

Har innvandrere lavere lønn og avkastning av utdanning enn majoritetsbefolkningen i Skandinavia?

- en analyse av forskjeller mellom innbyggere i Skandinavia

Bacheloroppgave i Politisk Økonomi

Veileder: Bjarne Strøm

Mai 2019

Har innvandrere lavere lønn og avkastning av utdanning enn majoritetsbefolkningen i Skandinavia?

- en analyse av forskjeller mellom innbyggere i Skandinavia

Bacheloroppgave i Politisk Økonomi
Veileder: Bjarne Strøm
Mai 2019

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Har innvandrere lavere lønn og avkastning av
utdanning enn majoritetsbefolkningen i
Skandinavia?

- en analyse av forskjeller mellom innbyggere i Skandinavia

Bacheloroppgave i Samfunnsøkonomi
SØK2901 Vår 2019

Antall ord: 6987

Abstrakt

Flyktningkrisen førte til en masseinnvandring i fra mange land. Den påfølgende innvandringspolitikk har fått mye kritikk for å være mislykket. «Svenske tilstander» har blitt et allment kjent begrep, en situasjon som betegner belastede områder preget av kriminalitet og vold, sammenbrudd i ulike offentlige tjenester og en etnisk og økonomisk segregasjon. For 20 år siden var derimot det mest effektive politiske argumentet en kunne bruke at vi ønsket å unngå «amerikanske tilstander». Er situasjonen virkelig så ille i Sverige, og hva med resten av Skandinavia?

Formålet med denne bacheloroppgaven er å se belyse integreringsprosessen i de tre skandinaviske landene. Har, og i så fall i hvor stor grad, innvandrere kommet dårligere ut enn majoritetsbefolkningen i de respektive landene? Oppgaven ser å hovedsakelig svare på om det er et gap i lønnsnivået og avkastningen på utdanning, samt å kontrollere for mulige effekter som kan påvirke disse to faktorene. Basert på teori og diskusjoner i mediebildet har oppgaven hovedhypotese vært at innvandrere vil komme dårligere ut, både ved lønnsnivå og avkastning på utdanning.

For å angripe problemet i oppgaven og for å finne mulige signifikante effekter ble minste kvadraters metode anvendt, og datasettet PIAAC for de tre landene benyttet. Funnene viser at innvandrere kommer dårligere ut, både ved lønn og avkastning på utdanning, enn det majoritetsbefolkningen i alle de tre Skandinaviske landene gjør. Konklusjonen stemte med antatte resultater og hypotese, og da analysen viser flere signifikante effekter kan mange av funnene konkludere med en signifikant og sterk korrelasjon. Altså stemmer de tre nullhypotesene og beholdes derav i oppgaven.

Innholdsfortegnelse

1. Innledning	5
1.2 Begrepsforklaring.....	5
1.2.1 Innvandrer	6
1.2.2 Avkastning	6
2. Teori	6
2.1 Hypoteser	9
3. Datamateriale	9
3.1 PIAAC.....	9
3.2 Variabler.....	10
3.3 Deskriptiv statistikk for avhengig variabel	11
3.4 Deskriptiv statistikk for kontrollvariabler	14
3.5 Datasettets svakheter	15
3.6 Datasettets styrker.....	16
4. Metode	16
4.1 Minste kvadraters metode (MKM).....	16
4.1.2 Forutsetninger for MKM.....	18
4.2 Mål på godhet - determinasjonskoeffisienten R^2	19
4.3 Funksjonsform	19
4.3.2 Mincer	20
4.4 Signifikantnivå.....	20
4.6 Heteroskedastisitet	21
4.6.1 Breusch-Pagan testen.....	21
5. Regresjonsanalysen	22
5.1 Hypoteser: det forskjell på lønnsnivået til innvandrere i Norge, Sverige og Danmark.....	22
5.2 Hypotese: utdanning, kjønn, arbeidserfaring og ferdigheter påvirker også lønnen i positiv grad.....	24
5.3 Hypotese: avkastning på utdanning er ulik for innvandrere og majoritetsbefolkningen	26
5.4 Teste for heteroskedastisitet.....	30

6. Konklusjon	32
7. Bibliografi	34

1. Innledning

Norge, Sverige og Danmark utgjør Skandinavia, et landområde bestående av tre like land. Det er grunnleggende fellestrekk mellom landene, som sterke kulturelle og historiske likheter. Mye er også ulikt. Herunder ser vi ulikheter i innvandringspolitikken, samt ved politikken gjennomføring og implementering. Forskjellen og omfanget er så stort at det er vel verdt, samt viktig, å studere. Der Sverige har stått for en liberal og åpen innvandringspolitikk, har Danmark stått som en streng motpol med høye krav og mindre innslipp. Norge har plassert seg i midten. Danmark har også distansert seg selv mer fra Norge og spesielt Sverige de siste tiårene. Dette har landet gjort ved å innføre mer restriktiv innvandring og strengere arbeidsmarkedsintegrasjonspolitik. Eksempler på politikkenring er reformene i 2002 da Danmark strammet innvandringspolitikken sin ved å forby familiegjenforening for ektefeller under 24 år og ved å redusere sosialhjelpen for innvandrere (Jakobsen, Korpi, & Lorentzen, 2018).

Da likestilling blant innbyggere og god integrering av innvandrere er en viktig politisk målsetting, har tema vekket stor interesse hos samfunnsforskere og politikere. Interessen har vokst med omfanget av innvandring til de skandinaviske landene. Jeg har i denne oppgaven valgt å fokusere på land i Skandinavia, da de som nevnt er like land, men med nokså ulik innvandringspolitikk. Dette er et grunnlag som en spennende analyse. Den liberale svenske politikken har ført til at Sverige har et langt høyere antall innvandrere i landet, spesielt flyktninger. Dette gjelder både i absolutte tall og i forhold til landets folketall. Har dette derimot gitt bedre resultater enn i de respektive nabolandene, eller har den liberale politikken hatt motsatt effekt?

Oppgavens problemstilling er hvorvidt innvandrere har lavere lønn og avkastning på utdanning enn majoritetsbefolkningen i Skandinavia.

1.2 Begrepsavklaring

Her er målet å unngå misforståelse og mistolkning av begreper brukt forløpende i oppgaven.

1.2.1 Innvandrere

Innvandring er et vidt og omfattende begrep som tilfaller ulike folkegrupper og ulik aktivitet. Dette har ført til ulike definisjoner og typer. Her går hovedskille mellom innvandrere født i utlandet som selv har innvandret til et land, og mennesker født i landet de bor i med to innvandrereforeldre. Denne oppgaven legger til grunn en definisjon om at innvandrere er førstegenerasjonsinnvandrere som er født i annet land enn de bor i nå, og som har minst en forelder født i utlandet. Det skilles ikke mellom hvilke land de opprinnelig kommer i fra, ei heller hvor deres foreldre har opphav.

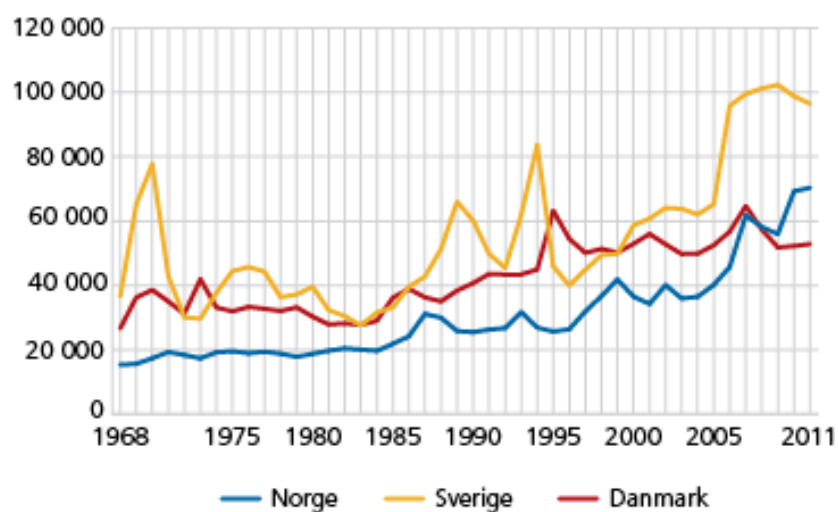
1.2.2 Avkastning

Avkastning brukes i denne oppgaven relatert til utdanning. Begrepsdefinisjonen er at avkastning betyr en lønnsvekst i prosent og derav at «avkastning på utdanning» viser til hvordan utdanning øker lønnen i prosent. Datasettet oppgir timelønn i nominelle verdier, men her vil jeg se på hvordan timelønnen til majoritetsbefolkningen og innvandrere endres prosentvis.

2. Teori

Som nevnt innledningsvis er Norge, Sverige og Danmark like land med like utfordringer. Landene skiller seg derimot på politisk front, med ulike lover og regler. Dette har vært med på å påvirke antall innvandrere i de ulike landene. Her troner Sverige øverst med flest bosatte innvandrere.

