

Magnus Holten Skånlund
Hilde Dørum Haug

Lønnsgapet mellom innvandrere og befolkningen ellers i Norge

En Oaxaca-Blinder dekomponering med fokus på heterogeniteter blant innvandrere

Masteroppgave i Samfunnsøkonomi

Veileder: Hildegunn Ekroll Stokke

Mai 2023

Magnus Holten Skånlund
Hilde Dørum Haug

Lønns-gapet mellom innvandrere og befolkningen ellers i Norge

En Oaxaca-Blinder dekomponering med fokus på heterogeniteter blant innvandrere

Masteroppgave i Samfunnsøkonomi
Veileder: Hildegunn Ekroll Stokke
Mai 2023

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Forord

Denne oppgaven markerer slutten på 5 års skolegang på NTNU. Arbeidet med oppgaven har vært utfordrende, men også spennende og lærerikt. Vi vil først og fremst takke vår veileder Hildegunn Ekroll Stokke for konstruktive innspill, nyttige tilbakemeldinger og god hjelp gjennom semesteret. I tillegg ønsker vi også å takke familiene våre for mye korrekturlesing og mange gode råd. Til slutt vil vi takke våre medstudenter for gode diskusjoner, og for å ha gjort studietiden vår uforglemmelig.

Trondheim 28.05.2023

Hilde Dørum Haug

Magnus Holten Skånlund

Sammendrag

I denne masteroppgaven har vi sett på lønnsgapet mellom innvandrere og befolkningen ellers i Norge. I oppgaven har vi brukt mikrodata på individnivå fra 2019, og videre brukt en Oaxaca-Blinder dekomponering for å skille og undersøke de forklarte og uforklarte effektene ved lønns-gapet. Vi finner et signifikant lønnsgap mellom innvandrere og befolkningen ellers, hvor innvandrere tjener gjennomsnittlig 11% mindre enn befolkningen ellers. Det meste av dette gapet skyldes at innvandrere jobber i dårligere betalte yrker. I tillegg har de dårligere avkastning på arbeidserfaring, som gjør at gapet blir større utover i arbeidslivet.

Disse resultatene gjelder alle innvandringsgruppene vi ser på, utenom innvandrere fra høyinntektsland, som gjennomsnittlig tjener mer enn befolkningen ellers. Vi finner at lønnsgapet er størst når vi ser på gruppen som har familie eller flukt som innvandringsgrunn. Her er lønns-gapet på hele 22%, men forskjeller i egenskaper forklarer en stor del av dette gapet. Denne studien fremhever at innvandrere ikke er en homogen gruppe, og at det er nyttig å se på effekter blant ulike innvandringsgrupper.

Abstract

In this master's thesis, we have looked at the wage gap between immigrants and the rest of the population in Norway. In the thesis, we used microdata at the individual level from 2019 and an Oaxaca-Blinder decomposition to separate and investigate the explained and unexplained effects of the wage gap. We find a significant wage gap between immigrants and the rest of the population, where immigrants earn, on average, 11% less. Most of this gap is due to immigrants working in lower-paid occupations. In addition, they have a lower return to work experience, which causes the gap to grow through their career.

These results apply to all the immigration groups we look at, except for immigrants from high-income countries, who earn more than the rest of the population on average. We find that the pay gap is most significant when we look at the group whose reason for immigration is family or refuge. Here the pay gap is as much as 22%, but differences in characteristics explain a large part of this gap. This study highlights that immigrants are not a homogeneous group and that it is useful to look at the effects among different immigration groups.

Innhold

1 Innledning	1
2 Teori og tidligere litteratur	3
2.1 Tidligere litteratur	3
2.2 Oaxaca-Blinder og lønnsgap	6
2.3 Oppsummering	8
3 Datagrunnlag	9
3.1 Microdata.no	9
3.2 Avgrensning av utvalget	10
3.3 Variabler	11
4 Metode	21
4.1 Oaxaca-Blinder - Dekomponering av lønnsgapet	22
4.2 Utvidelser av modellen - Heterogeniteter blant innvandrere	23
4.3 Økonometriske utfordringer	24
5 Resultater	28
5.1 Resultater fra OLS	28
5.2 Resultater fra Oaxaca-Blinder	31
5.3 Analyse av heterogeniteter	35
6 Diskusjon	42
6.1 Drøfting av resultatene	42
6.2 Utfordringer med datasettet og forslag til videre forskning	47
7 Konklusjon	49
Referanser	I
A Appendiks	III

1 Innledning

De siste tre tiårene har andelen innvandrere i befolkningen steget fra 4,5% til over 18% ved utgangen av 2019 (SSB, 2023). Majoriteten av dem er i jobb, og dermed er det ingen tvil om at innvandrere nå utgjør en betydelig del av arbeidsstyrken. Det er kjent at det er lønnsulikheter mellom innvandrere og befolkningen ellers i Norge, men det er forskjellige hypoteser og svar på hva dette faktisk skyldes. Denne oppgaven undersøker dette lønnsgapet med bruk av data fra Microdata.no på individnivå fra 2019. I oppgaven er en innvandrer definert som en person som har flyttet til Norge fra et annet land, eller som har minst en forelder som har gjort det. Lønnsgapet defineres som forskjellen i årlig gjennomsnittsinntekt mellom innvandrere og befolkningen ellers.

I denne oppgaven gir vi en analyse av lønnsgapet mellom innvandrere og befolkningen ellers i Norge med fokus på hvor mye som kan forklares av observerbare faktorer, og hvor mye som skyldes uobserverbare faktorer. Sistnevnte kan tyde på diskriminering i arbeidsmarkedet. Vi bruker en Oaxaca-Blinder dekomponering for å separere og analysere de forklarte og uforklarte delene av gapet, og se hva som skyldes forskjeller i egenskaper, og hva som skyldes forskjeller i avkastning på egenskaper. Videre ser vi på lønnsgapet for ulike innvandringsgrupper, separert på landbakgrunn og innvandringsgrunn for å identifisere heterogeniteter blant innvandrere.

Ulikhet i lønn kan ha mange årsaker. Et lønnsgap mellom innvandrere og befolkningen ellers kan skyldes forklarte faktorer som at mange innvandrere snakker dårlig norsk, som kan gjøre at de ikke kvalifiserer til noen jobber. I tillegg har innvandrere høyere sannsynlighet for ulike helseplager som diabetes, hjerte- og karsykdommer, psykiske helseplager og rygg- og nakkeplager, som kan gjøre at de må jobbe mindre enn befolkningen ellers (Folkehelseinstituttet, 2022). Videre kan det også skyldes at innvandrere har utdanning fra andre land som ikke godkjennes i Norge, og de ender opp i dårligere jobber enn utdanningsnivået skulle tilsi. Segregering i arbeidsmarkedet hvor innvandrere er overrepresentert i lavtlønnsyrker og -næringer kan også være en kilde til lønnsulikhet. I tillegg kan lønnsgapet skyldes uforklarte faktorer, som ulike typer diskriminering. Det er dette vi ønsker å se på.

Videre ser vi på om funnene vi gjør, gjelder på tvers av de ulike innvandringsgruppene. Trolig er ikke alle innvandrere én homogen gruppe, og det vil være store forskjeller på tvers av gruppene. En viktig del av analysen vår blir å belyse forskjeller og heterogeniteter blant innvandrere.

Denne oppgaven skal bidra til den allerede eksisterende forskningen på temaet om størrelsen, og mulige forklaringer, på lønnsgapet mellom innvandrere og befolkningen ellers, spesifikt i Norge. Bruken av registerdata for hele befolkningen gjør at vi kan få data med god kvalitet for et stort utvalg, med liten sjanse for store målefeil. Videre vil bruken av Oaxaca-Blinder gjøre at vi kan se mer spesifikt på hva lønnsgapet skyldes for innvandrere generelt, og for ulike innvandrergupper. Vår forventning til resultatene vil være at det er store forskjeller mellom de forskjellige innvandrerguppene. Det går ikke an å sette alle innvandrere inn i en gruppe ettersom det er såpass store variasjoner i landbakgrunn og begrunnelse for innvandringen. Derfor forventer vi å se store variasjoner mellom de forskjellige innvandringsgruppene.

Oppgaven er strukturert som følger. I kapittel 2 presenterer vi tidligere teori og litteratur rundt temaet. Kapittel 3 legger frem datagrunnlaget for oppgaven, med valg av variabler og deskriptiv statistikk. I kapittel 4 presenterer vi metoden vi bruker, og utfordringer med denne. Resultatene presenteres i kapittel 5, og blir diskutert videre i kapittel 6, hvor vi også drøfter videre forskning. Til slutt konkluderer vi og oppsummerer de viktigste funnene i kapittel 7.

2 Teori og tidligere litteratur

I dette kapitlet skal vi se på tidligere forskning rundt temaet lønn og innvandrere, og vi ser at det er en betydelig lønnsforskjell mellom innvandrere i Norge og befolkningen ellers. Andre analyser har vist at lignende lønnsgap kommer fram i de aller fleste land med lignende økonomisk situasjon som Norge. Innvandrere som kommer til Norge, tjener mindre enn befolkningen ellers (Bye, 2021). Dette kan komme av en rekke faktorer, spørsmålet som tidligere forskning har forsøkt å besvare blir dermed hva som ligger til grunn for denne forskjellen. Er det en konsekvens av systematisk diskriminering av innvandrere, eller er det andre forklarende faktorer som påvirker forhandlingskraften og lønnsgapet til innvandrere? Det finnes mange forskjellige metoder man kunne brukt for å analysere denne problemstillingen. Vi har valgt å bruke dekomponeringen fremstilt av Blinder (1973) og Oaxaca (1973). Denne økonometriske tilnærmingen er den mest hyppig brukte innen feltet om diskriminering og lønnsgap.

2.1 Tidligere litteratur

Tidligere litteratur har vist at vi har et lønnsgap selv når vi ser på andregenerasjonsinnvandrere, som vil si de som er født og oppvokst i Norge, men med minst én forelder som har innvandret. Utdanningsdirektoratets rapport peker på at elever med innvandrerbakgrunn generelt gjør det dårligere på nasjonale prøver enn befolkningen ellers (Utdanningsdirektoratet, 2017). Et slikt utgangspunkt kan være årsak til det lønnsgapet som forekommer senere i arbeidslivet. Vi skal nå presentere forskning og litteratur som peker på forskjellige faktorer til at innvandrere generelt tjener mindre enn befolkningen ellers.

2.1.1 Statistisk, preferansebasert og monopsonistisk diskriminering

Det finnes flere typer diskriminering, og i artikkelen til Kolsrud mfl. (2016) om innvandrere i det norske arbeidsmarkedet, skiller de mellom statistisk, preferansebasert og monopsonistisk diskriminering på arbeidsplassen. Dette er begreper som brukes for å forklare lønnsforskjeller mellom ansatte med innvandringsbakgrunn og resten av de ansatte.

Ved uregulerte markeder, der både arbeidsgiverne og arbeidstakerne er godt informert, vil arbeidstakere lønnes etter sin relative produktivitet. En arbeidstaker med høyere produktivitet vil da få høyere lønn. Dette er regnet som en legetim grunn til at noen får høyere lønn enn andre. Men der det ikke er full informasjon kan det oppstå det som kalles *statistisk diskriminering*.

Dette foregår når arbeidsgivere ikke har samme tilgang på informasjon om produktiviteten til jobbsøkere med innvandringsbakgrunn som de har om befolkningen ellers. En rasjonell arbeidsgiver vil dermed øke produktiviteten på tryggest mulig vis. Hen vil gi høyere lønn til de ansatte hen er mest sikker på har den høyeste produktiviteten, noe som ofte slår negativt ut for innvandrere. Dette er en form for legitim diskriminering der arbeidsgiveren baserer lønnen på en rasjonell oppfatning av en gruppes gjennomsnittsproduktivitet dersom hen mangler informasjon om produktiviteten til en arbeidstaker.

Det andre begrepet som nevnes i artikkelen til Kolsrud mfl. (2016) er *preferansebasert diskriminering*. Dette forekommer når grupper ender opp med en lavere stilling i arbeidsmarkedet fordi de ikke har de samme mulighetene eller alternativene som befolkningen ellers. Det handler ikke om at de får lavere lønn for å utføre den samme typen jobb, men at de oftere havner i jobber som har lavere gjennomsnittslønn. En slik diskriminering kan skje som følge av at innvandrere som kommer til Norge, har mindre kjennskap til arbeidsmarkedet og sine egne rettigheter. I tillegg til at de har færre jobbalternativer og dermed dårligere forhandlingskraft i møte med arbeidsgivere. Et annet aspekt kan være at folk fra forskjellige folkegrupper ønsker å jobbe sammen med personer med lik bakgrunn og dermed ender opp i et yrke med lavere lønn enn det deres individuelle produktivitet skulle tilsi. Dette vil hovedsakelig være problematisk for førstegenerasjonsinnvandrere. Generelt finner Kolsrud mfl. (2016) at innvandrere er overrepresentert i næringer med dårligere arbeidsvilkår, og underrepresentert i næringer med gode arbeidsvilkår. Dårlige arbeidsvilkår kan være lavere lønn, midlertidige ansettelser, lange arbeidsdager, og lignende (Kolsrud mfl., 2016).

Et annet problem for arbeidstakere som er innvandrere, er at utdanningen deres ikke nødvendigvis vil være gyldig i Norge. I sin artikkel om diskriminering i arbeidslivet i Norge, påpeker Dale-Olsen mfl. (2014) at arbeidsgivere trolig ikke mener å være diskriminerende, men at de kan stå overfor en situasjon hvor en potensiell arbeidstaker sin utdanning ikke er godkjent i Norge. Selv om utdanningen tilsynelatende er lik, er det ikke alle utdanninger som godkjennes. Dette er med på å gjøre det utfordrende for arbeidstakere å vise sin kompetanse, samtidig som det vil være høyere risiko for en arbeidsgiver å ansette en innvandrer med usikkert utdanningsgrunnlag enn en som har tatt den samme utdanningen i Norge.

Dale-Olsen mfl. (2014) tar også opp et tredje begrep, nemlig *monopsonistisk diskriminering*. En arbeidsgiver kan påvirke arbeidstilbudet gjennom lønnsfastsettelsen. Bedrifter bruker lønnen til

å få tak i den beste arbeidskraften (Manning, 2011). Under monopsonistisk diskriminering vil arbeidsgiveren hige etter å presse lønnen så lavt som mulig uten at det påvirker kvaliteten på arbeidskraft. Denne markedsmakten som arbeidsgiverne da innehar, vil åpenbart være sterkt avhengig av følsomheten til arbeidstilbudet. Hypotesen blir dermed at arbeidstilbudet til innvandrere er mindre følsomt enn for befolkningen ellers. Dette vil føre til at en arbeidsgiver med markedsmakt vil kunne utnytte dette til å presse ned lønningene relativt mer enn for befolkningen ellers. Følsomheten beregnes her gjennom parametere for sannsynlighet for å bli arbeidsledig, samt sannsynligheten for å få nye jobbtilbud. Det kommer fram at innvandrere generelt er mer tilbøyelige til å bli værende i jobb selv om stillingsprosenten blir redusert, og at de har en lavere smertegrense når det kommer til reduksjon i lønnen i forhold til befolkningen ellers.

2.1.2 Opprinnelse og integrering

Et gjentakende problem, spesielt for førstegenerasjonsinnvandrere, er at den utdanningen de har fra hjemlandet sitt, ikke nødvendigvis korresponderer med tilsvarende utdanningsgrad i Norge. Kvaliteten på utdanningen varierer mellom landene, og det vil generelt være enklere for en person fra et høyinntektsland å benytte utdanningen sin i Norge enn det vil være for en fra et lavinntektsland (Bratsberg mfl., 2014). Videre finner også Hardoy og Schøne (2011) avvik mellom avkastning på utdanningen til innvandrere og befolkningen ellers. De finner at utdanning en innvanderer tar i Norge, vil være like mye verdt som utdanningen noen fra befolkningen ellers tar, men at utdanning fra hjemlandet vil de få lavere avkastning på, spesielt med utdanning fra ikke-vestlige land. Generelt ser de også at ikke-vestlige innvandrere er overutdannet for jobbene de har, altså at deres utdanningsnivå er høyere enn det som kreves for jobben.

På bakgrunn av dette vil lønnsgapet mellom innvandrere og befolkningen ellers variere med hvilket land det innvandles fra, spesielt på grunn av kvaliteten på utdanningen. En arbeidsgiver kan ha forskjellig kunnskap om ulike utdanningsinstitusjoner rundt om i verden, og kan være mer skeptisk til en utdanning fra et universitet den ikke kjenner til, eller fra land den ikke er sikker på kvaliteten til utdanningen i. I tillegg er det utdanninger fra andre land som også reelt sett er dårligere enn utdanningen fra Norge. Som Hardoy og Schøne (2011) finner, er det spesielt de fra ikke-vestlige land, eller lavinntektsland, som vil ha størst vanskeligheter med å få utdanningen sin overført. Dette betyr at innvandrere fra høyinntektsland vil ha et konkurransefortrinn i forhold til resten av innvandrerne.

Det finnes mange eksempler på at landbakgrunnen og innvandringsgrunnen til en innvandrere har stor innvirkning på lønnen. Blant annet finner Kolsrud mfl. (2016) at det er et betydelig lønnsgap mellom innvandrere som kommer fra lavinntektsland (Asia, Afrika, Latin-Amerika og Øst-Europa) og befolkningen ellers, selv mellom individer med samme utdanning, kjønn og arbeidserfaring. Det eksisterer også et lønnsgap mellom innvandrere fra høyinntektsland (Nord-Amerika, resten av Europa og Oseania) og befolkningen ellers, men dette er vesentlig mindre, og ikke alltid negativt for innvandrerne. De finner også at lønnsgapet er størst for innvandrere som kommer som flyktninger, og mindre for de som kommer for arbeid. Lignende funn gjør også Barth mfl. (2012) i sin artikkel om lønnsgap mellom bedrifter. Der viser de at mye av lønnsgapet mellom innvandrere og befolkningen ellers kan forklares med at innvandrere generelt har dårligere betalte jobber og sliter med å komme seg opp til høyere betalte jobber, og dette gjelder spesielt for innvandrere fra utviklingsland.

Det er rimelig å anta at jo lengre en innvandrere har vært i Norge, jo mindre vil lønnsgapet bli, fordi de gradvis blir mer integrert. Flere artikler tyder imidlertid på at dette ikke alltid stemmer. Bratsberg mfl. (2011) ser på sysselsettingsrate og trygderate blant ulike innvandringskohorter som kom til Norge på ulike tidspunkt og med ulik innvandringsgrunn. De finner, som forventet, at de som kommer til Norge for arbeid har generelt høyere sysselsettingsrate og færre som går på trygd enn de som kommer for beskyttelse. Men samtidig finner de også at for nesten alle gruppene, uavhengig av innvandringsgrunn og innvandringstidspunkt, er det en tendens til at sysselsettingsratene synker, og trygderatene øker, etter at gruppen har vært i Norge i 10-15 år. Kolsrud mfl. (2016) finner også denne tendensen til at lønnsveksten avtar etter en periode på 10-15 år etter at de kom til landet, spesielt om de var flyktninger. De ser at for innvandrere fra Asia, Afrika og Latin-Amerika er botidseffekten negativ, så lønnsforskjellen mellom en innvandrere og en fra befolkningen ellers blir større jo lengre innvandreren har vært i Norge. I starten er det en jevn lønnsvekst, men lønnsveksten avtar betydelig mer for denne innvandringsgruppen etter 10-15 år. For innvandrere fra Øst-Europa er det først en positiv botidseffekt, men denne snur til negativ også 10-15 år etter innvandring.

2.2 Oaxaca-Blinder og lønnsgap

Ved analyser hvor man skal sammenligne to folkegrupper, herunder for å studere arbeidsmarkedseffekter, kan man bruke en Oaxaca-Blinder dekomponering. Den hjelper med å finne forskjeller mellom to grupper, som innvandrere og befolkningen ellers. Modellen deler opp lønns-

gapet mellom gruppene til en forklart og en uforklart komponent. Det forklarte er kontrollerbare faktorer som påvirker lønnsgapet, som for eksempel utdanningsnivå, arbeidserfaring, kjønn og lignende. Det uforklarte er det som ikke kan forklares av slike faktorer, og som ofte er et mål på diskriminering, men det inneholder også uobserverbare effekter (Jann, 2008).

Chiswick (1978) var den første som formulerte teorien rundt diskriminering på arbeidsplassen gjennom å bruke dekomponeringen av lønnsgap formulert av Oaxaca (1973). Etter denne artikkelen til Chiswick har Oaxaca-Blinder dekomponeringen blitt hyppig brukt for å studere effektene av blant annet lønnsgap mellom innvandrere og befolkningen ellers. Oaxaca-Blinder blir brukt til å determinere hva som egentlig inngår i lønnsgapet, om alle effektene kan forklares gjennom egenskaper, eller om det foreligger diskriminering.

