

Haraldsvik, Mathias
Skogvold, Ole-Andreas

En empirisk analyse av lønnsgapet mellom norske kvinner og menn

Bacheloroppgave i Samfunnsøkonomi
Veileder: Bjarne Strøm
Mai 2019

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

Haraldsvik, Mathias
Skogvold, Ole-Andreas

En empirisk analyse av lønnsgapet mellom norske kvinner og menn

Bacheloroppgave i Samfunnsøkonomi
Veileder: Bjarne Strøm
Mai 2019

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



SØK2901 – NTNU Trondheim

En empirisk analyse av lønnsgapet mellom norske kvinner og menn

Bacheloroppgave i samfunnsøkonomi

Kandidatnummer: 10021 & 10014

13.05.2019

1. INNLEDNING	2
1.1 MOTIVASJON FOR VALG AV TEMA.....	2
1.2 VÅRE PROBLEMSTILLINGER.....	3
1.3 SAMMENDRAG	4
2. TIDLIGERE LITTERATUR OG TEORETISK RAMMEVERK.....	4
2.1 KAPITTELINNLEDNING.....	4
2.2 TEORETISK RAMMEVERK.....	5
2.3 RESULTATER I TIDLIGERE LITTERATUR.....	8
2.3.1 <i>Relevant statistikk</i>	8
2.3.2 <i>Resultat fra tidligere rapporter og analyser</i>	10
2.4 FREMGANGSMÅTE OG UTFORMING AV ØKONOMETRISKE MODELLER I TIDLIGERE ANALYSER.....	11
2.4.1 <i>Overgang fra teori til estimerbare økonometrisk modeller</i>	12
2.5 OPPSUMMERING AV KAPITTEL 2.....	12
3. PRESENTASJON AV EMPIRI.....	13
3.1 OMTALE OG GJENNOMGANG AV DATA	13
3.2 BESKRIVELSER AV VARIABLENE BRUKT I ANALYSEN	13
3.3 DESKRIPTIV STATISTIKK FOR MATERIALET	15
3.4 FORDELER OG BEGRENSNINGER VED DATASETET	17
3.5 OPPSUMMERING AV KAPITTEL 3.....	18
4. REGRESJONSANALYSE	19
4.1. HOVEDSPØRSMÅL 1.....	19
4.2. HOVEDSPØRSMÅL 2.....	21
4.3. TILLEGGSPØRSMÅL.....	26
5. OPPSUMMERING OG KONKLUSJON.....	29
5.1. OPPSUMMERING.....	29
5.2. HOVEDRESULTATER FRA EMPIRISK ANALYSE - KONKLUSJON	29
5.3. BEGRENSNINGER VED ANALYSEN - FORSLAG TIL UTVIDELSER/VIDERE FORSKNING.....	31
6. LITTERATURLISTE	33
7. APPENDIX – KOMPLETTE REGRESJONSRESULTATER.....	34

1. Innledning

1.1 Motivasjon for valg av tema

Likestilling mellom kvinner og menn er en sentral del av den offentlige politiske diskusjon og har i lang tid vært et aktuelt tema. Innen området for likestilling finner vi flere viktige spørsmål, men ett av spørsmålene som har fått mest oppmerksomhet, både fra politikere og media, er spørsmålet om likelønn og lønnsgapet. I 2008, under Stoltenberg-regjeringen, utarbeidet likelønnskommisjonen en omfattende rapport for å gi en samlet beskrivelse av lønnsforskjellene mellom kvinner og menn i Norge. Etter den tid, uten at det nødvendigvis har noen sammenheng med rapporten, har spørsmål rundt likelønn blusset opp i media, og derav også blant politikere.

Fra likestillings- og diskrimineringsloven ser vi at Norge har tydelige lover når det kommer til likelønn:

§ 34. Lik lønn for arbeid av lik verdi

«Kvinner og menn i samme virksomhet skal ha lik lønn for samme arbeid eller arbeid av lik verdi. Lønnen skal fastsettes på samme måte, uten hensyn til kjønn.»

(Likestillings- og diskrimineringsloven, 2018, §34).

Til tross for denne loven kan det relativt ofte observeres påstander om at Norge har et problem når det kommer til likelønn - at kvinner og menn ikke får likt betalt. Et eksempel på en prominent driver bak påstander om dette er det årlige arrangementet "Equal Pay Day", som i 2018 påsto at fra 18. Oktober arbeidet kvinner «gratis ut året» (UDF, 2017). Med dette menes det at 18. Oktober har menn tjent det kvinner vil bruke hele året på å tjene. En viktig ting å poengtere er at i de aller fleste tilfeller hvor likelønn er diskutert i media og blant politikere, blir det er det gjennomsnittslønnen som blir sammenlignet, ikke lønn justert for ulikheter mellom menn og kvinner innad jobbmarkedet.. Dette kan skape forvirring blant folk flest.

Gjennom den empiriske analysen skal vi undersøke om det er forskjell på gjennomsnittslønn mellom kvinner og menn i Norge (gitt et datasett), og samtidig presentere det som kan tenkes å være forklaringene til den eventuelle forskjellen.

Som fremtidige samfunnsøkonomer er det av både relevans og interesse for oss å få en dypere forståelse for den økonomiske siden av likestillingsdiskusjonen. Ved å gjennomføre en empirisk analyse av lønnsforskjellene mellom menn og kvinner i Norge har vi som mål å tilordne oss analytiske ferdigheter som vil tillate oss å undersøke informasjon med statistiske metoder, for så å tolke resultatene gjennom økonomisk intuisjon. Vi ønsker også å oppnå en generelt dypere forståelse for temaet om likestilling - spesifikt likelønn og mulige forklaringer til at det stadig finnes forskjell i gjennomsnittslønn mellom kvinner og menn.

1.2 Våre problemstillinger

Vi skal i denne oppgaven se på ulike aspekter ved eventuelle lønnsforskjeller mellom kvinner og menn i Norge. Vi har konstruert spesifikke problemstillinger hvor svarene vil belyse de mest sentrale delene av temaet, gitt vårt datamateriale:

Hovedspørsmål:

- 1) Er det forskjell i gjennomsnittslønn mellom kjønnene i Norges voksne befolkning i vår data, og i så fall hvor stor er forskjellen?
- 2) Til hvilken grad kan størrelsen på lønnsgapet forklares av forskjeller mellom menn og kvinner i arbeidserfaring, utdanningsnivå, kognitive ferdigheter, sektor, yrke og industri?

Tilleggsspørsmål:

- 1) Får kvinner og menn ulik lønnsavkastning fra de tre kognitive ferdighetene undersøkt i PIAAC; tallforståelse, leseferdigheter og problemløsning i IKT-miljø?
- 2) Får kjønnene ulik lønnsavkastning av utdanning?
- 3) Er lønnsgapet i offentlig sektor forskjellig fra det i privat sektor?

Notat til problemstillinger: for å unngå seleksjonsproblemer på grunn av personer fortsatt under utdanning, vil vi avgrense analysen til personer som er 25 år og eldre.

For å finne svar på spørsmålene vil vi gjennomføre en empirisk analyse ved hjelp av statistiske og økonometriske metoder med grunnlag i data samlet inn fra PIAAC-undersøkelsen gjort i regi av OECD i perioden august 2011 til mars/april 2012.

Fremgangsmåten vil hovedsakelig være å utforme estimerbare økonometriske modeller ved

hjelp av datasett fra undersøkelsen, utføre regresjonsanalyser på modellene og deretter bruke statistiske metoder for hypotesetesting for å undersøke om spørsmål stilt til datamaterialet er statistisk signifikante. Hypotesetestene vil så etterfølges av kommentarer med forslag til statistisk tolkning av resultatene og mulige forklaringer basert på økonomisk intuisjon og teori.

1.3 Sammendrag

Den empiriske analysen finner at det rå lønnsgapet mellom norske kvinner og menn 25 år eller eldre er 14,7%. I forsøk på å undersøke hvilke faktorer som kan forklare denne forskjellen kontrollerer vi for erfaring, utdanning, kognitive ferdigheter, sektor, yrke og industri, men finner at selv etter denne kontrollen gjenstår et lønnsgap på 9,02%. Dette gapet har ikke våre modeller forutsetninger for å tilby en kvantitativ forklaring på. Analysen finner at forskjeller mellom kvinner og menn i antall år utdanning og arbeidserfaring ikke er en del av forklaringen til lønnsgapet, men at ulikheter i yrke, industri og kognitive ferdigheter er betydelige forklaringsfaktorer.

Videre finner analysen ingen statistiske bevis for at kvinner og menn får ulik lønnsavkastning fra kognitive ferdigheter eller utdanning, men vi observerer en signifikant forskjell på lønnsgapet i privat sektor sammenlignet med det i offentlig sektor; 11,4% i privat mot 6,08% i offentlig.

2. Tidligere litteratur og teoretisk rammeverk

2.1 Kapittelinnledning

Ved å anvende tidligere forskningslitteratur og rapporter knyttet til lønnsforskjeller ønsker vi å legge et solid teoretisk grunnlag for resten av oppgaven. Dette gjør vi ved å finne fellestrekk på tvers av tidligere studier i valg av fremgangsmåter og formuleringer av økonometriske modeller. Med dette vil vi få en innsikt i de vanligste begrunnelsene for valgene tatt i den tidligere litteraturen, og derav bli bedre rustet til å konstruere modeller. Tidligere litteratur som er relevant for oss inkluderer ikke bare økonomi og statistisk teori, men også statistisk litteratur som viser resultatene etter lang tid med ulike valg tatt av kvinner

og menn. Det vil være naturlig å anta at lønnsnivået et voksent menneske opplever i stor grad er et resultat av en rekke valg tatt gjennom kanskje først og fremst utdanningen, men også i løpet av karrieren. Det vil derfor være av stor relevans for vår analyse å først få en forståelse for fordelingen av kjønnene innad utdanninger og arbeid. Spesielt relevant vil være de områdene hvor ulike valg har resultert i de største skjevfordelinger av kvinner i forhold til menn. Med denne kunnskapen vil vi ha et bedre utgangspunkt for å gi korrekte tolkninger av resultat vi vil få gjennom analysearbeidet av utvalget vårt, da vi kan sette det opp mot det vi vet gjelder for populasjonen.

I dette kapittelet vil vi presentere relevant lønnsstatistikk i tillegg til resultater fra tidligere studier/rapporter som har sett på problemstillinger lik våre. Vi kommer også til å diskutere vanlige fremgangsmåter og økonometriske modellformuleringer i samme studier, og viser til slutt hvordan vi bruker teori til å konstruere estimerbare økonometriske modeller.