Figur 1. Innvandringer til Norge, Sverige og Danmark. 1968-2011.
Antall



Kilde: Befolkningsstatistikk: Statistisk sentralbyrå, Statistiska centralbyrån, Danmarks Statistik.

Det er vanskelig å lage gode analyser som sammenligner innvandring og integrering, da etablering av gode nok datagrunnlag har vist seg krevende. Ulike land definerer innvandring på ulike måter, som gjør sammenligningsgrunnlaget noe vagt. Sverige bruker begrepet innvandrer i sin offentlige statistikk om personer født i utlandet, mens Danmark og Norge også på foreldrenes opprinnelsesland i sin definisjon av en innvandrer (Pettersen & Østby, 2013). Et annet problem er at det ikke noe system for nummerering av landene i verden som vil tillate forskere å enkelt slå sammen varierte datasettene til ett datasett. Problemet oppstår fordi land er ulike, samt at land endrer sine territorielle grunnlag uten å endre navn (Cheibub, Gandhi, & Vreeland, 2010). Dette bidrar til problemer ved sammenligning, men er heldigvis noe enklere her da utvalget av land er begrenset.

Hvordan det går med innvandring henger ofte sammen med forskjeller i dens sammensetning og størrelse. Dette gjelder også for landene i Skandinavia. Samme problemer oppstår når en ønsker å se på effekter av integreringsarbeid, da integreringsfeltet ofte baserer seg på at man kan observere resultatene av et tiltak i etterkant. Dette gjør at det blir vanskelig å bestemme om det var de nye tiltakene som ga resultater, eller om det var andre forhold. Eller om resultatet muligens ville ha vært det samme uten tiltakene. Slike problemer oppstår da en ikke har kontrollgrupper som ikke har fått tiltakene å sammenligne med (Pettersen & Østby, 2013).

SSB har sett nøyere inn på tallene for lønn blant innvandring, spesielt i Norge. En rapport i fra 2017 (Strøm & Bye, 2017) viser at innvandrere og deres husholdninger har lavere inntekter enn hele befolkningen sett under ett. Rapporten konkluderer med at flertallet av innvandrere i Norge befinner seg i nedre del av inntektsfordelingen. Her er det derimot store forskjeller i inntektsnivå ut i fra opprinnelsesland, bakgrunn, botid og innvandringsgrunn. De fant også at innvandrere, på personnivå, har lavere lønn enn lønnstakere i resten av Norge. Igjen er det forskjeller blant innvandrere, som har går på variasjoner mellom yrke og landbakgrunn. Videre funn viser at norskfødte med innvandrerforeldre klarer seg økonomisk bedre enn sine foreldre jevnt over. Tallene for 25-39 år gamle norskfødte med innvandrerforeldre viser at inntektsnivået utgjorde 97 % av inntektsnivået til jevnaldrende i hele befolkningen. Tilsvarende hos innvandrere i samme aldersgruppe utgjorde 79 %, som er noe lavere. Her er det igjen forskjeller blant norskfødte med innvandrerforeldre, der barn med foreldre fra EU-land kommer best ut. Barn med foreldre med bakgrunn fra Asia og Afrika ligger noe under, med unntak av India.

Også i Sverige ser man tendenser til lavere lønn blant innvandrere, enn hos majoritetsbefolkningen. Institutet för arbetsmarknads- och utbildningspolitisk utvärdering (IFAU) er et forskningsinstitutt under Arbeidsdepartementet i Sverige og ser blant annet på tall og data for svensk befolkning. En rapport publisert i 2013 (Katz & Österberg, 2013) sammenligner innvandrere som har kommet til Sverige før fylte 16 år med personer som har svenskfødte foreldre. Hovedkonklusjonen er at innvandrere har lavere lønn enn majoritetsbefolkningen i landet. Rapporten viser også at innvandrere ofte har jobber som krever lavere utdanning enn det innfødte har. Funnet er gjort til tross for at begge gruppene har tatt høyere utdanning i Sverige, der innvandrere har lavere lønn og utbytte av sin eventuelle svenske utdanning enn svenskfødte. Her er det også forskjeller, der funnene hovedsakelig gjelder for mennesker som kommer i fra Asia, Afrika og Latin-Amerika. De største lønnsforskjellene er dog funnet hos innvandrere født i Tyrkia og Sør-Europa der foreldrene var arbeidsinnvandrere.

Oddbjørn Raaum for Frischsenteret (Raaum, 1999) har også sett på avkastningen av utdanningen til innvandrere, men kun i Norge. Han påpeker at innvandrere ikke er en homogen gruppe. Rapporten bringer frem at en rekke studier konkluderer med at vestlige innvandrere har opplevd suksess på arbeidsmarkedet, og en tilnærmet lik suksess som hos majoritetsbefolkningen. Studiene viser derimot motsatte resultat for ikke-vestlige innvandrere,

som generelt har lykket langt dårligere. Denne gruppen får lav avkastning for sin utdanning, der en høyt utdannet ikke-vestlig innvandrere har 20 % lavere inntekt enn en norskfødt person med samme, eller tilsvarende, utdanningsnivå. Dette står i kontrast til en vestlig innvandrere som kun tjener 5 % mindre enn en norskfødt med samme, eller tilsvarende, utdanningsnivå. Kjelsrud (Kjelsrud, 1996) fant videre at hver andre person i arbeid fra et ikke-vestlig land med høy utdanning har en jobb som normalt krever et lavere utdanningsnivå. Dette til motsetning til norskfødte og vestlige arbeidstakere der kun mellom hver fjerde og femte har en jobb som egentlig krever lavere utdanning.

Også i Danmark ser vi at innvandrere jevnt over tjener mindre enn en av dansk opprinnelse. I Danmark tjener en innvandrere fra et vestlig land omtrentlig likt som en danskfødt. Ser man heller på tallene for mennesker i fra ikke-vestlige land, ser man at personene tjener mindre enn danskfødte i alle yrker. Blant ikke-vestlige innvandrere er derimot lønnsforskjellen mellom kjønnene liten (Danmarks Statistikk, 2013).

Tidligere tall viser alle til at innvandrere tjener mindre enn majoritetsbefolkningen i landet de bor i. På grunnlag av dette vil hovedhypotesen være at innvandrere kommer dårligere ut enn majoritetsbefolkningen, både når en ser på lønningsnivå og avkastning på utdanning.

2.1 Hypoteser

De tre hypotesene som skal testes i denne oppgaven er som følger:

H1: Hypoteser: det forskjell på lønnsnivået til innvandrere i Norge, Sverige og Danmark

H2: Hypotese: effekten av å være innvandrere endres når man legger til flere kontrollvariabler

H3: Hypotese: avkastning på utdanning er ulik for innvandrere og majoritetsbefolkningen

3. Datamateriale

3.1 PIAAC

Denne oppgaven baserer seg på data hentet i fra en undersøkelse gjort av OECD fra august 2011 til mars/april 2012. Undersøkelsen er gitt navnet «*The Programme for the International Assessment of Adult Competencies*» (PIAAC) og baserer seg på innhentet materiale fra 24

land. I denne oppgaven er dataen fra de tre Skandinaviske landene hentet ut. Formålet med PIAAC var å se på den voksne befolkningens ferdighetsnivå innen leseferdigheter, tallforståelse og problemløsning i IKT-miljø. Disse tre ferdighetene har blitt omtalt av OECD som nøkkelferdigheter innen informasjonsbearbeiding, og danner grunnlaget for videre ferdighetsutvikling (OECD, 2012). Variablene brukt i denne analysen er alle hentet ut i fra PIAAC.

Innsamlingen er gjort ved bruk personlig intervju og selvutfylling, der sistnevnt ble gjennomført enten på PC eller ved hjelp av papir. Intervjuene ble gjennomført av Statistisk sentralbyrå (SSB), der de enten foregikk hjemme hos respondenten selv, på nøytral grunn eller i SSB sine lokaler. Godtgjørelse for deltakelse gjennom intervju og utfylling var et gavekort på 500 kr (SSB, 2013).

De forskjellige landene har blitt gitt nummerkoder. Her er Norge gitt koden 578, Sverige 752 og Danmark 208. Dataen for disse tre landene inneholder til sammen 28,722 observasjoner, der 5,218 er fra Norge, 8,938 fra Sverige og 14,656 fra Danmark. Det er likevel ikke slik at alle variabler inneholder samme antall observasjoner, fordi ikke alle informantene har svart på alle spørsmål. Kun 18,664 observasjoner i vårt datasett har svart på timelønn.

