

Forskjeller i lønnsavkastning av formell utdanning og ferdigheter mellom Norge og USA

En empirisk analyse med data fra PIAAC 2013

Bacheloroppgave i Samfunnsøkonomi

Veileder: Bjarne Strøm

Mai 2019

Forskjeller i lønnsavkastning av formell utdanning og ferdigheter mellom Norge og USA

En empirisk analyse med data fra PIAAC 2013

Bacheloroppgave i Samfunnsøkonomi
Veileder: Bjarne Strøm
Mai 2019

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

Innholdsfortegnelse

1. Innledning av problemstilling	2
2. Teoretisk rammeverk	2
2.1 Tidligere studier	3
2.2 Økonomisk teori.....	4
3. Presentasjon av datamateriale	4
3.1 PIAAC	5
3.2 Deskriptiv statistikk.....	7
4. Økonometrisk modell	8
4.1 Minste kvadraters metode	8
4.2 Multivariat semilogaritmisk regresjon.....	10
4.3 Forutsetninger for OLS	10
4.4 Multikollinearitet	11
4.5 Hypotesetesting	12
5. Regresjonsanalyse	13
5.1 Empirisk strategi.....	13
5.2 Empiriske resultater	14
5.3 Analyse av tilleggsspørsmål	19
5.4 Diskusjon.....	21
6. Oppsummering	22
6.1 Hovedfunn i oppgaven	23
7. Litteraturliste	24

1. Innledning av problemstilling

I denne oppgaven ønsker jeg å reise problemstillingen om *hvorvidt vi observerer forskjeller i lønnsavkastning på formell utdanning og ferdigheter mellom land med ulik grad av fagorganisering*. På basis av dette har jeg derfor valgt å se på forskjellene mellom USA og Norge, som på hver sin side representerer motpoler knyttet til fagorganisering.

Problemstillingen som reises i denne oppgaven oppfattes, etter min mening, som svært relevant ettersom vi ikke bare ser store forskjeller i lønnsavkastning mellom land, og da særlig mellom Norge og USA, men også fordi lønnsforskjeller i seg selv ofte er roten til en rekke andre tendenser. For eksempel indikeres det til at en lavere lønnsavkastning på utdanning og ferdigheter gir mindre inntektsforskjeller og lavere arbeidsledighet. Det teoretiske rammeverket vil i stor grad hvile på studien gjennomført av Hanushek, Schwerdt, Wiederhold og Woessmann (2015), som undersøker forskjeller i avkastning på ferdigheter ved bruk av data fra PIAAC undersøkelsen. Bakgrunnen for at vi inkluderer ferdigheter som et mål, fremfor å kun se på antall år med utdanning, er fordi de kvantitative målene for utdanning ignorerer det faktum at effekten av ett års ekstra utdanning, gitt ved kunnskapene og ferdighetene som tilegnes, er ulike mellom land som har forskjellig kvalitet på sitt utdanningssystem (Hanushek & Woessmann, 2008, p. 652). I tillegg kan det tenkes at sammenhengen mellom avhengig variabel, timelønn, og ferdigheter kan forklares av en viss korrelasjon mellom grad av utdanning og score på ferdigheter.

Interesseparameteren i oppgaven vil i hovedsak være avviket mellom lønnsavkastningen på utdanning i Norge, sammenliknet med USA. Videre vil jeg kontrollere for tallforståelse og leseferdigheter. Ferdighetsmålene har tidligere vist å ha en sterk effekt på timelønna, hvilket gjør den svært interessant for denne oppgaven. Oppgaven vil derfor i tillegg fokusere på forskjeller i lønnsavkastningen på tallforståelse og leseferdigheter mellom Norge og USA. Jeg vil til slutt kontrollere for innvandringsstatus, og hvorvidt førstegenerasjons innvandrere har en ulik lønnsavkastning enn innfødte, og om disse forskjellene varierer mellom Norge og USA. Til slutt vil jeg diskutere hvorvidt forskjellene i lønnsavkastning mellom de to landene kan knyttes til ulik grad av fagorganisering. Målene knyttet til antall sysselsatte som er omfattet av en tariffavtale og antall som er medlem i en fagforening vil med andre ord ikke inkluderes i regresjonsanalysen, men vil snarere benyttes som et bakgrunnsdatamateriale. Datamateriale som benyttes er fra PIAAC-undersøkelsen, gjennomført av OECD, som hadde som hensikt å kartlegge voksnes ferdighetsnivå. Ved bruk av dataprogrammet Stata vil jeg ved bruk av minste kvadraters metode kjøre en trinnvis regresjonsanalyse basert på åtte ulike lønnslikninger som vil presenteres senere i oppgaven.

2. Teoretisk rammeverk

I delkapittel 2.1 vil jeg presentere tidligere studier, og gi et kort innblikk i Norge og USAs lønns- og organisasjonsstruktur. Videre vil jeg i delkapittel 2.2 belyse hvordan humankapital, som i dette tilfellet måles ut i fra antall år utdanning og score på ferdighetstester, påvirker den enkeltes timelønn.

2.1 Tidligere studier

Det finnes en rekke grunner til at lønnsavkastningen av formell utdanning og ferdigheter er forskjellig mellom ulike land. Avkastningen kan for eksempel være preget av et lands relative tilbud av utdannet arbeidskraft, ulike former for studiefinansieringssystemer, eller andelen sysselsatte som er medlem av en fagforening eller omfattet av en tariffavtale. Hanushek, Schwerdt, Wiederhold og Woessmann (2015, p. 121) finner i sin studie at institusjonelle faktorer har stor betydning for avkastningen av ferdigheter på lønn. Studien viser at grad av fagorganisering, stillingsvern og andel sysselsatte i offentlig sektor har en signifikant negativ effekt på lønnsavkastningen av ferdigheter. Hanushek (2015) mfl. sin studie har imidlertid noen begrensninger. For det første benyttes variabelen *union density* som omfatter andelen fagorganiserte lønsmottakere. Forfatterne overser da lønsmottakerne som er omfattet av en tariffavtale uten å personlig være medlem i en fagforening. *Bargaining coverage* vil således være et bedre mål på fagforeningsinnflytelse. Videre tar ikke studien hensyn til lands studiefinansieringsordninger. Det er rimelig å anta at land med generøse studiefinansieringssystemer har en lavere avkastning på utdanning og ferdigheter, mens land som ikke tilbyr stipender eller subsidierte lån vil være avhengig av at lønnsavkastningen er høy for å kompensere for de marginale privatkostnadene knyttet til utdanning (Barth, Jon Marius, Schøne, von Simson, & Strøm, 2016, p. 56).