Mange tidligere studier av lønnsgap mellom innvandrere og befolkningen ellers har brukt en Oaxaca-Blinder dekomponering i sin analyse. Lehmer og Ludsteck (2011) ser på forskjellen i lønn for mannlige arbeidere fra ulike Øst- og Vest-Europeiske land i Tyskland. Der viser Oaxaca-Blinder dekomponeringen at lønnsforskjellene mellom arbeidere fra Øst-Europa og Vest-Europa trolig ikke skyldes diskriminering, men heller forskjeller i observerbare egenskaper.

Nielsen mfl. (2004) fant i sin artikkel at det først var en tydelig negativ lønnspremie av å være innvandrere, og så videre på hvordan dette lønnsgapet utviklet seg etterhvert som innvandrerne fikk tid til å integrere seg. Nielsen mfl. (2004) baserer seg på registerdata fra Danmark for å se på effekten av innvandring, om det foreligger diskriminering og hvordan effekten av integrering er over tid. De tok i bruk paneldata kombinert med dekomponeringsmodeller for å se på tre effekter; om det er et lønnsgap, om lønnsgapet avtar med erfaring (integrering) og om det foreligger diskriminering (dersom lønnsgapet ikke kan forklares). Gjennom bruk av en statisk dekomponeringsmodell (Oaxaca-Blinder) finner Nielsen mfl. (2004) at lønnsgapet mellom innvandrere og befolkningen ellers stort sett kan forklares gjennom forskjeller i kvalifikasjoner og egenskaper som språk og lignende. De peker også på at den aller største grunnen til at det i det hele tatt foreligger et lønnsgap er fordi at det er et skille i hvor innvandrere og befolkningen ellers ender opp med å jobbe. Disse funnene stemmer godt overens med de som ble presentert av Dale-Olsen mfl. (2014) og Kolsrud mfl. (2016) tidligere i kapittelet. Lønnsgapet handler mye om hvor folk ender opp med å jobbe, som igjen kommer an på hvor mye kunnskap hvert individ har om arbeidsmarkedet, kunnskapen arbeidsgiveren har om personen i form av utdanning og erfaring, samt individets faktiske kvalifikasjoner.

Det er mulig å bruke Oaxaca-Blinder dekomponering med flere forskjellige typer datasett, enten det er tversnitt eller paneldata. For vår analyse bruker vi tversnittsdata, ettersom vi skal se på individspesifikke egenskaper som ikke vil endre seg over tid, altså det å være innvandrere, og fordi Microdata.no best kan håndtere store mengder data når det er tversnitt.

2.3 Oppsummering

Forskningen vi har sett på, viser at det i de aller fleste tilfeller forekommer et tydelig lønnsgap mellom innvandrere og befolkningen ellers. Vi har vært innom forskjellige former for diskriminering av innvandrere med Dale-Olsen mfl. (2014) og Kolsrud mfl. (2016). De beskriver hvordan innvandrere ofte er et offer for både målrettet og utilsiktet diskriminering. De fleste analysene som sammenlikner individer med like forutsetninger med tanke på utdanning og erfaring, kommer fram til at det ikke er et tydelig lønnsgap dersom innvandrere og befolkningen ellers har samme stilling. Den store forskjellen, og grunnlaget for lønnsgapet ligger i ansettelsesprosessen og spesielt allokeringen av arbeidskraften.

Litteraturen om lønnsgap, spesielt mellom to grupper har gjort det tydelig at det er Oaxaca-Blinder dekomponering som er den beste tilnærmingen for vår analyse. I neste kapittel vil vi gå videre og beskrive datasettet og hvilke variabler vi benytter i vår analyse.

3 Datagrunnlag

3.1 Microdata.no

Microdata.no er et nettbasert analyseprogram som er utviklet og driftet av Statistisk Sentralbyrå og Sikt (kunnskapssektorens tjenesteleverandør) (SSB, 2022). Systemet gir tilgang til detaljerte mikrodata og gir mulighet til å hente ut informasjon på individnivå om blant annet utdanning, landbakgrunn, økonomi, trygd og sysselsetting. Datasettet inkluderer alle individer som har, eller har hatt et midlertidig eller permanent norsk fødselsnummer.

3.1.1 Konfidensialitetstiltak

Ettersom man i Microdata.no har tilgang til så detaljert og sensitiv persondata, er det ikke mulig å se på datasettet. Det er heller ikke mulig å eksportere datasettet til et annet program, eller å importere annen data inn i Microdata.no. Man kan kun bruke funksjonaliteter som deskriptiv statistikk eller regresjonsanalyser for å få beskrivelser av datasettet (Johansen, 2020).

Microdata.no har strenge konfidensialitetstiltak for å ivareta personvernet. Det er et eksplisitt krav i Microdata.no at det ikke skal være mulig å avsløre informasjon om enkeltindivider. Et av konfidensialitetstiltakene er at man ikke kan definere en populasjon med færre enn 1000 individer. Dette er problematisk i vår oppgave når vi ønsker spesifikk informasjon om for eksempel kun en yrkesgruppe. Da er det flere tilfeller hvor utvalget er for lite, og vi har vært nødt til å gruppere sammen yrker for å få informasjon, dette kommer vi mer tilbake til senere.

Videre benytter Microdata.no en 2% winsorisering, som betyr at de øverste 1% verdiene for hver variabel settes til 99-prosentilen, og de 1% laveste verdiene settes til 1-prosentilen. Dette påvirker alle deskriptive statistikker og lignende, og gjør at ekstremverdier ikke blir synlige. Dette vil føre til en mindre skjevfordeling, og samtidig vil vi få litt feilestimert gjennomsnitt og standardavvik. Winsoreringen omfatter ikke regresjonsanalyser ettersom det ikke gir personidentifiserende informasjon. Til slutt legges det til støy på operasjoner som viser antall enheter i et datasett. Ved alle positive visninger av antall i output vil det legges til, eller trekkes fra, 1-5 observasjoner, så for eksempel om det reelle tallet er 20, kan det vises som 15-25 i output. Det minste positive tallet som kan vises er 5, så verdiene 1-4 vil ikke kunne forekomme. Verdien 0 er mulig, og denne vil aldri støylegges. Støyen er alltid konstant, slik at en vil få det samme resultatet ved hver kjøring (SSB, 2022). Støyleggingen vil trolig ikke påvirke oppgaven mye ettersom populasjonen

er tilstrekkelig stor, og vi ikke ser på noe nøyaktig antall i analysen.

3.1.2 Begrensninger ved programmet

Som nevnt er det ikke mulig i Microdata.no å eksportere datasettet, eller å importere annen data. Man er derfor begrenset til å kun bruke de variablene som allerede eksisterer i Microdata.no, med den funksjonaliteten som finnes der. Dette byr på noen utfordringer. Man er avhengig av å gjøre all analysen i Microdata.no, men kan hente ut resultatene til Excel eller Word for videre bearbeiding. Dersom man ønsker mer oversiktlige regresjonsresultater eller figurer enn de som man får direkte fra programmet, er man nødt til å gjøre det manuelt i andre programmer.

Videre er variabelutvalget som finnes, noe begrenset. For denne oppgavens analyse kunne det vært interessant å se på et individs varighet i en jobb for å se om avkastningen av et ekstra år i jobb er forskjellig for innvandrere og befolkningen ellers. I tillegg kunne det vært nyttig å ha grunnskoleresultater for innvandrere fra andre land for å se om det er store forskjeller i resultater blant gruppene i så tidlig alder, som Carneiro mfl. (2005) skriver om i sin artikkel. I Microdata.no har man kun tilgang på grunnskoleresultater fra Norge, så om man skulle brukt disse resultatene, måtte man kun ha sett på innvandrere som kom til Norge før de begynte på skolen. Dette hadde gjort at vi hadde mistet alle innvandrere som har kommet til Norge i løpet av arbeidslivet, og hadde ført til en dårligere analyse ettersom år i det norske arbeidsmarkedet trolig er en stor kilde til lønnsulikhet.

3.2 Avgrensning av utvalget

Vi skal gjøre analysen med verdier fra perioden desember 2019 til januar 2020. De fleste av Microdata.no sine variabler gir tversnittsverdier fra enten den 16. hver måned, eller den 1. hver måned. Vi henter da de fleste verdiene fra enten 16. desember 2019 eller 1. januar 2020. Grunnen til at vi har valgt denne perioden er at det er så nylig som mulig, men vi ønsket å unngå problematikken med at datasettet var påvirket av perioden med koronapandemi. Vi startet med å lage et datasett hvor vi hentet registerdata fra 1. januar 2020 og hadde et utvalg med 8,5 millioner individer. Først fjernet vi alle observasjoner av individer som var enten døde eller utflyttede, dette fjernet 3,2 millioner individer. Videre begrenset vi utvalget til kun personer i alderen 25 til 54, som er en avgrensning brukt i flere artikler, i tillegg til at dette er en gruppering mye brukt av SSB. Vi har valgt å bruke denne aldersgrupperingen for å hovedsakelig inkludere personer

som er i jobb, og da fjerne pensjonister og studenter. Denne avgrensningen fjernet videre 2,2 millioner individer.

Deretter begrenset vi utvalget til å kun se på folk som er i fulltidsstilling. Vi bruker variabelen for avtalt stillingsprosent og fjerner alle som ikke har minst 100% stilling, altså alle som kun jobber deltid. Videre fjerner vi også alle som mottar arbeidsledighetstrygd, for å sørge for at vi kun gjør analysen basert på de som får inntekt fra jobb. Vi grupperer individene i utvalget etter innvandrerkategori, og vi fjerner alle personer som er født i utlandet med norske foreldre. Vi ser kun på de som bor i Norge i analyseåret. Disse begrensningene fjerner i underkant av en million individer, slik at vi sitter igjen med et datasett med 1 028 564 individer.

3.3 Variabler

Vi skal som nevnt se på forskjeller i lønn for innvandrere og befolkningen ellers. Vi velger i første omgang å skille mellom personer som har kommet til Norge som innvandrere selv, og personer som er barn av innvandrere, da henholdsvis førstegenerasjonsinnvandrer og andregenerasjonsinnvandrer. Vi har følgende variabler for ulike kategorier av innvandrere:

Innvandringskategori	Variabelnavn
- Førstegenerasjonsinnvandrere	<i>finnv</i>
- Andregenerasjonsinnvandrere	<i>ainnv</i>
- Samlevariabel for finnv og ainnv	<i>innv</i>
- Befolkningen ellers	<i>norsk</i>

Vi skiller mellom om personene kommer fra høy- eller lavinntektsland på *finnv* og *ainnv* med suffiksen *H* for høyinntektsland og *L* for lavinntektsland.

3.3.1 Avhengig variabel

Vår avhengige variabel i analysen er nominell årslønn, og vi bruker variabelen for lønnsinntekter fra Microdata.no, som gir oss kontantlønn, skattepliktige naturalytelser og syke- og fødsels- penger, i norske kroner, i løpet av kalenderåret. Microdata.no har flere variabler for lønn, men vi har valgt å bruke årslønn ettersom vi kun skal se på personer som er i fulltidsjobb. Vi tar også logaritmen til årslønn og bruker denne i analysene våre, både for å lettere kunne se på

prosjenter, men også for å gjøre hele lønnsvariabelen mindre venstreskjev og mer normalfordelt. Se histogram i appendiks A.1 for forskjellene. I tabell 1 har vi deskriptiv statistikk for hele utvalget. Vi ser at de fra befolkningen ellers generelt sett tjener mer enn innvandrere, og vi ser at det er størst forskjell i lønn mellom førstegenerasjonsinnvandrere fra lavinntektsland og personer fra befolkningen ellers.

VARIABLER	Observasjoner	Gj.Snitt	Std.avvik	1%	99%
Hele utvalget	1 028 564	645 501	240 073	319 000	1 580 000
norsk	819 834	658 943	241 928	325 000	1 600 000
innv	208 728	592 744	225 002	309 000	1 490 000
finnv_H	38 416	672 977	268 884	316 000	1 690 000
finnv_L	117 789	543 233	185 645	306 000	1 310 000
ainnv_H	32 905	670 062	255 290	320 000	1 650 000
ainnv_L	19 621	604 385	219 981	309 000	1 450 000

Tabell 1: Deskriptiv statistikk for årslønn for de ulike innvandringskategoriene

3.3.2 Forklaringsvariabler

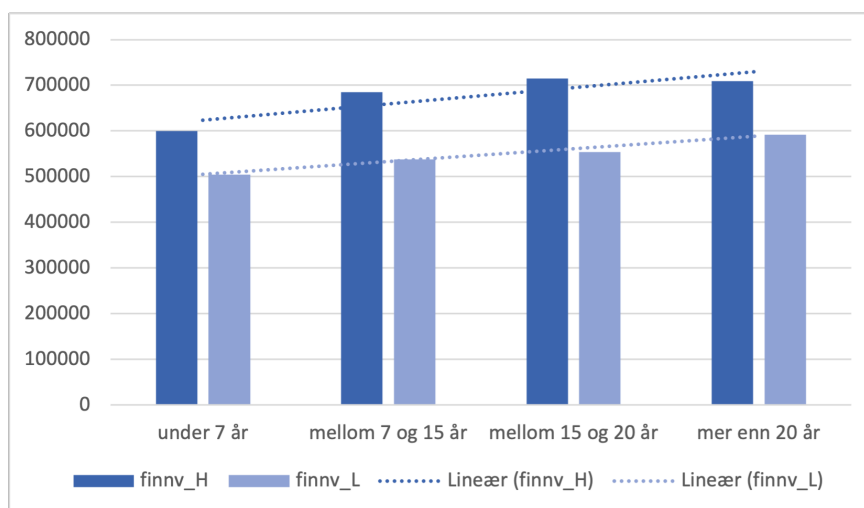
Vår analyse ser på forskjellene i lønn mellom innvandrere og befolkningen ellers. Vi definerer innvandrere som personer som har immigrert til Norge i løpet av sin levetid, eller som har minst en innvandret forelder. Vi bruker variabelen for innvandrerkategori fra Microdata.no for å gruppere dette.

Vi bruker variabelen landbakgrunn for å se hvor innvandrerne kom fra. Denne variabelen gir personens fødeland, for personer som ikke er født i Norge, og foreldres fødeland om personen er født i Norge. Vi bruker denne variabelen for å kunne få landbakgrunn også for andregenerasjonsinnvandrere, og se om det er noen effekter der. Vi bruker dette for å gruppere innvandrene etter om de, eller foreldrene, kom fra høyinntektsland eller lavinntektsland. Vi baserer oss på den samme inndelingen brukt i artikkelen av Hoen mfl. (2022), hvor vi definerer lavinntektsland som Øst-Europa, Asia, untatt Japan, Afrika og Sør Amerika, og høyinntektsland som Vest- og Nord-Europa, Nord Amerika, Japan og Oseania. Ettersom vi ser at det er større likheter mellom innvandringsgruppene avhengig av landbakgrunn enn om de er første- eller andregenerasjonsinnvandrere, velger vi å videre gruppere sammen disse til variablene *innv_H* og *innv_L*. Det er

disse variablene vi bruker videre i analysen.

Videre har vi laget variabler for innvandringsgrunn. Denne variabelen omfatter kun de som har immigrert til Norge fra ikke-nordiske land ettersom innvandrere fra norden ikke trenger oppholdstillatelse eller registreringsbevis for å bo eller jobbe i Norge, og da ikke trenger oppgi grunn for innvandring. Kategoriene som finnes er flukt, familie, utdanning, arbeid, ukjent og annet. Vi fjerner de siste to kategoriene fra datasettet ettersom det er for få observasjoner der. Vi har valgt å gruppere sammen utdanning og arbeid til variabelen *arb_utd* og flukt og familie til variabelen *fam_flu* ettersom vi så at det ikke var store forskjeller mellom de gruppene.

Til slutt lager vi variabler for hvor lenge en innvandrere har vært i landet. Vi lager grupper på innvandrere som har vært i Norge i under 7 år, mellom 7 og 15 år, mellom 15 og 20 år og mer enn 20 år. Dette er for å se om lønnen til innvandrere vil øke etter lengre botid i Norge. Grunnen til den spesifikke grupperingen er at som nevnt finner flere tidligere artikler tegn til at det er en reduksjon i sysselsetting blant innvandrere 10-15 år etter at de kom til Norge. Umiddelbart ser det ikke ut til at det er en slik trend i lønnsutviklingen i våre data, se figur 1. Vi ser at det er en relativt jevn stigende trend for lønn etter år i Norge både for innvandrere fra høyinntektsland og lavinntektsland. Samtidig er resultatene i figuren kun fra rådata, så det er ikke kontrollert for blant annet alder, som vil øke med antall år i Norge.



Figur 1: Gjennomsnittslønn for ulike innvandringsgrupper, avhengig av hvor lenge de har vært i Norge

3.3.3 Kontrollvariabler

I tabell 2 har vi deskriptiv statistikk for kontrollvariablene våre, separert på de ulike innvandringskategoriene vi skal fokusere på i oppgaven. For variablene våre som er dummyer får vi tilnærmet like standardavvik for alle variablene, for alle gruppene, som er på omtrent 0,49.

VARIABLER	Alder		Utd_høy	Storby	Kvinne	Gift	Barn
	Gj.snitt	Std.avvik	Gj.snitt	Gj.snitt	Gj.snitt	Gj.snitt	Gj.snitt
Totalt	40	8.1	0.53	0.51	0.42	0.42	0.65
norsk	40	8.2	0.53	0.48	0.43	0.39	0.66
innv	39	7.4	0.53	0.64	0.40	0.53	0.59
innv_H	40	7.7	0.63	0.63	0.41	0.42	0.61
innv_L	39	7.3	0.47	0.64	0.40	0.58	0.58
arb_utd	39	6.7	0.56	0.59	0.30	0.58	0.48
fam_flu	40	7.5	0.42	0.64	0.50	0.63	0.69

Tabell 2: Deskriptiv statistikk for kontrollvariablene

Alder

Dette er en variabel for alderen til personene i utvalget. I Microdata.no har vi ikke gode variabler for ansiennitet eller erfaring, så da kan vi bruke alder som en proxy-variabel for dette. Det betyr dermed ikke at aldersvariabelen direkte gir oss antall år med erfaring, men variabelen gir en god indikasjon på antall år i arbeid, gitt resten av forutsetningene i modellen. Det er generelt enighet om at man får høyere lønn jo eldre man er (Askvik, 2020). Vi har også med et andregradsledd av alder ettersom den positive effekten på lønn av økt alder trolig er avtakende. I vårt utvalg er gjennomsnittsalderen 40 år, og den er relativt lik for både innvandrere og befolkningen ellers, men jevnt over litt lavere for innvandrere. Vi oppretter også en dummy for om et individ er eldre eller yngre enn gjennomsnittsalderen vår, altså 40, gitt ved variabelen *ung* for de som er yngre enn 40.

Utdanning

Det er empirisk sett slik at høyere utdanning bidrar til høyere lønn, og derfor velger vi å kontrollere for dette. Variabelen for utdanning gir oss hvilken grad av utdanning et individ har. Gjennom denne variabelen kan vi betrakte forskjellene i grad av utdanning blant innvandrere og befolkningen ellers. Vi har for enkelhets skyld valgt å benytte en dummy for høyere utdanning, *utd_høy*. Det vil da si at vi har på den ene siden de med grunnskole og videregående opplæring, og på den andre siden de med høyere utdanning som bachelor, mastergrad og PhD. Vi ser at blant innvandrere og befolkningen ellers er det lik andel som har høyere utdanning, men det er større variasjoner mellom de ulike innvandringsgruppene.

Et problem med denne variabelen i vårt datasett er at vi har en del manglende verdier. Spesielt er det veldig mange manglende verdier for innvandrere. Totalt i datasettet mangler det verdier om utdanning for over 37 000 individer hvor nesten alle er innvandrere. Spesielt er det mange manglende verdier for førstegenerasjonsinnvandrere, hvor vi mangler nesten 1/5 av verdiene for denne gruppen, for innvandrere fra høyinntektsland og de fra lavinntektsland kombinert. Se tabell 3. Vi fjerner individene med manglende verdier fra utvalget. Dette kan gjøre at vi får skjevheter som fører til at utvalget vårt er mindre representativt for befolkningen.

	Antall	Andel av totalen	Andel av egen gruppe
Totalt	37 417	-	-
Norsk	546	1%	<1%
finnv_H	9 321	25%	19%
finnv_L	27 332	73%	18%
ainnv_H	98	<1%	<1%
ainnv_L	120	<1%	<1%

Tabell 3: Andelen manglende utdanningsverdier

By

By-variabelen skiller mellom individer basert på om de bor i en storby eller ikke. Vi bruker SSB sin rangering over de ti største byene i Norge som ramme for å skille mellom storbyer og distrikt (Høydahl, 2023), se appendiks A.2 for oversikt. Vi ser at for hele utvalget vårt er

personene ganske jevnt fordelt mellom storby og ikke, men andelen er høyere for innvandrere enn for de fra befolkningen ellers. Det er 64% av innvandrere som bor i storbyer, mens det for de fra befolkningen ellers bare er 48%.

Variabelen viser en persons bostedskommune i følge folkeregisteret, som gjør at den ikke vil fange opp arbeidskommune for personer som jobber i en annen kommune enn de bor i. Dette kan medføre noen feil, men trolig ikke store nok til at det vil påvirke analysen. Spesielt ettersom vi har gruppert Oslo og flere av de omkringliggende kommunene til Stor-Oslo. Da vil vi trolig få fanget opp de som pendler til Oslo for jobb, men som bor i en annen kommune. For videre forskning kunne man også gjort noe lignende for andre storbyer ved å bruke arbeidsmarkedsregioner i stedet for kommune som standard i inndelingen.