2.2 Teoretisk rammeverk

Det kan være utallige grunner til en eventuell forskjell i lønnsnivå mellom kvinner og menn. Mange av grunnene som vil foreslås under deriverer fra NOU 2008:6, da rapporten er den mest omfattende undersøkelsen av lønnsforskjeller mellom kjønnene i Norge.

Ulik utdanning og yrke

Menn og kvinner velger systematisk ulike utdanninger (Støren og Arnesen, 2003, s. 135). Dersom et av kjønnene er overrepresentert i utdanninger til yrker med lavere eller høyere lønn så vil dette gjenspeiles når man sammenligner lønnen mellom menn og kvinner.

Ulik arbeidstid

Ett av kjønnene kan være mer villig til å arbeide overtid, nattevakt, eller på andre tidspunkter som gjør gir lønnstillegg, noe som over tid vil føre til høyere gjennomsnittslønn for det kjønnnet.

Ulike stillinger innad yrker

Selv om et yrke kan ha lik fordeling av kvinner og menn, kan et av kjønnene være overrepresentert i stillinger innad yrket som gir høy lønn, for eksempel lederstillinger.

Ren kjønnsdiskriminering

Arbeidsgivere kan ha kognitive bias som gjør at de vurderer arbeidstakere galt basert på kjønn. Kjønnsdiskriminering kan også oppstå som resultat av tradisjoner og fordommer basert på disse.

Privat eller offentlig ansatt

Privat virksomhet betaler generelt sett mer enn offentlig virksomhet (Grini & Lien, 2011). Dersom kjønnsfordelingen i henholdsvis offentlig og privat virksomhet er forskjellig, vil dette gjenspeiles i lønnen.

Livsvalg, Karriere eller familie

Menn og kvinner tar forskjellige valg i livet. Å velge karriere foran familie, hobbyer eller andre ting, vil gjenspeile seg i lønnen til et individ. Derfor er det naturlig å tenke seg at dersom et av kjønnene er mer tenkende til å satse mer på karriere, så vil det være betydelig for lønnsstatistikken.

Økonomisk tilnærming til lønnsforskjeller

Ved å anta fullkommen konkurranse i arbeidsmarkedet kan man legge forklaringen på lønnsforskjellen mellom kjønn på forskjeller i produktivitet og verdiskapning. Selv om dette er en sentral del i den økonomiske tilnærmingen til lønnsforskjeller, vil det være en noe urealistisk antagelse ettersom ufullstendig informasjon, skjevfordeling av makt og bedrifters komplekse sammensetning av prioriteter utenfor ren profittmaksimering vil føre til at arbeidsmarkedet ikke kan beskrives som å være i fullkommen konkurranse. Videre vil en ren økonomisk tilnærming by på mangelfull tolkning av eventuelle lønnsforskjeller da økonomisk teori ofte baseres på sterke antagelser. Hvis man ønsker å forklare lønnsforskjeller i konteksten av tilbud og etterspørsel i arbeidsmarkedet, må man utover økonomisk teori ta hensyn til at aktørers rasjonelle valg, som baseres på preferanser og målsetninger, ikke alltid stemmer overens med den økonomiske antagelsen om at maksimering av produktivitet og lønn er den største prioriteten.

Ren økonomisk teori om lønnsdannelse går heller ikke perfekt overens med norsk lov om likestilling mellom kjønnene, ettersom loven ikke gir noen eksakt målestokk når det kommer til verdimåling av arbeidskraft, men heller vektlegger faktorer som kompetanse, anstrengelse, ansvar og arbeidsforhold. Generelt knytter loven om likestilling arbeidets verdi opp mot

innsatsen. Et økonomisk synspunkt avviker fra dette, fordi arbeidskraft ses på som en tjeneste/vare i markedet som vil bli prissatt med tanke på hvilken verdi den skaper. Dermed er det ikke tilstrekkelig å måle verdien av arbeid kun med tanke på innsats. Lønn brukes også som kompensasjon for ulemper og fordeler med arbeidet. Forhold som er truende for arbeidstakers fysiske og psykiske helse vil kompenseres med høyere lønn eller periodiske tillegg. Det er ulike oppfatninger om hvilke ulemper som kvalifiserer til kompensasjon, noe som kan føre til lønnsdiskrimineringer og videre i et kjønnsdelt arbeidsmarked utvikle seg til kjønnsbasert lønnsdiskriminering. Lønningen vil også gjenspeile markedsmakt. Under lønnsforhandlinger vil en arbeidstaker som er vanskeligere å erstatte ha bedre forutsetninger for å kunne sette krav i en lønnsforhandling. Menn og kvinner har forskjellig mobilitetsmønster, noe som kan føre til at kvinner totalt sett ikke krever like mye i lønnsforhandlinger som menn gjør.

Tre typer diskriminering

NOUs rapport fra 2008 beskriver tre typer diskriminering som passer den økonomiske tilnærmingen til lønnsforskjellene.

- **Preferansebasert diskriminering:** Det kan tenkes at en andel av arbeidsgivere har fordommer som gjør at de vil prioritere et kjønn foran et annet, og derav etterspørre det foretrukne kjønn framfor det andre. Et annet scenario er at forbrukere foretrekker en tjeneste levert av ett kjønn over samme tjeneste levert av det andre kjønn. Da vil etterspørselen etter kjønn som forbrukeren foretrekker være høyere.
- **Statistisk diskriminering:** Statistikk fører til at arbeidsgiver vil prioritere ett kjønn foran et annet når det kommer til ansettelse og lønn. For eksempel er det statistisk sett større sjans for at en mann vil prioritere karriere foran familie, og det vil derfor virke fornuftig for en bedrift å ansette en mann foran en kvinne.
- **Markedsmaktbasert diskriminering (monopsoni):** Arbeidstaker har innflytelse på egen lønn gjennom lønnsamtaler. Personer med spesialisert kompetanse som f.eks. utdanning innen omsorgssektoren, er ikke like fleksible når det kommer til å kunne bytte yrke, og vil dermed ha mindre forhandlingsstyrke under lønnsamtalen.

(NOU 2008:6, s.133)

2.3 Resultater i tidligere litteratur

Vi presenterer først statistikk som viser den ulike fordelingen av kvinner og menn innad utdanninger og arbeid, og går deretter videre til å legge frem resultater NOU (2008:6)-rapporten, og fra tidligere litteratur som har undersøkt problemstillingen lik våre.

2.3.1 Relevant statistikk

Høyere utdanning gir i de aller fleste tilfeller høyere lønn (Bye, 2018). Det kan derfor være verdt å merke seg at det er relativt stor forskjell på kvinner og menn når det kommer til valgt om å ta høyere utdanning eller ikke. Statistisk Sentralbyrå (SSB) har i en årrekke rapportert at det er flere kvinner som tar høyere utdanning enn det er menn.

Tabell 1: Antall norske studenter i høyere utdanning etter kjønn

Norske studenter i høyere utdanning i Norge og i utlandet			
Årstall	2008	2017	2018
Studenter i høyere utdanning			
I alt	225 469	293 123	293 287
Menn	87 911	118 928	118 809
Kvinner	137 558	174 195	174 478
Andel 19-24 år i høyere utdanning			
I alt	29,7	35,4	35,3
Menn	23,4	28,4	28,4
Kvinner	36,4	42,9	42,8
Andel 25-29 år i høyere utdanning			
I alt	14,5	15,9	15,9
Menn	12,6	14	14
Kvinner	16,4	17,9	17,9

Tall fra SSB (<https://www.ssb.no/utuvh>)

Tabellen viser at i 2008, 2017 og 2018 ligger andelen kvinner som tar høyere utdanning merkbart over andelen menn. (Selv om dette bør tilsi at kvinner tjener mer enn menn, vil vi vise senere i analysen at det paradoksalt nok ikke stemmer.)

Som nevnt, vil privat sektor typisk betale høyere lønninger enn offentlig sektor. Her finner vi en svært markant kjønnsdeling i Norge, hvor menn i mye større grad jobber i privat sektor, mens majoriteten av kvinner er ansatte i offentlig sektor.

Tabell 2: fordeling av kjønn på offentlig og privat sektor i Norge

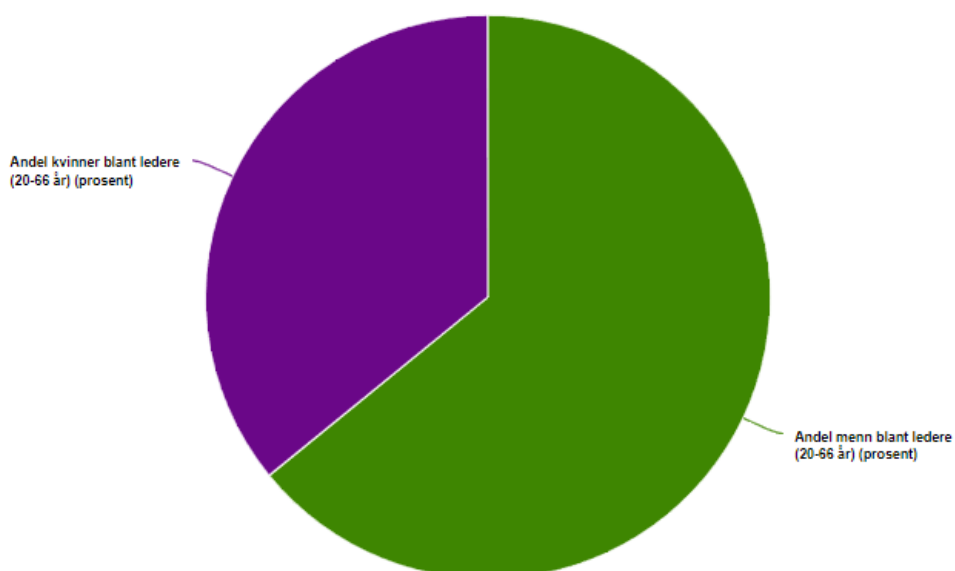
Kjønnsfordeling mellom privat og offentlig sektor i Norge (år 2017)		
	Menn	Kvinner
Andel i offentlig sektor	29,9 %	70,1 %
Andel i privat sektor	63,4 %	36,6 %

Tall fra SSB (<https://www.ssb.no/befolkning/faktaside/likestilling>)

Det er en logisk antagelse at ledere tjener bedre enn den gjennomsnittlige ansatte i en bedrift. Vi kan observere en tydelig kjønnsdeling blant ledere i Norge, hvor majoriteten er menn.