3.2 Variabler

<i>Tabell 1</i>	
Variabler	Beskrivelse av variabler
Innvandrер	Dummykodet, 1 om individet er innvandrер og 0 om individet er majoritetsbefolkning
Timelønn_PPP	Timelønn oppgitt i \$ og korrigert for PPP
Utdanning	Antall år utdanning
Erfaring	Antall år arbeidserfaring
Kvinne	Dummykodet, 1 om individet er kvinne og 0 om individet er mann
Leseferdigheter	Mål på individets leseferdigheter, standardisert
Tallferdigheter	Mål på individets tallferdigheter, standardisert

Her vil *Innvandrer* tjene som den sentrale forklaringsvariabelen. *Innvandrer* er variabelen jeg tar utgangspunkt i for analysen der jeg skal se på innvandrere, der den er dummykodet slik at *Innvandrer* = 1 hvis personen er førstegangsinnvandrer og 0 hvis noe annet. Annet vil her vise til det som kalles majoritetsbefolkningen eller etnisk skandinaviske.

Da denne oppgaven skal se på ulike lønns- og utdanningsnivå i Skandinavia, vil jeg se på lønnsvariabelen *Timelønn_PPP* og utdanningsvariabelen *Utdanning*. *Timelønn_PPP* viser timelønn i dollar, korrigert for kjøpekraftsparitet (PPP). Denne fungerer her som avhengig variabel (Y). Variabelen *Utdanning* viser utdanningsnivå og har verdier i fra 6 år til 21 år med utdanning. Denne gir derimot ingen informasjon om hvilken type utdanning dette er.

Datasettet inneholder to ulike variabler som lese-, tall- og problemløsningsferdigheter.

Variabelene jeg her har valgt å benytte meg av er de de «råe» testscorene standardisert til variable med gjennomsnitt 0 og standardavvik 1 for hvert land. Dette har betydning for senere tolkning av regresjonskoeffisienter og jeg har valgt de standardiserte testscorene fordi de er lettere å forstå og gir en sammenlignbar benevningsfri tolkning av koeffisienter på tvers av måleskalaer og studier.

3.3 Deskriptiv statistikk for avhengig variabel

Den deskriptive statistikken viser verdiene til de ulike variablene, som gjennomsnitt og minimum- og maksimumsverdier for alle. Tabellen viser også standardavviket til variablene, som beskriver det gjennomsnittlige avviket fra gjennomsnittet. For å unngå at summen av avvikene blir lik null, kvadreres gjennomsnittet.

Standardavvik =

$$\sqrt{\frac{\sum x_i^2}{n}}$$

Tabell 1: Lønn for innvandrere og majoritetsbefolkningen

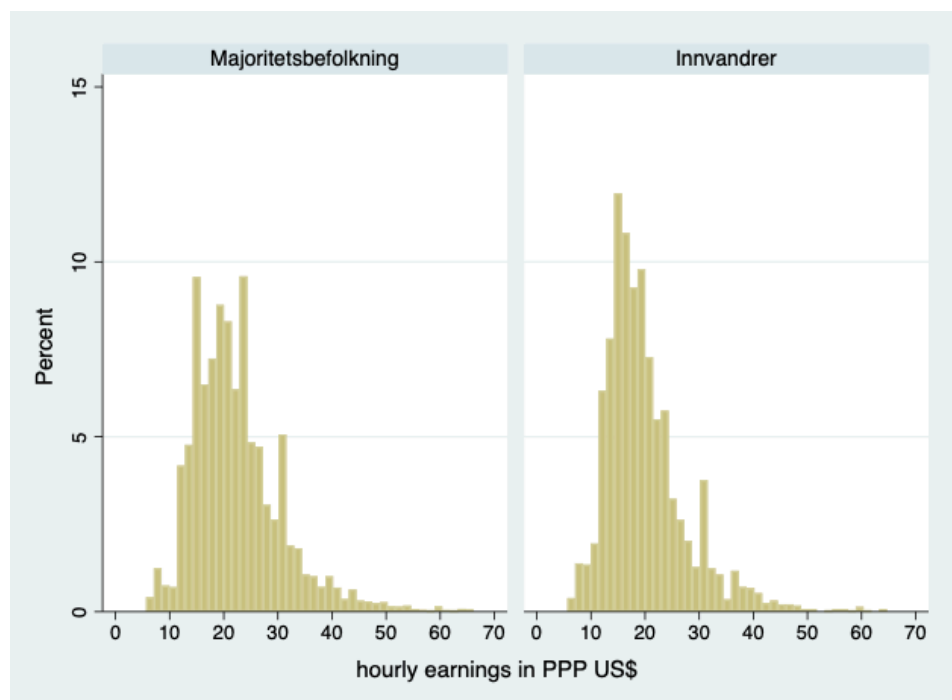
Skandinavia

Tabell 2

	Alle	Innvandrere	Majoritetsbefolkning
Gjennomsnittlig timelønn	22.03276	20.0059	22.40318
Minimum	5.632074	5.99309	5.632074
Maksimum	66.13428	63.83549	66.13428
Standardavvik	8.310652	7.677649	8.369716
Antall observasjoner	18,664	2818	15 804

Ser her at innvandrere har lavere lønn enn majoritetsbefolkningen, både når vi ser på gjennomsnitt og maksimumslønn. Innvandrere har derimot en høyere minstelønn. Majoritetsbefolkningen har et høyere standardavvik.

Lønnsfordelingen i Skandinavia kan også illustreres visuelt ved histogram:



Her ser vi at innvandrere jevnt over tjener mindre enn majoritetsbefolkningen, da den største prosentandelen av gruppen befinner seg på en 15-20 \$ per time. Hos majoritetsbefolkningen er lønnsspredningen større og en større prosentandel av befolkningen befinner seg over 20 \$ per time.

Videre kan vi se på tallene for hvert land isolert.

Norge (578)

Tabell 3

	Alle	Innvandrere	Majoritetsbefolkning
Gjennomsnittlig timelønn	24.1383	21,37751	24,49775
Minimum	5,632074	6,853756	5,632074
Maksimum	64,50208	60,65311	64,50208
Standardavvik	8.87392	8,341963	8,877403
Antall observasjoner	3558	414	3140

Sverige (752)

Tabell 4

	Alle	Innvandrere	Majoritetsbefolkning
Gjennomsnittlig timelønn	18.37923	17.35101	18.54121
Minimum	11.5	11.5	11.5
Maksimum	30.5	30.5	30.5
Standardavvik	5.371235	5.452753	5.339972
Antall observasjoner	5904	796	5086

Danmark (208)

Tabell 5

	Alle	Innvandrere	Majoritetsbefolkning
Gjennomsnittlig timelønn	23.56273	20.96701	24.12726
Minimum	5.812389	5.99309	5.812389
Maksimum	66.13428	63.83549	66.13428
Standardavvik	8.8612	8.113982	8.917205
Antall observasjoner	9202	1608	7578

Her blir antall observasjoner til sammen noe høyere enn hvis man ser på innvandrere og majoritetsbefolkningen hver for seg, men dette er fordi det oppstår noen «missing values» ved de to gruppene.

Ser av tabellene over at innvandrere har lavere gjennomsnittslønn i alle de tre landene. Med unntak av Sverige er også maksimumslønnen for innvandrere lavere enn for majoritetsbefolkningen, mens det er motsatt for minimumslønnen. Grunnen til at maks- og minlønnen er annerledes i Sverige er fordi dataen for landet er oppgitt i deciler, fremfor i kontinuerlig lønn. Det betyr at de 10% av observasjonene med høyest lønn er decil 10 og de 10% av observasjonene med lavest lønn er decil 1. Dette vil derimot ikke ha store utslag i regresjonen. Vi kan likevel ut i fra den deskriptive statistikken se med en gang at innvandrere jevnt over har lavere gjennomsnittslønn enn majoritetsbefolkningen i de tre landene.

Ser at det er store forskjeller på ekstremverdiene til timelønnen. Ser av tabellene at den laveste observerte timelønnen er på $\approx 5,6$ \$ timen og den høyeste ≈ 66 \$ timen.

Ekstremverdiene er beholdt i datasettet grunnet målet om å få en så virkelighetsnær analyse som mulig. Grunnet dette vil det oppstå enkeltunntak der enkelte individer kan ha en svært lav eller en svært høy timelønn.

3.4 Deskriptiv statistikk for kontrollvariabler

Presenterer her en tabell som viser deskriptiv for variablene samlet, samt tabeller som viser den deskriptive statistikken delt opp etter innvandrere og majoritetsbefolkning for å se på.

Dette er gjort ved å sortere verdiene til den dikotome etnisitetsvariabelen hver for seg.

Formålet er å kunne se forskjell på verdiene til de to folkegruppene.