Kjønn

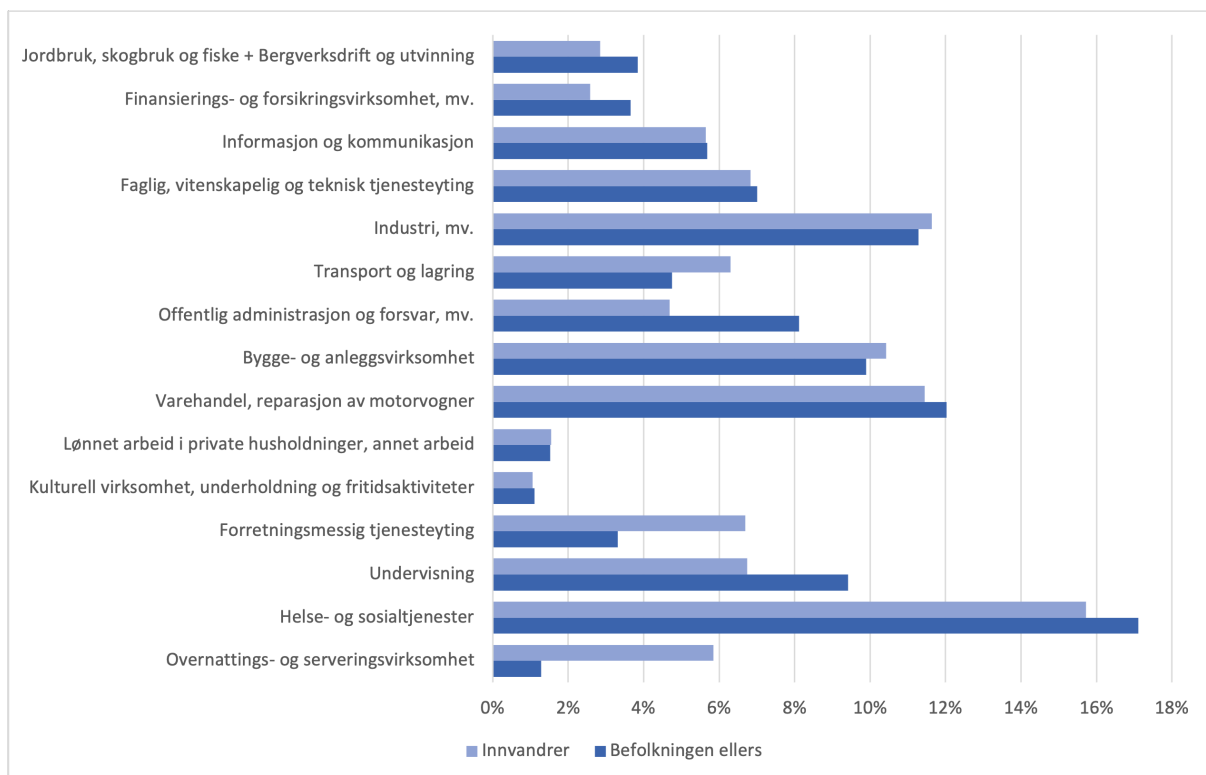
Dette er en dummyvariabel for om et individ er mann eller kvinne, som inkluderes for å kontrollere for kjønnseffekter, ettersom det empirisk sett er slik at menn tjener mer enn kvinner. I utvalget vårt er det 58% menn og, med unntak av de i gruppen *fam_flu*, hvor det er jevnt fordelt, er det blant innvandrere en enda større andel menn. Det er flere menn enn kvinner i befolkningen generelt i Norge, spesielt er det en skjevhet blant innvandrere. Kjønnforskjellen er mye større i vårt utvalg enn i befolkningen generelt. Grunnen til at vi har en såpass stor forskjell fra befolkningen i vårt utvalg skyldes trolig at vi kun ser på personer i fulltidsjobb, og ettersom det er flere kvinner enn menn som jobber deltid, hhv. 37% kvinner og 17% menn, mister vi nok en del her (Gram, 2022).

Familie

Videre kontrollerer vi også for familiesituasjon, altså et individs sivilstatus og om et individ har barn. Vi har opprettet dummyer for disse to variablene, hvor vi for barn kun kontrollerer for om individet har minst ett barn, ikke hvor mange barn. Her regnes barn som alle personer, uansett alder, som er registrert bosatt i familien. For variabelen gift har vi inkludert alle som er gift, har registrert samboer, er enke eller er gjenlevende partner. Vi ser at i utvalget vårt er 42% gift, og blant innvandrere så er det 53%, det er spesielt høyt blant de i gruppen *fam_flu*. For barn ser vi at det er omtrent to tredjedeler som har barn, men generelt en lavere andel som har barn blant innvandrerne.

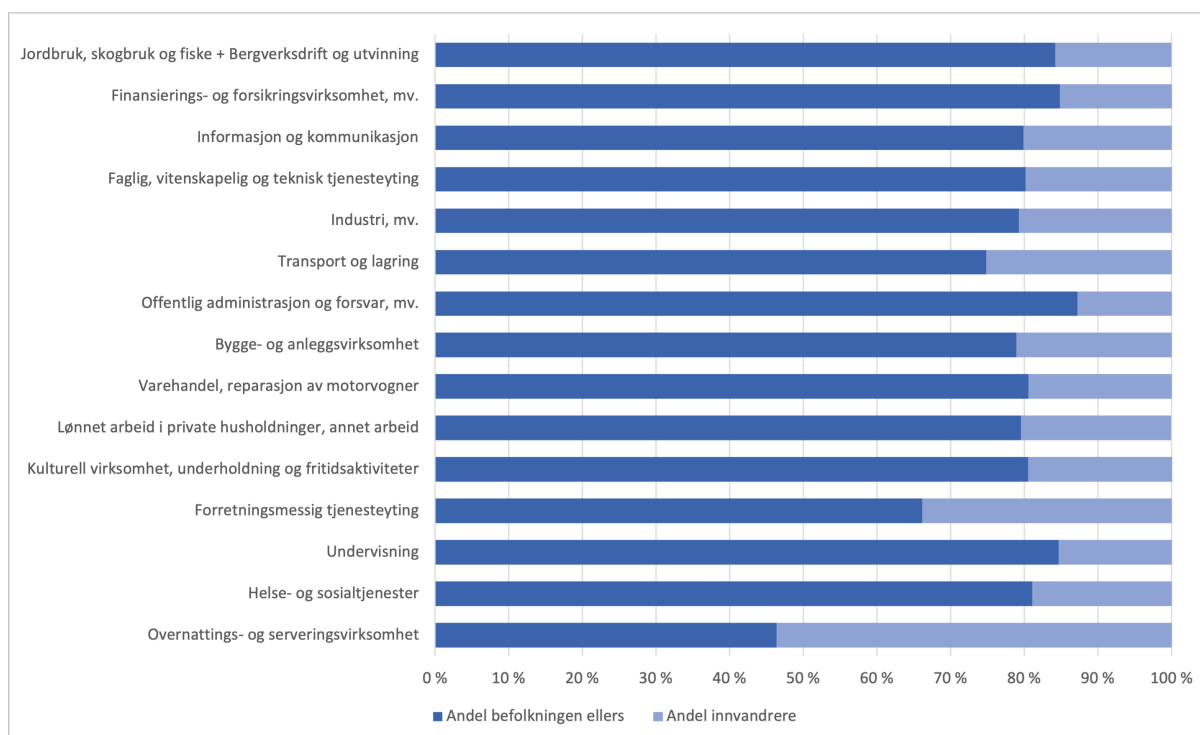
Næring

Vi kontrollerer for de ulike næringene som et individ kan jobbe i ved hjelp av en variabel som gir hovednæringen til virksomheten et individ jobber i. For personer med flere arbeidsforhold vil det være hovedarbeidsforholdet, da i hovedsak arbeidsforholdet med høyest avtalt arbeidstid for individet. Ettersom vi kun ser på heltidsansatte vil dette trolig ikke bli et problem. Vi har fordelt hovednæringene etter SSB sin standard for næringsgrupperinger, med en tosifferfordeling, slik at vi får 21 forskjellige næringer (Lightfoot, 2009). Vi har for få observasjoner i noen av næringene, så vi er nødt til å gruppere flere sammen for å fortsatt kunne inkludere de i datasettet. Se fullstendig oversikt i appendiks A.3 for alle næringene og hvordan de er gruppert. Vi får da 15 forskjellige næringer i modellen vår.



Figur 2: Fordeling av innvandrere og befolkningen ellers i de ulike næringene

I figur 2 ser vi fordelingen av innvandrere og befolkningen ellers i de ulike næringene. Næringene er sortert etter lønn, så de næringene med høyest lønn er øverst, og de med lavest lønn er nederst. I appendiks A.4 har vi den samme figuren separert på høy- og lavinntektsland. Vi ser at det er flest innvandrere som jobber i næringen for helse- og sosialtjenester fulgt av industri og varehandel, og det er færrest som jobber med kulturell virksomhet, underholdning og fritidsaktiviteter. Vi ser at det er flere næringer hvor en større andel av innvandrere enn befolkningen ellers jobber,



Figur 3: Andelen innvandrere i de ulike næringene

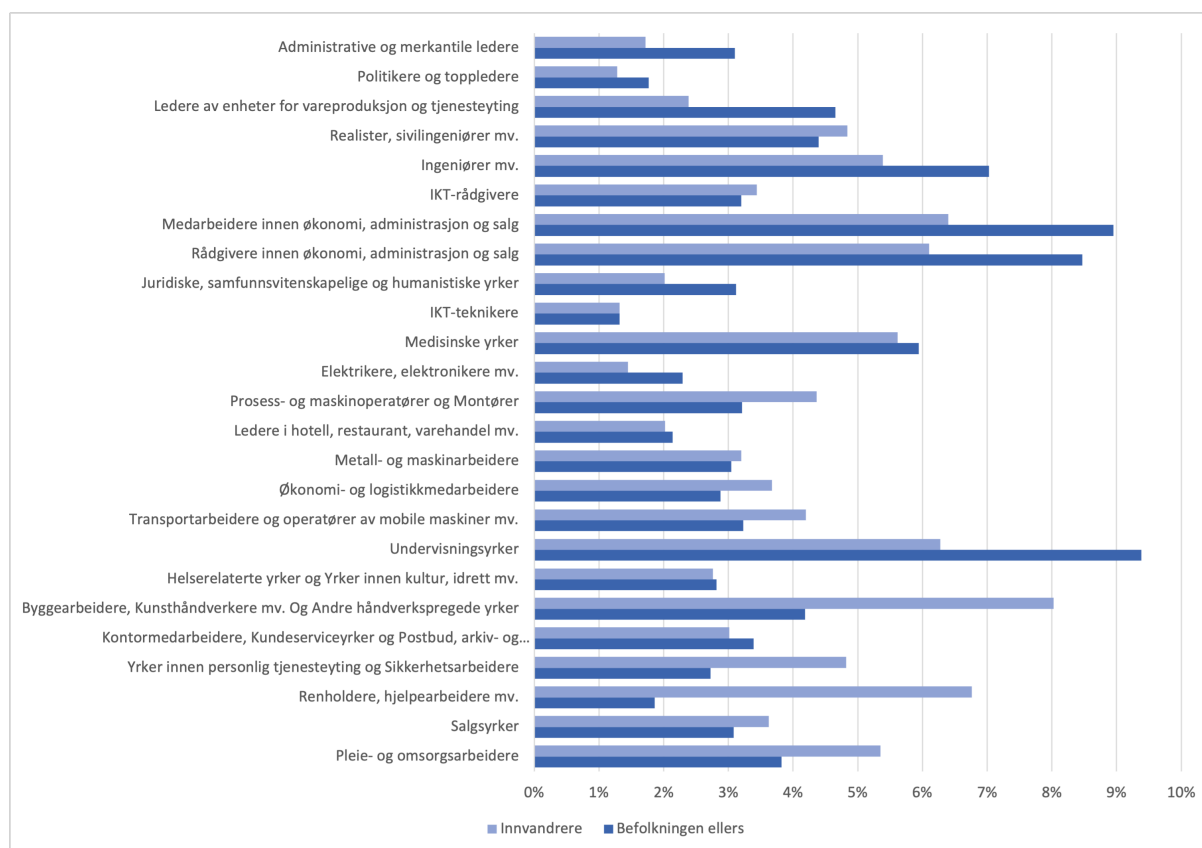
men det er varierende om dette er i høytlønnede eller lavtlønnede næringer. Vi ser for eksempel at det er en større andel innvandrere som jobber i industri, som har relativt høye lønninger, samtidig som det er en mindre andel innvandrere som jobber med helse, som har relativt lave lønninger, sammenlignet med befolkningen ellers. Samtidig kan vi se at det er en klart større andel innvandrere som jobber i overnatting og serveringsvirksomhet, som er den lavest betalte næringen.

I figur 3 ser vi andelen av de ansatte i hver næring som er henholdsvis innvandrere og befolkningen ellers, også sortert etter gjennomsnittslønn i næringene. Her ser vi at for de fleste næringer er det en ganske lik andel innvandrere som andelen innvandrere i vårt utvalg generelt, altså omtrent 20%. Men det er noen unntak. For overnatting og serveringsnæringen er over halvparten av alle arbeidere innvandrere, og for forretningsmessig tjenesteyting er det nesten 35%. Begge disse næringene er blant de med lavest lønninger.

Yrke

Av samme grunn som at næring påvirker lønn, vil også et individs yrke påvirke lønnen, og derfor kontrollerer vi for det. Vi bruker variabel for yrkeskode fra Microdata.no som bestemmes

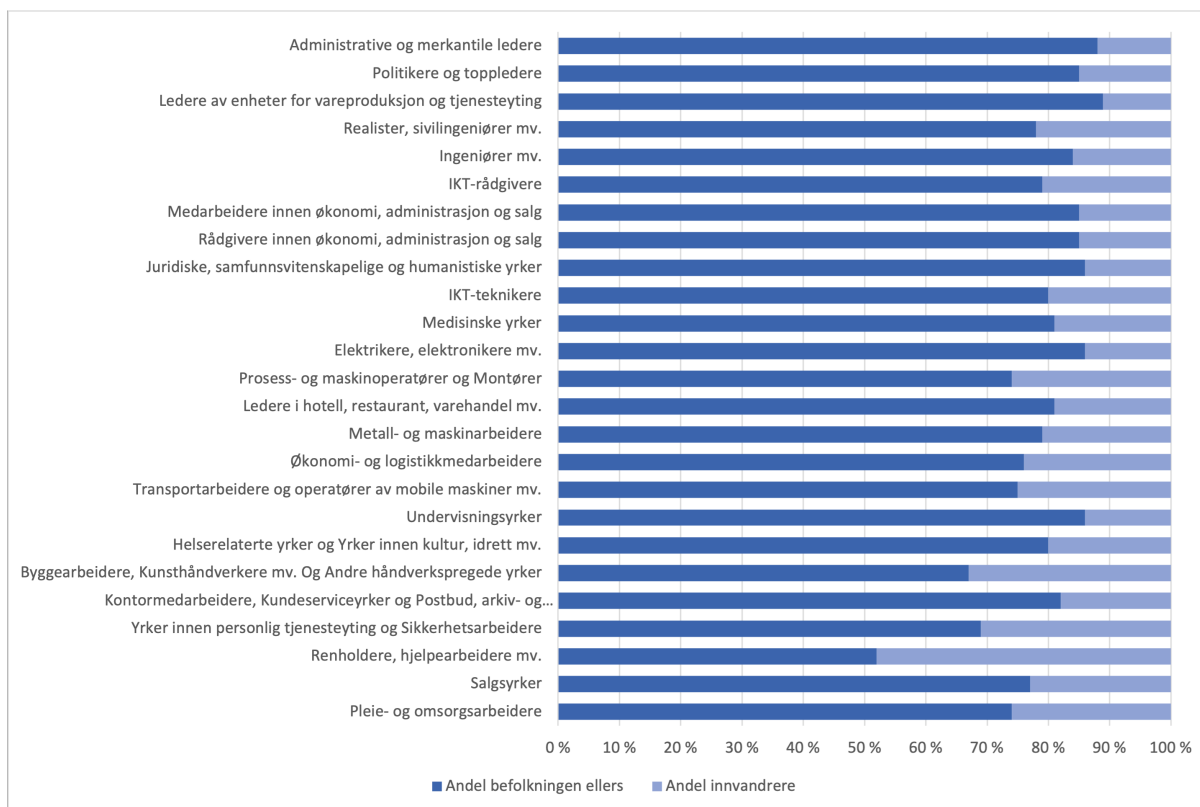
på samme måte som næring, etter et individs hovedarbeidsforhold. Yrkene grupperes sammen etter standard for yrkesklassifisering, og vi bruker her også en tosiffergruppering, slik at vi får 43 yrkesgrupper (SSB, 2011). Her også er det noen av yrkene som har for få observasjoner som grupperes sammen. For yrkene Bønder, fiskere mv. og Militære yrker og uoppgitt er det få individer, og ingen naturlig gruppering de kunne inkluderes i, så de blir fjernet fra datasettet. Vi får da totalt 25 forskjellige yrkesgrupper. Se appendiks A.5 for fullstendig oversikt over grupperingene.



Figur 4: Fordeling av innvandrere og befolkningen ellers i de ulike yrkesgruppene

I figur 4 ser vi fordelingen av innvandrere og befolkningen ellers i de ulike yrkene. Figuren er sortert etter gjennomsnittslønn, så yrkene med høyest gjennomsnittslønn er øverst, og de med lavest lønn er nederst. I appendiks A.6 har vi den samme figuren separert på høy- og lavinnteksland. Vi ser at det er flest innvandrere som jobber som byggearbeidere og andre håndverkspregede yrker, etterfulgt av renholdere og hjelpearbeidere. Det er færrest innvandrere som jobber som politikere og toppledere. I denne figuren ser vi en tendens til at yrkene hvor en større andel innvandrere jobber, sammenlignet med befolkningen ellers, er yrker med gjennomsnittlig lav lønn. I den nedre halvdel av figuren er det for de aller fleste yrker at søylen for innvandrere er større enn søylen for befolkningen ellers, og motsatt i den øvre halvdel av figuren. Her kan vi se

at innvandrere generelt sett jobber i dårligere betalte yrker enn befolkningen ellers, relativt til størrelse på gruppen.



Figur 5: Andelen innvandrere i de ulike yrkesgruppene

Vi ser videre på andelen innvandrere i ulike yrker, i figur 5. I denne figuren er også yrkene sortert etter gjennomsnittslønn. Her ser vi at den yrkesgruppen hvor det er størst andel innvandrere er renholdere, hjelpearbeidere mv., hvor nesten halvparten av alle ansatte er innvandrere. Videre er det også mange innvandrere som jobber i håndverksyrker og yrker innen personlig tjenesteyting, som alle er yrker med relativt lav gjennomsnittslønn. I yrkene med høy gjennomsnittslønn, som de forskjellige lederyrkene, er det bare rundt 10% av arbeiderne som er innvandrere.

4 Metode

I dette kapitlet vil vi presentere den økonometriske metoden vi har valgt å bruke, samt relevante økonometriske utfordringer vi har møtt på underveis. For å kunne betrakte forskjeller mellom de forskjellige folkegruppene brukes dekomponeringen først presentert av Blinder (1973) og Oaxaca (1973). Mot slutten av kapitlet vil vi gå nærmere inn på noen økonometriske utfordringer som er relevant ved bruk av minste kvadraters metode (OLS) og Oaxaca-Blinder dekomponering. Vi starter med å legge grunnlaget for Oaxaca-Blinder dekomponering gjennom å presentere regresjonene for utvalget.

OLS - Det initielle lønnsgapet

Utgangspunktet for en Oaxaca-Blinder dekomponering er separate OLS-regresjoner for de to gruppene vi ønsker å sammenligne, så vi starter med å se på det. Etter å ha avgrenset datasettet som forklart i kapittel 3.2, starter vi med å betrakte et rått lønns gap mellom innvandrere og befolkningen ellers. Dette initielle lønns gapet finner vi ved den enkle regresjonen i likning 1. Der $innv_i$ er en dummy lik 1 for innvandrere og 0 for befolkningen ellers, $\ln W_i$ gir logaritmen av lønnen for individ i og restleddet er gitt ved u_i . Koeffisienten α gir oss det lønns gapet som foreligger før man har inkludert noen andre variabler. Dersom $\alpha < 0$, er det et lønns gap i favør av befolkningen ellers, og α kan tolkes som prosentvis forskjell i lønn mellom innvandrere og befolkningen ellers. Det er nettopp dette initielle lønns gapet vi sikter på å dekomponere gjennom vår analyse.

$$\ln W_i = \alpha \cdot innv_i + u_i \quad (1)$$

$$\ln W_i = \alpha \cdot innv_i + \beta X_i + u_i \quad (2)$$

Likning 2 er en utvidelse av 1 som inkluderer en vektorvariabel X_i som inneholder kontrollvariablene *alder*, *utd_høy*, *storby*, *kvinne*, *gift* og *barn* samt dummyer for alle næringene og yrkene. Vi inkluderer vektorvariabelen for å se hvordan disse kontrollvariablene vil påvirke lønns gapet.