Figur 1: Kjønnsfordeling bland ledere

Kjønnsfordeling blant ledere (20 - 66 år). 2017



Kilde: Indikatorer for kjønnslikestilling, Statistisk sentralbyrå

Tall fra SSB (<https://www.ssb.no/likekom>)

Her ser vi at blant lederstillinger i Norge er kjønnsfordelingen ca. 3 til 1 med menn som majoritet.

2.3.2 Resultat fra tidligere rapporter og analyser

Flere rapporter og studier har blitt gjort i forsøk på å forklare lønnsforskjeller mellom kjønnene. Den kanskje mest prominente norske rapporten er arbeidet likelønnskommisjonen presenterte i 2008, etter at Regjeringen Stoltenberg II i 2006 utnevnte kommisjonen for å “gi en samlet beskrivelse av lønnsforskjellene mellom kvinner og menn, og vurdere hva som kan bidra til å redusere lønnsforskjellene.” (NOU 2008:6).

NOU-rapporten deler oppsummeringen av funnene sine i fem deler:

- 1) Forskjeller i utdanningslengde og alder forklarer svært lite av lønnsgapet i dag
- 2) Kvinner og menn har om lag lik lønn i samme stilling i samme virksomhet
- 3) Lønnsgapet følger det kjønnsdelte arbeidsmarkedet
- 4) Forhandlingssystemet opprettholder stabile lønnsrelasjoner, også mellom kvinner og menn
- 5) Lønnsforskjellene vokser i foreldrefasen

En annen Norsk rapport publisert av Institutt For Samfunnsforskning fra 2013 (Barth, Hardoy, Schøne & Østbakken, 2013) tar for seg lønnsforskjeller mellom kvinner og menn, og stiller spørsmålet “Hva har skjedd på 2000-tallet?”. Rapporten starter sammendraget sitt med “At kvinner tjener mindre enn menn er et gjennomgående funn”. Videre finner rapporten det at gjennomsnittlig tjener kvinner 12,7 prosent mindre enn menn (i 2011), og at ved sammenligning mellom kvinner og menn i samme yrke innen samme næring, er gapet på 7,1 prosent (også i 2011)

(Blau & Kahn, 2016) utførte en empirisk analyse med formål å finne forklaringer på lønnsforskjeller mellom kvinner og menn med utgangspunkt i data samlet inn i USA i perioden 1980-2010. Blant hvilke forklaringer som er prominente i lønnsgapet finner de blant annet at kjønnsforskjeller i yrke og industri kan forklare store deler av lønnsforskjellene, og at relativt nye forklaringer som psykologiske attributter og ikke-kognitive ferdigheter utgjør en liten til moderat porsjon av forklaringen på lønnsforskjeller.

(Stokke, 2016) studerte lønnsgapet og hvordan lønnsnivå utvikles ulikt for kvinner og menn gjennom karrieren. Studien finner at under halvparten av lønnsgapet kan forklares av forskjeller i observerbare faktorer, og at en viktig forklaring for det resterende lønnsgapet er at kvinner får lavere utbytte av arbeidserfaring.

2.4 Fremgangsmåte og utforming av økonometriske modeller i tidligere analyser

Vi finner i tidligere studier innen lønnsforskjeller at den vanligste fremgangsmåten er å utforme en eller flere økonometriske modeller, utføre regresjonsanalyser på modellene, og gjennomføre hypotesetester på enkelte forklaringsvariabler spesifikke for problemstillingen(e). Dette tolker vi som den mest naturlige og intuitive fremgangsmåten, og benytter oss dermed av den.

Videre kan det være av relevans å observere hvordan tidligere studier typisk utformer sine økonometriske modeller. Det ser ut til at trenden for å sammenligne lønnsnivå mellom menn og kvinner med ellers like forutsetninger (erfaring, utdanning osv.) er å utforme log-lineære modeller. Dette forutsetter gjerne at menn og kvinner har samme produktivitet.

Eksempel på vanlig utforming: $Y_i = \alpha + \beta_i X_i + \varepsilon$

Hvor Y_i er den naturlige logaritmen av lønn, X_i er en vektor av forklaringsvariabler med β_i som tilhørende koeffisient. ε er restleddet som inkluderer faktorer som ikke lar seg forklare av modellen.

Dynamikken i en log-lineær modell som dette:

Endres X_i med én enhet, endres Y_i med $100 \cdot \beta_i$ prosent. Dette er et gunstig resultat fordi det er uavhengig av målenivå på avhengig variabel, noe som muliggjør sammenligning av estimert lønnsgap på tvers av tidligere litteratur (og eventuelt land).

Det er to typer funksjonsformer som er vanlige i denne typen analyse, Lineær-lineær og log-lineær. Lineær-lineær dersom man ønsker å se på endringer i kroner, eller log-lineær dersom man ønsker å se på prosentvis endring. Når det kommer til lønnsgapet og hvordan forskjellige variabler endrer det, så er det lettest å utelukkende se på prosentvis endring, og vi vil derfor

utelukkende bruke log-lineære modeller. Det vil gjøre det lettere å tolke variablenes innvirkninger opp mot hverandre.

2.4.1 Overgang fra teori til estimerbare økonometrisk modeller

Ved å se på tidligere litteratur, finner vi variabler og faktorer relevant for lønnsforskjeller mellom menn og kvinner. Da kan vi finne lignende eller identiske variabler fra vårt datasett som kontrollvariabler for å se om vi får lignende resultater. Dette innebærer at vi vil utforme økonometriske modeller med lønn som avhengig variabel, kjønn som sentral forklaringsvariabel, og faktorer som industri, yrke, utdanning, sektor, arbeidserfaring og kognitive ferdigheter som kontrollvariabler.

Når vi skal undersøke om funn i datasettet vårt er statistisk signifikante, vil vi gjennom økonometri benytte oss av regresjonsanalyse i form av minste kvadraters metode (OLS), og baserer teknikkene våre på det beskrevet i “Using Statistics in Economics” av R. L. Thomas (2005).

OLS er en estimeringsmetode brukt for å finne korrelasjonen mellom en eller flere kontrollvariabler og en avhengig variabel. Estimeringen baserer seg på å finne en sammenheng mellom variablene som minimere det kvadrerte avviket mellom observasjonene og estimeringen, der estimatene er de ukjente parameterne i økonometriske modeller.

For å kunne gjøre konklusjoner basert på resultater fra regresjonsanalysene, må vi teste om resultatene er statistisk signifikante. Ved å utføre hypotesetester på de estimerte resultatene vil vi kunne fastslå hvor stor sannsynlighet det er for at utvalget i datamaterialet er representabelt for populasjonen (statistisk inferens). I løpet av oppgaven vil det bli brukt to former for hypotesetester; T-test når vi ønsker å undersøke en hypotese om én koeffisient, og F-test når vi undersøker flere koeffisienter.

2.5 Oppsummering av kapittel 2

Vi har i løpet av kapittel 2 presentert relevant økonomisk og delvis sosiologisk teori. Dette teorigrunnlaget vil være basisen for våre tolkninger av resultater senere i analysearbeidet. Videre har vi lagt frem relevant statistikk som viser kjønnsfordelinger innen utdanning, arbeid og lønn. Dette for å gi en illustrasjon av viktige forskjeller mellom kjønnene innen de

nevnte områdene. I kapittelet ble det også diskutert resultater fra tidligere rapporter og analyser, i tillegg til at vi kort så på hva som er typiske fremgangsmåter for analyser av lønnsforskjeller, og hvordan utforming av økonometriske modeller som er vanligst å benytte.

3. Presentasjon av empiri

For å gi leseren en bedre forståelse for hva den ser på ønsker vi å presentere datamaterialet vi skal arbeide med videre i oppgaven. I dette kapittelet vil vi forklare hvor dataen er hentet fra, og delvis hvordan den er innsamlet. De relevante variablene vil også bli definert, i tillegg til at det presenteres tabeller med deskriptiv statistikk for hovedmaterialet i vår data, da det er en sentral del i det å få et kvantitativt inntrykk av materialet.

3.1 Omtale og gjennomgang av data

Datamaterialet som brukes er fra den internasjonale PIAAC undersøkelsen om voksnes leseferdigheter. Undersøkelsen er den største av sitt slag, og måler hvor gode voksne i Norge og 30 andre land er i lesing, tallforståelse og problemløsning med IKT.

3.2 Beskrivelser av variablene brukt i analysen

Avhengig variabel - **earnhr**: timelønn målt i norske kroner.

Sentral forklaringsvariabel, kjønn, **female**:

- Dummyvariabel **female** hvor 1 = kvinne, 0 = mann.

Kontrollvariabler:

- Arbeidserfaring - **exper** (basert på **c_q09_c**): målt som antall år i lønnet arbeid
- Formelt utdanningsnivå - **educ** (basert på **yrsqual**): høyeste oppnådd utdanningsnivå, målt i antall år
- Leseferdigheter - **lit** (basert på **pvlit1**): testscore fra test av leseferdigheter
- Tallforståelse - **num** (basert på **pvnum1**): testscore fra test av tallforståelse
- Problemløsningsevne - **psolv** (basert på **pvpsl1**): testscore fra test av problemløsningsevne i IKT
- Sektor - **sector** (basert på **d_q03**): Dummyvariabel hvor:

- 0 = privat sektor
- 1 = offentlig sektor

- Yrke - **occupation** (basert på **isco1c**): indikator med verdier fra 2 til 11 hvor kategoriene er definerte som:

2	Legislators, senior officials and managers
3	Professionals
4	Technicians and associate professionals
5	Clerks
6	Service workers and shop and market sales workers
7	Skilled agricultural and fishery workers
8	Craft and related trades workers
9	Plant and machine operators and assemblers
10	Elementary occupations
11	No paid work for past 5 years

- Industri - **industry** (basert på **isi1c**): indikator med verdier fra 7 til 27 hvor kategoriene er definerte som:

7	Agriculture, forestry and fishing
8	Mining and quarrying
9	Manufacturing
10	Electricity, gas, steam and air conditioning supply
11	Water supply; sewerage, waste management and remediation activities
12	Construction
13	Wholesale and retail trade; repair of motor vehicles and motorcycles
14	Transportation and storage
15	Accommodation and food service activities
16	Information and communication
17	Financial and insurance activities
18	Real estate activities
19	Professional, scientific and technical activities
20	Administrative and support service activities
21	Public administration and defence; compulsory social security

22	Education
23	Human health and social work activities
24	Arts, entertainment and recreation
25	Other service activities
26	Activities of households as employers; undifferentiated goods- and services producing activ. of households for own use
27	Activities of extraterritorial organizations and bodies

Som nevnt tidligere, vil vi begrense analysen til observasjoner med personer 25 år eller eldre. Dette innebærer at variabelen female vil bli begrenset slik at alle observasjoner hvor $age < 25$ blir ekskludert.