Alle:

Tabell 6

Variabler	Obs	Gj.snitt	Std. Avvik	Min	Maks
Utdanning	28,461	12.87302	2.774096	6	21
Kvinne	28,722	0.4989555	0.5000076	0	1
Erfaring	27,326	21.34279	14.02384	0	47
Leseferdigheter	28,457	1.11e-07	0.9999297	-4.986001	2.818439
Tallferdigheter	28,457	-1.72e-07	0.9999297	-5.268829	3.264295
Problemløsningsferdighet	24,464	-7.91e-08	0.9999183	-5.512477	3.392345

Majoritetsbefolkningen:

Tabell 7

Variabler	Obs	Gj.snitt	Std. Avvik	Min	Maks
Utdanning	23,505	12.911	2.641227	6	21
Kvinne	23,506	0.4936187	0.4999699	0	1
Erfaring	22,738	22.36595	14.19256	0	47
Leseferdigheter	23,506	0.1574792	0.8473358	-4.986001	2.818439
Tallferdigheter	23,506	0.0671275	0.8585956	-4.076865	2.818439
Problemløsningsferdighet	21,230	0.0671275	0.965262	-4.28271	3.392345

Innvandrere:

Tabell 8

Variabler	Obs	Gj.snitt	Std. Avvik	Min	Maks
Utdanning	4,877	12.71417	3.325565	6	21
Kvinne	4,885	0.5295803	0.4991753	0	1
Erfaring	4,518	16.14166	11.85016	0	47
Leseferdigheter	4,885	-0.748014	1.290447	-4.847544	-2.603536
Tallferdigheter	4,885	0.7343977	1.265985	-5.268829	3.264295
Problemløsningsferdighet	3,188	-0.438101	1.104664	-5.512477	3.013169

3.5 Datasettets svakheter

Det er ulike svakheter ved PIAAC, som omfanget undersøkelsen fanger opp. Vi ser på data med 28,391 observasjoner fordelt på tre land. Når vi derimot ønsker å se på innvandrere i de tre skandinaviske landene, sitter vi igjen med 4885 observasjoner. Dette er et lite omfang, sett i lys av hvor mange innvandrere det er i de tre landene. Grunnet det lave antallet observasjoner vil det være en viss usikkerhet rundt konklusjonen i denne oppgaven.

Variabelen definerer også en innvandrer som «deltaker født i utlandet; minst en forelder født i utlandet». Denne definisjonen er noe ulik den brukt i offentlig statistikk om innvandrere i Skandinavia, der de tre landene igjen bruker ulike definisjoner.

En annen svakhet med datasettet, som er svært relevant for denne analysen, er at variabelen om innvandrere inneholder lite informasjon. Det er ikke mulig, gjennom benyttet datasett, å skille mellom innvandrere i fra ulike land. Variabelen for etnisitet ser på innvandrere som en stor homogen gruppe. For en bedre analyse hadde det vært hensiktsmessig å se på forskjellen mellom vestlige- og ikke-vestlige innvandrere. Dette er ikke mulig med PIAAC. Annen litteratur henvist til i denne oppgaven har derimot sett nøyere på dette skillet. Det er heller ikke mulig her å se på om innvandrerne som inngår i datasettet har tatt eventuell utdanning i Skandinavia eller i hjemlandet sitt. Denne forskjellen ville skapt en bedre og grundigere analyse som kunne gitt flere svar på utdanningsavkastning.

Det samme problemet kan overføres til andre variabler, herunder utdanning. Datasettet gir ingen svar på hvilket utdanningsnivå en informant innehar eller en mulighet til å skille informantene i fra hverandre. Her vil en person med bachelor og en person med tre ulike årsstudium komme likt ut, til tross for ulikt utdanningsnivå. Det samme gjelder for studieretning, da PIAAC ikke gir informasjon om hvilken grad informanten har. Dette kunne vært nyttig for å se på akkurat hvilke studieretninger som gir best avkastning. Variabelens svakhet kommer så videre frem når en ønsker å se på prestasjonsnivå. Datasettet gir ingen informasjon om hvor godt individene gjorde det i utdanningen, som kunne vært gunstig for å skille gode og mindre gode studenter i fra hverandre.

3.6 Datasettets styrker

Datasettet har også enkelte styrker, som at det inneholder variabler jeg trenger for å gjennomføre denne analysen. Det er data fra alle de tre skandinaviske landene, som er avgjørende for et sammenligningsgrunnlag. For analysen er det nødvendig at variabler som inntekt, utdanningsnivå, etnisitet og følger med.

PIAAC inneholder også en variabel som viser inntekt, samt en variabel som viser inntekt korrigert for kjøpekraftsparitet. Kjøpekraftsparitet, på engelsk purchasing power parities (PPP), er en omregning som gjør det lettere å sammenligne velstanden mellom ulike land fordi man jevner ut for forskjeller i pris og valuta. Jeg mener at denne variabelen, *Lønn*, derfor er den gunstigste, da jeg ønsker å sammenligne forskjeller mellom ulike land i Skandinavia.

Datasettet åpner også for å se på ulike grupper i samfunnet, da vi kan skille mellom kjønn, alder og om personene er etnisk skandinaviske eller ikke.

4. Metode

I dette kapittelet skal jeg gå gjennom viktige metoder som blir benyttet i oppgaven. Teorien om kapittelets statistiske metoder er hentet i fra boken til P Christian Hagen ved navn «*Innføring i sannsynlighetsregning og statistikk*» (Hagen, 2016) og fra boken til R. Leighton Thomas ved navn «*Using statistics in economics*» (Thomas, 2005).

4.1 Minste kvadraters metode (MKM)

Oppgavens analysedel benytter minste kvadraters metode (MKM). Dette er en anerkjent og mye benyttet estimeringsmetode brukt til behandling av data- og observasjonsmateriale. Målet er å finne en lineær modell som forklarer forholdet mellom en endogen variabel og en eller flere eksogene variable. Denne lineære modellen kalles for en regresjonslinje og beskriver hvordan en respondentvariabel, y , endres når en forklaringsvariable, x , skifter verdi. En rett linje kan skrives som:

$$y = \alpha + \beta x \quad (1)$$

der α sier hvor linjen krysser y-aksen og β er en konstant og linjens stigningstall. En økning i x på 1 vil da øke y med β . Linjen beskriver en modellmessig sammenheng mellom forklarings- og responsvariabelen. Hensikten med regresjon er da å finne et best mulig estimat til denne ukjente lineære linjen, som gjøres ved å estimere de ukjente størrelsene α og β på bakgrunn av innsamlet data. Den beste gjetningen på den ukjente rette linjen kalles for regresjonslinjen:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \quad (2)$$

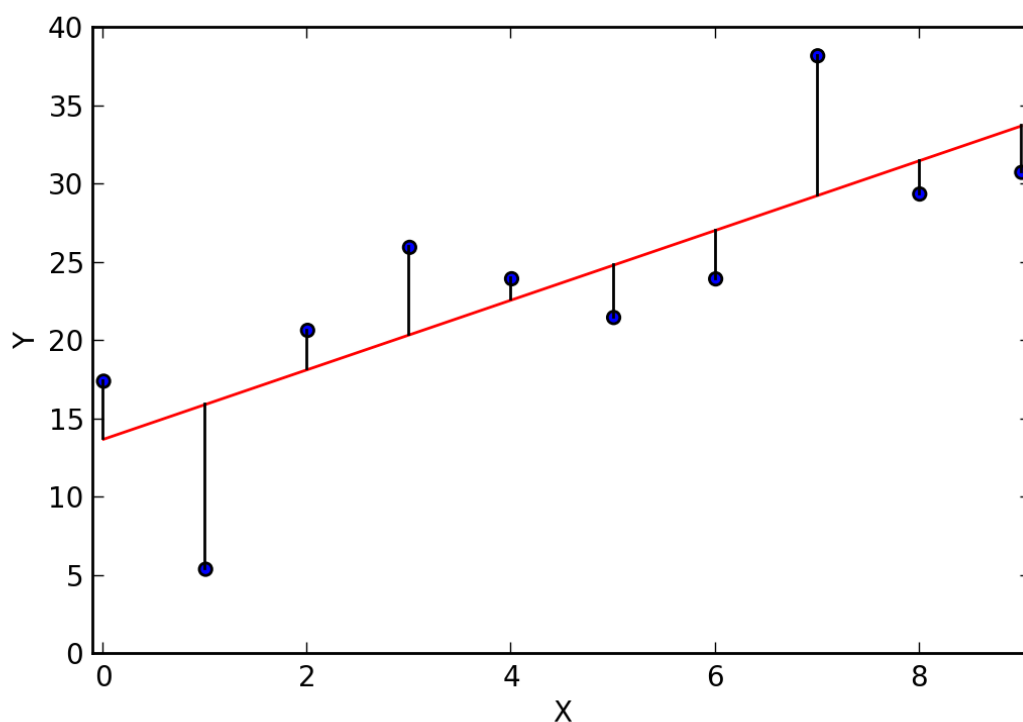
Generell regresjonsligning kan se slik ut:

$$Y_i = \alpha + \sum_{i=1}^n \beta_i X_i + \varepsilon \quad (3)$$

I ligningen over representerer Y utfallsvariabelen, α konstantleddet, X_i en eller flere forklaringsvariabler og β_i effekten av X_i på Y . Videre er ε restleddet, også kalt støyleddet, som her brukes til å fange opp andre effekter som påvirker Y . Slike effekter er uforklarte, samt ofte ikke-målbare faktorene som spiller inn på den forventede verdien til populasjonen. Restleddet kan være både positivt og negativt.