Regresjonene i likning 1 og 2 gir oss en indikasjon på lønns gapet mellom innvandrere og befolkningen ellers. Videre i neste kapittel vil vi gå nærmere inn på hvordan man kan dekomponere dette lønns gapet. Denne dekomponeringen vil være essensiell for den videre analysen av hvilke faktorer som påvirker lønns gapet.

4.1 Oaxaca-Blinder - Dekomponering av lønnsgapet

En Oaxaca-Blinder dekomponering baserer seg på lineære regresjonsmodeller og forsøker å estimere den gjennomsnittlige differansen til lønnen mellom to folkegrupper. Denne dekomponeringen vil kunne dele lønnsforskjellene mellom de to gruppene inn i en forklart og uforklart komponent (Jann, 2008).

Basert på Elder mfl. (2010) vil vi i dette kapittelet presentere tre måter å utføre en Oaxaca-Blinder dekomponering på. I første omgang har vi to dekomponeringer først presentert av Blinder (1973) og Oaxaca (1973), som vist i likning 3 og 4.

$$\ln W_i^N = \beta^N X_i^N + u_i^N \quad (3)$$

$$\ln W_i^I = \beta^I X_i^I + u_i^I \quad (4)$$

Likning 3 og 4 viser den prosentvise endringen i lønnen (W) der X_i er en vektor for kontrollvariablene i modellen med hensyn på individ i og u er et restledd. Toppskrift N og I representerer verdier for henholdsvis befolkningen ellers og innvandrere. Disse to regresjonene gir utgangspunkt for den opprinnelige Oaxaca-Blinder dekomponeringen, presentert i likning 5 og 6.

$$\ln \hat{W}^N - \ln \hat{W}^I = (\bar{X}^N - \bar{X}^I) \hat{\beta}^I + \bar{X}^N (\hat{\beta}^I - \hat{\beta}^N) \quad (5)$$

$$\ln \hat{W}^I - \ln \hat{W}^N = (\bar{X}^N - \bar{X}^I) \hat{\beta}^N + \bar{X}^I (\hat{\beta}^I - \hat{\beta}^N) \quad (6)$$

Likning 5 og 6 beskriver den forklarte og uforklarte delen av lønnsgapet slik Blinder (1973) og Oaxaca (1973) presenterte det. Her representerer venstresidevariablene den prosentvise differansen i gjennomsnittslønn mellom innvandrere (\hat{W}^I) og befolkningen ellers (\hat{W}^N). Høyresiden i begge likningene representerer de forklarte (første ledd) og uforklarte (andre ledd) komponentene av lønnsgapet. Det første leddet viser den delen av gapet som skyldes forskjeller i egenskaper, som for eksempel forskjellig utdanningsnivå, yrkesvalg og kjønn. Det andre leddet derimot viser forskjell i avkastning dersom individer har det samme nivået på egenskapene, som feks samme utdanningsnivå, samme yrke eller samme kjønn. \bar{X}^I og \bar{X}^N gir gjennomsnittsverdiene til alle kontrollvariablene for henholdsvis innvandrere og befolkningen ellers. $\hat{\beta}^I$ og $\hat{\beta}^N$ representerer koeffisienten til kontrollvariablene hentet fra likning 4 og 3. De to variantene av Oaxaca-Blinder

i likning 5 og 6 beskriver mye av de samme effektene, og ved analyser er det naturlig å vise resultater fra begge ettersom de gir forskjellige resultater (Elder mfl., 2010).

Det er flere som har utformet egne varianter av Oaxaca-Blinder dekomponeringen, men den som er mest brukt er dekomponeringen til Neumark (1988), vist i likning 7 (Elder mfl., 2010). Det er også denne vi kommer til å fokusere på i analysen. Neumark (1988) sin tilnærming varierer fra de andre dekomponeringene. Den inkluderer et ledd basert på en samlet regresjon (pooled OLS) for hele utvalget, fra likning 2. Når man kombinerer verdiene fra likningene 2, 3 og 4 får man dekomponeringen i likning 7.

$$\ln\hat{W}^I - \ln\hat{W}^N = (\bar{X}^N - \bar{X}^I)\hat{\beta}^* + \bar{X}^I(\hat{\beta}^I - \hat{\beta}^*) + \bar{X}^N(\hat{\beta}^* - \hat{\beta}^N) \quad (7)$$

Fra dekomponeringen i likning 7 vil man kunne skille mellom lønnsfordelen og ulempen ved å være fra befolkningen ellers og innvandrerguppen. Som i de opprinnelige dekomponeringene til Blinder (1973) og Oaxaca (1973) vil det første leddet på høyresiden av likning 7 gi oss den forklarte delen, men i denne varianten til Neumark (1988) vil summen av de to siste leddene være et estimat på den uforklarte delen av lønnsgapet. Det er denne varianten som vist i likning 7 som vil legge grunnlaget for analysen vår. Fra resultatene vil man se at denne dekomponeringen gir det mest nyanserte bildet på forklaringen av lønnsgapet. I appendiks A.7 finnes resultatene for de to alternative dekomponeringene.

Estimering av modellen - begrensninger ved Microdata.no

Når vi har estimert både resultatene fra regresjonene og spesielt dekomponeringen fra Oaxaca-Blinder, eksporterer vi resultatene fra Microdata.no til Excel for å foreta utregningene. Dette vil si at vi har eksportert ut alle gjennomsnittsverdiene og betakoeffisientene til de forskjellige modellene. Deretter har vi foretatt dekomponeringene slik de er lagt fram i dette kapittelet. I Microdata.no eksisterer det en kommando for Oaxaca, men denne gir foreløpig kun informasjon om hvor stor den forklarte og uforklarte delen av gapet er, altså ingen informasjon om relativ betydning for de ulike variablene.

4.2 Utvidelser av modellen - Heterogeniteter blant innvandrere

I denne delen presenterer vi de utvidelsene vi skal ha med i analysen vår. Vi vil fokusere på effekten av hvilket land innvandrere kommer fra, samt bakgrunnen for innvandringen. Poenget

med utvidelsene er å ta høyde for at innvandrere ikke er en homogen gruppe, og dermed betrakte heterogeniteter blant innvandrere.

Måten disse utvidelsene foretas på er ved å bruke den initielle samlede regresjonen fra likning 2 som utgangspunkt. Deretter setter vi inn verdier hentet fra regresjoner for de 4 forskjellige innvandringsgruppene vi skal utvide for. Disse gruppene er innvandrere fra høy- og lavinntektsland, innvandrere som kommer på grunn av arbeid eller utdanning samt de som kommer på grunn av familie eller flukt. Likningene som brukes i denne Oaxaca-Blinder dekomponeringen er bygd opp på samme måte som i kapittel 4.1. For utregning se appendiks A.8. Utregningen følger de samme stegene som vist i likning 3-7, og vil kunne vise eventuelle forskjeller i lønns GAP og grad av diskriminering mellom innvandringsgruppene.

4.3 Økonometriske utfordringer

Vi skal i dette kapitlet gå gjennom de økonometriske utfordringene vi har møtt på, eller som kan ha vært relevante for vår analyse. Vi ser på endogenitet og andre utfordringer som er relevante for analysen.

4.3.1 Endogenitetsproblemet

Når en OLS regresjon settes opp, er det viktig at kontrollvariablene er eksogent gitt i modellen. Det vil si at de ikke er korrelert med restleddet u . Vi kan forklare dette gjennom et lite eksempel. La oss si at vi har en helt enkel regresjon som i likning 8.

$$W = \alpha_0 + \beta_1 innv + u \quad (8)$$

I dette restleddet (u) er det en rekke variabler som potensielt kan påvirke innvandringsvariabelen sin effekt på lønnen (W). Vi får da et uttrykk for effekten av innvandring på lønnen:

$$\frac{dW}{dinnv} = \beta_1 + \frac{du}{dinnv} \quad (9)$$

Som vi ser fra ligning 8 er vi opprinnelig ute etter å finne kun verdien til β_1 , men siden $innv$ i hovedsak er endogent gitt og korrelert med restleddet vil resultatet bli påvirket av det siste leddet i ligning 8 (Cameron & Trivedi, 2005). Derfor utvider vi modellen til å inkludere flere kontrollvariabler, men trolig vil det være noen endogene effekter vi ikke får fanget opp.

Vi ser i hovedsak på hvordan det å være innvandrere vil påvirke lønnen, men man kan tenke seg at det også er omvendt kausalitet. Dersom en innvandrere skal velge et land å innvandre til, vil trolig lønningene i det landet påvirke beslutningen, spesielt for arbeidsinnvandrere. Så lønnsvariabelen vil kunne påvirke innvandringsvariabelen. På samme måte vil lønnen kunne bli påvirket av at folk innvandrere til Norge, dermed er det her grunnlag for omvendt kausalitet.

For de fleste innvandrere er som sagt variabelen *innv* endogent gitt, men dette gjelder ikke nødvendigvis for alle innvandringsgruppene. For innvandrere med arbeid, utdanning eller familie som innvandringsgrunn tar de selv et aktivt valg om å flytte til Norge, mens for gruppen med flukt som innvandringsgrunn er ikke dette nødvendigvis gitt, og valget om å innvandre kan dermed være eksogent. Vi har tidligere funnet at gruppene for familie og flukt er veldig like med tanke på gjennomsnittslønn, samt annen deskriptiv statistikk, og har derfor valgt å gruppere de sammen. Vi finner også at forskjellene i det predikerte gapet for disse to gruppene separert er relativt lite (hhv. -22,3% for familie og -21,8% for flukt). Dette tilsier at vi trolig har lite skjevhet mellom de to innvandringsgruppene på grunn av endogenitet.

Vi ser også forskjell i endogeniteten i modellene til første- og andregenerasjonsinnvandrere. Andregenerasjonsinnvandrere har ikke selv tatt et valg om å flytte til Norge, så deres status som innvandrere er ikke nødvendigvis avhengig av de samme elementene som førstegenerasjonsinnvandrere. Hvis vi sammenligner alle første- og andregenerasjonsinnvandrere finner vi flere forskjeller mellom gruppene. Førstegenerasjonsinnvandrere har lavere gjennomsnittslønn, og det å være innvandrere har en større negativ effekt på lønnen enn for andregenerasjonsinnvandrere. Samtidig kan vi se fra tabell 1 at vi finner større ulikheter mellom innvandringsgruppene basert på landbakgrunn enn om de er første- eller andregenerasjonsinnvandrere. Derfor har vi valgt å gruppere sammen på denne måten.

4.3.2 Multikollinearitet

Multikollinearitet er når man har høy korrelasjon mellom to eller flere uavhengige variabler (Wooldridge, 2015). Multikollinearitet vil føre til en høyere varians, som kan gjøre at estimatene blir mindre statistisk signifikante. Vi ser på en korrelasjonsmatrise, vedlagt i appendiks A.9, for alle de uavhengige variablene, og ser at det ikke er noen store korrelasjoner. Vi kan også bekrefte dette med å kjøre en VIF-test, resultater for denne er også i appendiks A.9. VIF brukes som et mål på multikollinearitet mellom de uavhengige variablene i en modell. Dersom man får en høy

verdi, vil man måtte korrigere for dette i modellen ved å enten inkludere interaksjonsledd, utelate variabelen eller finne en annen måte å løse problemet på. Problemet her er dersom korrelasjonen mellom to av variablene er for stor så vil dette kunne ha utslag på resultatet fra modellen. Vi ser fra testen at det ikke er noen tilfeller av for høy korrelasjon, med unntak av *alder* og *alder2*, og dette er å forvente. Vi får da en gjennomsnittlig korrelasjon på 7,5, og hvis vi ikke tar med *alder2* har vi gjennomsnittlig korrelasjon på 1,7, som er langt under den omdiskuterte grensen på 10 (Wooldridge, 2015).

4.3.3 Heteroskedastisitet

Homoskedastisitet er når variansen til det uobserverbare leddet u , er konstant, altså $\text{var}(u_i|x) = \sigma^2$. Heteroskedastisitet skjer da når u varierer for ulike deler av befolkningen (Wooldridge, 2015). Konsekvensene av heteroskedastisitet er at vi kan få feilberegnete og biased estimatorer, i tillegg til ukorrekte standardavvik og konfidensintervall. Vi har trolig heteroskedastisitet i vår oppgave, da det er lite sannsynlig at alle gruppene med innvandrere har like personlige egenskaper, og vil derfor trolig ha forskjellig variasjon i lønningene. Vi så fra den deskriptive statistikken at det var forskjell mellom standardavvikene til de ulike innvandringsgruppene. Vi løser denne problematikken ved å bruke robuste standardavvik.

4.3.4 Utelatte relevante variabler

Utelatte relevante variabler er et problem som oppstår når en variabel som påvirker både den avhengige og de uavhengige variablene ikke er inkludert i modellen. Dette kan føre til over- eller underestimering av koeffesientene, og det kan føre til at det er vanskeligere å trekke pålitelige konklusjoner fra analysen. Vår modell vil trolig være noe underspesifisert. Ettersom vi bruker Microdata.no, og de har et noe begrenset variabelutvalg, vil det være flere variabler, spesielt de knyttet til personlige egenskaper og annen individuell karakteristikk som ikke er mulig å inkludere. Her er det nok flere ting som kan påvirke lønn. I tillegg kan det være flere av de personlige egenskapene som ikke bare påvirker lønnen, men også beslutningen om å flytte til Norge, så både den uavhengige og avhengige variabelen kan bli påvirket av restleddet. Vi forsøker å motvirke mye av problemene med utelatte variabler ved å inkludere mange kontrollvariabler som forhåpentligvis kan gi noen av de samme effektene som de utelatte variablene hadde gitt, slik at vi fortsatt kan trekke relevante konklusjoner.

4.3.5 Målefeil

Målefeil er når de observerte verdiene til en variabel ikke er helt lik de faktiske verdiene til variabelen. Vi skiller mellom målefeil i den avhengige variabelen og målefeil i de uavhengige variablene. Målefeil i den avhengige variabelen kan føre til bias i OLS dersom den er systematisk relatert til en eller flere av de uavhengige variablene, men hvis den ikke er det, vil den ikke føre til noen skjevheter. Målefeil i de uavhengige variablene kan være mer problematisk, og dette fører også til skjevheter og biased estimatorer, og det kan videre påvirke validiteten til analysen (Wooldridge, 2015).

Man får målefeil i den avhengige variabelen dersom det er forskjell i faktisk årslønn og innrapportert årslønn. Dette kan skje dersom arbeidsgiver rapporterer inn feil arbeidstid eller lignende. Men så lenge disse avvikene ikke er systematiske, er dette ikke å anse som et problem for oppgaven. For målefeil i de uavhengige variablene er det lite som tyder på at vi vil få problemer her heller. Det kan være at vi har noen feil i utdanningsvariabelen vår, spesielt for innvandrere, da det kan hende at det er noe utdanning som kan ha blitt rapportert feil i overføringen til norske utdanningsverdier. De fleste datasett vil ha noe målefeil, men det er ikke noe som tilsier at det vil være et signifikant problem for oss.

4.3.6 utfordringer med Oaxaca-Blinder og datasettet

Vi bruker som nevnt Oaxaca-Blinder til vår hovedanalyse, og dette er generelt en god metode for å dekomponere et lønns-gap mellom to grupper, men det er også noen problemer knyttet til denne metoden. Et problem med Oaxaca-Blinder er at selv om den tar hensyn til forskjeller i de observerbare egenskapene til to individgrupper, som forskjell i utdanning og lignende, kan den ikke se på de uobserverbare forskjellene, som personlige egenskaper eller motivasjon. Videre vil det med Oaxaca-Blinder, som med OLS, være problemer dersom vi har målefeil. Her gjelder det samme som ble diskutert tidligere i kapitlet. Oaxaca-Blinder metoden antar også eksogenitet i de uavhengige variablene, altså at de ikke er avhengige av lønnsvariabelen. I realiteten vil det trolig være flere av de uavhengige variablene som er avhengige av lønn, som for eksempel er det rimelig å anta at et individs bosted påvirkes av lønnen. Generelt er det ingen av disse problemene som vil gjøre resultatene våre ugyldige, men det er viktig å ta hensyn til og være klar over at resultatene vi får kan skyldes mange forskjellige faktorer.

5 Resultater

I dette kapitlet skal vi se på resultatene fra analysen vår. Vi skal først presentere resultatene fra en enkel OLS-modell og deretter utvidelser med kontrollvariabler. Vi går så videre til hovedmodellen for oppgaven som er resultatene ved Oaxaca-Blinder dekomponering. Her starter vi med å sammenligne innvandrere generelt og de fra befolkningen ellers, og deretter sammenligne ulike innvandringsgrupper.

5.1 Resultater fra OLS

Vi starter først med en helt enkel OLS-modell som betrakter de initielle effektene av å være innvandrer på lønnen. For alle OLS-modellene vi ser på, kommer vi til å bruke samlevariabelen *innv*. Vi kommer ikke til å skille mellom ulike grupper av innvandrere i første omgang. Resultatene fra den første modellen er gitt i kolonne 1 i tabell 4. Vi ser at vi har en relativt stor, negativ, beta-koeffisient for innvandrer, som betyr at en innvandrer gjennomsnittlig vil tjene mindre enn en fra befolkningen ellers. Dette stemmer overens med tidligere litteratur, og vår hypotese.

Vi antar at det er flere variabler som også påvirker lønnen, så vi utvider modellen til å inkludere flere kontrollvariabler, presentert i kapittel 3.3.3, og disse resultatene finnes i kolonne 2 i tabell 4. Vi hadde initielt et rått observert gap på 10,8%, og vi ser at inkluderingen av kontrollvariablene har gjort at vi nå har et justert gap på 7,9%. Vi skal videre se på de ulike koeffisientene i modell 2 og hvilken effekt de har på lønnen. Vi ser at alle koeffisientene våre er statistisk signifikante, dette er å forvente ettersom vi har et såpass stort utvalg. Alle koeffisienter tolkes *ceteris paribus*.

Alder

Vi ser som forventet at alder har en positiv, men avtakende effekt på lønnen. Endringen i lønn vil avhenge av hvor gammel man er, og vi kan se at for gjennomsnittsalderen, som er 40, vil et år ekstra i alder føre til en økning i lønn på 0,9431%¹. Vi finner også at alderen der lønnen er på det høyeste, er når individet er 53,5 år gammelt. Vi har som nevnt ikke med noen variabler for erfaring i vår modell, så det å være et år eldre tolkes som et år ekstra i jobb.

Vi ønsker å sjekke om innvandrere har lik avkastning på erfaring som befolkningen ellers, og kjører da to regresjoner til hvor vi skiller mellom de som er unge og gamle, her eldre eller yngre

¹Finner dette ved å ta $(\delta \ln W)/(\delta \text{alder}) = 0,037351 - 2 * 0,000349 * \text{alder}$, her er $\text{alder}=40$, så $(\delta \ln W)/(\delta \text{alder}) = 0,009431$

VARIABLER	(1) OLS lnW	(2) OLS lnW	(3) OLS lnW	(4) OLS lnW
innv	-0,108147 (0,000807)***	-0,078641 (0,000653)***	-0,063571 (0,000865)***	-0,08597 (0,000958)***
alder		0,037351 (0,000353)***	0,05162 (0,001437)***	0,038349 (0,00228)***
alder2		-0,000349 (0,000004)***	-0,000536 (0,000022)***	-0,000352 (0,000024)***
kvinne		-0,127722 (0,000598)***	-0,114181 (0,000828)***	-0,132596 (0,000841)***
storby		0,038179 (0,000527)***	0,025168 (0,000739)***	0,04898 (0,000733)***
utd_høy		0,096905 (0,000688)***	0,054314 (0,000997)***	0,130191 (0,000937)***
barn		0,017544 (0,00058)***	-0,012859 (0,000773)***	0,033706 (0,000859)***
gift		0,024425 (0,000564)***	0,02146 (0,000846)***	0,024993 (0,000746)***
konstant	13,34202	12,682918	12,333934	12,682573
Næringsdummier	Nei	Ja	Ja	Ja
Yrkesdummier	Nei	Ja	Ja	Ja
Utvalg	Hele	Hele	Unge	Gamle
Observasjoner	1 035 576	1 035 576	490496	545081
R-squared	0,017031	0,424144	0,360042	0,434521

Standardavvik i parenteser
 *** p<0.01, ** p<0.05, * p<0.1

Tabell 4: Estimerte effekter av å være innvandrere på lønnen med OLS, separert på alder i de siste to kolonnene.

enn 40, og viser resultatene i de to siste kolonnene i tabell 4. Vi ser at blant gruppen som er unge, er det en mindre negativ effekt på lønnen av å være innvandrere enn det er blant gruppen som er gamle. Dette tilsier at når en innvandrere blir eldre vil hen få dårligere avkastning på erfaring sammenlignet med en fra befolkningen ellers, enn når hen var yngre.