Norge, som i høy grad er preget av en sammenpresset lønnsstruktur grunnet sentraliserte lønnsforhandlinger, har som følge en relativt lav lønnsavkastning på utdanning. Norges inntektspolitikk har dog medført små inntektsforskjeller, lav arbeidsledighet og gjennomgående høy reallønnsvekst (NOU, 2013, p. 12). I motsatt ende finner vi USAs lønnsystem som i større grad er preget av en desentralisert forhandlingsmodell, der lønnsforhandlinger tar sted på lokalt nivå snarere enn på sektor- eller nasjonalt nivå. Mangelen på en enhetlig inntektspolitikk gir rom for større variasjoner i inntekt mellom bedrifter og sektorer, hvilket har resultert i en økende inntektsulikhet. Dette gir også grunnlaget for en større lønnsavkastning på utdanning. Både USA og Norge har opplevd en nedgang i både andel sysselsatte som er medlem av en fagforening og andel sysselsatte omfattet av en tariffavtale siden tidlig 2000-tallet. Dataen presentert i tabell 2.1.1 reflekterer den store forskjellen mellom Norge og USAs inntektspolitikk, der organisasjonsgraden er langt større ved førstnevnte enn ved sistnevnte. Der omkring halvparten av alle sysselsatte i Norge er omfattet av en tariffavtale og medlem av en fagforening, er det kun litt over 10% som er det samme i USA.

	Andel sysselsatte omfattet av tariffavtale		Andel sysselsatte medlem av fagforening	
	2002	2014	2001	2016
Norge	70,5%	67%	55%	52%
USA	14,5%	12,3%	12,9%	10,3%

Tabell 2.1.1: Fagforeningsopplutning og kollektiv forhandlingsdekning i Norge og USA

Kilde: Data hentet fra OECD. Stat 21.04.2019

2.2 Økonomisk teori

Når vi snakker om avkastningen av utdanning og ferdigheter kan det være lurt å skille mellom den privatøkonomiske og samfunnsøkonomiske. Vi kan for eksempel si at den private kostnaden ved utdanning er gitt ved alternativkostnaden av å ikke være i arbeid, eller kostnaden knyttet til skolepenger og lærebøker. De private gevinstene kan enkelt sammenliknes med en aksjeinvestering. Marginalavkastningen henviser da gjerne til økningen i lønnen av ett års ekstra utdanning. Ytterligere privatøkonomiske gevinster ved høyere utdanning omfatter høyere personlig og profesjonell mobilitet og bedret livskvalitet. De sosiale, eller samfunnsøkonomiske, kostnadene og gevinstene knyttet til utdanning reflekteres som oftest i eksternaliteter, som vil si tredjepartsvirkninger som ikke inngår i enkeltindividets nytte- og kostnadsberegninger. Samfunnets kostnader kan da være de direkte kostnadene staten bruker på å subsidiere skolesystemet. De positive eksternalitetene er gjerne vanskeligere å beregne, og kan inkludere alt fra lavere kriminalitet, bedre helse, mindre inntektsulikhet og nasjonal stabilitet (World Bank, 1995, pp. 19-20). Hovedfokuset i denne oppgaven vil være den privatøkonomiske avkastningen av utdanning, målt i inntekt.

Økonomiske analyser av avkastningen på utdanning hviler i stor grad på det teoretiske rammeverket lansert av Becker (1964). Teorien går ut på at utdanning sees på som en investering, der enkeltindividers valg av utdanning er et resultat av at de ønsker å maksimere nåverdien av sin inntekt. Dette følger av antakelsen om at gevinstene ved en utdannelse overgår kostnadene i form av høyere produktivitet og følgelig høyere lønn. Mincer (1974) var den første til å lansere en statistisk lønnsfunksjon, som hadde som formål å studere effekten av ett ekstra utdanning på lønna:

$$\ln Y_i = \beta_0 + \beta_1 X_{1i} + \beta_n X_{ni} + \varepsilon_i \quad (2.2.1)$$

Der $\ln Y$ er logaritmen til lønna til individ i , X_{1i} er antall år utdanning og X_{ni} er en vektor av andre variable som påvirker inntekten. Dette kan være alt fra arbeidserfaring, kjønn, arbeidssektor, etnisitet og bosted. ε_i er et stokastisk restledd, som representerer alle andre mulige forstyrrelser. Lønnsfunksjonen vil være grunnlaget for oppgavens analyse og problemstilling.

3. Presentasjon av datamateriale

I delkapittel 3.1 vil jeg presentere datamateriale som vil være grunnlaget for oppgavens analyser. Jeg vil først gi et kort innblikk i selve undersøkelsen, og deretter presentere og definere de utvalgte variablene. I delkapittel 3.2 vil det presenteres deskriptiv statistikk for variablene i Norge og USA.

3.1 PIAAC

Analysen vil gjennomføres ved bruk av den internasjonale undersøkelsen om lese- og tallforståelse (PIAAC), som i regi av OECD pågikk fra august 2011 til april 2012 i til sammen 24 land (Bjørkeng, 2013, p. 7). Undersøkelsen hadde som målsetning å kartlegge voksnes ferdighetsnivå, gitt ved leseferdigheter, tallforståelse og problemløsning innen IKT. Ytterligere bakgrunnsspørsmål knyttet til utdanningsnivå, arbeidserfaring, helsetilstand, innvandringsstatus og så videre ble og besvart av respondentene. Jeg vil i denne oppgaven undersøke hvorvidt forskjellene i lønnsavkastningen av utdanning mellom Norge og USA kan forklares av ulik grad av fagorganisering. Grunnet tekniske hindringer, vil imidlertid ikke målene på fagorganisering inkluderes i regresjonsanalysen, men heller fungere som en underliggende antakelse. Tabell 3.1.1 gir oss samlet frekvenstabell for henholdsvis Norge og USA. Vi ser at antall observasjoner ligger på 4079, med en prosentfordeling på 49,42% for Norge, og at antall observasjoner i USA ligger på 4174, hvilket gir en prosentfordeling på 50,58.

	Observasjoner	Prosentfordeling	Kumulativ prosentfordeling
Norge	4079	49,42	50,58
USA	4174	50,58	100
Total	8253	100	

Tabell 3.1.1: Frekvenstabell for Norge og USA

Følgende variabler er inkludert i datasettet:¹

- Timelønn, PPP justert US\$
- Antall år utdanning
- Leseferdigheter, ikke-standardisert
- Tallforståelse, ikke-standardisert
- Andel av sysselsatte omfattet av tariffavtale
- Andel av sysselsatte som er medlem i en fagforening
- Dummyvariabel for land (Norge=1, USA=0)
- Dummyvariabel for innvandringsstatus, (Førstegenerasjons innvandrere=1, Annet=0)

Variabelen *earnhrppp* måler timelønna notert i amerikanske dollar, og er justert for prisnivå og kjøpekraften i hvert av landene. Dette gjør oss i stand til å sammenlikne det reelle lønnsnivået for hvert av landene. Videre har vi *yrsqual* som måler høyest utdannelsesnivå oppnådd beregnet ut i fra antall år utdanning.