3.3 Deskriptiv statistikk for materialet

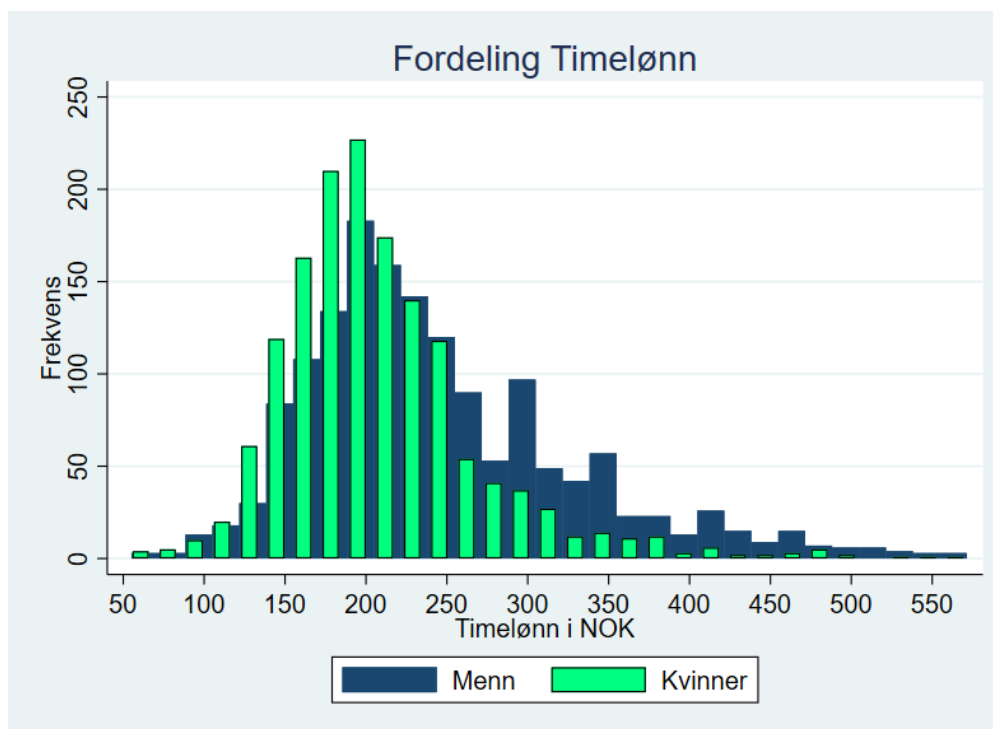
Tabell 3. Deskriptiv statistikk for timelønn

Deskriptiv statistikk - lønn (earnhr for age >= 25)			
	Alle	Menn	Kvinner
Gjennomsnittslønn	226.4945	244.3118	208.0413
Standardavvik	76.3264	83.70927	62.73888
Minimum	52.30769	55.05092	52.30769
Maksimum	572.6316	571.4286	572.6316
Antall observasjoner	3,023	1,538	1,485

Fra den deskriptive statistikken for timelønnvariabelen earnhr, finner vi at:

- Den kvinnelige gjennomsnittslønnen er ca 85.2% av den mannlige
- Både den høyeste og den laveste observerte lønnen er hos kvinner
- Utvalget har en relativt lik fordeling av kjønnene.
 - Utvalget består av 50.9% menn, 49.1% kvinner.
- Menn har betydelig større variasjon i timelønn enn kvinner
(Vi illustrerer forskjellen på variasjonen i timelønn under)

Figur 2: variasjon i timelønn etter kjønn



Tabell 4: Deskriptiv statistikk forklaringsvariabler

	Deskriptiv statistikk kontrollvariabler (age >= 25)					
	Alle		Menn		Kvinner	
	Gjennomsnitt	Standardavvik	Gjennomsnitt	Standardavvik	Gjennomsnitt	Standardavvik
Utdanning (educ)	14,66	2,45	14,63	2,46	14,69	2,43
Erfaring (exper)	21,69	11,61	23,20	12,01	20,06	10,91
Sektor (sector) prosent	36 %		26 %		53 %	
Tallforståelse (num)	284,2	54,3	291,3	54,4	276,6	53,1
Tekstforståelse (lit)	282,9	46,6	284,7	47,1	281,0	46,0
Problemløsningssevne (psolv)	286,0	40,5	289,5	41,0	282,3	39,6

Fra den deskriptive statistikken til de sentrale kontrollvariablene finner vi at:

- Kvinner har gjennomsnittlig marginalt lengre utdanning
- Menn har gjennomsnittlig noe over 3 år mer arbeidserfaring
- 26% av menn jobber i offentlig sektor, mot 53% av kvinner.

- Menn scorer høyere enn kvinner på alle tre av de kognitive testene, men markant høyere på tallforståelse.
- Menn har større variasjon i alle kontrollvariablene (ekskludert sektor)

Tabell 5. Gjennomsnittlig timelønn etter sektor

	Gjennomsnittslønn etter sektor og kjønn		
	Menn	Kvinner	Alt
Privat Sektor	247,0769	207,3962	215,4363
Offentlig Sektor	238,9722	208,6378	212,9842

- Menn i privat sektor har høyest gjennomsnittslønn
- Kvinner i privat sektor har lavest gjennomsnittslønn
- Privat sektor har høyere gjennomsnittslønn enn offentlig sektor

3.4 Fordeler og begrensninger ved datasettet

Det vil være urealistisk og naivt å anta at datasettet vårt inneholder absolutt all relevant informasjon. Det realistiske vil derimot være å gå ut fra at datasettet har mangler, for så å identifisere disse manglene. På den måten oppnår vi et mer realistisk inntrykk av resultatene fra vårt arbeid ved å være klare over begrensninger satt av datasettets innhold, og mangel derav. Med mangler mener vi ufullstendig informasjon og/eller total mangel på informasjon som ville vært relevant for vår analyse, og som derav begrenser analysemulighetene så vel som nøyaktigheten av resultatene. Vi identifiserer at de mest prominente begrensninger ved vårt datasett er som følgende (en mer utdypende gjennomgang av begrensninger finnes i kapittel 5.3):

- Ingen informasjon om ansvarsnivå på ansatte. Vil være relevant å vite hvilke observasjoner som er av personer i lederstillinger ettersom ledere typisk tjener betydelig mer enn ansatte uten lederansvar.
- Ingen informasjon om eventuelle barn. Tidligere litteratur viser at kvinners gjennomsnittslønn blir sterkt påvirket av det å få barn (NOU 2008:6, s.178).

- Ufullstendig informasjon om utdanning. Har kun antall år utdanning, men ingen informasjon om hvilken type utdanning - om den er/var relevant for nåværende jobb osv.
- Lite detaljert informasjon om lønn. Vi vet ikke om timelønnen oppgitt er en fastsatt timelønn, eller om det er en gjennomsnittlig timelønn som inkluderer individenes overtidsarbeid/tillegg/bonuser osv. Dette kan skape unøyaktighet i analysen hvis enkelte observasjoner har oppgitt sin faste timelønn uten tillegg og bonuser, mens andre har inkludert dette.
- Enkelte respondenter har ikke besvart alle spørsmålene, dette fører til en reduksjon i gyldige observasjoner når flere kontrollvariabler introduseres til de økonometrisk modellen.

I tillegg til begrensninger, vil det være like naturlig å anta at datasettet kan inneholde fordeler. Noen fordeler kan være at:

- Datasettet inneholder mål på kognitive ferdigheter, da dette kan gi indikasjoner på intelligensen personene observert, og intelligens er knyttet til sannsynlighet for karrieresuksess (Schmidt & Hunter, 1998).
- Utvalget er stort, noe som muliggjør mer nøyaktige konfidensintervall enn om utvalget var betydelig mindre.

3.5 Oppsummering av kapittel 3

Det foregående kapittelet har presentert og forklart datamaterialet vi skal bruke i resten av oppgaven. Det har også blitt presentert hovedtrekk ved datamaterialet i form av deskriptiv statistikk, i tillegg til at vi har diskutert fordeler og begrensninger ved datasettet.

4. Regresjonsanalyse

4.1. Hovedspørsmål 1

Er det forskjell i gjennomsnittslønn mellom kjønnene i Norges voksne befolkning i vår data, og i så fall hvor stor er forskjellen?

Vi vil bruke besvarelsen av dette spørsmålet for å demonstrere vår fremgangsmåte for regresjonsanalyse og hypotesetesting (med t-test). Dette for å etablere hva vi legger i hypotesetesting slik at vi i de senere spørsmålene kan referere mer generelt til resultat fra testing utført i Stata, og dermed forenkle notasjonen.

$$\text{Økonometrisk modell: } \mathit{learnhr} = \alpha + \delta \mathit{female} + \varepsilon$$

Hvor $\mathit{learnhr}$ er den logaritmiske funksjonen av variabelen earnhr – timelønn.

Tolkning av modellen:

For å undersøke forskjell i gjennomsnittslønn mellom kvinner og menn benytter vi oss av en log-lineær modell med lønn ($\mathit{learnhr}$) som avhengig variabel, og dummyvariabelen female som sentral forklaringsvariabel. Ettersom forklaringsvariabelen er en dummy som indikerer at observasjonen er av det mannlige kjønn når variabelen har verdi 0, vil referansegruppen være menn. Dette innebærer at etter regresjon av denne enkle modellen vil estimatet av koeffisienten (δ) til forklaringsvariabelen (female) fortelle oss hva lønnsgapet mellom kvinner og menn i vårt datamateriale er. Intuisjonen bak dette er at når forklaringsvariabelen har verdi null, observerer vi gjennomsnittslønnen til menn, og når forklaringsvariabelen har verdien 1, observerer vi kvinner, og endringen i lønnsnivået med utgangspunkt i den mannlige gjennomsnittslønnen blir gitt av koeffisienten til forklaringsvariabelen. Ettersom vi bruker logaritmen til timelønnen, tolkes endringen i lønn som $100 * \delta\%$.

