Snijders, T. A. B., og Bosker, R. J. (2012). *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling.* (2nd. Edition). Thousand Oaks, CA: SAGE Publications Inc.
- Statistisk sentralbyrå. (2012). *Samboere, 2011*. Hentet 28.05.16 fra <https://www.ssb.no/befolkning/statistikker/samboer/aar/2012-05-31>
- Statistisk sentralbyrå. (2016). *Nøkkeltall for likestilling*. Hentet 28.05.16 fra <https://www.ssb.no/befolkning/nokkeltall/likestilling>
- Støren, L. A., og Arnesen, C. Å. (2003). Et kjønnsdelt utdanningssystem. *Utdanning 2003 – Ressurser, rekruttering og resultater*. Statistiske analyser nr. 60. Statistisk sentralbyrå.
- Swallow, W. H., og Monahan, J. F. (1984). Monte Carlo Comparison of ANOVA, MIVQUE, REML, and ML Estimators of Variance Components. *Technometrics*. Vol. 26(1). 44-57.
- Taratukhina, M. S., Polyakova, M. N., Berezina, T. A., Notkina, N. A., Sheraizina, R. M., og Borovkov, M, I. (2006). *Early childhood care and education in the Russian Federation*. Background paper for the Education for All Global Monitoring Report 2007. Strong Foundations: early childhood care and education.
- The World Bank. (1999-2014). *World Development Indicators: GINI index* [Data fil] <http://data.worldbank.org/indicator/SI.POV.GINI>
- UNDP HDR Statistics (udatert). United Nations Development Programme. Gender Inequality
- UNESCO. (2010). Europe and North America. *Early childhood care and education regional report*.
- UNESCO International Bureau of Education (IBE). (2006a). Philippines Early Childhood Care and Education (ECCE) programmes. *Country profile prepared for the Global Monitoring Report 2007. Strong Foundations: Early Childhood Care and Education*.
- UNESCO International Bureau of Education (IBE). (2006b). South Africa Early Childhood Care and Education (ECCE) programmes. *Country profile prepared for the Global Monitoring Report 2007. Strong Foundations: Early Childhood Care and Education*.
- UNESCO International Bureau of Education (IBE). (2006c). Ukraine Early Childhood Care and Education (ECCE) programmes. *Country profile prepared for the Global Monitoring Report 2007. Strong Foundations: Early Childhood Care and Education*.
- UNESCO International Bureau of Education (IBE). (2006d). Uruguay Early Childhood Care and Education (ECCE) programmes. *Country profile prepared for the Global Monitoring Report 2007. Strong Foundations: Early Childhood Care and Education*.
- UNESCO. (2015). *Regional overview: Sub-Saharan Africa*.
- Warner, U. (2009). *Die messung von Einkommen in der international vergleichenden Umfragenforschung*. Mannheim, Tyskland: Forschung Raum und Gesellschaft. e.V.
- World Economic Forum. (udatert). *The Global Gender Gap Index 2014: Measuring the Global Gender Gap*. Hentet 10.05.16 fra <http://reports.weforum.org/global-gender-gap-report-2014/part-1/the-global-gender-gap-index-2014/>

7. Vedlegg

Merk at all syntaks i vedlegg ble gjort før datasettene var slått sammen. Det er grunnen til at jeg ikke har sortert etter årstall samtidig som land der det skjer.