Ferdighetsmålene, gitt ved leseferdigheter og tallforståelse, estimeres ut i fra en skala som strekker seg fra 0 til 500 poeng. Deltakerne tok ulike tester, der hensikten var å beregne de

¹ Variablene som presenteres i den deskriptive statistikken, er ikke nødvendigvis de samme som vil brukes i regresjonsanalysen. Variabelen for fagorganisering vil ikke bli benyttet i regresjonsanalysen, videre vil regresjonsanalysen benytte seg av de standardiserte målene for leseferdigheter og tallforståelse.

direkte poengene, samt sannsynligheten for at personer med liknende egenskaper er i stand til å gjennomføre oppgavene. Variabelen *pvlit1* måler voksnes leseferdigheter, hvilket innebærer «evnen til å forstå, vurdere og bruke skrevet tekst for å delta aktivt i samfunnet, oppnå egne mål og utvikle egen kunnskap og egne evner» (Bjørkeng, 2013, p. 9).

Variabelen *pvnum1* måler voksnes tallforståelse, hvilket omhandler «evnen til å tilegne seg, bruke, tolke og formidle matematiske informasjon og ideer, for å kunne håndtere hverdagslige situasjoner som krever matematiske beregninger» (Bjørkeng, 2013, p. 10).

Den prosentvise andelen av sysselsatte som er medlem i en fagforening, gitt ved *union_density_*, og den prosentvise andelen av sysselsatte som er omfattet av en tariffavtale, gitt ved *bargaining_coverage_OECD_* er hentet fra datamateriale benyttet i artikkelen *Return to skills around the world: Evidence for PIAAC (2015)* utført av Hanushek, Schwerdt, Wiederhold og Woessmann. Variablene vil bli benyttet i den deskriptive statistikken, for å få et overblikk over forskjellene i graden av fagforeningsinnflytelse i Norge og USA. Variablene vil imidlertid ikke benyttes i regresjonsanalysen grunnet tekniske kodeforskjeller i datasettene.

Videre har vi to **dummyvariabler** som produseres med utgangspunkt i kategoriske variabler, der kategoriene i variablene tildeles verdier på 1 eller 0. I vårt tilfelle vil dummyvariabelen for land kategoriseres slik at Norge tildeles verdien 1, og USA tildeles verdien 0. På slik måte vil USA fungere som en referansekategori, hvilket gjør det mulig å sammenlikne forskjeller i lønnsavkastning mellom landene i én regresjonslikning. Variabelen *migrant_first* er kategorisert slik at alle som går under betegnelsen ‘førstegenerasjonsinnvandrere’ får verdien 1, mens alle andre får verdien 0. Definisjonene på en førstegenerasjonsinnvandrer er en person som er født i utlandet, altså født utenfor enten Norge eller USA, og som selv har innvandret til de respektive landene (Tønnessen, 2018). Under definisjonen til variabelen innebærer dette også at minst én av foreldrene er født utenlands.

I tillegg til variablene nevnt overfor vil jeg benytte meg av en rekke **interaksjonsledd**, som er produktet av to (eller fler) uavhengige variabler. Følgende interaksjonsledd vil bli generert:

- Antall år utdanning * Dummyvariabel for Norge (= *yrsqual_norway*)
- Leseferdigheter, standardisert * Dummyvariabel for Norge (= *litscore1_norway*)
- Tallforståelse, standardisert * Dummyvariabel for Norge (= *numscore1_norway*)
- Dummy variabel for migrantstatus * Dummyvariabel for Norge (= *migrant_first_norway*)

Bakgrunnen for at vi generer de overnevnte interaksjonsleddene er for å se på hvorvidt utdanning og ferdigheter i Norge gir høyere eller lavere avkastning på lønn sammenliknet med referansekategorien, som i dette tilfellet er USA. I tillegg undersøkes det om innvandringsstatus har forskjellig innvirkning på lønna i de to landene.

3.2 Deskriptiv statistikk

Variabel	Obs.	Gjennomsnitt	Std. avvik	Min.	Maks.
Timelønn, PPP justert US\$	2959	25,62	8,56	5,89	64,50
Antall år utdanning	3901	14,67	2,45	7	21
Leseferdigheter	3897	282,72	46,67	67,46	411,38
Tallforståelse	3897	284,15	54,34	50,44	419,91
Innvandrer	3893	0,13	0,33	0	1
%-andel omfattet av tariffavtale	1	0,74		0,74	0,74
%-andel medlem i fagforening	1	0,54		0,54	0,54

Tabell 3.2.1: Deskriptiv statistikk for Norge

Variabel	Obs.	Gjennomsnitt	Std. avvik	Min.	Maks.
Timelønn, PPP justert US\$	2438	22,30	13,05	7,4	51,5
Antall år utdanning	3565	13,91	3,02	6	21
Leseferdigheter	4085	272,16	49,14	101,54	417,98
Tallforståelse	4085	256,09	56,88	53,40	433,54
Innvandrer	4062	0,13	0,34	0	1
%-andel omfattet av tariffavtale	1	0,13		0,13	0,13
%-andel medlem i fagforening	1	0,11		0,11	0,11

Tabell 3.2.2: Deskriptiv statistikk for USA

Tabell 3.2.1 og 3.2.2 viser en oversikt over den deskriptive statistikken for de overnevnte variablene for henholdsvis Norge og USA. Tabellene gir oversikt over antall observasjoner, gjennomsnittsverdiene, standardavvik (som gir verdienes gjennomsnittlige avvik fra gjennomsnittet), samt minimums- og maksimumsverdiene. Vi ser i første omgang at Norge har en gjennomsnittlig høyere timelønn på litt over 3 US\$ mer enn USA. Et annet interessant aspekt ved timelønnen, er at USA har et høyere standardavvik, hvilket kan indikere at lønnsforskjellene er større i USA enn i Norge. Resultatet kan i stor grad underbygges av tidligere teori som nevnt innledningsvis. Vi finner videre at gjennomsnittlig år utdanning er høyere i Norge enn i USA, og at Norge scorer høyere på testene knyttet til både leseferdigheter og tallforståelse. Også innenfor ferdighetsmålene finner vi et noe høyere standardavvik i USA, sammenliknet med Norge. Bjørkeng (2013, p. 54) poengterer også i sin evaluering av PIAAC-undersøkelsen at differansen mellom de svakeste og sterkeste leserne er størst i USA. Om dette skyldes ulik kvalitet i utdanningssystemet, større økonomiske barrierer innen høyere utdanning, eller lavere gjennomsnittlig utdanning, er ikke noe vi kan fastslå i denne oppgaven. Nivåmessig ligger Norge på nivå 3 både innen leseferdigheter og tallforståelse, mens USA ligger på nivå 2 (Bjørkeng, 2013, pp. 14, 23). Dummyvariabelen for innvandringsstatus forteller oss ikke mye, annet enn at 13% av utvalget består av førstegenerasjonsinnvandrere i både Norge og i USA. Til slutt ser vi at forskjellene i grad av fagforeningsinnflytelse er svært stor. Den prosentvise andelen av sysselsatte som er omfattet

av en tariffavtale ligger på 74% for Norge og 13% i USA. Videre er 54% av norske og 11% av amerikanske arbeidere medlem i en fagforening. Vi kan med dette fastslå at Norge og USA er ekstreme motpoler når det kommer til grad av fagforeningsinnflytelse.