MKM tar utgangspunkt i en vilkårlig linje, der man for hvert punkt tar avstanden mellom punktet og linjen og kvadrerer den avstanden. Da finner man avvikskvadratet. Så finner man summen K av arealet til disse avvikskvadratene, der MKM finner linjen som gir minst mulig kvadratsum K . Det finner ingen alternativ som gjør denne summen mindre enn 0.110.

Figur (1)



Figur (1) viser en grafisk fremstilling av regresjonslinjen og tilhørende restledd

4.1.2 Forutsetninger for MKM

Restleddet er forskjellen mellom observert verdi og predikert verdi, der residual = observert y - predikert y . Små residualer tolkes dithen at variablene forklarer mye av endringen i den endogene variabelen. Residualene representerer mulige feil ved modellen. Ved bruk av MKM må ulike forutsetninger oppfylles. Disse tar utgangspunkt i restleddet, ε .

1. $E(\varepsilon_i) = 0$, Restleddene har forventning lik null
2. $\text{Var}(\varepsilon_i) = \sigma^2 < \infty$, Variansen til restleddene er konstant, samt uendelig for alle verdier.
3. $\text{Cov}(\varepsilon_i), (\varepsilon_j) = 0$, Restleddene er uavhengige av hverandre.
4. $\varepsilon_i \sim N(0, \sigma^2)$, Restleddet er normalfordel
5. Det er ikke sammenheng mellom restleddene og tilhørende x -verdi

Blir ikke forutsetningene oppfylt kan man få feil resultat ved modellen og den blir da ikke best mulig.

4.2 Mål på godhet - determinasjonskoeffisienten R²

R² er et mål som viser hvor passende eller god modellen vår er i forhold til dataen og hvor mye variablene forklarer variasjonen i modellen. R² forklarer andelen variasjon i y som forklares av x. I utgangspunktet passer modellen godt om residualene er små, da avvikene i fra modellen til hver observasjon vil være liten. R² forteller da videre hvor mye forklaringsvariablene forklarer denne endringen i modellen. Matematisk kan vi uttrykke determinasjonskoeffisienten slik:

$$R^2 = 1 - \frac{SSR}{SST} \quad (4)$$

I ligningen over representerer SSR summen av de kvadrerte residualene. SST representerer summen av de kvadrerte avvikene i fra gjennomsnittet. SSR/SST tolkes derav som andelen av variasjonen restleddet kan forklare. Den resterende andelen $1 - SSR/SST$ viser variasjonen som forklares av modellen. Determinasjonskoeffisienten R² ligger mellom 0 og 1, der 1 betyr at 100 prosent av variasjonen i modellen forklares ved de eksogene forklaringsvariablene og det representerer en god modell.

4.3 Funksjonsform

Denne oppgaven benytter seg av: $\text{Timelønn_PPP} = f(\text{Innvandrer}, x)$ der Innvandrer representerer interessevariabelen innvandrer og tar verdien 1 hvis individet er fra et annet land og 0 hvis personen er skandinaver. Her er x videre et sett av mulige kontrollvariabler som kan være med på å påvirke resultatet, som utdanning.

I denne oppgaven vil jeg benytte meg av en log-lineær modell. En generell log-lineære modellen kan uttrykkes slik:

$$\ln Y = \alpha + \sum_{i=1}^n \beta_i X_i + \varepsilon \quad (5)$$

Her er venstresiden er logaritmisk og høyresiden er lineær. Her vil en enhets økning i X_i øke $\ln Y$ med β_i . Hva som derimot blir effekten på Y finner man ved å finne verdien på e^{β_i} . Ved

små verdier av β_i kan man gjennom en approksimativ tilnærming si at $e_i^\beta \approx 1 + \beta_i$.

Multipliserer man da β_i med 100 finner man en omtrentlig prosentvis vekst i Y .

Oppgaven benytter seg av en log-lin funksjon fordi det er en vanlig funksjonsform i litteratur og lignende tidligere studier. Formen gjør at man kan sammenligne estimerte lønnsgap på tvers av land og forskjellige studier. Dette er mulig siden estimert lønnsforskjell kan tolkes som et tilnærmet prosentvis lønnsgap og uavhengig av målenivå på den avhengige variabelen.

4.3.2 Mincer

Jacob Mincer, som var en amerikansk samfunnsøkonom, utledet i 1974 Mincer-Modellen (Mincer, 1974). Modellen har siden da blitt svært innflytelsesrik i økonometrien og i forskningen på humankapital, samt at den også er på log-lin form. Mincers modell relaterer lønn til befolkningens utdanningsnivå og arbeidserfaring. Modellen gir også mulighet for å legge til andre variabler, som etnisitet og kjønn. Denne lønnsfunksjonen er derfor relevant for denne oppgaven og vil bli brukt som inspirasjon til valget av en log-lin funksjonsform.

Mincerfunksjonen blir uttrykt slikt:

$$\ln Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i \quad (6)$$

I ligningen over representerer notasjonen Y_i timelønnen, S_i antall år utdanning, X_i antall år med arbeidserfaring, X_i^2 antall år med arbeidserfaring kvadrert for å få med avtakende økning og ε_i er restleddet.

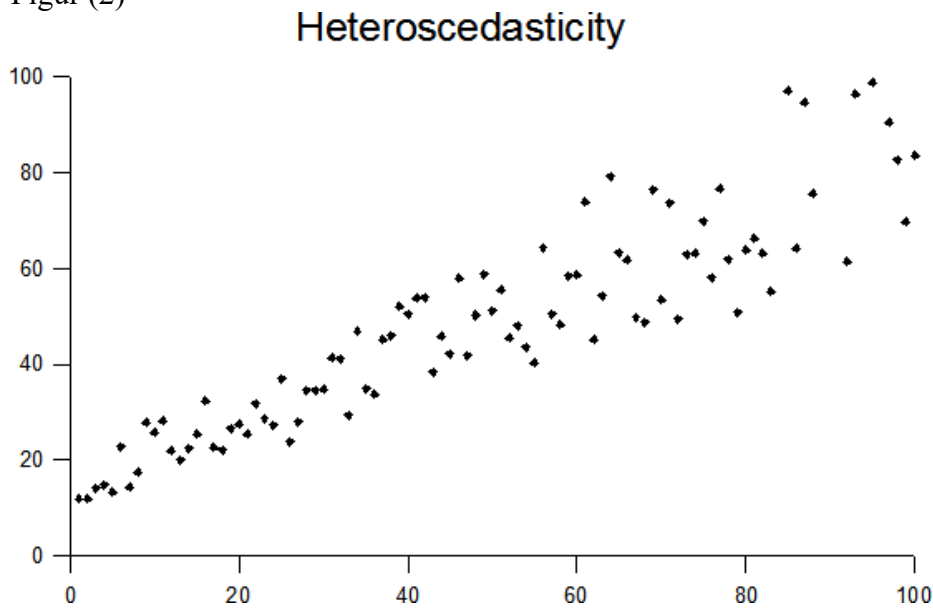
4.4 Signifikantnivå

Videre har man med et signifikantnivå, α , som er et mål på hvor sannsynlig det er for at dataene man har skyldes tilfeldigheter. Nivået settes ut fra hvor stor sjanse en ønsker å ta for feilaktig å forkaste en gyldig nullhypotese. Her settes signifikantnivået til $\alpha = 5\%$ for alle hypotesetestene. $\alpha = 5\%$ betyr derfor at vi her godtar en sannsynlighet på 5% for at vi kan feilaktig forkaste H_0 når H_0 stemmer. P-verdier er relatert til signifikantnivået og sier noe om hvor stor sannsynlighet det er for at H_0 er feil. Da vi her har satt $\alpha = 5\%$ må vi godta at H_0 er sann om $p > 0.05$. Er derimot $p < 0.05$ må H_0 forkastes og H_A godtas.

4.6 Heteroskedastisitet

Ved brudd på forutsetning 2. for MKM der det ikke er konstant variasjon, oppstår heteroskedastisitet. Dette er til motsetning til homoskedastisitet som er ønskelig, der variansen er konstant. Enda en forutsetning for MKM vil derfor være et fravær av heteroskedastisitet.

Figur (2)



Grafen over viser heteroskedastisitet hvor man ser at variansen øker systematisk med variabelen.

4.6.1 Breusch-Pagan testen

Heteroskedastisitet kan testes. Forkastes H_0 har man heteroskedastisitet, så med andre ord vil man ved testing av heteroskedastisitet ha at nullhypotesen er homoskedastisitet. Det finnes flere måter å teste en modell for heteroskedastisitet, som Breusch-Pagan- og Lagrange-multiplikator testen. Her benyttes Breusch-Pagan testen, da den kan utføres i Stata. Testen antar at variansen til restleddene er en funksjon av et lineært forhold mellom W_i , som representerer variabler som kan innvirke på $V(\varepsilon_i)$. Disse variablene kan inneholde variablene i fra regresjonslinja.

$$V(\varepsilon_i) = f(a_1 + a_2W_2 + a_3W_3 + a_4W_4 + \dots + a_mW_m) \quad (8)$$

der a_i er konstanter.