Kvinne

Det er som forventet en negativ effekt på lønnen av å være kvinne. Her vil en person som er kvinne tjene gjennomsnittlig over 12% mindre enn en mann med samme egenskaper for hele utvalget. Vi ser at dette gjelder for alle aldersgrupper, men at den negative effekten er større for de som er gamle. Altså vil effekten av å være kvinne, på lønnen, være større jo eldre man er.

Bosted

Vi ser av koeffisienten for storby at det er en positiv effekt av å bo i en storby for alle gruppene. Man vil her tjene gjennomsnittlig 4% mer dersom man bor i en av storbyene. Her ser vi at denne effekten er større dersom man er eldre, altså vil en person som er eldre ha større utbytte av å bo i storby enn en som er yngre. En gjennomsnittlig høyere lønn av å bo i storbyer kan skyldes agglomerasjonseffekter, altså fordelene som kommer av samlokalisering av varer og tjenester (Glaeser, 2010). Næringslivsklynger bidrar til lavere transportkostnader, kunnskapsdeling og andre kostnadsfordeler. Dette er med på å gjøre lønningene relativt høyere enn i distriktene. Samtidig er det også verdt å nevne at selv om lønningene i storbyer er gjennomsnittlig høyere, er også ofte levekostnadene høyere, spesielt knyttet til bolig.

Høyere utdanning

Om et individ har høyere utdanning vil dette ha en positiv effekt på lønnen. En person med høyere utdanning vil tjene nesten 10% mer gjennomsnittlig, enn et individ uten høyere utdanning for hele utvalget. Her ser vi at denne effekten er mye større blant utvalget med de gamle, hvor det å ha høyere utdanning vil øke lønnen med gjennomsnittlig 13%, mens for de unge vil det kun øke lønnen med gjennomsnittlig 5%. Det at høyere utdanning fører til høyere lønn stemmer overens med annen empiri og tidligere litteratur. Det er også rimelig å anta at denne effekten er avtakende, slik at det vil være en positiv effekt på lønnen av å ta et ekstra år med utdanning, men denne effekten vil bli mindre for hvert år ekstra.

Familiesituasjon

Videre ser vi at variablene for om et individ er gift og har barn, for hele utvalget, er begge positive, altså er det en positiv effekt på lønnen av å være gift og/eller ha barn. Samtidig ser vi at for det unge utvalget er det faktisk en negativ effekt av det å ha barn. Det vil gjennomsnittlig redusere lønnen med 1,3% for en ung person, mens det vil øke lønnen med gjennomsnittlig 3,4% for en eldre person. Dette gir mening ettersom vi har satt skillet mellom gammel og ung på 40 år, og påvirkningen av å ha barn på det å jobbe vil trolig være størst når barnet er ungt, og gjerne før de selv er 40 år. Det å være gift har en positiv effekt på lønnen for alle aldersgrupper, og det er en relativt lik effekt for alle, med gjennomsnittlig 2% økning i lønnen.

Næringer

For næringer er alle koeffisientene for de ulike næringene negative. Dette betyr at de har en negativ effekt på lønnen sammenlignet med referansenæringen vår, som er Bergverksdrift og utvinning, slått sammen med Jordbruk, skogbruk og fiske grunnet få observasjoner i sistnevnte. Innenfor næringen for Bergverksdrift og utvinning finner vi blant annet det meste av oljesektoren, som har høye lønninger, så det gir mening at de andre næringene har lavere lønn enn denne. Selv om alle næringskoeffisientene er negative er det varierende hvor store de er. De varierer fra å skulle redusere lønnen med gjennomsnittlig 14% for Finansierings- og forsikringsvirksomhet, mv. til å redusere lønnen med 38% for Overnattings- og serveringsvirksomhet. Som vi ser fra figur 2 er dette de næringene som har henholdsvis høyest og lavest gjennomsnittslønn (foruten referansenæringen).

Yrker

For de forskjellige yrkesvariablene finner vi at de fleste yrker har en negativ koeffisient, og noen få har en positiv koeffisient sammenlignet med referanseyrket vårt, som er Politikere og toppledere. Ved de fleste yrker vil det ha en negativ effekt på lønnen å jobbe i de yrkene, sammenlignet med referansen. Som vi ser i tabell 4 er referanseyrket vårt et av de yrkene med høyest lønn, så dette gir mening. Samtidig kan vi se en tendens der yrkene som har en stor andel innvandrere, som renholdsyrker og serviceyrker, har den største negative effekten på lønnen. Og de yrkene det er en liten innvandrerandel, som lederyrker, har en positiv effekt på lønnen relativt til referanseyrket.

5.2 Resultater fra Oaxaca-Blinder

Vi ser fra kapittel 5.1 at vi har et predikert lønns-gap mellom innvandrere og befolkningen ellers på omtrent 11%. Vi er nå interessert i å dekomponere lønns-gapet for å se hvor mye som kan forklares av forskjeller i egenskaper, og hvor mye som skyldes forskjeller i avkastning på egenskaper. Vi utleder Oaxaca-Blinder som presentert i kapittel 4.1 og får resultatene i tabell 6. Dekomponeringen er basert på regresjonene som finnes i tabell 4, kolonne (2) for vår samlede regresjon, og i tabell 5 for de separate regresjonene for innvandrere og befolkningen ellers. Vi går fra lønns-gap på 10,8% til 7,9% når vi har med kontrollvariabler, som impliserer at litt over en fjerdedel av gapet kan forklares av forskjeller i egenskaper, resten skyldes ulik avkastning på egenskaper mellom innvandrere og de fra befolkningen ellers. Som vi ser fra tabell 6 vil det

VARIABLER kategori	lnW Innvandrere	lnW Norsk
alder	0,034337 (0,000819)***	0,040402 (0,000390)***
alder2	-0,000324 (0,000010)***	-0,000383 (0,000005)***
kvinne	-0,115658 (0,001346)***	-0,128472 (0,000682)***
storby	0,029178 (0,001175)***	0,040167 (0,000577)***
utd_høy	0,065458 (0,001476)***	0,110455 (0,000848)***
barn	0,001476 (0,001243)***	0,010591 (0,000647)***
gift	-0,005586 (0,001229)***	0,036618 (0,000629)***
konstant	12,526629	12,668853
Næringsdummier	Ja	Ja
Yrkesdummier	Ja	Ja
Observasjoner	210 937	824 639
R-squared	0,424575	0,424322

Standardavvik i parenteser
*** p<0.01, ** p<0.05, * p<0.1

Tabell 5: Lønnsregresjon for innvandrere og de fra befolkningen ellers separat.

forventede lønns-gapet variere noe fra OLS-resultatene. Grunnen til denne variasjonen er at ved OLS så inkluderes automatisk flere variabler enn ved Oaxaca-Blinder, som er noe mer restriktiv. Gjennom analysen og metoden fremvist i kapittel 4 har vi kommet fram til at dette er en akseptabel variasjon mellom de to prediksjonene av lønns-gapet, på hhv. 10,8% og 11,4% (fra tabell 4 og 6).

Vi ser fra tabell 6 at det som forklarer den største delen av lønns-gapet, er forskjeller i yrker mellom innvandrere og de fra befolkningen ellers. Vi ser at 38% av det predikerte lønns-gapet skyldes forskjeller i yrker. Som nevnt tidligere vet vi at innvandrere ofte jobber i yrker med lavere lønn. De yrkene det er ansatt flest innvandrere i er pleie- og omsorgsarbeidere og salgstryker, som har gjennomsnittslønn på henholdsvis 442 000 og 484 000, dette er langt under snittet vi så i tabell 1. Et interessant aspekt å legge merke til i panel A er at innvandrere er overrepresentert i lavtlønnsyrker samtidig som de er overrepresentert i høytlønnede næringer. Som vi ser så har

	Hele befolkningen
Forventet lønnsgap	0,114
<i>Panel A: Forskjeller i egenskaper</i>	
Alder	0,006
Kjønn	-0,007
Storby	-0,006
Utdanning	0,000
Barn	0,001
Gift	-0,003
Næring	-0,007
Yrke	0,043
Totalt	0,029 (26,5%)
<i>Panel B: Forskjeller i avkastning</i>	
Alder	0,144
Kjønn	-0,005
Storby	0,007
Utdanning	0,024
Barn	-0,012
Gift	0,02
Næring	-0,085
Yrke	-0,149
Konstant	0,142
Totalt	0,085

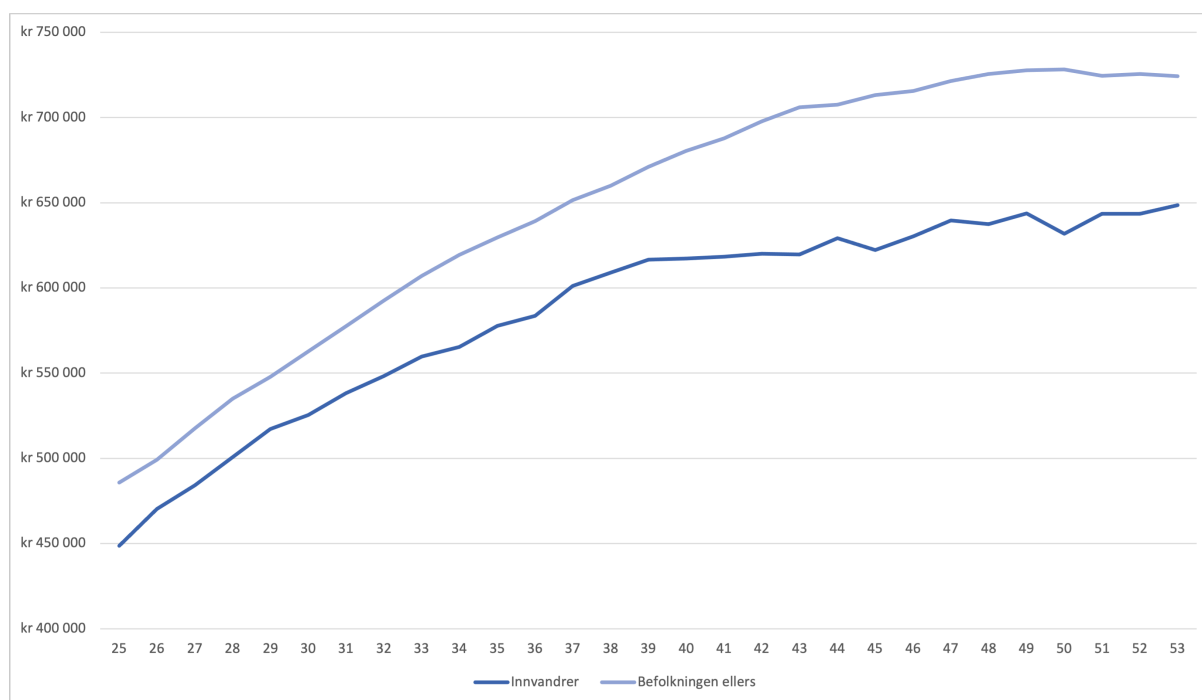
Tabell 6: Dekomponering av lønnsgapet mellom innvandrere og personer fra befolkningen ellers

næring en liten, men negativ verdi, altså bør den trekke i retning av høyere lønn, for innvandrere relativt til resten av befolkningen. Fra figur 2 ser vi at det her er flere av de høytlønnede næringene hvor det er en større andel innvandrere, som for eksempel industri og transport og lagring. I tillegg ser vi at det også er flere av de lavtlønnede næringene med en større andel av befolkningen ellers, som undervisning og helse. Vi ser generelt fra figur 3 at det er en jevnere fordeling mellom innvandrere og befolkningen ellers blant lavt- og høytlønnsnæringer. Vi skal komme tilbake til hvorfor det kan være sånn at yrke bidrar positivt til lønnsgapet mens næring bidrar negativt i kapittel 6.

Videre forklarer forskjeller i alder, som er vår proxy-variabel for arbeidserfaring, omtrent 5% av gapet. Dette stemmer overens med det vi fant i kapittel 3.3 med deskriptiv statistikk, at innvandrere har litt lavere gjennomsnittsalder enn befolkningen ellers.

I utvalget vårt er andelen menn høyere blant innvandrere, det er en større andel innvandrere som bor i storbyer, og det er en større andel innvandrere som er gift, enn blant befolkningen ellers. Alle disse egenskapene skal, ifølge koeffisientene fra tabell 5, trekke i retning av høyere lønn. Lønnsgapet eksisterer da til tross for at innvandrerne har disse fordelaktige egenskapene. Utdanning påvirker ikke den forklarende delen av lønnsgapet vårt. Vi vet at det er samme andel som har høyere utdanning blant innvandrere og befolkningen ellers, fra tabell 2, så dette gir mening ut ifra analysen.

26,5% av lønnsgapet forklares av forskjeller i egenskaper, så de resterende 73,5% forklares av forskjeller i avkastning på egenskaper, som vi finner i panel B i tabell 6. Denne delen er ofte sett på som diskrimineringsandelen. Vi ser at den største delen av gapet her skyldes alder. Dette betyr at innvandrere får generelt sett dårligere avkastning på et år mer i arbeid enn det befolkningen ellers gjør. Vi ser fra tabell 5 at en person fra befolkningen ellers vil få 0,97% høyere lønn ved et år ekstra dersom hen er 40 år gammel, mens en innvandrer vil få 0,84% høyere lønn. Vi er interessert i å se hvordan dette utvikler seg, så vi ser på gjennomsnittlig lønn for alle aldre for både innvandrere og befolkningen ellers og får resultatene i figur 6.



Figur 6: Lønnsbanen for innvandrere og befolkningen ellers.

Vi ser at i starten av yrkeskarrieren stiger lønnen til innvandrere og de fra befolkningen ellers relativt likt, men etterhvert begynner lønnen til innvandrere å sakke mer. Det virker som at lønnen

til innvandrere tidligere når et punkt hvor den begynner å slake seg ut, mens lønnen til de fra befolkningen ellers fortsetter å vokse med mer per år. I årene fra 38 til 53 stiger den gjennomsnittlige innvandrerlønnen med omtrent 40 000kr, mens på de samme årene stiger gjennomsnittslønnen til befolkningen ellers med 65 000kr. I tillegg ser vi at befolkningen ellers når sitt toppunkt når de er 50, og så går lønnen litt ned de neste årene, mens for innvandrere er lønnen høyest det siste året. Videre ser vi også at for befolkningen ellers er lønnen stigende hele tiden, med unntak av de siste tre årene, mens for innvandrere er det mindre jevnt. Dette kan også skyldes at vi har mindre utvalg for innvandrere, så det er større sjanser for at det blir tydeligere svingninger.

Videre ser vi, av tabell 6, også at utdanning har en relativt stor effekt på den uforklarte delen av gapet, som vil bety at innvandrere får dårligere avkastning på å ta høyere utdanning. Denne relativt lavere effekten av utdanning for innvandrere bekrefter resultatene vi fikk ved OLS fra tabell 5. Hvis vi ser på resultatene fra denne regresjonen, ser vi at innvandrere vil få gjennomsnittlig 6,5% høyere lønn ved å ha høyere utdanning, mens befolkningen ellers vil få hele 11% høyere lønn, altså nesten dobbelt så mye. I tillegg ser vi at innvandrere også får dårligere avkastning på å bo i en storby og å være gift, mens det å ha barn eller være kvinne får de bedre avkastning på. En interessant ting vi ser fra tabell 5 er at det å være gift bidrar til en positiv effekt på lønnen for befolkningen ellers, men en negativ effekt på lønnen for innvandrere.

5.3 Analyse av heterogeniteter

Analysen utvides for å ta hensyn til heterogeniteter mellom innvandrere. Det er stor variasjon mellom hvor innvandrere kommer fra, og bakgrunnen for hvorfor de velger å innvandre. Variasjonene mellom innvandringsgrupper skal vi se har potensielt stor innvirkning på lønnsgapet. I analysen av heterogeniteter fokuseres det på om innvandrere kommer fra et høy- eller lavinntektsland, samt om innvandringen er på grunn av enten familie eller flukt eller utdanning eller arbeid. Denne delen av analysen presenterer resultatene fra Oaxaca-Blinder dekomponeringen i kapittel 4.2.

5.3.1 Landbakgrunn som heterogenitet

Av de utvidelsene vi foretar, er det utvidelsen som omhandler landbakgrunn som gir den største variasjonen i lønnsgap. Lønnsgapet mellom innvandrere og befolkningen ellers er på 17% dersom de er fra lavinntektsland, mens det faktisk er en liten og positiv lønnspremie på 1% for de

fra høyinntektsland. Dette betyr at store deler av det initielle lønns-gapet fra tabell 4 er drevet av innvandrere fra lavinntektsland. Vi vil derfor, på bakgrunn av at resultatet fra høyinntektsland er tilnærmet lik null, se bort ifra dette resultatet i diskusjonen. Regresjonene for denne analysen finnes i appendiks A.10. Regresjonen for utvalget med *innv.H* og befolkningen ellers består av 896 481 individer, og for *innv.L* og befolkningen ellers er det 963 733 individer. Vi kjører regresjonene fra likning 14 og 15 fra kapittel A.8.1. Gjennom disse likningene kjøres separate regresjoner for å se forskjellen i lønn mellom innvandrere fra høyinntektsland mot befolkningen ellers og innvandrere fra lavinntektsland mot befolkningen ellers. Resultatene etter Oaxaca-Blinder dekomponeringen presenteres i tabell 7. Grunnet for få observasjoner i noen yrker og næringer må vi gruppere disse sammen for å kunne bruke dem i analysen ².

Fra panel A i tabell 7 ser vi at cirka 30% av lønns-gapet mellom innvandrere fra lavinntektsland og befolkningen ellers kan forklares gjennom de variablene som er inkludert i modellen. Av det forklarte gapet skyldes hele 64% valget av yrke, som vil si at den største faktoren som forklarer det predikerte lønns-gapet er yrkesvalg. Altså at innvandrere fra lavinntektsland velger eller ender opp med å jobbe i yrker hvor de tjener generelt mindre enn befolkningen ellers. Denne faktoren som beskriver innvandreres yrkesvalg utgjør 30,5% av det totale predikerte lønns-gapet.

Til forskjell fra resultatene i kapittel 5.2 så finner vi her at utdanning er en forklarende faktor på det predikerte gapet, som utgjør 3,5% av det forventede lønns-gapet. Dette funnet betyr at den mekanismen vi fant i kapittel 5.2, der utdanning ikke hadde noen effekt på lønns-gapet for den initielle innvandringsgruppen, er drevet av den motsatte effekten for de innvandrerne som kommer fra høyinntektsland. Det er da en mindre andel innvandrere fra lavinntektsland som har høyere utdanning enn befolkningen ellers. Dette ser vi fra tabell 2 at det er 53% av befolkningen ellers som har høyere utdanning, mens det bare er 47% av innvandrere fra lavinntektsland som har det. I likhet med hovedgruppen for innvandrere i kapittel 5.2 forklares cirka 10% av det forklarte gapet av alderen på individet, altså denne innvandrerguppen er gjennomsnittlig yngre enn befolkningen ellers.

De resterende 70% av lønns-gapet finner vi i panel B, tabell 7. For innvandrere fra lavinntektsland er avkastning på alder, og dermed erfaring, noe som påvirker det predikerte gapet mye. Som vi kan se fra lønnsbanen presentert i figur 7 så vil denne innvandringsgruppen ikke bare ha et lavere

²Vi grupperer sammen yrkene *Administrative og merkantile ledere* og *Politikere og toppledere*, og vi grupperer næringen for *Kulturell virksomhet mv.* inn i næringen for *Lønnet arbeid i private husholdninger og annet arbeid* for innvandringsgruppen fra lavinntektsland

	Høyinntektsland	Lavinntektsland
Forventet lønnsgap	-0,01	0,17
<i>Panel A: Forskjeller i egenskaper</i>		
Alder	-0,002	0,008
Kjønn	-0,002	-0,003
Storby	-0,006	-0,007
Utdanning	-0,011	0,006
Barn	0,001	0,001
Gift	-0,001	-0,006
Næring	-0,001	-0,001
Yrke	-0,008	0,052
Totalt	-0,031	0,051 (29,9%)
<i>Panel B: Forskjeller i avkastning</i>		
Alder	-0,115	0,15
Kjønn	-0,005	-0,009
Storby	-0,003	0,007
Utdanning	0,016	0,026
Barn	-0,005	-0,009
Gift	0,002	0,022
Næring	-0,036	-0,148
Yrke	-0,079	-0,169
Konstant	0,247	0,25
Totalt	0,021	0,119

Tabell 7: Dekomponering av lønnsgapet etter landbakgrunn

utgangspunkt med tanke på gjennomsnittslønn, men og en mer avtakende lønnsutvikling over tid i forhold til både befolkningen ellers og innvandrere fra høyinntektsland. Dette resultatet samsvarer med den initielle regresjonen for de to innvandringsgruppene, se appendiks tabell A.10.1.