	Gjennomsnitt (earnhrppp)	Standardavvik (earnhrppp)	Min (earnhrppp)	Maks (earnhrppp)	N (earnhrppp)
Innfødt	24.50	10.83	5.89	64.50	4700
Innvandrer	21.66	11.41	7.04	60.65	682

Tabell 3.2.3: Deskriptiv statistikk – forholdet mellom innvandringsstatus og timelønn justert for kjøpekraft

Tabell 3.2.3 gir oss deskriptiv statistikk for timelønna justert for kjøpekraft i amerikanske dollar fordelt ut i fra innvandringsstatus. Vi ser umiddelbart at førstegenerasjons innvandrere har en lavere gjennomsnittlig timelønn enn innfødte. Standardavviket i timelønningen er også høyere blant innvandrere, hvilket indikerer at lønnsforskjellene er større blant innvandrere enn blant innfødte. Det er dog viktig å bemerke seg at antall observasjoner varierer stort mellom to de grupperingene. Jeg vil senere i oppgaven forsøke å belyse hvorfor innvandrere har en lavere gjennomsnittlig timelønn enn innfødte.

4. Økonometrisk modell

I delkapittel 4.1 presenteres estimeringsmetoden som benyttes i oppgaven, nemlig minste kvadraters metode. Videre vil delkapittel 4.2 ta for seg funksjonsform, og da nærmere multipel semilogaritmisk regresjon. I 4.3 går vi igjennom forutsetningen som ligger til grunn for at estimeringsmetoden, OLS, skal være den beste metoden tilgjengelig. Til slutt vil 4.4 belyse problemet med multikollinearitet, og dets relevans for min oppgave, mens delkapittel 4.5 kort vil gå gjennom hypotesetesting.

4.1 Minste kvadraters metode

Oppgaven vil benytte seg av minste kvadraters metode, heretter betegnet som OLS (ordinary least squares), for å estimere relasjoner mellom flere økonomiske variabler. Metoden går ut på at man ut i fra observerte verdier estimerer en regresjonslinje ved å minimere summen av de kvadrerte avvikene. Løsningen blir der spredningen mellom observasjonene og regresjonslinjene er minst mulig (Thomas, 2005, p. 266).

Det lineære forholdet mellom den endogene variabelen Y og eksogene variabelen X, kan forklares ut i fra følgende likning:

$$\hat{Y} = \alpha + \beta X \quad (4.1.1)$$

Sett ut i fra vår oppgave vil den forventede timelønna justert for kjøpekraft (*earnhrppp*), \hat{Y} , være gitt ved X som er antall år utdanning (*yrsqual*) og parameteren β som angir hvor mye timelønna øker når utdanning økes med ett år, *ceteris paribus*.

$$Y_i = \alpha + \beta X_i + \varepsilon \quad (4.1.2)$$

$$\hat{Y}_i = a + bX_i \quad (4.1.3)$$

Likning 4.3 gir den predikerte verdien av Y , gitt ut i fra de estimerte verdiene på α og β gitt ved henholdsvis a og b . Støyleddet ε representerer differansen mellom forventet og faktisk verdi på Y . Den inkluderer alle andre faktorer som påvirker timelønna, samt andre tilfeldige faktorer som kan ha en viss påvirkning. Det empiriske motstykket til støyleddet er gitt ved residualen e_i , hvilket vil være negativ dersom forventet Y overgår faktisk Y , og motsatt. Får dermed:

$$Y_i = \hat{Y}_i + e_i \rightarrow e = Y_i - \hat{Y}_i \quad (4.1.4)$$

Som nevnt innledningsvis, vil OLS forsøke å minimere summen av de kvadrerte residualene $\sum e_i^2$, ved å velge estimatorene a og b slik at differansen mellom predikert og faktisk Y blir så liten som mulig (Thomas, 2005, p. 266).

Hvor mye den estimerte regresjonslikningen faktisk forklarer, og hvor godt modellen forklarer variasjonen i avhengig variabel Y , uttrykkes i forklaringskoeffisienten R^2 . Den sier hvor mye av variasjonen i den avhengige variabelen som kan tilskrives variasjonen i den uavhengige variabelen. I vårt tilfelle vil den gi oss en prosentvis andel som forteller oss hvor mye av variasjonene i timelønn som kan tilskrives variasjonen i utdanning. R^2 vil med andre ord ligge mellom 0 og 1, der 1 tilsier at det er perfekt korrelasjon mellom avhengig og uavhengig variabel, mens 0 indikerer ingen korrelasjon (Thomas, 2005, p. 278). Vi finner variansen ved å finne den totale samlede variasjonen (SST) og den samlede forklarte variasjonen (SSE). Førstnevnte, som er den totale summen av kvadrater, viser til avviket mellom observert avhengig variabel og dets gjennomsnitt.

$$SST = \sum (Y_i - \bar{Y})^2 \quad (4.1.5)$$

Den samlede forklarte variasjonen viser til avviket mellom predikert avhengig variabel og dets gjennomsnitt.

$$SSE = \sum (\hat{Y}_i - \bar{Y})^2 \quad (4.1.6)$$

Forklaringskoeffisienten gir oss dermed andelen av variasjonen i Y som kan tilskrives variasjonen i X :

$$R^2 = \frac{SSE}{SST} = \frac{b^2 \sum X_i^2}{\sum Y_i^2} \quad (4.1.7)$$

4.2 Multivariat semilogaritmisk regresjon

Når den lineære regresjonsfunksjonen inkluderer flere uavhengige variabler betegner vi det som en multippel eller multivariat regresjonsanalyse. Vi inkluderer flere uavhengige variable fordi det er høyst tenkelig at timelønna påvirkes av en rekke andre forhold. Dersom vi utelater variabler som kan tenkes å ha enn innvirkning på avhengig variabel, vil vi få utelatt variabelskjevhet, hvilket medfører forventningsskjevne estimater.

$$\ln(\text{earnhrppp}) = \beta_0 + \beta_1 \text{yrsqual}_i + \beta_i X_i + \varepsilon_i \quad (4.2.1)$$

Likning 4.1.1 illustrerer en log-lineær, eller semilogaritmisk, funksjonsform der $\ln(\text{earnhrppp})$ er logaritmen til earnhrppp . Vi justerer for kjøpekraft for å kunne sammenlikne lønnsnivået mellom Norge og USA. Som alltid representerer β_0 konstantleddet, som gir verdien av den avhengige variabelen når verdien av de uavhengige variablene holdes lik 0. Siden vi har en semilogaritmisk regresjonsfunksjon, vil β_1 fremstille den *prosentvise* endringen i timelønna når utdanning endres med én enhet, til gitt nivå på kontrollvariablene. Bakgrunnen for at vi har valgt en log-lineær funksjonsform, er fordi det gjør det mulig å sammenlikne lønnsforskjeller på tvers av land, hvilket i dette tilfellet består av Norge og USA, uavhengig av målenivå på avhengig variabel. Foruten de øvrige kontrollvariablene, gitt ved X_i , har vi i tillegg et restledd (ε) som representerer alle øvrige forhold som kan påvirke den avhengige variabelen.