Breusch-Pagan testen finner predikerte verdier (\hat{Y}) og restleddene (ε). Restleddene blir kvadrert og omgjort slik at gjennomsnittet blir lik 1. Videre blir de kvadrerte restleddene regressert for \hat{Y} . Om nullhypotesen er sann finnes som forkart ingen heteroskedastisitet og testen har en kji-kvadratfordeling med én frihetsgrad (Williams, 2015). Testen vil bli benyttet til å teste modellene for heteroskedastisitet i kapittel 5.

5. Regresjonsanalysen

Ser av den deskriptive statistikken i kapittel 3.3 at innvandrere har lavere lønn enn majoritetsbefolkninga i alle de tre skandinaviske landene. Da dette kan leses rett ut av tabellene velger jeg å ikke se næyere inn på dette ved bruk av regresjon.

5.1 Hypoteser: det forskjell på lønnsnivået til innvandrere i Norge, Sverige og Danmark

Ønsker å teste om lønnsnivået er ulikt mellom innvandrere i de tre ulike skandinaviske landene. Setter opp en modell:

$$\ln W = \beta_0 + \beta_1 \text{Nor} + \beta_2 \text{Swe} + \beta_3 \text{INNV} + \beta_4 \text{NOR} * \text{inn} + \beta_5 \text{SWE} * \text{inn} + \varepsilon \quad (9)$$

der $\ln W$ er logaritmen til timelønn, NOR en dummyvariabel for Norge, SWE en dummyvariabel for Sverige, INNV er en dummyvariabel for etnisitet, NOR*inn en variabel for et samspilledd mellom Norge og innvandrere, SWE*INNV et samspilledd mellom Sverige og innvandrere og ε restleddet. En generell metode for å teste om det er effektforskjeller i dataen på er å gjennomføre en regresjonsanalyse med et samspilledd, også kalt interaksjonsledd. Derfor opprettes det her samspillvariabler mellom innvandrere og land for å se på effekten av det å være innvandrere på lønnen i de ulike skandinaviske landene. Danmark brukes her som referansekategori.

VARIABLER	(1) lnTimelønn PPP
Innvandrere	-0.140***

	(-14.74)
NOR	0.0223***
	(3.04)
SWE	-0.233***
	(-37.17)
NOR*inn	-0.00227
	(-0.11)
SWE*inn	0.0689***
	(4.25)
Konstantledd	3.114***
	(785.27)
Observasjoner	18,622
R-squared	0.099

t verdier i parentes

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

Ser først at alle funnene, foruten om samspilletleddet i mellom Norge og innvandrere, er signifikante funn og kan derfor generaliseres til populasjonen. R^2 er 0,099, som tilsvarer en forklaringsgrad på 9,9 %.

NOR viser at befolkningen i Norge vil ha 2,2% høyere lønn enn i Danmark og SWE viser at befolkningen i Sverige vil ha 23% lavere lønn enn i Danmark.

Ser videre at det å være innvandrere i Danmark påvirker lønnen negativt, fremfor å være majoritetsbefolkningen. Dette stemmer overens med funnene gjort ved den deskriptive statistikken og som nevnt innledningsvis i kapitlet. En innvandrere i Danmark vil ha 14% lavere forventet lønn enn majoritetsbefolkningen. I Norge vil innvandrere også ha tilnærmet 14% lavere forventet lønn, da effekten av å være innvandrere i Norge er liten. I Sverige vil effekten på forventet lønn av å være innvandrere være 7,11 %.

Kan predikere lønnen for å se på forskjellen mellom innvandrere landene, med utgangspunkt i modellen. Ettersom summen av restleddet alltid vil være null ved MKM-regresjonsanalyse kan vi se bort ifra dette leddet i prediksjonen:

$$\text{Norge: } \ln W = 3,114338 + 0,0223046 + (-0,1397137) + (-0,0022698) = \underline{2,9946591}$$

$$\text{Sverige: } \ln W = 3,114338 + (0,232621) + (-0,1397137) + 0,0688854 = \underline{2,8108887}$$

$$\text{Danmark: } \ln W = 3,114338 + (-0,1397137) = \underline{2,9746243}$$

Da vi ser på logaritmen til lønn må vi opphøye de predikerte verdiene i e:

$$NOR: e^{2,9946591} = \underline{19,99}$$

$$SWE: e^{2,8108887} = \underline{16,62}$$

$$DKK: e^{2,9746243} = \underline{19,60}$$

Ser av de predikerte tallene at innvandrere i Sverige kommer dårlig ut, samt at det er et lønnsgap i mellom de tre Skandinaviske landene. Ser også i fra modellen over at dummyvariabelen for Sverige er negativ, som viser til en negativ effekt av å bo i Sverige fremfor de to andre Skandinaviske landene. Gapet betyr at det er forskjeller i lønnsnivået til innvandrere, det mest gunstige er å bo i Norge, så Danmark og til slutt Sverige.

Dette kan tyde på at den svenske liberale innvandringspolitikken fører til lavere lønn i Sverige da landets mål å ta inn mange vil skape et press på arbeidsmarkedet, der innvandrere i fra ulike land vil måtte konkurrere om arbeidsplassene i høyere grad. Basert på økonomisk teori vil dette bidra til å presse lønnsnivået deres ned, og de vil derav muligens oppleve en lavere utdanningsavkastning også.

5.2 Hypotese: utdanning, kjønn, arbeidserfaring og ferdigheter påvirker også lønnen i positiv grad

Ønsker her å teste om effekten av å være innvandrer på lønn endres om man legger til flere kontrollvariabler i modellen og om kontrollvariablene påvirker lønnen positivt. Utvider derfor modellen her til å også inneholde variabler for utdanning, kjønn, arbeidserfaring og de tre ulike ferdighetsnivåene.

$$\ln W = \beta_0 + \beta_1 D_{NOR} + \beta_2 D_{SWE} + \beta_3 innv + \beta_4 NOR * innv + \beta_5 SWE * innv + \beta_6 utd + \beta_7 kj\ddot{o}nn + \beta_8 Arb + \beta_9 Lese + \beta_{10} tall + \beta_{11} IKT + \varepsilon \quad (10)$$

Notasjon er som tidligere, men her med tillegg av kontrollvariablene. Kontrollvariabelene er denotert med utd for utdanning, kj\ddot{o}nn for kjønn, arb for arbeidserfaring, lese for leseferdigheter, tall for tallferdigheter og IKT for problemløsning i et IKT-miljø. Estimerer modellen og får:

VARIABLER	(1) lnTimelønnPPP	(2) lnTimelønnPPP
Innvandrer	-0.140*** (-14.74)	-0.0310*** (-3.41)
NOR	0.0223*** (3.04)	0.00459 (0.73)
SWE	-0.233*** (-37.17)	-0.175*** (-33.67)
NOR*innv	-0.00227 (-0.11)	-0.0239 (-1.29)
SWE*innv	0.0689*** (4.25)	0.00769 (0.52)
Utdanning		0.0537*** (55.94)
Kvinne		-0.0922*** (-21.00)
Erfaring		0.0101*** (54.19)
Leseferdighet		0.0155*** (2.79)
Tallferdighet		0.0387*** (7.20)
Problemløsnings- ferdighet		0.0119*** (2.95)
Konstantledd	3.114*** (0.00397)	2.193*** (0.0131)
Observasjoner	18,622	16,839
R-squared	0.099	0.428

t verdier i parentes
*** p<0.01, ** p<0.05, * p<0.1

Forklaringsgraden har økt i fra 9.9 % til 42%, som tilsier at modell 2 her forklarer langt mer av endringen i timelønn enn modell 1 uten kontrollvariabler. Antall observasjoner har blitt redusert med 1,783, grunnet at ikke alle variabelen inneholder tilstrekkelig med informasjon og mangler verdier.

Effekten av å være innvandrer i Danmark på lønnen er mindre etter et tillegg av flere kontrollvariabler, men den har fortsatt en negativ påvirkning. Påvirkningen av å være innvandrer i Danmark på forventet lønn har derimot blitt redusert til 3,1 %, fra tidligere 14 % i modell (1). I Norge er samme effekt redusert til 5,4 % og i Sverige er den nå 2,3%. Ved å

utvide modellen med flere kontrollvariabler forsvinner derimot noen signifikante effekter. Her er ikke lengre dummyvariabelen for Norge signifikant og heller ikke samspillvariabel mellom Sverige og innvandrere. Effekten på forventet lønn for innvandrere i Sverige er derfor ikke signifikant lengre. Samspillvariabelen mellom Norge og innvandrere er fortsatt ikke signifikant. Effekten av de resterende variabelen er derimot signifikant og kan derav generaliseres til populasjonen.