Resultat fra regresjon:

VARIABLES	(1) learnhr
female	-0.147*** (0.0112)
Constant	5.444*** (0.00787)
Observations	3,023
R-squared	0.053

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Fra regresjon gjennom Stata finner vi at variabelen *female* har den estimerte koeffisient $d = -0,147$ med standardavvik $S_d = 0,0112$.

Dette kan vi tolke som at det "rå" lønnsgapet (siden modellen ikke har kontrollvariabler) er -14,7%. Kvinner tjener gjennomsnittlig 14,7% mindre enn menn i vårt datasett når vi ikke kontrollerer for potensielle forklaringsvariabler.

Hypotese:

$$H_0: \delta = 0 \quad , \quad H_A: \delta \neq 0$$

Hvis δ er null, betyr det at lønn ikke blir påvirket av om dummyvariabelen *female* er 0 eller 1; kjønn har ingenting å si for gjennomsnittslønn. Hvis δ ikke er lik null, blir lønnsnivået påvirket av kjønn.

Når vi finner testobservatoren manuelt vil vi inkludere flere desimaler enn det som er gitt ovenfor slik at vi får et mer nøyaktig resultat.

Testobservator:

$$TS = \frac{d}{S_d} \sim t(n-2) \quad \Rightarrow \quad TS = \frac{-0,1466391}{0,0112236} = -13,06524664$$

Testobservatoren har en t-fordeling med

Kritisk verdi ved 5%-signifikansnivå: $t_{0,25,3023} = \pm 1,96$

$$\text{Vi har at } |TS| = 13,06524664 > 1,96$$

Absoluttverdi for testobservator overstiger kritisk verdi, og vi kan forkaste nullhypotesen med et 5% signifikansnivå. Dette impliserer at kjønn har en statistisk signifikant innvirkning på lønnsnivået.

Resultatene fra regresjonen viser faktisk at vi kan forkaste nullhypotesen med over 99% sikkerhet. I tillegg til dette, kan vi se på regresjonsresultatet at forklaringskoeffisienten (R^2) er over null, og derav at modellen har noe forklaringskraft. Stata bekrefter med sin automatiske F-test at R^2 er positiv med over 99% sikkerhet.

Vi kontrollerer resultatet ved å utføre en t-test i Stata. Resultat fra testen forteller at vi kan forkaste nullhypotesen til fordel for en alternativhypotese om at koeffisienten er negativ, med svært stor sikkerhet (over 99%). Dette er i tråd med våre funn, og vi kan med stor sikkerhet svare på spørsmålet i problemstillingen; Datasettet vårt viser et "rått" lønnsgap mellom kvinner og menn i Norge (25 år eller eldre) på 14,7%. Resultatet er statistisk signifikant.

4.2. Hovedspørsmål 2

I hvilken grad kan størrelsen på lønnsgapet forklares av forskjeller mellom menn og kvinner i arbeidserfaring, formelt utdanningsnivå, kognitive ferdigheter, sektor, yrke og industri?

For å undersøke effekten av de nevnte kontrollvariablene på lønn, konstruerer vi flere log-lineære modeller hvor vi legger til én kontrollvariabel for hver ny modell kronologisk i henhold til formuleringen av problemstillingen (foruten i modell m4 hvor vi introduserer alle tre variablene for de kognitive ferdighetene). Dette lar oss observere endringen i lønnsgapet etterhvert som modellene tar hensyn til flere og flere faktorer.

For modell (m6) og (m7), hvor yrke (occupation) og industri (industry) blir introdusert, har vi generalisert uttrykket for de nevnte variablene og deres tilhørende koeffisienter. Dette er fordi begge variablene er basert på kategorier, noe som forutsetter at det opprettes en dummyvariabel for hver kategori når regresjon skal utføres. For yrke, betyr dette at den senkede skriften på den tilhørende koeffisienten er *occ*, hvor *occ* representerer indikatorverdien for hver yrkeskategori ($occ = 2, 11$). Den samme intuisjonen gjelder for industri ($ind = 7, 27$).

Økonometriske modeller:

(m1)

$$\text{learnhr} = \alpha + \delta_1 \text{female} + \varepsilon$$

(m2)

$$\text{learnhr} = \alpha + \delta_1 \text{female} + \beta_1 \text{exper} + \beta_2 \text{exper}^2 + \varepsilon$$

(m3)

$$\text{learnhr} = \alpha + \delta_1 \text{female} + \beta_1 \text{exper} + \beta_2 \text{exper}^2 + \beta_3 \text{educ} + \varepsilon$$

(m4)

$$\text{learnhr} = \alpha + \delta_1 \text{female} + \beta_1 \text{exper} + \beta_2 \text{exper}^2 + \beta_3 \text{educ} + \beta_4 \text{lit} + \beta_5 \text{num} + \beta_6 \text{psolv} + \varepsilon$$

(m5)

$$\text{learnhr} = \alpha + \delta_1 \text{female} + \beta_1 \text{exper} + \beta_2 \text{exper}^2 + \beta_3 \text{educ} + \beta_4 \text{lit} + \beta_5 \text{num} + \beta_6 \text{psolv} + \delta_2 \text{sector} + \varepsilon$$

(m6)

$$\text{learnhr} = \alpha + \delta_1 \text{female} + \beta_1 \text{exper} + \beta_2 \text{exper}^2 + \beta_3 \text{educ} + \beta_4 \text{lit} + \beta_5 \text{num} + \beta_6 \text{psolv} + \delta_2 \text{sector} + \delta_{occ} \text{occupation}_i + \varepsilon$$

(m7)

$$\text{learnhr} = \alpha + \delta_1 \text{female} + \beta_1 \text{exper} + \beta_2 \text{exper}^2 + \beta_3 \text{educ} + \beta_4 \text{lit} + \beta_5 \text{num} + \beta_6 \text{psolv} + \delta_2 \text{sector} + \delta_{occ} \text{occupation}_i + \delta_{ind} \text{industry}_i + \varepsilon$$

Tabell 6: Resultat fra regresjon av modeller m1-m7*

	m1	m2	m3	m4	m5	m6	m7
Erfaring		x	x	x	x	x	x
Utdanning			x	x	x	x	x
Kognitive ferdigheter				x	x	x	x
Sektor					x	x	x
Yrke						x	x
Industri							x
Lønns-gap	14,70 %	14,70 %	15,20 %	13,40 %	12 %	11,20 %	9,02 %
R²	0,053	0,105	0,255	0,287	0,298	0,359	0,421

* Mer detaljert tabell finnes i appendix

Fra tabellen (og fra hovedspørsmål 1) ser vi at det «rå» lønns-gapet ligger på 14,7%. det endres etter hvert som flere faktorer tas hensyn til, og ender til slutt opp på 9,02% når erfaring, utdanning, kognitive ferdigheter, sektor, yrke og industri er introdusert som kontrollvariabler. Det er en klar trend at jo flere kontrollvariabler som tas hensyn til, desto lavere vil lønns-gapet være.

Erfaring (m2)

Vi ønsker først å undersøke om erfaring har en statistisk signifikant innvirkning på timelønn. Dette gjør vi ved å utføre en F-test med følgende hypoteseformulering:

$$H_0: \beta_1 = \beta_2 = 0$$

$$H_A: \beta_1 \text{ eller } \beta_2 \neq 0$$

F-test viser at vi kan forkaste nullhypotesen ved et signifikansnivå under 0.1%.

Ved å inkludere en kvadrert versjon av erfaringsvariabelen, tester vi samtidig om lønnsavkastningen av erfaring er lineær. Hvis koeffisienten til den kvadrerte variabelen var lik null, vil de implisere at avkastningen er lineær. Hypotesetesten viser at $\beta_2 \neq 0$ og dermed at lønnsavkastningen fra erfaring ikke er lineær.

Vi ser at selv om arbeidserfaring har en statistisk signifikant virkning, reflekterer ikke en kontroll for erfaring noen betydelig endring i lønns-gapet ettersom det forblir på 14,7%, og vi kan ikke si at ulikheter i erfaring er en forklaringsfaktor for lønns-gapet. Dette strider med funnene gjort av Stokke (2016) ettersom den studien fant at kvinner fikk mindre lønnsavkastning av erfaring enn menn og at det var en viktig forklaringsfaktor for lønns-gapet.

Hadde kvinner fått lavere lønnsavkastning av erfaring burde resultatene fra modell m2 vist at lønnsgapet økte når vi kontrollerte for erfaring, ettersom kontrollen innebærer at erfaringsnivå holdes likt for menn og kvinner (hvis vi legger et interaksjonsledd mellom *exper* og *female* til modell m2 finner vi at vi *ikke* kan forkaste en nullhypotese om at koeffisienten til interaksjonsleddet er lik null).

Utdanning (m3)

I tredje modell kontrolleres det for utdanning. T-test viser en testobservator for *educ* på 24,6, noe som er godt over kritisk verdi for 1%-signifikansnivå i t-fordelinga. Vi kan dermed forkaste nullhypotesen om at $\beta_3 = 0$, og konkluderer med at utdanningsnivå har en statistisk signifikant effekt på timelønn. Vi ser at etter kontroll for utdanningsnivå *øker* lønnsgapet fra 14,7% til 15,2%, noe som kan indikere at kjønnene får ulik lønnsavkastning av samme antall år utdanning (dette testes i tilleggsspørsmål 2). Dette betyr at forskjell i antall år utdanning ikke ser ut til å være en del av forklaringen til lønnsgapet, dessuten så vi i den deskriptive statistikken (s. 17) at kvinner gjennomsnittlig har flere år utdanning enn menn i vårt datasett. Resultatet stemmer overens med hva NOU fant i sin rapport (2008:6).

Kognitive ferdigheter (m4)

Fjerde modell kontrollerer for de tre kognitive ferdighetene lese- og tallforståelse og problemløsningsevne i IKT-miljø. F-test viser at vi kan forkaste nullhypotesen om ingen innvirkning ($\beta_4 = \beta_5 = \beta_6 = 0$) ved et signifikansnivå under 0.1%. Menn i vårt datasett scorer høyere enn kvinner på samtlige tester, og betydelig høyere på tallforståelse. Tidligere nevnte vi kort at intelligens kan knyttes til karrieresuksess (s. 20). Derfor kan en mulig tolkning av resultatet være at hvis vi tolker de tre kognitive testene som test av intelligens, vil menn i vårt datasett ha en fordel og de kognitive ferdighetene vil dermed bidra til lønnsgapet. Når vi kontrollerer for de tre kognitive ferdighetene ser vi at lønnsgapet faller fra 15,2% til 13,4%.