4.3 Forutsetninger for OLS

For at OLS skal være den beste estimatoren tilgjengelig for regresjonsanalysen, må flere forutsetninger rundt forklaringsvariablene og restleddene være oppfylt.

Følgende forutsetninger rundt de *eksogene variablene* må være oppfylt for at OLS skal kunne benyttes som estimeringsmetode:

1. Hver eksogen variabel antas å være ikke-stokastisk
2. Hver eksogen variabel antas å ha fastsatte verdier i hvert utvalg
3. Variansen til hver eksogen variabel har en øvre grense
4. Fravær av multikollinearitet: ingen av de eksogene variablene må korrelere sterkt med hverandre

De to første forutsetningene henviser til X som en ikke-tilfeldig variabel. Verdiene for X -ene er fastsatt av en etterforsker, og verdiene vil være like for hvert utvalg (Thomas, 2005, p. 394). Avhengig variabel vil derimot være stokastisk, ettersom den ikke bare avhenger av X , men også av ukjente forstyrrelser gitt av restleddet. Dette produserer lineære OLS estimatorer. Den tredje forutsetningen forteller oss at variansen i til X -verdiene, $V(X)$, ikke vil øke uendelig når utvalgsstørrelsen øker (Thomas, 2005, p. 358). Dette er for å unngå spuriøse korrelasjoner. Forutsetningen om fravær av multikollinearitet er særlig viktig for multippel

regresjonsanalyse med flere uavhengige variabler. Dersom to eller flere uavhengige variabler er sterkt korrelert, vil ikke OLS estimatoren være i stand til å skille mellom disse, hvilket resulterer i feil estimerte koeffisienter for minst en av regressorene (Thomas, 2005, p. 404). I vårt eksempel er det trolig at ferdighetsmålene, lese- og regneferdigheter, har en sterk korrelasjon, hvilket jeg senere vil undersøke i en korrelasjonsmatrise.

Følgende forutsetninger rundt *restleddet* må være oppfylt for at OLS skal kunne benyttes som estimeringsmetode:

1. $E(\varepsilon_i)$ er lik null for alle i
2. Kovariansen til (ε_i, X_i) er lik null
3. Støyleddets varians er konstant: $V(\varepsilon_i) = \sigma^2$
4. Kovariansen til $(\varepsilon_i, \varepsilon_j)$ er lik null
5. ε_i er normalfordelt

Den første forutsetningen konstaterer at den forventede verdien til restleddet må være lik null, som med andre ord betyr at det estimerte avviket mellom observert og estimert verdi på Y er lik null. Gitt forutsetningen om at X er ikke-stokastisk og har fastsatte verdier for hvert utvalg, vil forutsetninger om at $E(\varepsilon_i)=0$ bety at a og b er forventingsrette estimatorene for henholdsvis α og β . Fraværet av korrelasjon mellom restledd og forklaringsvariabel, er i likhet med forutsetningen om multikollinearitet, inkludert for å unngå at OLS estimatorene tillegger forklaringsvariabelen en del av variasjonen i Y som egentlig stammer fra feilleddet. Dette kan bli et problem dersom vi opplever utelatt variabelskjevhet. Restleddet vil i et slikt tilfelle fange opp noe av effekten av den utelatte variabelen, hvilket medfører at enkelte forklaringsvariabler feil estimeres og tillegges mer forklaringskraft enn det som faktisk forklares av modellen (Thomas, 2005, p. 498). Videre vil forutsetningen om en konstant varians ved feilleddet, dvs. fraværet av heteroskedastisitet, bety at variansen til feilleddets fordeling ikke endres fra observasjon til observasjon (Thomas, 2005, p. 360). Forutsetning nummer fire henviser til fraværet av autokorrelasjon, som vil si at restleddene ikke påvirker hverandre i de ulike utvalgene (Thomas, 2005, pp. 360-361). Til slutt antas det at restleddene følger en normalfordeling, hvilket betyr at Y har en normalfordeling (grunnet forutsetning 1 og 2 om forklaringsvariablene) (Thomas, 2005, p. 361).

4.4 Multikollinearitet

Som tidligere nevnt er det trolig at ferdighetsmålene knyttet til leseferdigheter og tallforståelse er sterkt korrelert, hvilket bryter med forutsetningen om fravær av multikollinearitet. Tabell 4.3.1 viser en korrelasjonsmatrise mellom ferdighetsmålene, *litscore1*, *numscore1* og *yrsqual*. Det er en sterk korrelasjon mellom leseferdigheter og tallforståelse tilsvarende 0,8996. Dette kan gjøre det vanskelig å vurdere hvorvidt resultatene er feilaktig estimert. Som en delvis løsning på problemet vil jeg derfor bygge flere modeller der de sterkt korrelerte variablene holdes separerte. Jeg vil derimot ikke utelate variablene, ettersom jeg anser de som svært relevante for både oppgaven og for betydningen av lønn. En eventuell utelatelse kan medvirke til det tidligere nevnte problemet med utelatt

variabelskjevhet. En annen løsning for å unngå multikollinearitet er å øke utvalgsstørrelsen, men dette er ikke mulig i mitt tilfelle.

Videre finner vi en relativ sterk korrelasjon mellom utdanning og tallforståelse, på 0.6096, og mellom utdanning og leseferdigheter tilsvarende 0.4998. Hanushek m.fl. (2015) finner liknende tendenser, og mener at sammenhengen mellom lønn og ferdigheter kan forklares av det faktum at høy ferdighetsgrad ofte kan knyttes til høy utdanningsgrad.

	Leseferdigheter, standardisert	Tallforståelse, standardisert	Utdanning
Leseferdigheter, standardisert	1		
Tallforståelse, standardisert	0,8996	1	
Utdanning	0,4998	0,5096	1

Tabell 4.4.1: Korrelasjonsmatrise mellom ferdighetsmål og avhengig variabel

4.5 Hypotesetesting

En statistisk hypotese er en antakelse eller påstand om egenskaper ved en eller flere populasjoner. Vi presenterer da en nullhypotese, H_0 , og en alternativhypotese, H_1 , for å teste en gitt påstand. Nullhypotesen uttrykkes som oftest som en antitese, der det antas å ikke finnes noen korrelasjon eller sammenheng mellom de valgte variablene, mens alternativhypotesen fastslår en gitt verdi for en variabel eller en gitt sammenheng mellom flere variabler. Siden det antas at vi ikke med 100 prosent sikkerhet kan bekrefte en påstand, er det i hovedsak nullhypotesen som testes, slik at en eventuell avkreftelse av nullhypotesen indirekte gir støtte til alternativhypotesen. Vi formulerer med andre ord tosidige hypotesetester.

$$H_0: \beta_i = 0 \quad \text{vs.} \quad H_1: \beta_i \neq 0 \quad (4.5.1)$$

For å kunne vurdere hvorvidt resultatene er statistisk signifikante, det vil si hvor sannsynlig det er at for at man forkaster en gyldig nullhypotese, må man velge et signifikansnivå. Jeg vil i denne oppgaven benytte meg av det vanligste nivået, hvilket er et 5% signifikansnivå, $\alpha = 0,05$. Videre velges et konfidensintervall, her på 95%, som angir hvor gode estimatene av de ukjente parameterne er. I oppgaven vil det utledes fem hypoteser, som vil testes ved en t-test eller en F-test.