Utdanning har en positiv effekt på timelønnen med 5,3 % økning for hvert år ekstra med utdanning. Det samme gjelder for arbeidserfaring, som her viser en positiv effekt på timelønnen lik 1,01 % per år med ekstra arbeidserfaring. Å være kvinne er derimot negativt på lønnen, fremfor å være mann. Effekten måles her til lik 9,2% lavere forventet lønn for kvinner enn for menn.

Ser videre at effekten av de ulike ferdighetene alle er positive, som tilsier at et høyere nivå av lese-, tall- og problemløsningsferdigheter alle vil virke positivt inn på timelønnen. Tallforståelse har størst påvirkning, lik 3,8%. Videre har problemløsning i et IKT-miljø en lønnsnivåpåvirkning lik 1,19% og leseferdigheter til slutt lik 1,5%.

5.3 Hypotese: avkastning på utdanning er ulik for innvandrere og majoritetsbefolkningen

Ønsker her å teste om innvandrere har lik avkastning på utdanning enn majoritetsbefolkningen. Ser her på innvandrere i Skandinavia som helhet og lar være å skille mellom innvandrere i de tre ulike landene.

Starter med å benytte en Mincer modell uten å se på de ulike samfunnsgruppene:

$$\ln Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 \quad (11)$$

der $\ln W$ er logaritmen til timelønnen, S er antall år med utdanning og X er antall år med arbeidserfaring. Legger til erfaring kvadrert for å avsløre eventuelle kurvelineære mønstre i regresjonsanalysen. Kvadrerte variabler får man ved å gange variabelen med seg selv.

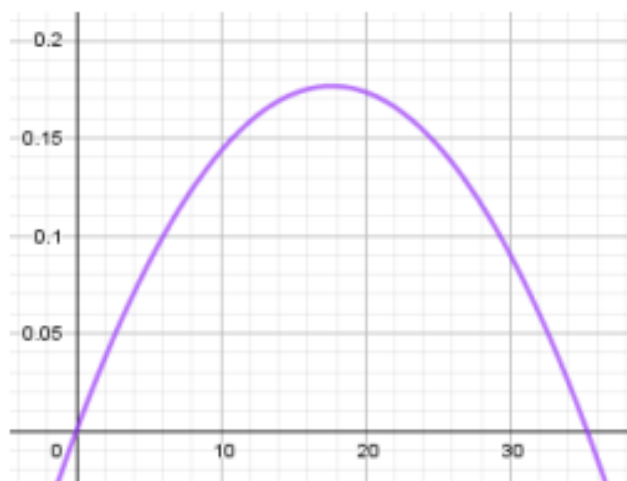
Ved å estimere modellen får vi følgende resultat:

VARIABLER	(3) lnTimelønn PPP
Utdanning	0.0585*** (69.07)
Erfaring	0.0251*** (39.19)
Erfaring^2	-0.000342*** (-24.43)
Konstantledd	1.924*** (161.42)
Observasjoner	18,647
R-squared	0.350

t verdier i parentes
 *** p<0.01, ** p<0.05, * p<0.1

Modellen estimerer en signifikant og positiv avkastning på utdanning. Har at ett ekstra år med høyere utdanning gir en årlig avkastning på 5,58%. Det er også en positiv og signifikant avkastning av arbeidserfaring der ett ekstra år med erfaring øker avkastningen med ytterligere 2,5 %. Erfaring^2, det kvadrerte leddet, måler en avtakende effekt av arbeidserfaring lik 0,0342% per år.

Figur (3):



Figur (3) viser at når det legges til ett år ekstra med arbeidserfaring så vil logaritmen til timelønn først stige, før den så stiger i avtagende grad og slutt gir en negativ effekt slik vi har i regresjonsmodellen over. Kurvens form avhenger av hvor sterk førstegradskoeffisienten er i forhold til den for andregradsleddet.

For å finne formen på kurven kan man regne ut vendepunktet med Thranes formel (Ringdal, 2018):

$$-\beta_1/(2 * \beta_2) \tag{12}$$

Ved å fylle inn får vi:

$$-(0,0251)/(2 * -0,000342) = 36$$

Arbeidserfaring vil ha et toppunkt på 36 år, før kurven så snur og går nedover.

Utvider så modellen til å inneholde en dummyvariabel som ser på om individet er innvandrere eller majoritetsbefolkning:

$$\ln Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \delta_0 D_m \tag{13}$$

Her er D_m dummyvariabelen for innvandrere. Estimerer modellen og får:

VARIABLES	(3) lnTimelønn PPP	(4) lnTimelønn PPP
Utdanning	0.0585*** (69.07)	0.0587*** (69.38)
Erfaring	0.0251*** (39.19)	0.0251*** (39.34)
Erfaring^2	-0.000342*** (-24.43)	-0.000351*** (-25.05)
Innvandrere		-0.0745*** (-12.29)
Konstantledd	1.924*** (161.42)	1.938*** (162.11)
Observasjoner	18,647	18,605
R-squared	0.350	0.355

t verdier i parentes

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

Sammenlignet med den vanlige Mincer-modellen i modell (3) så estimeres denne en minimal økning i avkastningen på utdanning på 0,0002%, mens det er nøyaktig likt for avkastning av arbeidserfaring. Det kvadratiske leddet av erfaring er her noe høyere og vil her gi en årlig avtakende effekt på 0,0351% sammenlignet med tidligere 0,0342%. Alle variablene er fortsatt

signifikante. Ser at dummyvariabelen også er signifikant, men negativ. Tolkningen av dette vil være at majoritetsbefolkningen, 0, vil starte med en høyere timelønn enn innvandrere og at innvandrere vil ha et dårligere lønnsutgangspunkt. Dette stemmer med tidligere funn, forskning og tall presentert i kapittel 3.3. Her vil en innvandrer ha 7,4% lavere forventet timelønn.

Kan videre utvide modellen igjen for å se på samspillet mellom innvandrer og utdanning:

$$\ln Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \delta_0 D_m + \delta_1 SD_m \quad (14)$$

hvor SD_m er en samspillsvariabel opprettet på lik måte som i kapittel 5.1. Samspillet viser her effektforskjellene av å være innvandrer og majoritetsbefolkningen på avkastning på utdanning.

Modellen estimeres:

VARIABLES	(3) lnTimelønn PPP	(4) lnTimelønn PPP	(5) lnTimelønn PPP
Utdanning	0.0585*** (69.07)	0.0587*** (69.38)	0.0609*** (64.24)
Erfaring	0.0251*** (39.19)	0.0251*** (39.34)	0.0248*** (38.70)
Erfaring ²	-0.000342*** (-24.43)	-0.000351*** (-25.05)	-0.000343*** (-24.44)
Innvandrer		-0.0745*** (-12.29)	0.0694** (2.43)
Utdanning*innv.			-0.0106*** (-5.16)
Konstantledd	1.924*** (161.42)	1.938*** (162.11)	1.910*** (146.11)
Observasjon	18,647	18,605	18,605
R-squared	0.350	0.355	0.356

t verdier i parentes
 *** p<0.01, ** p<0.05, * p<0.1

Den fullstendige modellen estimerer igjen kun signifikante effekter så vi kan generalisere alle funnene til populasjonen. R^2 har økt gradvis for hver modell og vi ser at modell (5) her har en

forklaringskraft på 35%. 35 % av variasjonen i timelønnen kan derfor forklares ved endringer i forklaringsvariablene i denne modellen.

Utdanning påvirkes her forventet timelønn med 6%, som er en liten økning i fra de to tidligere utvidede Mincermodellene. Arbeidserfaring påvirker også lønnen i positiv grad fortsatt, men i noe lavere grad enn ved de to tidligere Mincermodellene. Arbeidserfaring påvirker forventet lønn med 2,4%.

Ser at det er en forskjell på avkastningen på utdanning for innvandrere og majoritetsbefolkningen. Innvandrere har en lavere avkastning på sin utdanning, presentert ved det negative fortegnet foran samspillsvariabelen.