Videre ser vi fra panel B at det er flere andre egenskaper som innvandrere fra lavinntektsland får dårligere avkastning på enn de fra befolkningen ellers. I likhet med alle innvandrere får denne gruppen dårligere avkastning på høyere utdanning enn de fra befolkningen ellers. Som vi ser fra tabell A.10.1, vil innvandrere få gjennomsnittlig 6% høyere lønn av å ha høyere utdanning, mens de fra befolkningen ellers vil få 11%. I tillegg ser vi at innvandrere fra lavinntektsland får

en mindre positiv effekt av å bo i en storby. For denne gruppen ser vi også den samme trenden som med alle innvandrere, at det å være gift vil gi en negativ effekt på lønnen, mens det for befolkningen ellers gir en positiv effekt. Vi kan nok si at gruppen med innvandrere fra lavinntektsland bidrar mest til lønnsforskjellene mellom befolkningen ellers og innvandrere generelt.

5.3.2 Innvandringsgrunn som heterogenitet

Den andre utvidelsen betrakter bakgrunnen for innvandring som en heterogenitet. Resultatene her vil dermed vise om det har noe å si for lønnsulempen om man har kommet til landet på grunn av arbeid eller utdanning, eller familie eller flukt. Regresjonene for denne analysen finnes i appendiks A.11. Regresjonen for utvalget med *arb_utd* og befolkningen ellers består av 898 007 individer, og for *fam_flu* og befolkningen ellers er det 881 529 individer. Grunnet få observasjoner blant noen av variablene for disse gruppene er vi nødt til å gruppere sammen noen næringer og yrker³.

Som presentert i tabell 8 er lønnsgapet nesten dobbelt så stort for innvandrere som kommer på grunn av familie eller flukt, enn det er for innvandrere som kommer på grunn av arbeid eller utdanning, med et predikert gap på henholdsvis 22,2% for familie eller flukt, og 11,7% for arbeid eller utdanning.

Arbeid eller utdanning som innvandringsgrunn

Fra panel A i tabell 8 ser vi at modellen ikke kan forklare mye av lønnsgapet for arbeids- og utdanningsinnvandrere. Dette betyr at lønnsgapet mellom denne innvandringsgruppen og befolkningen ellers kun skyldes uobserverbare faktorer eller diskriminering. Vi ser fra første kolonne, panel B i tabell 8, at arbeids- og utdanningsinnvandrere har, som tidligere grupper, tydelig dårligere avkastning på et år ekstra i arbeid i forhold til befolkningen ellers, samt dårligere avkastning på høyere utdanning. Disse to variablene utgjør til sammen 62% av det predikerte lønnsgapet.

Et annet aspekt vi ser her i panel A er at den negative verdien til utdanningsvariabelen tilsier at denne innvandringsgruppen har generelt høyere utdanning enn den øvrige befolkningen, som beveger lønnsgapet i favør av innvandrerne. Vi ser og at det er en større andel menn, en større

³Grupperer *Kulturell virksomhet mv.* og *Lønnet arbeid i private husholdninger og annet arbeid* sammen, og grupperer alle lederyrkene sammen. I tillegg grupperer vi *IKT-teknikere* og *ingeniører mv.* sammen for gruppen med familie eller flukt som innvandringsgrunn

	Arbeid eller utdanning	Familie eller flukt
Forventet lønnsgap	0,117	0,222
<i>Panel A: Forskjeller i egenskaper</i>		
Alder	-0,001	0,002
Kjønn	-0,016	0,009
Storby	-0,005	-0,007
Utdanning	-0,004	0,012
Barn	0,003	0,000
Gift	-0,006	-0,008
Næring	-0,019	0,014
Yrke	0,040	0,063
Totalt	-0,008	0,085 (38,2%)
<i>Panel B: Forskjeller i avkastning</i>		
Alder	0,051	0,416
Kjønn	-0,006	-0,018
Storby	0,002	0,013
Utdanning	0,022	0,029
Barn	-0,019	-0,003
Gift	0,017	0,03
Næring	-0,099	-0,233
Yrke	-0,129	-0,1
Konstant	0,286	0,002
Totalt	0,125	0,137

Tabell 8: Dekomponering av lønnsgapet etter innvandringsgrunn

andel som er gift, en større andel som bor i storby og en større andel som er sysselsatt i næringer med høyere snittlønn. Alle disse faktorene skulle tilsi at de tjener mer enn befolkningen ellers, men slik er det altså ikke. Dette skyldes faktorene i panel B, som viser forskjellene i avkastning på egenskaper.

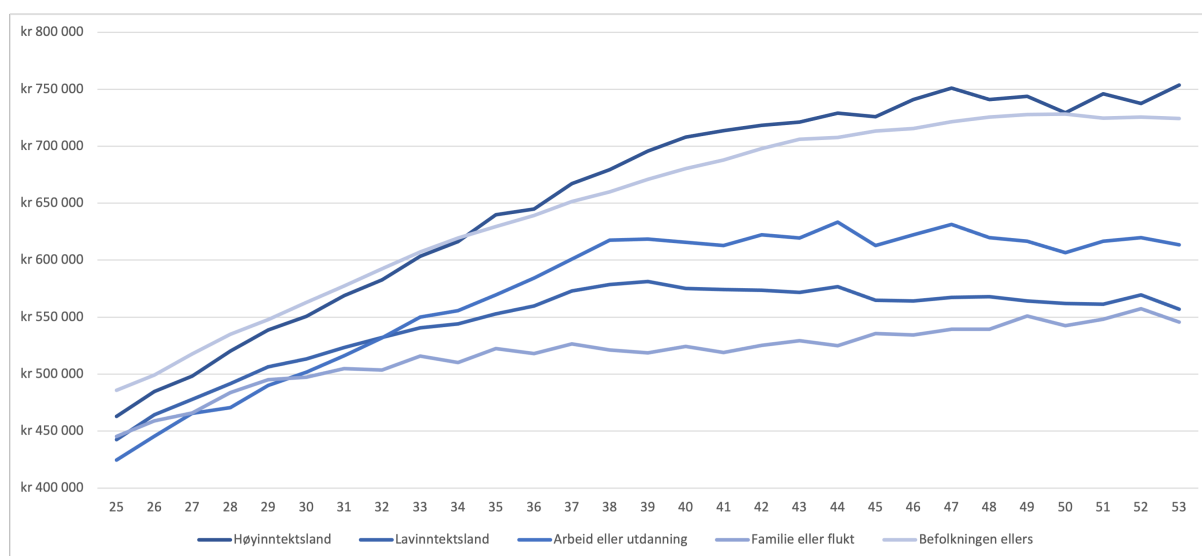
Valget av yrke blir igjen en viktig faktor å ta med i betraktning. Fra panel A later arbeids- og utdanningsinnvandrere til å jobbe i yrker som tjener mindre enn den øvrige befolkningen. De gjør dette til tross for at de i utgangspunktet har tilsynelatende bedre forutsetninger enn resten av befolkningen til å tjene mer.

Familie eller flukt som innvandringsgrunn

For innvandrere med familie eller flukt som innvandringsgrunn kan 38,2% av lønnsgapet forklares av de inkluderte variablene, altså forskjeller i egenskaper. Et interessant aspekt i tabell 8 er at for denne gruppen så utgjør 5,5% av det predikerte gapet variasjoner i utdanning i forhold til befolkningen ellers. Dette er til stor kontrast fra de andre innvandringsgruppene, der utdanning ikke later til å ha noen forsterkende innvirkning på det forklarte lønnsgapet (se panel A, tabell 6). Denne innvandringsgruppen er den som har størst andel kvinner som vi ser i tabell 2. Dette ser vi fra panel A i tabell 8 at er med på å forklare 4% av det predikerte lønnsgapet. De største faktorene som bidrar til lønnsgapet er igjen nærings- og yrkesvariablene. Hvilken næring og yrke en innvandrer velger utgjør, for innvandrere med familie eller flukt som innvandringsgrunn, hele 34,7% av det predikerte lønnsgapet. De positive verdiene for nærings- og yrkesvariablene forteller oss at denne innvandrergruppen generelt jobber innenfor næringer og yrker som tjener gjennomsnittlig mindre enn befolkningen ellers. Fra panel A ser det ut til at dette kan skyldes mangel på arbeidserfaring og dårligere, eller manglende, utdanning i forhold til resten av befolkningen.

I panel B ser vi at innvandrere som kommer på grunn av familie eller flukt har drastisk dårligere avkastning på spesielt erfaring, det utgjør faktisk 50% av det uforklarte gapet. Dette kan skyldes at det potensielt blir vanskeligere å bruke aldersvariabelen som en proxy-variabel for erfaring på denne gruppen. Det kan være mer usikkerhet rundt tidligere arbeidserfaring, spesielt for de som kommer som flyktninger. I tillegg vil de trolig også bruke lengre tid på å komme seg i jobb etter innvandring, i motsetning til andre innvandringsgrupper. Fra regresjonsresultatene i tabell A.11.1 ser vi også at innvandrere fra gruppen for familie eller flukt ikke har den samme høye lønnskurven med tanke på alder/erfaring som befolkningen ellers. Dette kan og ses grafisk gjennom lønnsbanene i figur 7.

Vi ser at avkastning på erfaring er noe som bidrar til å forklare store deler av lønnsgapet for alle innvandringsgruppene, så vi sammenligner lønnsbanene for de ulike gruppene i figur 7. Vi ser at alle de ulike innvandringsgruppene starter med relativt lik lønn, som er lavere enn befolkningen ellers, men de ender på veldig forskjellige steder. Vi ser at de fra høyinntektsland krysser befolkningen ellers etter omtrent ti år, og blir værende over resten av arbeidslivet. De resterende gruppene har alle lavere lønn hele arbeidslivet, men med noe forskjellig utvikling. De fra lavinntektsland og de som kom for arbeid eller utdanning stiger de første årene ganske



Figur 7: Lønnsbanen for de ulike innvandringsgruppene og befolkningen ellers.

jevnt med de fra befolkningen ellers, men så slaker det seg ut rundt slutten av 30-årene, hvor de for arbeid og utdanning ender nederst, og får en gradvis reduksjon i lønn etter dette. For de som kom for familie eller flukt, stiger lønnen det meste av tiden, men lønnsutviklingen er mye slakere enn alle de andre gruppene, og de har jevnt over den laveste lønnen nesten hele arbeidslivet. Som med figur 6 ser vi at innvandringsbanene er mindre jevne og varierer mer fra år til år, dette skyldes trolig at det er mye mindre utvalg, så større sjanse for variasjoner.

6 Diskusjon

I dette kapittelet skal vi starte med å diskutere resultatene fra de tre Oaxaca-Blinder dekomponeringene vi gjorde i kapittel 5. Vi skal diskutere dem separat og sammenligne dem, i tillegg til å diskutere resultatene i lys av tidligere forskning og teori vi har lagt frem i kapittel 2. Videre vil vi gå inn på litt kritikk av datamaterialet vårt, og til slutt skal vi snakke litt om videre forskning på dette temaet.

6.1 Drøfting av resultatene

Vi skal nå se på resultatene vi fant i kapittel 5.2-5.3 og diskutere disse. Vi skal se på hva resultatene kan skyldes, knytte dem opp mot tidligere forskning, samt sammenligne de tre modellene våre.

Innvandrere og befolkningen ellers

Det eksisterer et lønnsgap mellom innvandrere og befolkningen ellers, og selv når vi separerer gapet i det forklarte og uforklarte, ser vi at det uforklarte fortsatt trekker i retning av lavere lønn for innvandrere. For hele innvandringsgruppen er det nesten tre fjerdedeler som er uforklart, så det skyldes enten uobserverbare faktorer, eller ren diskriminering. Den største kilden til lønnsgap blant det uforklarte er forskjell i avkastning på alder. Vi bruker som nevnt alder som en proxyvariabel for erfaring, så her kan man si at innvandrere trolig får dårligere avkastning på erfaring enn en fra befolkningen ellers.

Denne relativt dårligere avkastningen kan skyldes flere ting. For førstegenerasjonsinnvandrere kan det handle om tidspunktet de har innvandret til Norge. Dersom de har kommet etter at de var 25, altså den yngste aldersgruppen i vårt utvalg, vil de ha måttet starte sin yrkeskarriere senere enn en jevnaldrene fra befolkningen ellers. Så dersom de enten ikke har jobberfaring fra sitt hjemland, eller den jobberfaringen de har ikke er overførbar, vil de ha et lavere utgangspunkt mtp. lønn. Vi ser fra datasettet vårt at nesten 70% av innvandrerne i Norge har kommet før de fylte 30 år, og majoriteten av dem kommer når de er mellom 20 og 30, så det vil være noen som må starte sin jobbkarriere senere enn de fra befolkningen ellers, men det vil trolig ikke være den eneste effekten. Den lavere avkastningen på erfaring kan også skyldes diskriminering eller andre faktorer som påvirker innvandrere negativt. Det skal vi se mer på når vi diskuterer forskjeller i yrker.

Fra den forklarte delen av gapet fant vi at innvandrere som jobber i yrker med dårligere lønn forklarer det meste av gapet. Som nevnt i kapittel 2 finner flere artikler, blant annet (Barth mfl., 2012), at innvandrere ofte er i dårligere betalte jobber, og sliter med å komme seg til høyere betalte jobber, vi har funnet de samme tendensene. Dette kan skyldes en kombinasjon av statistisk, preferansebasert og monopsonistisk diskriminering som lagt frem av Kolsrud mfl. (2016) og Dale-Olsen mfl. (2014) og forklart i kapittel 2. Trolig vil innvandrere slite med å komme seg til høyere betalte jobber dersom arbeidsgiver driver med statistisk diskriminering. Arbeidsgiveren vil anta at innvandreren har lavere produktivitet enn en arbeider som er fra befolkningen ellers, og vil velge å enten ikke gi lønnsøkning, eller bare ikke ansette i det hele tatt.

Videre er også preferansebasert diskriminering et problem for innvandrere. Som nevnt har innvandrere, og spesielt førstegenerasjonsinnvandrere, mindre informasjon om arbeidsmarkedet og egne rettigheter. De kan da lettere havne i jobber som er dårligere betalt, eller jobber med dårlige arbeidsvilkår. Dette fører oss videre til den siste typen diskriminering som innvandrere trolig også vil påvirkes av, nemlig monopsonistisk. Når en innvandrer har mindre informasjon om arbeidsmarkedet og om sin egen verdi, er det lettere for en arbeidsgiver å presse ned lønningene uten at hen mister produktivitet, fordi en innvandrer da er mindre følsom for endringer i lønn eller arbeidsvilkår. Innvandrere vil nok bli værende i dårligere betalte jobber både fordi de ikke selv er klar over at de er verdt mer, og fordi de trolig vil ha vanskeligheter med å bli ansatt, eller få høyere lønn, i andre yrker. Så selv om forskjell i valg av yrker er i den forklarte delen av lønnsgapet, kan man ikke se bort ifra at diskriminering spiller en viktig rolle her også.

Som vi så tidligere, jobber innvandrere i dårligere betalte yrker, men bedre betalte næringer. En forklaring på dette er at innenfor de næringsgrupperingene vi bruker så finnes det flere mindre næringsgrupperinger, som kan ha forskjellig lønn og forskjellig andel innvandrere. For eksempel for næringen Bergverksdrift og utvinning finnes det flere næringer om bryting av ulike typer stein, samtidig som utvinning av råolje og naturgass også er der. Oljesektoren har generelt veldig høye lønninger, men samtidig veldig få innvandrere. Dermed, om det er flere innvandrere i de andre næringsgruppene, med lavere lønn, kan dette forklare hvorfor innvandrere generelt er i mer høylønnsnæringer, men lavtlønnsyrker. Næringen generelt får høyere gjennomsnittslønn grunnet høy lønn i noen av undergruppene, men den lønnen vil ikke gjelde for innvandrere dersom de jobber i de andre undergruppene. Dette kan tyde på at det kunne vært lurt å bruke en mer separert næringsgruppering for å se om dette stemmer, eller om det er andre forklaringer.

Vi kan også se fra panel B tabell 6 at utdanning bidrar til en stor del av den uforklarte delen av lønnsgapet. Som nevnt i kapittel 5.2 vil en fra befolkningen ellers få nesten dobbelt så mange prosentpoeng høyere lønn enn en innvandrer ved å ha høyere utdanning. Mye av dette vil trolig skyldes, som nevnt i kapittel 2, at ikke all utdanning vil bli sidestilt med en norsk utdanning, selv i den situasjonen hvor den er like god. Dette kan skyldes bias fra arbeidsgivere som er mer skeptiske til utdanningsinstitusjoner de ikke kjenner til. Dette kan også skyldes at innvandrere jevnt over har færre år med høyere utdanning enn befolkningen ellers. Ettersom *utd_høy* kun er en dummy som tar verdien 1 om individet har minst et år universitetsutdanning, eller lignende, vil den ikke fange opp forskjeller i antall år med høyere utdanning.

Landbakgrunn

Som vi så fra resultatene i kapittel 5.3.1, har det mye å si på lønnsgapet hvilket land en innvandrer kommer fra. Skillet mellom om en innvandrer kommer fra et høy- eller lavinntektsland er stort og påvirker lønnen i stor grad. Det kan være flere grunner til dette, som vi skal gå nærmere inn på nå.

Fra panel A i tabell 7 ser vi at også for innvandrere fra lavinntektsland vil yrke ha en stor påvirkning på lønnen, ettersom også denne gruppen innvandrere oftere jobber i dårligere betalte yrker. Dette vil nok skyldes mye av det samme som ble diskutert i forrige delkapittel. Videre ser vi at alder har mye å si her, denne gruppen er jevnt over yngre enn befolkningen ellers, og har da trolig mindre arbeidserfaring. Denne aldersskjevheten kan skyldes flere ting. Det kan være at vi får noen av de samme tendensene som Bratsberg mfl. (2011) og Kolsrud mfl. (2016) finner i sine artikler, altså at sysselsettingsraten synker og trygderaten øker blant innvandrere 10-15 år etter at de kom til landet. Om vi har de samme trendene, kan det bidra til å forklare hvorfor denne gruppen har lavere gjennomsnittalder, ettersom vi har fjernet alle som mottar trygd og som ikke jobber fulltid. Som vi ser av figur 1, er det en stigende sammenheng mellom år i Norge og lønn, så vi finner ikke helt den samme trenden. Samtidig kan vi se at for innvandrere fra lavinntektsland er det en mye mindre økning mellom gruppene for de som har vært i Norge i mellom 7 og 15 år og de som har vært i Norge i 15 og 20 år, enn de første to gruppene. Så dette kan bidra til å forklare denne ulikheten.

Fra panel B, kolonne 2 i tabell 7 kan vi se hvordan graden av utdanning vil påvirke innvandrere ut ifra hvor de kommer fra. Som vi var innom i kapittel 2.1.2, vil deler av differansen i utdan-

ningsvariabelen her trolig skyldes kvaliteten på utdanningsinstitusjonene i hjemlandet. Det vil trolig være lettere for en fra et høyinntektsland å overføre utdanningen sin til en norsk tilsvarende utdanning. Dermed vil det være lettere for en arbeidsgiver å være trygg på kompetansen til jobbsøkere fra høyinntektsland enn lavinntektsland. Innvandrere fra lavinntektsland vil ha vanskeligere for å unngå den statistiske diskrimineringen som ofte oppstår når det er usikkerhet rundt kvaliteten på utdanningen deres. Vi ser også fra panel B, at for innvandrere fra lavinntektsland, vil også avkastningen på erfaring, det å være gift, og bosted være dårligere enn for befolkningen ellers. Dette tyder igjen på at disse variablene har forskjellig effekt for innvandringsgruppen og befolkningen ellers.

For gruppen fra lavinntektsland er det også interessant å se at til tross for at det er en større andel menn i denne gruppen, at det er en større andel som bor i storbyer og at det er en større andel som er gift, enn befolkningen ellers, eksisterer det et lønns-gap. Alle disse egenskapene bør trekke i retning av høyere lønn, så det er interessant å se at lønns-gapet fortsatt er såpass stort.

Innvandringsgrunn

Vi så videre på forskjellige grupper innvandrere avhengig av deres innvandringsgrunn, og sammenlignet det med befolkningen ellers gjennom Oaxaca-Blinder. Vi fant da at det var tydelige forskjeller mellom de to innvandringsgruppene. Som nevnt i kapittel 5.3.2, ser vi en stor forskjell i lønns-gapet blant disse gruppene, hvor de som har arbeid eller utdanning som innvandringsgrunn har et lønns-gap på 11%, altså omtrent det samme vi fant for alle innvandrere, og de som kom for familie eller flukt har et lønns-gap på over 22%. Det er tydelig at innvandringsgrunnen påvirker lønnen mye. Dette er det samme som ble funnet av Kolsrud mfl. (2016) og som ble snakket om i kapittel 2. De fant også at lønns-gapet var størst mellom innvandrere som kommer som flyktninger og befolkningen ellers.

Vi ser først på gruppen med arbeid eller utdanning som innvandringsgrunn, som vi finner i kolonne 1 i tabell 8. Vi ser at den delen av lønns-gapet som skyldes forskjeller i egenskaper, altså panel A, nesten bare har negative verdier, og det resulterer i at den totale forklarte delen av gapet er negativt. Det betyr at denne innvandringsgruppen har egenskaper som bør gjøre at de tjener mer enn de fra befolkningen ellers. De er gjennomsnittlig eldre, det er en større andel menn, det er en større andel med høyere utdanning og så videre. Forskjellene i egenskaper tilsier at de egentlig bør tjene mer enn de fra befolkningen ellers, men det gjør det altså ikke. Det skyldes da

forskjeller i avkastning på egenskaper.