H1: Tester følgende nullhypotese om at det ikke finnes noen forskjell i lønnsavkastningen på utdanning mellom Norge og USA

$$H_0: \beta_3 = 0, \quad H_A: \beta_3 \neq 0$$

H2: Tester følgende nullhypotese om at tallforståelse, leseferdigheter og innvandringsstatus ikke har noen innvirkning på timelønna

$$H_0: \beta_4 = \beta_5 = \beta_6 = 0, \quad H_A: \beta_4 \neq 0, \beta_5 \neq 0 \text{ eller } \beta_6 \neq 0$$

H3: Tester nullhypotesen om lønnsavkastningen av utdanning og leseferdigheter ikke er forskjellig mellom Norge og USA.

$$H_0: \beta_3 = \beta_7 = 0, \quad H_A: \beta_3 \neq 0 \text{ eller } \beta_7 \neq 0$$

H4: Tester nullhypotesen om lønnsavkastningen av utdanning og tallforståelse ikke er forskjellig mellom Norge og USA.

$$H_0: \beta_3 = \beta_8 = 0, \quad H_A: \beta_3 \neq 0 \text{ eller } \beta_8 \neq 0$$

H5: Tester nullhypotesen om lønnsavkastningen av utdanning og innvandringsstatus ikke er forskjellig mellom Norge og USA.

$$H_0: \beta_3 = \beta_9 = 0, \quad H_A: \beta_3 \neq 0 \text{ eller } \beta_9 \neq 0$$

5. Regresjonsanalyse

I delkapittel 5.1 vil de ulike lønnslikningene som har blitt formulert presenteres. Dette vil legge grunnlaget for den empiriske strategien knyttet til oppgavens hypotese. Videre vil vi i delkapittel 5.2 og 5.3 gi oversikt over de empiriske resultatene. Regresjonen vil først vise estimert forskjell i lønnsavkastning på utdanning mellom landene basert på hovedmodellene. Deretter vil tilleggsspørsmålene analyseres i en ny regresjonstabell. Til slutt vil det føres en kort diskusjon rundt til resultatene, samtidig som oppgavens svakheter belyses, i delkapittel 5.4.

5.1 Empirisk strategi

Vi ønsker primært å estimere størrelsen på forskjellen i lønnsavkastningen av utdanning mellom Norge og USA, med ulike sett av kontrollvariabler. Vi ser på logaritmen til timelønna, hvilket betyr at vi undersøker den prosentvise forskjellen i avkastningen på utdanning og ferdigheter. Med utgangspunkt i likning (4.2.1) presentert tidligere, har jeg utarbeidet totalt åtte lønnslikninger. Siden vi hovedsak ønsker å se på individer som er sysselsatt, og eventuelt i fulltidsarbeid, har jeg valgt å ekskludere alle som er under 25 år gammel i datasettet. Interesseparameteren er i hovedsak β_3 , som angir avviket i lønnsavkastningen på utdanning mellom Norge og referansepunktet USA.

Modell A: $\text{learnhrppp} = \beta_0 + \beta_1 \text{yrsqual} + \beta_2 \text{norway}$

Modell B: $\text{learnhrppp} = \beta_0 + \beta_1 \text{yrsqual} + \beta_2 \text{norway} + \beta_3 \text{yrsqual_norway}$

Modell C: $\text{learnhrppp} = \beta_0 + \beta_1 \text{yrsqual} + \beta_2 \text{norway} + \beta_3 \text{yrsqual_norway} + \beta_4 \text{litscore}$

Modell D: $\text{learnhrppp} = \beta_0 + \beta_1 \text{yrsqual} + \beta_2 \text{norway} + \beta_3 \text{yrsqual_norway} + \beta_5 \text{numscore}$

Modell E: $\text{learnhrppp} = \beta_0 + \beta_1 \text{yrsqual} + \beta_2 \text{norway} + \beta_3 \text{yrsqual_norway} + \beta_4 \text{litscore} + \beta_5 \text{numscore} + \beta_6 \text{migrant_first}$

Modell F: $\text{learnhrppp} = \beta_0 + \beta_1 \text{yrsqual} + \beta_2 \text{norway} + \beta_3 \text{yrsqual_norway} + \beta_4 \text{litscore} + \beta_6 \text{migrant_first} + \beta_7 \text{litscore_norway}$

Modell G: $\text{learnhrppp} = \beta_0 + \beta_1 \text{yrsqual} + \beta_2 \text{norway} + \beta_3 \text{yrsqual_norway} + \beta_5 \text{numscore} + \beta_6 \text{migrant_first} + \beta_8 \text{numscore_norway}$

Modell H: $\text{learnhrppp} = \beta_0 + \beta_1 \text{yrsqual} + \beta_2 \text{norway} + \beta_3 \text{yrsqual_norway} + \beta_4 \text{litscore} + \beta_5 \text{numscore} + \beta_6 \text{migrant_first} + \beta_9 \text{migrant_first_norway}$

Figur 5.1.1: Modelloversikt med utgangspunkt i Mincers lønnsfunksjon

Avhengig variabler er logaritmen til timelønn, juster etter kjøpekraft. Modell A gir oss den grunnleggende prosentvise lønnsavkastningen av utdanning, gitt ved β_1 , og den prosentvise forskjellen i lønnsnivået mellom Norge og USA, gitt ved β_2 . I modell B har vi utelatt variablene for USA for å unngå perfekt multikollinearitet. Koeffisientene for interaksjonsleddene, der *norway* er inkludert, representerer da avviket fra referansepunktet som er USA. β_3 , som er vår interesseparameter i denne oppgaven, representerer da avviket i lønnsavkastning på utdanning i Norge i forhold til USA. I resten av modellen vil β_1 gi oss avkastningen på utdanning i USA. I modell C-D har jeg valgt å inkludere et sett av kontrollvariabler for å undersøke hvorvidt disse har noen innvirkning på timelønna, og for å se på hvordan forskjellen i avkastningen på utdanning mellom Norge og USA endres når vi stegvis legger til variablene. I modell F-H har jeg i tillegg generert en rekke interaksjonsledd, der hensikten er å se på hvordan kontrollvariablene i samspill med dummyvariabelen for Norge avviker fra USA.