Syntaks for yrkesinndeling

```
gen yrkesgrupper=.
replace yrkesgrupper=1 if ISCO88>=1 & ISCO88<=999
replace yrkesgrupper=2 if ISCO88>=1100 & ISCO88<=1199
replace yrkesgrupper=3 if ISCO88>=1200 & ISCO88<=1299
replace yrkesgrupper=4 if ISCO88>=1300 & ISCO88<=1399
replace yrkesgrupper=5 if ISCO88>=2100 & ISCO88<=2199
replace yrkesgrupper=6 if ISCO88>=2200 & ISCO88<=2299
replace yrkesgrupper=7 if ISCO88>=2300 & ISCO88<=2399
replace yrkesgrupper=8 if ISCO88>=2410 & ISCO88<=2419
replace yrkesgrupper=9 if ISCO88>=2420 & ISCO88<=2429
replace yrkesgrupper=10 if ISCO88>=2430 & ISCO88<=2439
replace yrkesgrupper=11 if ISCO88>=2440 & ISCO88<=2449
replace yrkesgrupper=12 if ISCO88>=2450 & ISCO88<=2459
replace yrkesgrupper=13 if ISCO88>=2460 & ISCO88<=2469
replace yrkesgrupper=14 if ISCO88>=2470 & ISCO88<=2479
replace yrkesgrupper=15 if ISCO88>=3100 & ISCO88<=3199
replace yrkesgrupper=16 if ISCO88>=3200 & ISCO88<=3299
replace yrkesgrupper=17 if ISCO88>=3300 & ISCO88<=3399
replace yrkesgrupper=18 if ISCO88>=3410 & ISCO88<=3419
replace yrkesgrupper=19 if ISCO88>=3420 & ISCO88<=3429
replace yrkesgrupper=20 if ISCO88>=3430 & ISCO88<=3439
replace yrkesgrupper=21 if ISCO88>=3440 & ISCO88<=3449
replace yrkesgrupper=22 if ISCO88>=3450 & ISCO88<=3459
replace yrkesgrupper=23 if ISCO88>=3460 & ISCO88<=3469
replace yrkesgrupper=24 if ISCO88>=3470 & ISCO88<=3479
replace yrkesgrupper=25 if ISCO88>=3480 & ISCO88<=3489
replace yrkesgrupper=26 if ISCO88>=4000 & ISCO88<=4099
replace yrkesgrupper=27 if ISCO88>=4100 & ISCO88<=4199
replace yrkesgrupper=28 if ISCO88>=4200 & ISCO88<=4299
replace yrkesgrupper=29 if ISCO88>=5100 & ISCO88<=5199
replace yrkesgrupper=30 if ISCO88>=5200 & ISCO88<=5299
replace yrkesgrupper=31 if ISCO88>=6100 & ISCO88<=6199
replace yrkesgrupper=32 if ISCO88>=6200 & ISCO88<=6299
replace yrkesgrupper=33 if ISCO88>=7100 & ISCO88<=7199
replace yrkesgrupper=34 if ISCO88>=7200 & ISCO88<=7299
replace yrkesgrupper=35 if ISCO88>=7300 & ISCO88<=7399
replace yrkesgrupper=36 if ISCO88>=7410 & ISCO88<=7419
replace yrkesgrupper=37 if ISCO88>=7420 & ISCO88<=7429
replace yrkesgrupper=38 if ISCO88>=7430 & ISCO88<=7439
replace yrkesgrupper=39 if ISCO88>=7440 & ISCO88<=7449
replace yrkesgrupper=40 if ISCO88>=7500 & ISCO88<=7599
replace yrkesgrupper=41 if ISCO88>=7900 & ISCO88<=7999
replace yrkesgrupper=42 if ISCO88>=8000 & ISCO88<=8099
replace yrkesgrupper=43 if ISCO88>=8100 & ISCO88<=8199
replace yrkesgrupper=44 if ISCO88>=8200 & ISCO88<=8299
replace yrkesgrupper=45 if ISCO88>=8300 & ISCO88<=8399
replace yrkesgrupper=46 if ISCO88>=9100 & ISCO88<=9199
replace yrkesgrupper=47 if ISCO88>=9200 & ISCO88<=9299
replace yrkesgrupper=48 if ISCO88>=9300 & ISCO88<=9399
```

```
label define yrkesgrupper 1 "Armed forces" 2 "Legislators, senior officials and
managers" 3 "Corporate managers" 4 "General managers" 5 "Physical, mathematical
```

and engineering science professionals" 6 "Life science and health professional"
7 "Teaching professionals" 8 "Business professionals" 9 "Legal professionals" 10
"Archivists, librarians and related information professionals" 11 "Social
science and related professionals" 12 "Writers and creative performing artists"
13 "Religious professionals" 14 "Public service administrative professionals" 15
"Physical and engineering science associate professionals" 16 "Life science and
health associate professionals" 17 "Teaching associate professionals" 18
"Finance and sales associate professionals" 19 "Business services agents and
trade brokers" 20 "Administrative associate professionals" 21 "Customs, tax and
related government associate professionals" 22 "Police inspectors and
detectives" 23 "Social work associate professionals" 24 "Artistic, entertainment
and sports associate professionals" 25 "Religious associate professionals" 26
"Clerks" 27 "Office Clerks" 28 "Customer service clerks" 29 "Personal and
protective services workers" 30 "Models, salespersons and demonstrators" 31
"Market-oriented skills agricultural and fishery workers" 32 "Subsistence
agricultural and fishery workers" 33 "Extraction and building trade workers" 34
"Metal, machinery and related trades workers" 35 "Precision, handicraft,
printing and related trades workers" 36 "Food processing and related trades
workers" 37 "Wood treaters, cabinet-makers and related trades workers" 38
"Textile, garment and related trades workers" 39 "Pelt, leather and shoemaking
trades workers" 40 "Metal worker general" 41 "Master craftsman, supervisor" 42
"Plant and machine operators and assemblers" 43 "Stationary plant and related
operators" 44 "Machine operators and assemblers" 45 "Drivers and mobile plant
operators" 46 "Sales and services elementary occupations" 47 "Agricultural,
fishery and related labourers" 48 "Labourers in mining, construction,
manufacturing and transport"

label values yrkesgrupper yrkesgrupper

*lage missingvariabel

gen missvar=1

replace missvar=. if yrkesgrupper==.n | yrkesgrupper==.a | yrkesgrupper==. |
SEX==.a | SEX==.n | SEX==.

*lage prosentvariabel innad i yrkesgruppene

bysort land yrkesgrupper SEX missvar, sort: gen freq = _N if missvar==1

bysort land yrkesgrupper missvar: generate pc_yrkesgrupper=100*freq/_N if
missvar==1

replace pc_yrkesgrupper=. if missvar==.