Vi ser fra panel B at det er mye likt som de tidligere gruppene. Den største effekten er alder, som igjen betyr at denne gruppen innvandrere får dårligere avkastning på erfaring enn de fra befolkningen ellers. Videre ser vi også at denne gruppen får ganske mye dårligere avkastning på å ha høyere utdanning, og litt dårligere avkastning på å bo i storby. Dette skyldes nok mye det samme som er diskutert tidligere for de andre analysene. Samtidig ser vi at denne gruppen faktisk får bedre avkastning på å ha barn, enn det en fra befolkningen ellers gjør.

Videre ser vi på gruppen som har familie eller flukt som innvandringsgrunn, som vi har i kolonne 2 i tabell 8. Vi ser for det første, at for denne gruppen er det en mye større andel av lønnsgapet som skyldes forskjeller i egenskaper, nesten 40%. Yrkesvalg bidrar til den største delen av det forklarte gapet. For denne gruppen, i likhet med tidligere grupper, er det en større andel som bor i storby og som er gift, som bør gi høyere lønn, samtidig som vi fortsatt får et lønnsgap. Den største forskjellen for denne gruppen sammenlignet med de andre er at det er en større andel kvinner sammenlignet med befolkningen ellers, som da bidrar til å forklare lønnsgapet.

Når vi ser på forskjeller i avkastning i panel B, ser vi at det er mye likt som gruppen fra lavinntektsland, men at denne innvandringsgruppen har veldig mye dårligere avkastning på erfaring enn det befolkningen ellers har. Som nevnt tidligere, utgjør dette halvparten av det uforklarte gapet. Fra regresjonsresultatene, som finnes i appendiks A.11, ser vi at en person fra befolkningen ellers vil få gjennomsnittlig 1% høyere lønn av et år ekstra erfaring, mens en fra innvandrergruppen familie eller flukt kun vil få 0,6% mer. Dette kan igjen skyldes de ulike diskrimineringstypene som vi har diskutert tidligere, og trolig spesielt monopsonistisk diskriminering. Som nevnt, er det lettere for en arbeidsgiver å redusere lønninger, eller unngå å øke lønninger, uten å miste arbeidstakere, dersom arbeidstakeren har mindre informasjon om arbeidsmarkedet og sin egen verdi. Eller dersom den vet at den vil ha vanskeligheter med å bli ansatt andre steder. Barth mfl. (2012) finner at innvandrere fra fattigere land oftere må gjennom lengre perioder med arbeidsledighet før de finner en jobb enn befolkningen ellers, og ettersom mange av de som kommer for familie og flukt også kommer fra lavinntektsland, er det rimelig å anta at det vil gjelde denne gruppen også.

Sammenligning

Som vi har sett er det mange forskjeller mellom de ulike innvandringsgruppene, og resultatene vi fant fra utvidelsene av Oaxaca-Blinder, gjør det tydelig at man ikke kan se på innvandrere som en homogen gruppe. Blant de ulike gruppene er det noen hvor befolkningen ellers tjener over 20 prosent mer, men også noen innvandringsgrupper som faktisk tjener mer enn befolkningen ellers. Samtidig ser vi også at det er stort sammenfall mellom noen av gruppene, spesielt mellom innvandrere fra lavinntektsland og de som kommer for familie eller flukt, og noe sammenfall mellom innvandrere fra høyinntektsland og de som kommer for arbeid eller utdanning.

Selv om det er forskjell i størrelsen de ulike egenskapene har på lønns-gapet, ser vi som nevnt de samme tendensene mellom gruppene, med hensyn på hva de får bedre og dårligere avkastning på, og hva som bidrar til den forklarte delen av lønns-gapet. For gruppene arbeid eller utdanning og høyinntektsland er dette ekstra interessant. Både arbeid eller utdanning og høyinntektsland har en total negativ sum av forskjeller i egenskaper, altså de observerbare egenskapene til disse innvandringsgruppene bør tilsi at de har høyere lønn enn befolkningen ellers, men dette stemmer kun for gruppen fra høyinntektsland. De som kommer for arbeid eller utdanning får dårligere avkastning på flere av egenskapene som de fra høyinntektsland får positiv avkastning på, sammenlignet med de fra befolkningen ellers. Fra tabellen i appendiks A.12 ser vi at for de som har arbeid eller utdanning som innvandringsgrunn, er det mange fra både høy- og lavinntektsland, men flere fra lavinntektsland. Og hvis vi ser på gjennomsnittslønn blant de som er i denne gruppen, ser vi at det er en betydelig forskjell mellom landbakgrunnene. Det virker da kanskje som at det ikke er så mye innvandringsgrunnen som har påvirket lønnen, men heller landbakgrunnen, og da er nok igjen mye av forklaringen ulike typer diskriminering.

6.2 utfordringer med datasettet og forslag til videre forskning

Det er noen svakheter og utfordringer med vårt datasett. De fleste av dem er knyttet til begrensninger som følger av å bruke Microdata.no, men det er andre svakheter også. Kanskje en av de største svakhetene med vårt datasett er at vi kun ser på individer som er i fulltidsjobb. Det er en større andel innvandrere som ikke jobber fulltid enn fra befolkningen ellers, og ettersom vi fjerner alle som ikke jobber fulltid, fjerner vi da en større andel innvandrere fra utvalget vårt. Det kunne vært interessant å se på alle som er i jobb uavhengig av stillingsprosent, og bruke timelønn som avhengig variabel. Dette hadde gjort at vi ikke måtte fjerne så mange innvandrere,

i tillegg til at man kunne fått en mer nyansert analyse av effektene av arbeidstimer på timelønn.

Videre er det også en svakhet med oppgaven at vi kan ha utelatte relevante variabler. Spesielt ville det vært nyttig å ha en variabel for et individs arbeidserfaring. I oppgaven har vi kun hatt variabel for alder, og brukt denne som en proxy-variabel. Dette har nok fungert greit ettersom vi kun har sett på arbeidere i fulltidsjobb, men denne variabelen vil ikke kunne fange opp tilfeller hvor individer bytter yrke helt, og må starte på nytt med erfaringen sin. I tillegg vil den ikke fange opp andre tilfeller med hull i erfaringen, som for eksempel om et individ går ut i barselpermisjon eller av andre personlige grunner må være borte fra jobb over lengre perioder. Inkluderingen av år i jobb hadde trolig gitt oss en bedre analyse.

Denne oppgaven har brukt Microdata.no, som er en relativt ny forskningsplattform og som stadig utvikles, det har ført til flere begrensninger og utfordringer. I analysen har vi kun sett på ett år og sett på de individuelle effektene dette året. For videre forskning kunne det vært interessant å ha en lengre tidshorisont på analysen, og følge en gruppe over flere år, for å se hvordan deres lønn endrer seg etterhvert som de er lengre i arbeidsmarkedet. Ved bruk av paneldata i Microdata, vil man måtte ha et mindre utvalg, i tillegg til at det vil være flere variabler som ikke vil være mulig å bruke, avhengig av årene inkludert i analysen. Ved videreutvikling av Microdata.no og deres funksjonaliteter og variabler, vil dette trolig være noe man kan se nærmere på.

I tillegg kunne det vært interessant å sammenligne forskjellige innvandrergupper fra forskjellige år for å se på om det er store forskjeller i tendenser fra ulike år. Man kan, som Bratsberg mfl. (2011) ser på i sin artikkel, sammenligne resultater fra flere innvandringsgrupper, og kanskje spesielt se på år hvor det har vært en stor økning i andelen innvandrere, om det så er på grunn av flukt eller arbeidsinnvandring.

7 Konklusjon

I denne oppgaven har vi sett på lønnsgapet mellom innvandrere og befolkningen ellers. Vi har benyttet analyseplattformen Microdata.no og gjennomført analysen med et tverrsnittsdatasett med omtrent en million individer i fulltidsjobb. Vi startet med en enkel OLS-modell for å se på det rå lønnsgapet, og gikk så videre til en Oaxaca-Blinder dekomponering for å se hva lønnsgapet består av. Som tidligere litteratur fant vi også fra både OLS og Oaxaca-Blinder at det eksisterer et betydelig lønnsgap mellom innvandrere og befolkningen ellers, i innvandrernes disfavør. Om vi kontrollerer for andre relevante variabler som kan påvirke lønnen, reduseres gapet noe, men vi finner at det fortsatt er et betydelig gap som er uforklart.

Vi bygger videre på analysen ved å dele opp innvandrerguppen etter landbakgrunn, altså om de kommer fra et høyinntektsland eller lavinntektsland, og innvandringsgrunn, altså om de innvandrete på grunn av familie eller flukt, eller arbeid eller utdanning. Vi fant da at for alle gruppene, utenom innvandringsgruppen fra høyinntektsland, eksisterer det et lønnsgap i innvandrernes disfavør. Dette gapet varierer mye på tvers av gruppene, og det er størst for de som har familie eller flukt som innvandringsgrunn. Etterfulgt av innvandrere fra lavinntektsland, og deretter innvandrere som har kommet for arbeid eller utdanning. Innvandrere fra høyinntektsland tjener faktisk gjennomsnittlig mer enn de fra befolkningen ellers, men ikke med en stor margin.

Generelt ser vi at valg av yrke er den faktoren som påvirker lønnsgapet mest, på tvers av alle innvandringsgruppene. Vi har funnet, som mye tidligere forskning også har funnet, at innvandrere har en tendens til å havne i yrker som er dårligere betalt enn de befolkningen ellers havner i. Videre er også avkastning på erfaring og utdanning det som forklarer de største delene av det uobserverbare gapet. Så innvandrere får generelt mindre økning i lønn av å ha høyere utdanning, eller å være et år eldre enn det en fra befolkningen ellers gjør. Her har vi også gått inn på flere mulige forklaringer på disse trendene.

Til slutt har vi sett at det er viktig i slike analyser å ikke se på innvandrere som en homogen gruppe. Vi har sett at det er store forskjeller blant de ulike innvandringsgruppene, hvor noen tjener veldig mye mindre enn befolkningen ellers, mens andre faktisk tjener mer. Dersom man har analyser hvor disse kun er én gruppe, kan det føre til at man får resultater hvor man kun får et gjennomsnitt og ikke får frem de faktiske effektene på lønn for de ulike innvandringsgruppene.

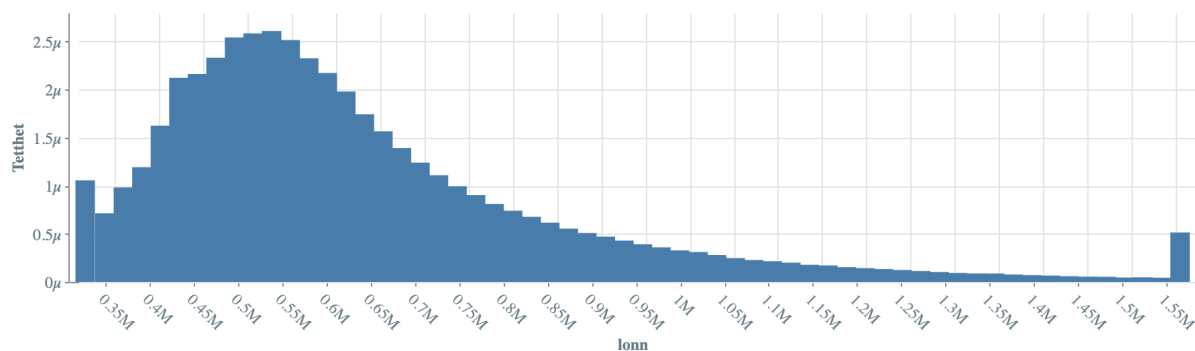
Referanser

- Askvik, T. (2020). Alder og yrke påvirker lønnsgapet. <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/alder-og-yrke-pavirker-lonnsgapet>
- Barth, E., Bratsberg, B., & Raaum, O. (2012). Immigrant wage profiles within and between establishments. *Labour Economics*, 19(4), 541–556.
- Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. *Journal of Human resources*, 436–455.
- Bratsberg, B., Røed, K., & Raaum, O. (2011). Yrkesdeltaking på lang sikt blant ulike innvandrergrupper i Norge. *Rapport, 1*, 2011.
- Bratsberg, B., Raaum, O., Røed, M., & Schøne, P. (2014). Immigration wage effects by origin. *The Scandinavian Journal of Economics*, 116(2), 356–393.
- Bye, K. S. (2021). Utdanning og Lønnsnivå Hos innvandrere. <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/utdanning-og-lonnsniva-hos-innvandrere>
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.
- Carneiro, P., Heckman, J. J., & Masterov, D. V. (2005). Labor market discrimination and racial differences in premarket factors. *The Journal of Law and Economics*, 48(1), 1–39.
- Chiswick, B. R. (1978). The effect of Americanization on the earnings of foreign-born men. *Journal of political Economy*, 86(5), 897–921.
- Dale-Olsen, H., Røed, M., & Schøne, P. (2014). Makt til å diskriminere? *Søkelys på arbeidslivet*, 31(4), 311–328.
- Elder, T. E., Goddeeris, J. H., & Haider, S. J. (2010). Unexplained gaps and Oaxaca–Blinder decompositions. *Labour Economics*, 17(1), 284–290.
- Folkehelseinstituttet. (2022). Innvandrere får dårligere helse av Å bo lenge i norge, Hvis de Ikke lærer norsk. <https://www.fhi.no/nyheter/2022/innvandrere-far-darligere-helse-av-a-bo-lenge-i-norge-hvis-de-ikke-larer-no/>
- Glaeser, E. L. (2010). *Agglomeration economics*. University of Chicago Press.
- Gram, K. (2022). Likestillingsutfordringer I Deltidsarbeid og Utdanningsnivå. <https://www.ssb.no/befolkning/likestilling/statistikk/indikatorer-for-kjonnlikestilling-i-kommunene/artikler/likestillingsutfordringer-i-deltidsarbeid-og-utdanningsnivalg>
- Hardoy, I., & Schøne, P. (2011). Returns to pre-immigration education for non-western immigrants: why so low? *Education Economics*, 22(1), 48–72.

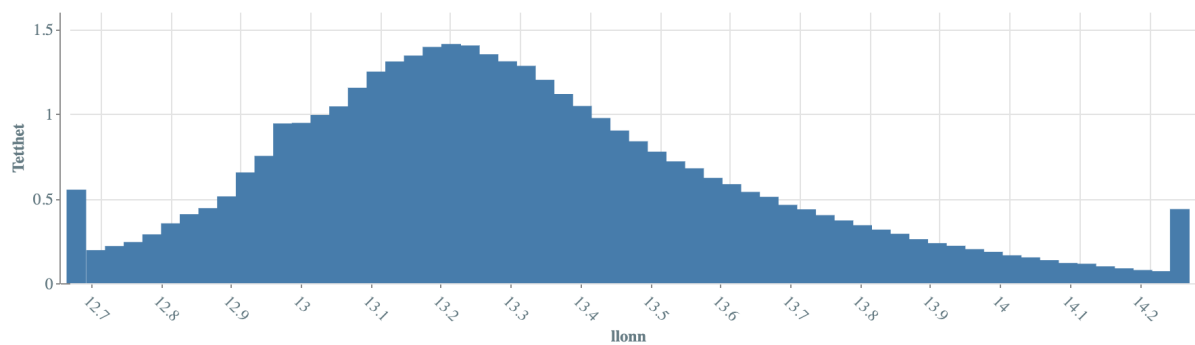
-
- Hoen, M. F., Markussen, S., & Røed, K. (2022). Immigration and economic mobility. *Journal of Population Economics*, 35(4), 1589–1630.
- Høydahl, E. (2023). De 10 Største Byene og Tettstedene i Norge. <https://www.ssb.no/befolkning/folketall/artikler/de-storste-byene-og-tettstedene-i-norge>
- Jann, B. (2008). The Blinder–Oaxaca decomposition for linear regression models. *The Stata Journal*, 8(4), 453–479.
- Johansen, S. (2020). microdata.no–søknadsfri tilgang til registerdata om boforhold. *Tidsskrift for boligforskning*, 3(1), 87–98.
- Kolsrud, D., Røed, M., Schøne, P., & von Simson, K. (2016). Bidrag og belønning: Om innvandrere i det norske arbeidsmarkedet. *Rapport-Institutt for samfunnsforskning*.
- Lehmer, F., & Ludsteck, J. (2011). The immigrant wage gap in Germany: Are East Europeans worse off? *International migration review*, 45(4), 872–906.
- Lightfoot, H. (2009). Standard for næringsgruppering. <https://www.ssb.no/klass/klassifikasjoner/6/koder>
- Manning, A. (2011). Imperfect competition in the labor market. I *Handbook of labor economics* (s. 973–1041). Elsevier.
- Neumark, D. (1988). Employers' discriminatory behavior and the estimation of wage discrimination. *Journal of Human resources*, 279–295.
- Nielsen, H. S., Rosholm, M., Smith, N., & Husted, L. (2004). Qualifications, discrimination, or assimilation? An extended framework for analysing immigrant wage gaps. *Empirical Economics*, 29(4), 855–883.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693–709.
- SSB. (2011). Standard for yrkesklassifisering (STYRK-08). *Notater 17/2011*.
- SSB. (2022). Brukermanual for microdata. <https://www.microdata.no/wp-content/uploads/2022/04/brukermanual-no-1.pdf>
- SSB. (2023). 13880: Innvandrere og norskfødte med innvandrerforeldre, etter kjønn, statistikkvariabel og år. *Statistikkbanken*. [Database]. <https://www.ssb.no/statbank/table/13880/tableViewLayout1/>
- Utdanningsdirektoratet. (2017). Barn, unge og voksne med innvandrerbakgrunn i grunnsopplæringen. *Utdanningsdirektoratet*.
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

A Appendiks

A.1 Ln-transformering av avhengig variabel



Figur A.1.1: Fordelingen til den lineære lønnsvariabelen



Figur A.1.2: Fordelingen til den ln-transformerte lønnsvariabelen

A.2 Storbyer

Storby
Stor-Oslo
4601 - Bergen
5001 - Trondheim
1103 - Stavanger
1108 - Sandnes
3004 - Fredrikstad
3003 - Sarpsborg
3005 - Drammen
3807 - Skien
3806 - Porsgrunn
4204 - Kristiansand
1507 - Ålesund
3803 - Tønsberg

Tabell A.2.1: Kommunene som regnes som storbyer

Stor Oslo
0301 - Oslo
3024 - Bærum
3025 - Asker
3030 - Lillestrøm
3029 - Lørenskog
3027 - Rælingen
3031 - Nittedal
3020 - Nordre Follo
3049 - Lier

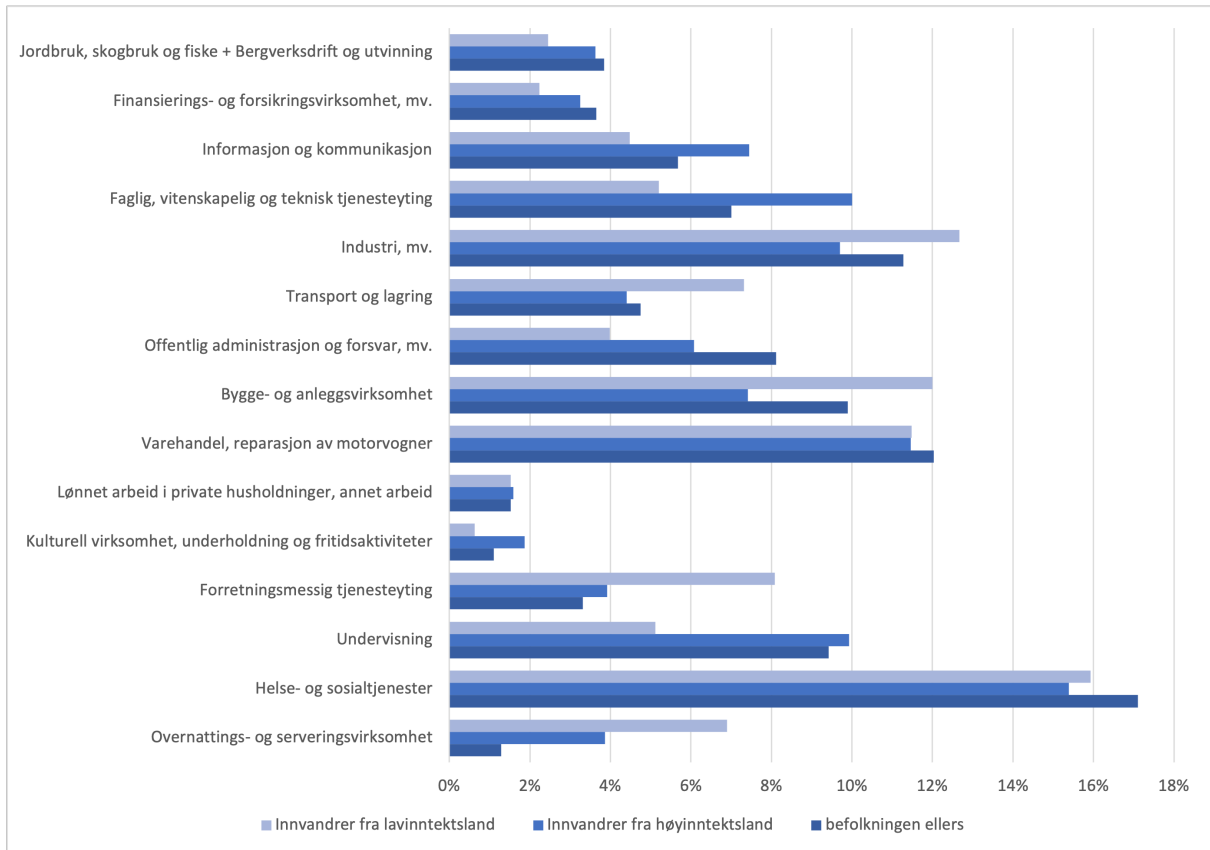
Tabell A.2.2: Kommunene som inngår i Stor-Oslo regionen.