5.2 Empiriske resultater

Vi har ved bruk av dataprogrammet Stata kjørt en trinnvis regresjonsanalyse som presentert i *Tabell 5.2.1* under. Regresjonen viser estimert forskjell i lønnsavkastning av utdanning mellom Norge og USA basert på modellene A-E som ble presentert tidligere. Estimeringsmetoden som har blitt benyttet er minste kvadraters metode.

Modell	A	B	C	D	E
Variabler	<i>learnhrppp</i>	<i>learnhrppp</i>	<i>learnhrppp</i>	<i>learnhrppp</i>	<i>learnhrppp</i>
<i>yrsqual</i>	0,0716*** (0,00215)	0,0921*** (0,00290)	0,0755*** (0,00312)	0,0702*** (0,00309)	0,0719*** (0,00311)
<i>norway</i>	0.185*** (0.0114)	0.833*** (0.0635)	0.803*** (0.0625)	0.791*** (0.0617)	0.782*** (0.0617)
<i>yrsqual_norway</i>		-0.0442*** (0.00426)	-0.0413*** (0.00420)	-0.0404*** (0.00415)	-0.0399*** (0.00414)
<i>litscore1</i>			0.0835*** (0.00634)		-0.0396*** (0.0120)
<i>numscore1</i>				0.111*** (0.00634)	0.138*** (0.0121)
<i>migrant_first</i>					-0.0583*** (0.0169)
Konstant	1,937*** (0,0319)	1,644*** (0,0424)	1,869*** (0,0451)	1,939*** (0,0446)	1,923*** (0,0447)
Observasjoner	5087	5087	5087	5087	5087
R ²	0,231	0,247	0,271	0,289	0,292

Standardavvik i parentes

*** p<0,01, ** p<0,05, * p<0,1

Tabell 5.2.1: Estimert forskjell i lønnsavkastning av utdanning mellom Norge og USA basert på modellene A-E. Estimeringsmetode: OLS.

I **modell A** får vi den generelle avkastningen på utdanning, gitt ved *yrsqual*, som ligger på 0.0716. Det betyr at i Norge og USA vil ett ekstra år utdanning, alt annet likt, gi en gjennomsnittlig økning i lønna tilsvarende 7.16%. Videre forteller dummyvariabelen *norway* oss at lønnsnivået i Norge er 18.5% høyere enn lønnsnivået i USA. Funnene er statistisk signifikante på 1% signifikansnivå.

I **modell B** har vi inkludert interaksjonsleddet, *yrsqual_norway*, som viser avkastningen av ett års ekstra utdanning i Norge med USA som referansepunkt. Den estimerte koeffisienten foran interaksjonsleddet forteller oss at effekten av utdanning på lønn er $(-0.0442 \cdot 100) = 4.4\%$ lavere i Norge enn i USA. β_3 angir med andre ord avviket mellom landenes lønnsavkastning på utdanning. Avkastningen av utdanning i Norge finner vi på følgende måte:

$$\frac{\partial \text{learnhrppp}}{\partial \text{yrsqual}} \text{ if norway} = \beta_1 + \beta_3 \quad (5.2.3)$$

Dette gir en lønnsavkastning på $(-0.0442+0.0921) = 0.0479$. Det betyr at ett års ekstra utdanning i Norge gir en økning i lønna på 4.8%. I modell B og utover vil USAs lønnsavkastning gis av β_1 , altså *yrsqual*, og tilsvarer her 9.2%. Vi ser altså at effekten av utdanning på lønna er nesten dobbelt så høy i USA sammenliknet med Norge. Dette er et svært interessant funn for vår oppgave, og kan i stor grad underbygges av tidligere litteratur. Vi må imidlertid teste om funnet er statistisk signifikant gjennom en t-test. Vi tester følgende hypotese om at det ikke finnes noen forskjell i lønnsavkastningen på utdanning mellom Norge og USA, mot alternativhypotesen om at det finnes en forskjell i lønnsavkastningen på utdanning mellom landene.

$$H_0: \beta_3 = 0, \quad H_A: \beta_3 \neq 0 \quad (5.2.4)$$

Benytter oss av testobservatoren:

$$TS = \frac{b_3 - \beta_3}{s_{b_3}} \sim t_{n-k} \quad (5.2.5)$$

Ved 5% signifikansnivå vil kritisk verdi tilsvare 1.960 når vi har 5101 observasjoner og fire parametere.

$$t_{5087,0.025} = |1.960| \quad (5.2.6)$$

Vi tar i første omgang utgangspunkt i modell B, og får da fra *tabell 5.2.1*:

$$TS = \frac{-0.0442}{0.00426} = |10.38| \quad (5.2.7)$$

Ser at $TS > \text{Kritisk verdi}$, hvilket betyr at vi forkaster nullhypotesen ved 5% signifikansnivå. Ved å kjøre en t-test direkte i Stata, finner vi at vi kan avvise nullhypotesen ved 1% signifikansnivå. Kjører vi den samme t-testen for modell C, D og E får vi en testobservator på henholdsvis 9.8, 9.7 og 9.6, hvilket betyr at vi kan forkaste nullhypotesen i alle modellene. Det er altså en forskjell i lønnsavkastningen på utdanning mellom Norge og USA. Om forskjellen i lønnsavkastningen på utdanning mellom de to landene kan skyldes ulik grad av fagorganisering, som vi så i *tabell 3.2.1* og *3.2.2*, er uvisst men en trolig forklaring.

I modell C, D og E inkluderer vi tre kontrollvariabler, som ansees som relevante for sammenhengen mellom lønn og utdanning for å unngå spuriøse sammenhenger. I **modell C** har jeg lagt til variabelen *litscore1*, som er den standardiserte versjonen av leseferdigheter. Den estimerte koeffisienten til β_4 forteller oss hvor mye lønna øker når vi scorer 10 poeng ekstra på testen for leseferdigheter. Økte leseferdigheter, alt annet likt, gir altså en økning i lønna tilsvarende 8.4% på generell basis. Inkludering av *litscore1* gir en svakere absoluttverdi til β_3 , som nå er lik -0.0413. Dette tyder på at koeffisienten for leseferdigheter fanger opp noe av variasjonen som ble forklart av utdanning i modell B. Dette er et eksempel på utelatt-

variabelskjevhet. Vi fant tidligere at leseferdigheter og utdanning har en relativ høy korrelasjon. Dermed vil deler av effekten leseferdigheter har på timelønna plukkes opp av utdanning gjennom restleddet, hvilket gjør deler av sammenhengen mellom utdanning og lønn spurios. Lønnsavkastningen på utdanning i Norge vil i denne modellen estimeres til å være $(-0.0413+0.0755) = 0.0342$, altså 3.4%. USAs lønnsavkastning på utdanning har fra modell B til C falt fra 9.2 til 7.5%, hvilket igjen tyder på at leseferdigheter fanger opp deler av variasjonen i lønn som tidligere ble forklart av utdanning. Igjen må vi teste om funnet er statistisk signifikant. Før jeg gjør dette vil jeg først presentere de øvrige kontrollvariablene, for så å gjennomføre en multippel hypotesetest ved bruk av en f-test. For å være i stand til å gjøre dette har jeg derfor ekskludert enkelte observasjoner i modellene A-D (ved bruk av følgende kommando `if!missing(migrant_first)`) slik at vi har like mange observasjoner i alle modellene.