*lage totalprosentvariabel totalt i utvalget.

*totalt antall kvinner og menn i land dersom de ikke er missing i yrkesvar

bysort land SEX missvar: gen B=_N if missvar==1

*totalt antall folk i land dersom de ikke er missing i yrkesvar

bysort missvar land : gen BB=_N if missvar==1

replace BB=. if missvar==.

*dele antall menn og kvinner i land på antall folk i land

bysort missvar land yrkesgrupper : gen BBB=B/BB if missvar==1

replace BBB=BBB*100 if missvar==1

*trekke prosentvar fra totalprosent, kvadrere og kvadratrot for å fjerne
minustegn, dele på 10 og avrunde til 2 desimaler.

gen y=pc_yrkesgrupper-BBB

replace y=y^2

replace y=sqrt(y)

replace y=y/10

replace y=round(y, .01)

Syntaks for sletting av doble utvalg.

```
bysort landaar: gen ID1=_N
bysort landrunde: gen ID2=_N
gen skille=1
replace skille=. if ID2<ID1 & issprunde!=DATEYR

*rette opp i falske positive
replace skille=1 if land==100 & issprunde==2007
replace skille=1 if land==372 & issprunde==2002
replace skille=1 if land==484 & issprunde==2005
replace skille=1 if land==616 & issprunde==2004
replace skille=1 if land==616 & issprunde==2008
replace skille=1 if land==616 & issprunde==2012
drop if skille==.
```

Syntaks for konstruksjon av inntektsvariabler.

```
gen income=.
gen rincome=.
foreach x of varlist *_INC {
    replace income=`x' if !missing(`x')
}
foreach z of varlist *_RINC {
    replace rincome=`z' if !missing(`z')
}
```

Syntaks for å sette missing til logiske verdier i inntektsvariablene.

```
*income=rincome hvis ugift og ikke bor med partner og bor alene.
replace income=rincome if missing(income) & cohab==2 & missing(spwrkst) &
marital!=1 & marital!=4 & !missing(rincome)
*rincome=income hvis man er ugift/samboer og bor alene.
replace rincome=income if cohab==2 & missing(spwrkst) & marital!=1 & marital!=4
& missing(rincome) & !missing(income)
*inc=rinc hvis gift/samboer og partner ikke har arb
replace income=rincome if missing(income) & marital==1 & !missing(rincome) &
spwrkst!=1 & spwrkst!=2 & spwrkst!=3 & spwrkst!=. & spwrkst!=.n & spwrkst!=.a &
hhcycle>=5 & !missing(hhcycle)
replace income=rincome if income==. & cohab==1 & !missing(rincome) & spwrkst!=1
& spwrkst!=2 & spwrkst!=3 & !missing(spwrkst) & hhcycle>=5 & !missing(hhcycle)

*sette personlig inntekt til 0 hvis man ikke er i arbeid.
replace rincome=0 if wrkst!=1 & wrkst!=2 & wrkst!=3 & wrkst!=.n & !
missing(wrkst) & missing(rincome)
*sette husholdningsinntekt til 0 hvis man er gift og partneren er arbeidsledig.
replace income=0 if wrkst!=1 & wrkst!=2 & wrkst!=3 & !missing(wrkst) &
missing(income) & spwrkst!=1 & spwrkst!=2 & spwrkst!=3 & !missing(spwrkst) &
marital==1
replace income=0 if wrkst!=1 & wrkst!=2 & wrkst!=3 & !missing(wrkst) &
missing(income) & spwrkst!=1 & spwrkst!=2 & spwrkst!=3 & !missing(spwrkst) &
cohab==1

*kjøre dette igjen i tilfelle man får flere 0-verdier.
*income=rincome hvis ugift og ikke bor med partner og bor alene.
replace income=rincome if missing(income) & cohab==2 & missing(spwrkst) &
marital!=1 & marital!=4 & !missing(rincome)
*rincome=income hvis man er ugift/samboer og bor alene.
replace rincome=income if cohab==2 & missing(spwrkst) & marital!=1 & marital!=4
& missing(rincome) & !missing(income)
*inc=rinc hvis gift/samboer og partner ikke har arb
replace income=rincome if missing(income) & marital==1 & !missing(rincome) &
spwrkst!=1 & spwrkst!=2 & spwrkst!=3 & spwrkst!=. & spwrkst!=.n & spwrkst!=.a &
hhcycle>=5 & !missing(hhcycle)
replace income=rincome if income==. & cohab==1 & !missing(rincome) & spwrkst!=1
& spwrkst!=2 & spwrkst!=3 & !missing(spwrkst) & hhcycle>=5 & !missing(hhcycle)
```

Syntaks for å lage kvartiler av inntektsvariablene innad i utvalg.

```
egen hinntektk = xtile(income), by(land) n(4)
egen pinntektk = xtile(rincome), by(land) n(4)
```