A.3 Næringsgrupperinger

Gruppe	Næring
A+B	Jordbruk, skogbruk og fiske + Bergverksdrift og utvinning
C+D+E	Industri + Elektrisitets-, gass-, damp- og varmtvannsforsyning + Vannforsyning, avløps- og renovasjonsvirksomhet
F	Bygge- og anleggsvirksomhet
G	Varehandel, reparasjon av motorvogner
H	Transport og lagring
I	Overnattings- og serveringsvirksomhet
J	Informasjon og kommunikasjon
K+L	Finansierings- og forsikringsvirksomhet + Omsetning og drift av fast eiendom
M	Faglig, vitenskapelig og teknisk tjenesteyting
N	Forretningsmessig tjenesteyting
O	Offentlig administrasjon og forsvar, og trygdeordninger underlagt offentlig forvaltning
P	Undervisning
Q	Helse- og sosialtjenester
R	Kulturell virksomhet, underholdning og fritidsaktiviteter
S+T+U	Annen tjenesteyting + Lønnet arbeid i private husholdninger + Internasjonale organisasjoner og organer

Tabell A.3.1: Næringsgrupperinger

A.4 Fordeling av innvandrere og befolkningen ellers i de ulike næringene, separert etter landbakgrunn



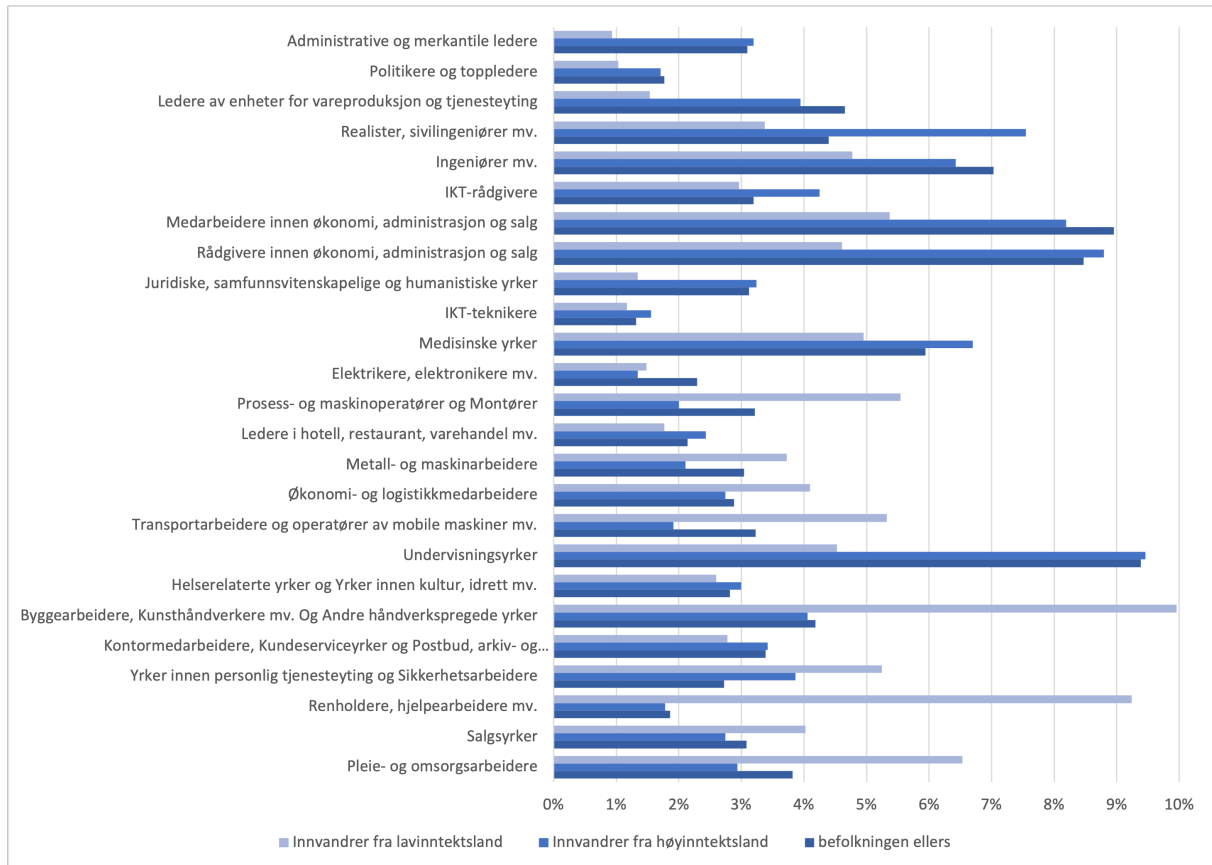
Figur A.4.1: Fordelingen av innvandrere og befolkningen ellers i de ulike næringene, separert etter landbakgrunn

A.5 Yrkesgrupperinger

Gruppe	Yrke
1 - Ledere	
11	Politikere og toppledere
12	Administrative og merkantile ledere
12	Ledere av enheter for vareproduksjon og tjenesteyting
14	Ledere i hotell, restaurant, varehandel mv.
2 - Akademiske yrker	
21	Realister, sivilingeniører mv.
22	Medisinske yrker
23	Undervisningsyrker
24	Rådgivere innen økonomi, administrasjon og salg
25	IKT-rådgivere
26	Juridiske, samfunnsvitenskapelige og humanistiske yrker
3 - Høyskoleyrker	
31	Ingeniører mv.
32+34	Helserelaterte yrker + Yrker innen kultur, idrett mv.
33	Medarbeidere innen økonomi, administrasjon og salg
35	IKT-teknikere
4 - Kontoryrker	
41+42+44	Kontomedarbeidere + Kundeserviceyrker + Postbud, arkiv- og personalkontomedarbeidere
43	Økonomi- og logistikkmedarbeidere
5 - Salgs- og serviceyrker	
51+54	Yrker innen personlig tjenesteyting + Sikkerhetsarbeidere
52	Salgsyrker
53	Pleie- og omsorgsarbeidere
7 - Håndverkere	
71+73+75	Byggearbeidere + Presisjonsarbeidere, kunsthåndverkere, grafiske arbeidere mv. + Andre håndverkspregede yrker
72	Metall- og maskinarbeidere
74	Elektrikere, elektronikere mv.
8 - Prosess- og maskinoperatører, transportarbeidere mv.	
81+82	Prosess- og maskinoperatører + Montører
83	Transportarbeidere og operatører av mobile maskiner mv.
9 - Renholdere, hjelpearbeidere mv.	
91+92+93+94+95+96	Renholdere mv. + Hjelpearbeidere i jordbruk, skogbruk og fiske + Hjelpearbeidere i bergverk, industri, bygg og anlegg og transport + Kjøkkenassistenter + Reklamedistributører mv. + Renovasjons- og gjenvinningsarbeidere mv.
0	Militære yrker og uoppgitt
6	Bønder, fiskere mv.

Tabell A.5.1: Yrkesgrupperinger

A.6 Fordeling av innvandrere og befolkningen ellers i de ulike yrkene, separert etter landbakgrunn



Figur A.6.1: Fordeling av innvandrere og befolkningen ellers i de ulike yrkene, separert etter landbakgrunn

A.7 Resultater fra Oaxaca-Blinder ved alternative dekomponeringer

	Alternativ 1	Alternativ 2
	Hele befolkningen	Hele befolkningen
Forventet lønnsgap	0,109	0,109
<i>Panel A: Forskjeller i egenskaper</i>		
Alder	0,005	0,004
Kjønn	-0,003	-0,003
Storby	-0,007	-0,005
Utdanning	0,000	0,000
Familie	-0,004	0,003
Næring	-0,003	0,002
Yrke	0,031	0,048
Totalt	0,019 (17,9%)	0,05 (45,6%)
<i>Panel B: Forskjeller i avkastning</i>		
Alder	0,144	0,145
Kjønn	-0,005	-0,005
Storby	0,007	0,005
Utdanning	0,024	0,024
Familie	0,01	0,003
Næring	-0,085	-0,091
Yrke	-0,148	-0,164
Konstant	0,142	0,142
Totalt	0,089	0,059

Tabell A.7.1: Oaxaca-Blinder dekomponering med alternativ 1 og 2

A.8 Utvidelser av modellen

Vi presenterer her de utvidelsene vi har med i analysen vår. Vi vil da fokusere på effekten av hvilket land innvandrere kommer fra, samt bakgrunnen for innvandringen. For enkelhets skyld vil vi nå kun fokusere på modellen presentert av Neumark, 1988 som vist ved i likning 7.

A.8.1 Høy- vs lavinntektsland

Når vi skal tolke effektene av å komme fra et høy- eller lavinntektsland estimerer vi to nye simple regresjoner (likning 10 og 11) for å betrakte det rå lønnsgapet, to nye pooled OLS-regresjoner med påfølgende Oaxaca-Blinder dekomponeringer. Variabler med toppskrift H og L representerer innvandrere fra hhv. høy- og lavinntektsland.

$$\ln W_i^H = \alpha \cdot \text{innv}_i^H + u_i \quad (10)$$

$$\ln W_i^L = \alpha \cdot \text{innv}_i^L + u_i \quad (11)$$

$$\ln W_i^H = \alpha \cdot \text{innv}_i^H + X_i^H + u_i^H \quad (12)$$

$$\ln W_i^L = \alpha \cdot \text{innv}_i^L + X_i^L + u_i^L \quad (13)$$

I de to regresjonene estimeres lønnsgapet gjennom en variabel for om en innvandrer kommer fra et høy- (likning 12) eller lavinntektsland (likning 13). I begge regresjonene vil dermed den andre innvandrergruppen være ekskludert fra utvalget. Ut ifra disse regresjonene kan vi dermed lage dekomponeringene for de to innvandringsgruppene.

$$\ln \hat{W}^H - \ln \hat{W}^N = (\bar{X}^N - \bar{X}^H) \hat{\beta}^* + \bar{X}^H (\hat{\beta}^H - \hat{\beta}^*) + \bar{X}^N (\hat{\beta}^* - \hat{\beta}^N) \quad (14)$$

$$\ln \hat{W}^L - \ln \hat{W}^N = (\bar{X}^N - \bar{X}^L) \hat{\beta}^* + \bar{X}^L (\hat{\beta}^L - \hat{\beta}^*) + \bar{X}^N (\hat{\beta}^* - \hat{\beta}^N) \quad (15)$$

Disse dekomponeringene følger samme intuisjon som i kapittel 4.1, og vil gi oss muligheten til å betrakte den forklarte og uforklarte delen av lønnsgapet mellom innvandrere fra høy- og lavinntektsland og befolkningen ellers.

A.8.2 Innvandringsgrunn

Når vi skal estimere effekten av innvandringsgrunn gjør vi det på samme måte som i forrige delkapittel og bytter ut notasjonen. Vi følger inndelingen fra kapittel 3.3.2, der bakgrunnen for innvandring delt inn i to grupper. Disse to gruppene er innvandring på grunn av arbeid eller utdanning (*arb_utd*), samt innvandring på grunn av familie eller flukt (*fam_flu*). Notasjonen blir da henholdsvis *A* og *F*.

$$\ln W_i^A = \alpha \cdot \text{arb_utd}_i + u_i \quad (16)$$

$$\ln W_i^F = \alpha \cdot \text{fam_flu}_i + u_i \quad (17)$$

$$\ln W_i^A = \alpha \cdot \text{arb_utd} + X_i^A + u_i^A \quad (18)$$

$$\ln W_i^F = \alpha \cdot \text{fam_flu} + X_i^F + u_i^F \quad (19)$$

$$\ln \hat{W}^A - \ln \hat{W}^N = (\bar{X}^N - \bar{X}^A) \hat{\beta}^* + \bar{X}^A (\hat{\beta}^A - \hat{\beta}^*) + \bar{X}^N (\hat{\beta}^* - \hat{\beta}^N) \quad (20)$$

$$\ln \hat{W}^F - \ln \hat{W}^N = (\bar{X}^N - \bar{X}^F) \hat{\beta}^* + \bar{X}^F (\hat{\beta}^F - \hat{\beta}^*) + \bar{X}^N (\hat{\beta}^* - \hat{\beta}^N) \quad (21)$$

Vi viser her estimeringen av det initielle lønnsgapet ved likning 16 og 17, de samlede regresjonene til Oaxaca-Blinder estimeringen ved likning 18 og 19 og til slutt Oaxaca-Blinder dekomponeringene i likning 20 og 21.

A.9 Multikollinearitet

	innv	alder	alder2	kvinne	storby	utd_hoy	barn	gift	næring	yrkeT
llohn	-0,129731	0,278394	0,268882	-0,221349	0,095582	0,202221	0,112304	0,147932	-0,170503	0,004003
innv	-	-0,046636	-0,053695	-0,017576	0,132488	-0,001041	-0,057751	0,107609	-0,026881	-0,017678
alder	-0,046636	-	0,995836	0,001846	-0,054947	-0,046541	0,229581	0,304956	0,008062	0,006955
alder2	-0,053695	0,995836	-	0,001422	-0,057388	-0,053274	0,209194	0,293002	0,005768	0,007136
kvinne	-0,017576	0,001846	0,001422	-	0,045438	0,278345	0,069313	0,016204	0,413378	0,012769
storby	0,132488	-0,054947	-0,057388	0,045438	-	0,186244	-0,093797	-0,006503	0,096785	-0,000760
utd_hoy	-0,001041	-0,046541	-0,053274	0,278345	0,186244	-	0,024057	0,065455	0,406856	-0,004753
barn	-0,057751	0,229581	0,209194	0,069313	-0,093797	0,024057	-	0,345327	0,042543	0,005538
gift	0,107609	0,304956	0,293002	0,016204	-0,006503	0,065455	0,345327	-	0,028691	-0,003384
næring	-0,026881	0,008062	0,005768	0,413378	0,096785	0,406856	0,042543	0,028691	-	0,012658

Tabell A.9.1: Korrelasjonsmatrise for alle variablene i modellen

	VIF	1/VIF
innv	1,103402	0,906288
alder	130,28827	0,007675
alder2	128,75433	0,007767
kvinne	1,398133	0,71524
storby	1,118448	0,894096
utd_hoy	1,850606	0,540363
barn	1,238502	0,807427
gift	1,248464	0,800984
næring	2,163940	0,521476
yrke	1,594645	0,661766
Gj.snitt	7,541096	-

	VIF	1/VIF
innv	1,095751	0,912616
alder	1,174832	0,851185
kvinne	1,397151	0,715742
storby	1,11806	0,894406
utd_hoy	1,847739	0,541202
barn	1,188695	0,841259
gift	1,245944	0,802604
næring	2,163893	0,521487
yrke	1,594438	0,661850
Gj.snitt	1,717927	-

Tabell A.9.2: VIF-tester for forklaringsvariablene, med og uten *alder2*

A.10 Regresjonsresultater for landbakgrunn

VARIABLER kategori	lnW høyinnteksland	lnW norsk	lnW lavinnteksland	lnW norsk
alder	0,045065 (0,001474)***	0,040402 (0,000390)***	0,036945 (0,000962)***	0,040668 (0,000392)***
alder2	-0,000426 (0,000018)***	-0,000383 (0,000005)***	-0,000389 (0,000012)***	-0,000386 (0,000005)***
kvinne	-0,117568 (0,002308)***	-0,128472 (0,000682)***	-0,105110 (0,001638)***	-0,126806 (0,000685)***
storby	0,044811 (0,002142)***	0,040167 (0,000577)***	0,030216 (0,001385)***	0,041239 (0,000579)***
utd_høy	0,084696 (0,002865)***	0,110455 (0,000848)***	0,059681 (0,001674)***	0,113989 (0,000851)***
barn	0,018747 (0,002324)***	0,010591 (0,000647)***	0,025457 (0,001453)***	0,011008 (0,000650)***
gift	0,032852 (0,002301)***	0,036618 (0,000629)***	-0,002576 (0,001465)*	0,037219 (0,000631)***
konstant	12,422111	12,668853	12,494672	12,744537
Næringsdummier	Ja	Ja	Ja	Ja
Yrkesdummier	Ja	Ja	Ja	Ja
Observasjoner	71 849	824 639	139 088	824 639
R-squared	0,414123	0,424322	0,388998	0,420016

Standardavvik i parenteser
 *** p<0.01, ** p<0.05, * p<0.1

Tabell A.10.1: Regresjonsresultater for landbakgrunn, separert ⁴

⁴Regresjonene for *norsk* i kolonne 2 og 4 er forskjellige ettersom det er forskjellig gruppering av næringer og yrker mellom *innv.H* og *innv.L*, og vi har kjørt separate regresjoner slik at det er de samme variablene i regresjonene som sammenlignes.

VARIABLER	OLS lnW	OLS lnW	OLS lnW	OLS lnW
innv_H	0,010949 (0,001289)***	-0,019628 (0,001061)***		
innv_L			-0,169662 (0,000947)***	-0,118930 (0,000823)***
konstant	13,342020	12,65381	13,342020	12,65381
Kontrollvariabler	nei	ja	nei	ja
Observasjoner	896 481	896 481	963 733	963 733
R-squared	0,000080	0,422145	0,032202	0,423039

Standardavvik i parenteser
 *** p<0.01, ** p<0.05, * p<0.1

Tabell A.10.2: Regresjonsresultater for landbakgrunn og befolkningen ellers, pooled regresjon, med og uten kontrollvariabler

A.11 Regresjonsresultater for innvandringsgrunn

VARIABLER kategori	lnW arbeid_ utdanning	lnW norsk	lnW familie_ flukt	lnW norsk
alder	0,040328 (0,001594)***	0,040689 (0,000392)***	0,022775 (0,001417)***	0,040760 (0,000393)***
alder2	-0,000409 (0,000020)***	-0,000386 (0,000005)***	-0,000205 (0,000018)***	-0,000387 (0,000005)***
kvinne	-0,106971 (0,002534)***	-0,126912 (0,000685)***	-0,091525 (0,002357)***	-0,127232 (0,000686)***
storby	0,038302 (0,001973)***	0,041137 (0,000579)***	0,019638 (0,002025)***	0,040616 (0,000580)***
utd_høy	0,073719 (0,002467)***	0,113964 (0,000852)***	0,046182 (0,002503)***	0,113816 (0,000854)***
barn	0,049120 (0,002091)***	0,011026 (0,000650)***	0,015037 (0,002300)***	0,011349 (0,000651)***
gift	0,005901 (0,002074)***	0,037223 (0,000631)***	-0,011864 (0,002231)***	0,037311 (0,000632)***
konstant	12,464145	12,750038	12,751358	12,753406
Næringsdummier	Ja	Ja	Ja	Ja
Yrkesdummier	Ja	Ja	Ja	Ja
Observasjoner	73 372	824 639	56 885	824 639
R-squared	0,435557	0,419971	0,373646	0,418353

Standardavvik i parenteser
 *** p<0.01, ** p<0.05, * p<0.1

Tabell A.11.1: Regresjonsresultater for innvandringsgrunn, separat⁵

⁵Regresjonene for *norsk* i kolonne 2 og 4 er forskjellige ettersom det er forskjellig gruppering av næringer og yrker mellom *arb_utt* og *fam_flu*, og vi har kjørt separate regresjoner slik at det er de samme variablene i regresjonene som sammenlignes.

VARIABLER	OLS lnW	OLS lnW	OLS lnW	OLS lnW
arb_utd	-0,116129 (0,001272)***	-0,123986 (0,001130)***		
fam_flu			-0,221880 (0,001419)***	-0,136980 (0,001170)***
konstant	13,342020	12,65381	13,342020	12,65381
Kontrollvariabler	nei	ja	nei	ja
Observasjoner	898 007	898 007	881 529	881 529
R-squared	0,009199	0,419327	0,026997	0,425819

Standardavvik i parenteser
 *** p<0.01, ** p<0.05, * p<0.1

Tabell A.11.2: Regresjonsresultater for innvandringsgrunn, pooled regresjon, med og uten kontrollvariabler

A.12 Gjennomsnittslønn og fordeling blant innvandringsgruppen for arbeid eller utdanning

	Andel
finnv_H	22%
Finnv_L	78%

Tabell A.12.1: Fordeling av landbakgrunn for innvandrere med arbeid eller utdanning som innvandringsgrunn

	gjennomsnittslønn
finnv_H	692 665
Finnv_L	558 237

Tabell A.12.2: Gjennomsnittslønn for innvandrere med arbeid eller utdanning som innvandringsgrunn, separert på landbakgrunn