5.4 Teste for heteroskedastisitet

Ønsker å teste modellene som er jobbet med gjennom oppgaven:

$$\ln W = \beta_0 + \beta_1 \text{Nor} + \beta_2 \text{Swe} + \beta_3 \text{INNV} + \beta_4 \text{NOR} * \text{inn} + \beta_5 \text{SWE} * \text{inn} + \varepsilon \quad (9)$$

$$\ln W = \beta_0 + \beta_1 D_{\text{NOR}} + \beta_2 D_{\text{SWE}} + \beta_3 \text{innv} + \beta_4 \text{NOR} * \text{innv} + \beta_5 \text{SWE} * \text{innv} + \beta_6 \text{utd} + \beta_7 \text{kjønn} + \beta_8 \text{Arb} + \beta_9 \text{Lese} + \beta_{10} \text{tall} + \beta_{11} \text{IKT} + \varepsilon \quad (10)$$

$$\ln Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 \quad (11)$$

$$\ln Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \delta_0 D_m \quad (13)$$

$$\ln Y_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \delta_0 D_m + \delta_1 SD_m \quad (14)$$

For å teste disse modellene formuleres hypotesene:

$$H_0: a_2 = a_3 = a_4 = a_5 = \dots = a_m = 0$$

$$H_A: a_2 = a_3 = a_4 = a_5 = \dots = a_m \neq 0$$

Der restleddene er homoskedastiske i nullhypotesen og heteroskedastiske alternativhypotesen. Testobservatoren har X^2 -fordeling med $m - 1$ frihetsgrader, der m er antall konstanter (a) i antakelsen til Breusch-Pagan. Skrives som:

$$TS = nR^2 \sim X^2(m - 1)$$

Tester her en kji-kvadratfordelt nullhypotese om at variansen er konstant og at vi ikke har heteroskedastiske restledd. Ved en kji-kvadratverdi lik 11,96 får vi en p-verdi på 0,0005, som er valgt signifikant i denne oppgaven. Da vil nullhypotesen forkastes til et 5% signifikansnivå og konkluderer med at variansen er ulik.

Ved å utføre Breusch-Pagan testen i Stata finner jeg at de to første modellene inneholder heteroskedastisitet. Forkaster derfor nullhypotesen om at restleddene er homoskedastiske ved disse to modellene og de må derfor korrigeres. Gjør dette ved å estimere robuste standardfeil ved hjelp av Stata og før da følgende informasjon:

Modell (1)		Resultater fra Stata med og uten robust verdi				
Variabel	t-verdi	t-verdi r.	endring	koeffisient	koeffisient r.	endring
Innvandrere	-14.74	-13.69	1,05	-0.14	-0.14	0
NOR	3.04	2.90	0,14	0.02	0.02	0
SWE	-37.17	-39.95	2,78	-0.23	-0.23	0
NOR_innv	-0.11	-0.10	0,01	-0.002	-0.002	0
SWE_innv	4.25	4.65	0,4	0.070	0.070	0

Modell (2)		Resultater fra Stata med og uten robust verdi				
Variabel	t-verdi	t-verdi r.	endring	koeffisient	koeffisient r.	endring
Innvandrere	-3.41	-3.21	0.20	-0.031	-0.031	0
NOR	0.73	0.71	0.02	0.004	0.004	0
SWE	-33.67	-34.08	0.41	-0.17	-0.17	0
NOR_innv	-1.29	-1.25	0.04	-0.023	-0.023	0
SWE_innv	0.52	0.54	0.02	0.008	0.008	0
Utdanning	55.94	51.13	4,81	0.053	0.053	0
Kvinne	-21.00	-20.77	0.23	-0.092	-0.092	0
Erfaring	54.19	50.64	3.55	0.010	0.010	0
Leseferdighet	2.79	2.76	0.0	0.015	0.015	0
Tallferdighet	7.20	7.27	0.03	0.038	0.038	0

Problemløsning	2.95	2.86	0.09	0.012	0.012	0
----------------	------	------	------	-------	-------	---

Heteroskedastisitet skal av betydelig størrelse før den gir et betydelig utslag på standardfeilene. Vi ser derimot av tabellene at det er har små endringer i t-verdiene og ingen endring i koeffisientene. Ser derfor bort i fra endringene og beholder modellene.

6. Konklusjon

Jeg har med denne oppgaven forsøkt å belyse forskjeller ved lønn og avkastning på utdanning mellom innvandrere og majoritetsbefolkningen i de tre skandinaviske landene. Dette er gjort ved å benytte PIAAC, en internasjonal undersøkelse om voksnes ferdigheter.

Funnene i denne oppgaver viser seg å være i tråd med hypotesene: innvandrere kommer dårligere ut, både ved lønn og avkastning på utdanning, enn det majoritetsbefolkningen i alle de tre Skandinaviske landene gjør. Altså stemmer de tre nullhypotesene og beholdes derav.

Jeg ser en moderat forskjell på gapet mellom folkegruppene i landene, men generelt viser det seg ugunstig å ikke være etnisk skandinaver på arbeidsmarkedet. At innvandrere har lavere lønn enn majoritetsbefolkningen er en signifikant effekt i Danmark, men ikke i de to andre landene i alle modellene. Dette begrenser muligheten til å generalisere alle funnene til populasjonen, og er en begrensning med modellene.

Det er derimot en signifikant effekt at innvandrere har en lavere avkastning av utdanning, som samsvarer med tidligere forskning og teori. Dett konkluderes derfor her likt som Raaum (Raaum, 1999) konkluderte med i 1999 da han fant at flertallet av innvandrere opplever en lavere avkastning på utdanningen sin. Det er derimot verdt å nevne at begrensninger i datasettet fører til at det ikke går an å skille innvandrere som har tatt utdanning i utlandet fra de som har tatt utdanning i Skandinavia, og heller ikke skille innvandrere i fra hverandre basert på bakgrunn. Dette er problematisk i form av at konklusjonen kan være basert på mangelfull data og jeg kan ikke konkludere med at denne effekten gjelder innvandrere i fra alle land.

Videre funn viser at utdanning og arbeidserfaring har en positiv effekt på timelønn, der begge vil gi en økt timelønn ved én enhets økning. Utdanning har størst påvirkningskraft av de to og vil stimulere timelønnen i høyere grad enn arbeidserfaring. Jeg finner også en positiv korrelasjon mellom de ulike ferdighetene og forventet timelønn. Her vil en person med gode leseferdigheter ha en stor sannsynlighet for også har gode tallferdigheter, god problemløsningsevne i IKT-miljø og derav en høyere forventet timelønn enn personer som ligger på et lavere ferdighetsnivå.

Verdt å nevne er derimot at forventet timelønn er satt sammen av flere faktorer og vil påvirkes av langt flere variabler enn inkludert her. Vi ser at dette også kommer frem i fra modell (2), en fullstendig utvidet modell med alle kontrollvariablene i oppgaven. Til tross for utvidelsen forklares ikke mer enn 42% av endringen i timelønn. Derfor oppfordres det avslutningsvis til å forske videre på lønnsdannelsen til innvandrere i Skandinavia, der en ser på andre faktorer som vil påvirke timelønn. Da vi ser en så negativ, signifikant effekt av å være innvandrer, oppfordres det til en videre forskning også på nettopp hvorfor det har blitt slik. Hva kan gjøres for å utjevne dette gapet?

7. Bibliografi

- Cheibub, J. A., Gandhi, J., & Vreeland, J. R. (2010). *Democracy and Dictatorship Revisited*, *Public Choice*, vol. 143, no. 2-1. Springer US.
- Danmarks Statistik. (2013). *Befolkningens løn*. København: Danmarks Statistik.
- Hagen, P. C. (2016). *Innføring i sannsynlighetsregning og statistikk*. Cappelen Damm Akademisk.
- Jakobsen, V., Korpi, T., & Lorentzen, T. (2018). *Immigration and Integration Policy and Labour Market Attainment Among Immigrants to Scandinavia*. *European Journal of Population*.
- Katz, K., & Österberg, T. (2013). *Unga invandrare – utbildning, löner och utbildningsavkastning*. Uppsala: Institutet för arbetsmarknads- och utbildningspolitisk utvärdering (IFAU).
- Kjelsrud, M. (1996). *Høyt utdannet, men lite etterspurt arbeidskraft*. Oslo: Samfunnsspeilet 2/96, SSB.
- Mincer, J. (1974). *Schooling, experience, and earnings*. New York: National Bureau of Economic Research.
- OECD. (2012). *The Programme for the International Assessment of Adult Competencies*. OECD.
- Pettersen, S. V., & Østby, L. (2013, Desember 10). *Innvandrere i Norge, Sverige og Danmark*. Hentet fra SSB: <https://www.ssb.no/befolkning/artikler-og-publikasjoner/innvandrere-i-norge-sverige-og-danmark>
- Raaum, O. (1999). *Inntektseffekter av utdanning i Norge – en litteraturoversikt*. Frischsenteret.
- Ringdal, K. (2018). *Enhet og Mangfold*. Fagbokforlaget.
- SSB. (2013). *Ferdigheter i voksenbefolkningen Resultater fra den internasjonale undersøkelsen om lese- og tallforståelse (PIAAC)*. SSB.
- Strøm, F., & Bye, K. S. (2017). *INNTEKT OG LØNN- Norskfødte med innvandrerforeldre har høyere inntekter enn sine foreldre*. Statistisk Sentralbyrå - SSB.
- Thomas, R. L. (2005). *Using statistics in economics*. McGraw-Hill Education.
- Williams, R. (2015). *Heteroskedasticity*. Hentet fra <https://www3.nd.edu/~rwilliam/stats2/l25.pdf>