Sektor (m5)

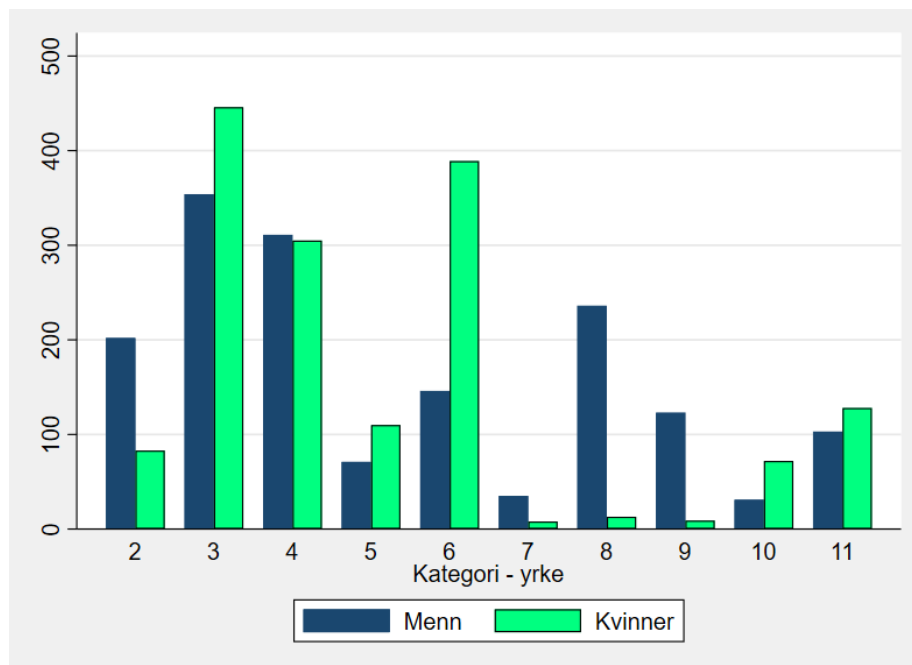
Når sektor blir kontrollert for ser vi at verdien på den tilhørende koeffisienten $\delta_2 = -0,0128776$. *Sektor* har en testobservator med absoluttverdi på 5,56, noe som overstiger kritisk verdi i t-fordelinga ved 0.1%-signifikansnivå. Vi forkaster dermed nullhypotesen om at sektor ikke har innvirkning på lønnsnivå. Tall fra SSB viste en betydelig skjevfordeling av

kjønn mellom offentlig og privat sektor, hvor majoriteten av menn arbeidet i privat sektor, mens majoriteten av kvinner befinner seg i offentlig sektor, noe som også var tilfellet i vårt datasett. Vi viste også at i vårt datasett har privat sektor høyere gjennomsnittslønn enn offentlig sektor. Kontroll for sektor får lønnsgapet til å falle fra 13,4% til 12%.

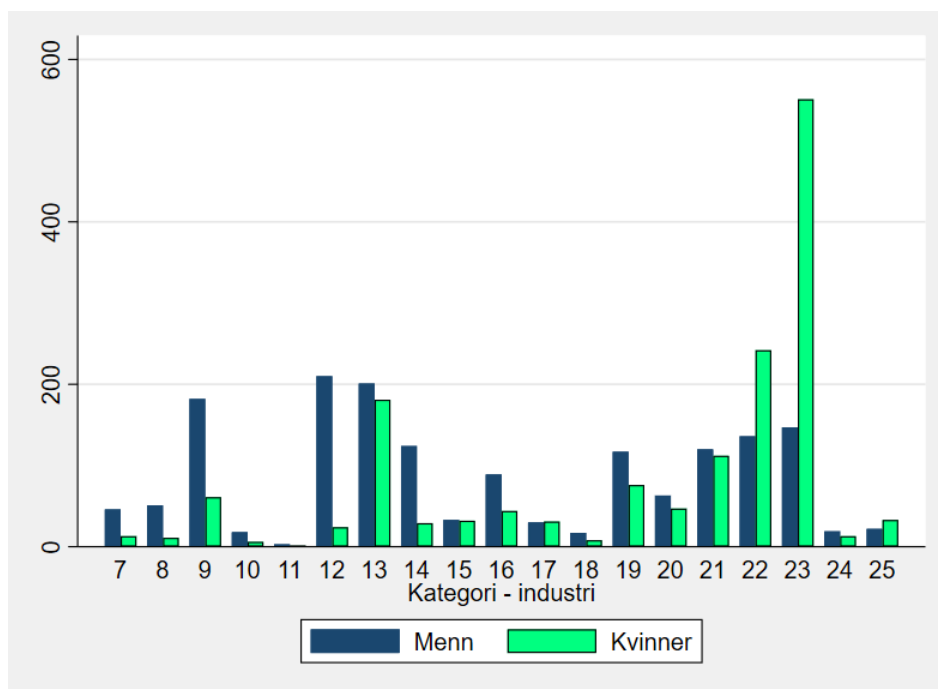
Yrke (m6) og Industri (m7)

Resultat fra F-test tilsier at vi kan forkaste nullhypotesene for yrke og industri ved et signifikansnivå under 0.1%, og vi kan konkludere med at forskjeller mellom kvinner og menn i yrke og industri har en statistisk signifikant innvirkning på lønnsgapet. Dette stemmer overens med hva Blau & Kahn (2016) skriver er viktige forklaringsfaktorer for lønnsgapet. I vårt datasett finner vi at både yrke og industri er svært kjønnsdelt (se side 14-15 for kategoridefinisjoner):

Figur 3: Kjønnsfordeling etter yrke



Figur 4: Kjønnfordeling etter industri



Etter kontroll for yrke og industri faller lønnsgapet fra 12% til 9,02%. Dette tyder på at menn arbeider i yrker og industrier som jevnt over gir bedre betalt enn yrkene og industriene kvinner typisk arbeider i.

4.3. Tilleggsspørsmål

Tilleggsspørsmål 1:

- **Får kvinner og menn ulik lønnsavkastning fra de tre ferdighetene undersøkt i PIAAC; tallforståelse, leseferdigheter og problemløsning i IKT-miljø?**

For å undersøke om kvinner og menn får ulik lønnsavkastning av de tre kognitive ferdighetene må vi konstruere interaksjonsledd mellom den avhengige variabelen *female* og de tre variablene som representerer testscore på kognitive tester, *lit*, *num* og *psolv*. Vi bruker en utvidet og transformert (log-log) versjon av modell m7.

Økonometrisk modell:

$$\begin{aligned}
 \text{learnhr} = & \alpha + \delta_1 \text{female} + \beta_1 \text{exper} + \beta_2 \text{exper}^2 + \beta_3 \text{educ} + \beta_4 \ln(\text{lit}) + \beta_5 \ln(\text{num}) \\
 & + \beta_6 \ln(\text{psolv}) + \delta_2 \text{sector} + \delta_{occ} \text{occupation} + \delta_{ind} \text{industry} + \gamma_1 (\ln(\text{lit}) \\
 & * \text{female}) + \gamma_2 (\ln(\text{num}) * \text{female}) + \gamma_3 (\ln(\text{psolv}) * \text{female}) + \varepsilon
 \end{aligned}$$

Vi bruker en log-log-modell for å kunne tolke resultatene som prosentvis endring i timelønn gitt en endring i testscorene på én prosent. Referansegruppen når dummyvariabelen *female* er null, er menn. Menns prosentvise lønnsavkastning av ett proSENTS økning i testscore på de kognitive ferdighetene leseferdigheter, tallforståelse og problemløsningsevne vil være henholdsvis β_4 , β_5 og β_6 . Kvinnens avkastning vil da være $(\beta_4 + \gamma_1)$, $(\beta_5 + \gamma_2)$ og $(\beta_6 + \gamma_3)$ ettersom γ_i vil representere den ytterligere effekten en endring i testscoren har på lønn, hvis observasjonen er en kvinne.

Når vi undersøker om funnene fra en regresjon er statistisk signifikante vil nullhypotesen være $\gamma_i = 0$ ($i=1,2,3$), mot alternativhypotesen $\gamma_i \neq 0$. Intuisjonen bak dette er at hvis en koeffisient er over eller under null, får kvinner mer eller mindre lønnsavkastning fra en økning i testscoren til den ferdigheten sammenlignet med menn. Hvis koeffisienten er null, får kvinner samme avkastning som menn. Når vi bruker t-test for hver enkel kognitiv ferdighet, viser testene at ingen av testobservatorene overstiger kritisk verdi ved noe betydelig signifikansnivå i t-fordelinga. Vi kan derfor ikke med sikkerhet si at kvinner og menn får ulik lønnsavkastning fra de tre nevnte kognitive ferdighetene. En F-test for alle koeffisientene samtidig viser til samme resultat.

Tilleggsspørsmål 2:

- **Får kjønnene ulik lønnsavkastning av utdanning?**

Det er et etablert faktum at høyere utdanning vil gi høyere lønn i de aller fleste tilfeller. Vi ønsker derfor å undersøke om kvinner og menn får ulik lønnsavkastning av å ta flere år utdanning. For å undersøke dette konstruerer vi en utvidelse av modell m7, hvor vi legger til et interaksjonsledd mellom utdanning og kjønn.

Økonometrisk modell:

$$\begin{aligned} \text{learnhr} = & \alpha + \delta_1 \text{female} + \beta_1 \text{exper} + \beta_2 \text{exper}^2 + \beta_3 \text{educ} + \beta_4 \text{lit} + \beta_5 \text{num} + \beta_6 \text{psolv} \\ & + \delta_2 \text{sector} + \delta_{occ} \text{occupation} + \delta_{ind} \text{industry} + \gamma_1 \text{educ} * \text{female} + \varepsilon \end{aligned}$$

Ved å sette dummyvariabelen *female* lik null, lar dette oss observere lønnsavkastningen menn får av ett år ekstra utdanning ved å se på β_3 . $(\beta_3 + \gamma_1)$ vil derfor være lønnsavkastningen kvinner får av å ta ett år ekstra utdanning.

Regresjonsanalyse av modellen viser at $\beta_3 = 0,0299657$, og $\gamma_1 = 0,0013601$. Dette impliserer at når alle andre kontrollvariabler holdes konstant, får menn i vårt datasett en

gjennomsnittlig lønnsøkning på ~2,997% for hvert år ekstra utdanning de tar, mens kvinner får en økning på ~3,133%. Men, ved å utføre en hypotesetest med nullhypotesen om ved å ta ett år ekstra utdanning får kvinner og menn samme effekt på lønn ($\gamma_1 = 0$), finner vi at vi *ikke* kan forkaste nullhypotesen.

Vi observerte i hovedspørsmål 1 at å kontrollere for utdanning økte lønns-gapet, noe som skulle tilsi at menn får høyere lønnsavkastning enn kvinner av samme antall år utdanning. Dette finner vi altså at *ikke* er statistisk signifikant.

Tilleggsspørsmål 3:

- **Er lønns-gapet i offentlig sektor forskjellig fra det i privat sektor?**

For å undersøke om lønns-gapet i offentlig sektor er forskjellig fra i privat sektor utvider vi modell m7 ved å inkludere et interaksjonsledd mellom *female* og *sector*. Dette vil gi oss en referansegruppe med menn i privat sektor når dummyvariablene holdes lik null.

I tidligere kapittel om deskriptiv statistikk for datamaterialet ser vi at den gruppen som har høyest gjennomsnittslønn er menn i privat sektor, og at gruppen med lavest er kvinner i privat sektor.

Økonometrisk modell:

$$\begin{aligned} \text{learnhr} = & \alpha + \delta_1 \text{female} + \beta_1 \text{exper} + \beta_2 \text{exper}^2 + \beta_3 \text{educ} + \beta_4 \text{lit} + \beta_5 \text{num} + \beta_6 \text{psolv} \\ & + \delta_2 \text{sector} + \delta_{occ} \text{occupation} + \delta_{ind} \text{industry} + \gamma_1 \text{sector} * \text{female} + \varepsilon \end{aligned}$$

Regresjonsresultatene viser en estimert koeffisient til interaksjonsleddet lik $\gamma_1 = 0,0536293$. Dette impliserer at lønns-gapet mellom kvinner og menn i offentlig sektor er 5,36% mindre enn i privat sektor i vårt datamateriale. I denne modellen kan vi tolke koeffisienten δ_1 som lønns-gapet mellom kvinner og menn i privat sektor. Legger vi til γ_1 får vi naturligvis lønns-gapet i offentlig sektor. Estimaten viser at privat sektor har et lønns-gap mellom kvinner og menn på -11,4%, og at gapet i offentlig sektor er -6.04%. Resultatet er statistisk signifikant målt med t-test, og vi kan konkludere med at lønns-gapet i privat sektor er betydelig større enn det vi finner i offentlig sektor. I kapittelet om teoretisk rammeverk påpekte vi at privat sektor har en mye større andel menn enn kvinner, noe som muligens kan være med på å forsterke lønns-gapet mellom kjønnene.

5. Oppsummering og konklusjon

5.1. Oppsummering

Oppgaven ble åpnet med en beskrivelse av motivasjonen vår for valg av tema, og temaet ble aktualisert. Innenfor temaet dannet vi oss en problemstilling som var delt opp i to hovedspørsmål og tre tilleggsspørsmål. Ved å studere tidligere litteratur i form av rapporter og studier dannet vi oss et teoretisk rammeverk for å støtte opp bak forklaringer og tolkninger senere i oppgaven.

Datamaterialet som skulle brukes i analysen ble så presentert og forklart med definisjoner av relevante variabler i tillegg til deskriptiv statistikk som sammen ga en god oversikt over materialet. Eventuelle fordeler og begrensninger ved datasettet ble også utpekt.

I analysedelen av oppgaven brukte vi en relativt enkel fremgangsmåte; spørsmålene i problemstillingen ble konstruert som økonometrisk estimerbare modeller som vi kunne bruke for å formulere hypotesetester til. Etter regresjonsanalyse og hypotesetester på estimeringene var gjort, tolket vi resultatene med grunnlag i det teoretiske rammeverket.

Resten av kapittel 5 vil inneholde våre konklusjoner og endelige svar på problemstillingene, basert på det empiriske analysearbeidet gjort i oppgaven.

5.2. Hovedresultater fra empirisk analyse - konklusjon

Hovedspørsmål 1:

- Vi observerer en forskjell i gjennomsnittslønn mellom norske kvinner og menn 25 år eller eldre. Det "rå" lønnsgapet er 14,7%. Dette innebærer at kvinner gjennomsnittlig tjener 14,7% mindre enn menn når vi ikke kontrollerer for potensielle forklaringsfaktorer.

Hovedspørsmål 2:

- Vi finner at antall år erfaring har en signifikant effekt på lønnsnivå, men at forskjellen i erfaring mellom kvinner og menn *ikke* er en betydelig del av forklaringen til lønnsgapet.

- Forskjellen i antall år utdanning ser *ikke* ut til å være en forklaringsfaktor for lønnsgapet. Vi observerer at kvinner i vårt datasett gjennomsnittlig har noe mer utdanning enn menn, men når vi kontrollerer for utdanningsnivå øker lønnsgapet.
- Ulikhet i kognitive ferdigheter mellom kvinner og menn er en betydelig forklaringsfaktor for lønnsgapet. Våre observasjoner viser at menn scorer marginalt høyere på to av tre kognitive tester, og markant høyere på en tredje. Når vi kontrollerer for denne ulikheten faller lønnsgapet.
- Vi finner at den ulike fordelingen mellom kvinner og menn i offentlig og privat sektor er en signifikant del av forklaringen til lønnsgapet.
- Forskjell mellom kvinner og menn i yrke er en viktig forklaringsfaktor for lønnsgapet.
- Forskjell mellom kvinner og menn i industri er en betydelig forklaringsfaktor for lønnsgapet.
- Når alle kontrollvariable (i henhold til våre problemstillinger) er tatt hensyn til, er lønnsgapet 9,02%.

Tilleggsspørsmål 1:

- Vi finner *ikke* bevis for at kvinner og menn får ulik lønnsavkastning fra de tre kognitive ferdighetene leseferdigheter, tallforståelse og problemløsningsevne i IKT-miljø.

Tilleggsspørsmål 2:

- Vi finner *ikke* bevis for at kvinner og menn får ulik lønnsavkastning av utdanningsnivå (antall år utdanning).

Tilleggsspørsmål 3:

- Vårt datamateriale viser en tydelig forskjell i lønnsgapet innad privat sektor sammenlignet med det i offentlig sektor. I privat sektor er gapet 11,4%, mot kun 6,04% i den offentlige sektor.

5.3. Begrensninger ved analysen - forslag til utvidelser/videre forskning

Grunnet begrensninger ved datasettet kan det være nyttig å gjennomføre ytterligere analyser som har mer informasjon om faktorer som potensielt kan være med på å forklare lønnsgapet. Vår analyse finner at etter kontrollvariablene er introduserte, gjenstår det fortsatt et lønnsgap mellom kvinner og menn. Vår endelige forklaringskoeffisient er bare 0,421; våre kontrollvariable kan kun forklare 42,1% av variasjonene i lønnsnivå. Vi oppsummerer de sentrale begrensningene ved vår analyse som resultat av våre valg og begrensninger ved datasettet:

- Analysen tar ikke hensyn til eventuelle forskjeller mellom kvinner og menn med tanke på deltid og heltidsarbeid.
- Analysen tar ikke hensyn til eventuell innvandrerbakgrunn på observasjonene
- Analysen tar ikke hensyn til størrelsen på selskap observasjonene er ansatte i (inkludert om observasjonene er selv-ansatt eller ikke).
- Analysen tar ikke hensyn til eventuelle forskjeller i helsetilstand mellom observasjonene.
- Ingen informasjon om ansvarsområde til ansatte. Vi har analysen brukt variabler for antall år utdanning, erfaring, kognitive ferdigheter, sektor, yrke og industri. Dette sier ingenting om hvilken rolle og ansvarsområde arbeiderne har i bedriften, noe som oftest vil påvirke lønnen betydelig. En lederstilling tjener som oftest betydelig mer enn ansatte uten lederansvar (Statistisk sentralbyrå, 2019)
 - Det kan også være relevant å inkludere informasjon om arbeidsmengde (f.eks. gjennomsnittlig antall timer jobbet i uken).
- Ingen informasjon om eventuelle barn. Tidligere litteratur viser at når en får barn påvirker dette ikke bare lønnen, men også yrkesmuligheter samt muligheter til å oppnå lederstillinger. Det vises også av tidligere litteratur at det å få barn får større

økonomiske konsekvenser for kvinner enn menn (Kleven, Landais & Sjøgaard, 2018). For å få en mer fullstendig analyse ville det dermed vært interessant å inkludere informasjon om eventuelle barn. Det kan også være relevant å inkludere informasjon om antall sammenhengende år i arbeid.

- Ufullstendig informasjon om utdanning. Analysen inneholder kun informasjon om antall år utdanning, men mangler informasjon om hvilken type utdanning og om den er relevant for stillingen arbeidstakeren har.
Det vil også være relevant å vite hvilket nivå utdanningen er tatt på; det kan for eksempel tenkes at en person med bachelorgrader i to helt forskjellige fag opplever annet lønnsnivå enn en person med doktorgrad.
- Lite detaljert informasjon om lønn. Vi vet ikke om timelønnen oppgitt er en fastsatt timelønn, eller om det er en gjennomsnittlig timelønn som inkluderer individenes overtidsarbeid/tillegg/bonuser osv. Dette kan skape unøyaktighet i analysen hvis enkelte observasjoner har oppgitt sin faste timelønn uten tillegg og bonuser, mens andre har inkludert dette. Videre kan det være relevant å justere timelønnen for risiko arbeidet eventuelt medfører, da det ofte kommer risikotillegg for farlig arbeid (militær, politi, gruvearbeid eller lignende).
- Å kontrollere for faktisk produktivitet kan tenkes å ha innvirkning på resultatene. Modellene i vår analyse forutsetter lik produktivitet for alle observasjonene, men det kan antas at en mer produktiv ansatt over tid vil se en mer positiv lønnsutvikling enn mindre produktiv ansatt i samme type arbeid.
- Analysen kontrollerer ikke for eventuell ren kjønnsdiskriminering i arbeidsmarkedet/på arbeidsplassen.

Våre forslag til videre forskning vil rett og slett være å gjennomføre analyser lik vår, men med flere og mer detaljerte kontrollvariabler.

6. Litteraturliste

- Barth, E., Hardoy, I., Schøne, P. & Østbakken, K. M. (2013:7). *Lønnsforskjeller mellom kvinner og menn – hva har skjedd på 2000-tallet?*
- Blau, F. D. & Kahn, L. M. (2016). The gender wage gap: Extent, Trends, and Explanation. *The National Bureau of Economic Research*, vol 55(3), sider 789-865. <https://doi.org/10.3386/w21913>
- Bye, K. S. (21.11.2018). Mastergrad gir høyere lønn. Hentet fra: <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/mastergrad-gir-hoyere-lonn>
- NOU 2008:6 (2008). *Kjønn og lønn – fakta, analyser og virkemidler for likelønn*. Oslo: Regjeringen.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, vol 124(2), sider 262-274. <https://dx.doi.org/10.1037/0033-2909.124.2.262>
- Statistisk Sentralbyrå. (21.12.2018). Indikatorer for kjønnslikestilling i kommunene. Hentet fra: <https://www.ssb.no/likekom>
- Statistisk Sentralbyrå. (28.03.2019). Studenter i høyere utdanning. Hentet fra: <https://www.ssb.no/utuvh>
- Stokke, H. E. (2016). The gender wage gap and the early-career effect. *The Society of Labor Economists*.
- Thomas, R. L. (2005). *Using Statistics in Economics*. Berkshire: McGraw Hill Education
- Utdanningsforbundet Oslo. (18.10.2017). Equal Pay Day. Hentet fra: <https://www.utdanningsforbundet.no/fylkeslag/oslo/nyheter/felles/equal-pay-day/>
- Likestillings- og diskrimineringsloven. (2018). Lov om likestilling og forbud mot diskriminering (LOV-2017-06-16-51). Hentet fra: <https://lovdata.no/dokument/NL/lov/2017-06-16-51?q=lov%20om%20likestilling%20og%20forbud>

7. Appendix – komplette regresjonsresultater

1. Hovedspørsmål 1

VARIABLES	(1) learnhr
female	-0.147*** (0.0112)
Constant	5.444*** (0.00787)
Observations	3,023
R-squared	0.053

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2. Hovedspørsmål 2 (2 sider)

VARIABLES	m1	m2	m3	m4	m5	m6	m7
female	- 0.147*** (0.0112)	-0.147*** (0.0110)	-0.152*** (0.0100)	-0.134*** (0.0105)	-0.120*** (0.0109)	-0.112*** (0.0119)	-0.0902*** (0.0120)
exper		0.0243*** (0.00198)	0.0241*** (0.00181)	0.0224*** (0.00190)	0.0223*** (0.00191)	0.0212*** (0.00199)	0.0198*** (0.00196)
exper_sq		- 0.000452*** (4.21e-05)	- 0.000398*** (3.85e-05)	- 0.000330*** (4.09e-05)	- 0.000319*** (4.11e-05)	- 0.000320*** (4.30e-05)	- 0.000294*** (4.24e-05)
educ			0.0540*** (0.00220)	0.0427*** (0.00256)	0.0466*** (0.00264)	0.0312*** (0.00305)	0.0305*** (0.00307)
lit				-0.000265 (0.000287)	-0.000164 (0.000289)	-0.000400 (0.000299)	-0.000595** (0.000295)
num				0.000959*** (0.000242)	0.000887*** (0.000243)	0.000967*** (0.000252)	0.000850*** (0.000247)
psolv				0.000967*** (0.000238)	0.000882*** (0.000240)	0.000500** (0.000250)	0.000585** (0.000246)
sector					-0.0620*** (0.0110)	-0.0797*** (0.0118)	0.0115 (0.0173)
3.occupation						-0.0419** (0.0206)	-0.0364* (0.0209)
4.occupation						-0.0566*** (0.0206)	-0.0941*** (0.0207)
5.occupation						-0.218*** (0.0277)	-0.213*** (0.0278)
6.occupation						-0.207*** (0.0229)	-0.201*** (0.0228)
7.occupation						-0.476*** (0.128)	-0.235 (0.157)
8.occupation						-0.157*** (0.0266)	-0.196*** (0.0283)
9.occupation						-0.157*** (0.0337)	-0.212*** (0.0356)
10.occupation						-0.259*** (0.0388)	-0.257*** (0.0391)
8.industry							0.595*** (0.105)
9.industry							0.311*** (0.101)

10.industry							0.356*** (0.113)
11.industry							0.412** (0.172)
12.industry							0.283*** (0.101)
13.industry							0.184* (0.100)
14.industry							0.238** (0.103)
15.industry							0.169 (0.106)
16.industry							0.351*** (0.102)
17.industry							0.314*** (0.105)
18.industry							0.194* (0.114)
19.industry							0.277*** (0.102)
20.industry							0.216** (0.103)
21.industry							0.172* (0.103)
22.industry							0.0719 (0.102)
23.industry							0.172* (0.101)
24.industry							0.0939 (0.119)
25.industry							0.234** (0.110)
Constant	5.444*** (0.00787)	5.185*** (0.0211)	4.354*** (0.0389)	4.038*** (0.0532)	4.015*** (0.0534)	4.530*** (0.0697)	4.372*** (0.121)
Observations	3,023	3,021	3,020	2,734	2,687	2,350	2,235
R-squared	0.053	0.105	0.255	0.287	0.298	0.359	0.421

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Tillegg til hovedspørsmål 2 - kommentar om ulik avkastning på erfaring i modell m2:

VARIABLES	(1) learnhr
female	-0.135*** (0.0241)
exper	0.0248*** (0.00213)
exper_sq	-0.000456*** (4.28e-05)
experfemale	-0.000535 (0.00101)
Constant	5.178*** (0.0248)
Observations	3,021
R-squared	0.105

Standard errors in parentheses

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

3. Tilleggsspørsmål 1

VARIABLES	(1) learnhr
female	0.538 (0.440)
exper	0.0195*** (0.00197)
exper_sq	-0.000290*** (4.26e-05)
educ	0.0306*** (0.00307)
llit	-0.206* (0.115)
lnum	0.345*** (0.100)
lpsolv	0.143 (0.0917)
sector	0.0113 (0.0173)
3.occupation	-0.0362* (0.0209)
4.occupation	-0.0950*** (0.0207)
5.occupation	-0.214*** (0.0278)
6.occupation	-0.204*** (0.0229)
7.occupation	-0.226 (0.157)
8.occupation	-0.195*** (0.0284)
9.occupation	-0.210*** (0.0357)
10.occupation	-0.256*** (0.0392)
8.industry	0.600*** (0.105)
9.industry	0.318*** (0.101)
10.industry	0.361*** (0.113)
11.industry	0.416** (0.172)
12.industry	0.290*** (0.101)
13.industry	0.191* (0.100)

14.industry	0.245** (0.103)
15.industry	0.181* (0.106)
16.industry	0.356*** (0.102)
17.industry	0.323*** (0.105)
18.industry	0.202* (0.114)
19.industry	0.283*** (0.102)
20.industry	0.224** (0.103)
21.industry	0.178* (0.103)
22.industry	0.0783 (0.102)
23.industry	0.178* (0.101)
24.industry	0.104 (0.119)
25.industry	0.239** (0.110)
llitfemale	0.0709 (0.160)
lnumfemale	-0.200 (0.135)
lpsolvfemale	0.0184 (0.129)
Constant	3.011*** (0.352)

Observations	2,235
R-squared	0.423

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

4. Tilleggsspørsmål 2

VARIABLES	(1) learnhr
female	-0.111 (0.0759)
exper	0.0198*** (0.00196)
exper_sq	-0.000295*** (4.24e-05)
educ	0.0300*** (0.00370)
lit	-0.000597** (0.000295)
num	0.000850*** (0.000247)
psolv	0.000586** (0.000246)
sector	0.0114 (0.0173)
3.occupation	-0.0364* (0.0209)
4.occupation	-0.0940*** (0.0207)
5.occupation	-0.213*** (0.0279)
6.occupation	-0.200*** (0.0230)
7.occupation	-0.237 (0.157)
8.occupation	-0.196*** (0.0284)
9.occupation	-0.212*** (0.0357)
10.occupation	-0.256*** (0.0392)
8.industry	0.594*** (0.105)
9.industry	0.310*** (0.101)
10.industry	0.356*** (0.113)
11.industry	0.409** (0.172)
12.industry	0.282*** (0.101)
13.industry	0.183* (0.100)

14.industry	0.236** (0.103)
15.industry	0.168 (0.106)
16.industry	0.350*** (0.103)
17.industry	0.314*** (0.105)
18.industry	0.192* (0.114)
19.industry	0.276*** (0.102)
20.industry	0.214** (0.103)
21.industry	0.171* (0.103)
22.industry	0.0709 (0.102)
23.industry	0.171* (0.101)
24.industry	0.0924 (0.119)
25.industry	0.233** (0.111)
educfemale	0.00136 (0.00489)
Constant	4.381*** (0.126)

Observations	2,235
R-squared	0.422

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

4. Tilleggsspørsmål 3

VARIABLES	(1) learnhr
female	-0.114*** (0.0157)
exper	0.0197*** (0.00196)
exper_sq	-0.000292*** (4.24e-05)
educ	0.0309*** (0.00307)
lit	-0.000588** (0.000294)
num	0.000865*** (0.000247)
psolv	0.000583** (0.000246)
sector	-0.0209 (0.0220)
3.occupation	-0.0347* (0.0209)
4.occupation	-0.0931*** (0.0207)
5.occupation	-0.206*** (0.0280)
6.occupation	-0.195*** (0.0229)
7.occupation	-0.249 (0.157)
8.occupation	-0.199*** (0.0283)
9.occupation	-0.215*** (0.0356)
10.occupation	-0.256*** (0.0390)
8.industry	0.586*** (0.105)
9.industry	0.300*** (0.101)
10.industry	0.353*** (0.113)
11.industry	0.398** (0.172)
12.industry	0.273*** (0.101)
13.industry	0.178* (0.100)

14.industry	0.231** (0.102)
15.industry	0.163 (0.106)
16.industry	0.340*** (0.102)
17.industry	0.309*** (0.105)
18.industry	0.187* (0.114)
19.industry	0.269*** (0.102)
20.industry	0.210** (0.103)
21.industry	0.171* (0.103)
22.industry	0.0663 (0.102)
23.industry	0.165 (0.101)
24.industry	0.0974 (0.119)
25.industry	0.237** (0.110)
sectorfemale	0.0536** (0.0227)
Constant	4.377*** (0.120)

Observations	2,235
R-squared	0.423

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