I **modell D** har jeg lagt til variabelen *numscore1*, samt ekskludert *litscore1*, for å unngå eventuelle problemer med multikollinearitet. Også her ser vi på den standardiserte versjonene av tallforståelse. Den estimerte koeffisienten til β_5 forteller oss hvor mye lønna øker når vi scorer 10 poeng ekstra på testen for tallforståelse. Vi finner en svært sterk effekt av tallforståelse på lønna, der økt tallforståelse medfører at lønna øker med 11.1%. Igjen ser vi at inkludering av *numscore1* gir en svakere absoluttverdi til β_3 , som nå er lik -0.0404 . Dette kan gi indikasjoner på at koeffisienten for tallforståelse fanger opp noe av variasjonen som ble forklart av utdanning i modell B. Vi får en større nedgang i β_3 fra modell B til modell D, sammenliknet med nedgangen fra modell B og modell C. Dette, sammen med den høyre koeffisienten til β_5 , indikerer at tallforståelse har stor innvirkning på lønna. Her vil lønnsavkastningen på utdanning i Norge estimeres til å være $(-0.0404+0.0702) = 0.0298$, altså 2.9%, mot USAs lønnsavkastning på 7%

Til slutt har jeg i **modell E** inkludert begge ferdighetsmålene, i tillegg til den siste kontrollvariabelen *migrant_first*. Estimeringskoeffisienten til β_6 gir oss effekten av å være førstegenerasjons innvandrere på timelønna. Denne ligger på -0.0583 , hvilket betyr at innvandrere tjener 5.8% mindre enn en innfødt nordmann eller amerikaner. Som nevnt er både *litscore1* og *numscore1* lagt til, der estimeringskoeffisienten til sistnevnte øker til 13.8%. Et noe urovekkende tegn er at estimatorene til *litscore1* endres fra 0.0835 til -0.0396 fra modell C til modell E. Effekten av leseferdigheter har med andre ord gått fra å ha en positiv effekt på timelønna, til å ha en negativ effekt, tilsvarende -3.9% , på lønna. Selv om funnet er statistisk signifikant på 1% signifikansnivå, anser jeg det som lite plausibelt at en 10 poeng økning i leseferdigheter gir en nedgang i lønna på 3.9%. Det er derfor verdt å spørre om estimeringskoeffisienten er feilberegnet som følge av problemet med multikollinearitet. Vi ser til slutt at vi får nok en nedgang i absoluttverdien til estimeringskoeffisienten β_3 som i denne modellen tilsvarer 0.0399. Lønnsavkastningen på utdanning i Norge blir da $(-0.0399+0.0719) = 0.032$, altså 3.2%, mot USAs lønnsavkastning på 7.2%. Forskjellen i landenes lønnsavkastning på utdanning har minket når vi inkluderer alle kontrollvariablene.

Ønsker nå å teste hvorvidt tallforståelse, leseferdigheter og innvandringsstatus har noen innvirkning på timelønna. Tester følgende nullhypotese om at tallforståelse, leseferdigheter og innvandringsstatus ikke har noen innvirkning på timelønna, mot alternativhypotesen om at variablene har en innvirkning på timelønna:

$$H_0: \beta_4 = \beta_5 = \beta_6 = 0, \quad H_A: \beta_4 \neq 0, \beta_5 \neq 0 \text{ eller } \beta_6 \neq 0 \quad (5.2.8)$$

Vi bruker følgende testobservator:

$$TS = \frac{R_U^2 - R_R^2 / h}{1 - R_U^2 / (n-k)} \sim F_{h, n-k} \quad (5.2.9)$$

Forklaringskoeffisienten til den restriktive modellen, R_R^2 , får vi fra modell B, mens forklaringskoeffisienten til den ikke-restriktive modellen, R_U^2 , gis av modell E. Vi har 5087 observasjoner, en restriksjon om at tre parametre er lik 0, totalt 7 parametre og et signifikansnivå lik 5%. Fra f-fordelinga finner vi kritisk verdi på 2.6049. Får da:

$$TS = \frac{(0.292 - 0.247) / 3}{(1 - 0.292) / (5087 - 7)} = 107.63 > 2.6049 \quad (5.2.10)$$

Vi forkaster nullhypotesen ved 5% signifikansnivå, hvilket vil si at tallforståelse, leseferdigheter og innvandringsstatus har en innvirkning på timelønna.

5.3 Analyse av tilleggsspørsmål

Modell	F	G	H
Variabler	<i>learnhrppp</i>	<i>learnhrppp</i>	<i>learnhrppp</i>
<i>yrsqual</i>	0.0695*** (0.00349)	0.0643*** (0.00345)	0.0715*** (0.00311)
<i>norway</i>	0.646*** (0.0706)	0.644*** (0.0699)	0.797*** (0.0621)
<i>yrsqual_norway</i>	-0.0302*** (0.00483)	-0.0297*** (0.00479)	-0.0402*** (0.00415)
<i>litscore1</i>	0.110*** (0.0102)		
<i>migrant_first</i>	-0.0701*** (0.0170)	-0.0573*** (0.0168)	-0.00860 (0.0246)
<i>litscore1_norway</i>	-0.0561*** (0.0129)		
<i>numscore1</i>		0.137*** (0.0101)	0.103*** (0.00662)
<i>numscore1_norway</i>		-0.0545*** (0.0130)	
<i>migrant_first_norway</i>			-0.0791** (0.0327)
Konstant	1.960*** (0.0497)	2.026*** (0.0491)	1.922*** (0.0450)
Observasjoner	5087	5087	5087
R ²	0.276	0.293	0.291

Standardavvik i parentes

*** p<0,01, ** p<0,05, * p<0,1

Tabell 5.3.1: Estimert forskjell i lønnsavkastning av utdanning mellom Norge og USA basert på modellene F-H. Fullstendige estimeringsresultater i Appendiks A1. Estimeringsmetode: OLS.

Som nevnt har vi i **modell F** inkludert interaksjonsleddet *litscore1_norway*, som gir oss lønnsavkastningen av leseferdigheter i Norge med USA som referansepunkt. I dette tilfellet vil *litscore1* gi oss lønnsavkastningen av leseferdigheter i USA. Estimeringskoeffisienten β_7 forteller oss at lønnsavkastningen av leseferdigheter i Norge ligger på 5.6% lavere enn i USA. Lønnsavkastningen på leseferdigheter i Norge finner vi ved addere koeffisienten for β_4 og β_7 , hvilket gir følgende $(-0.0561+0.110) = 0.0539$, altså 5.4% mot USAs avkastning på 11%. Vi får i likhet med avkastningen på utdanning, et liknende resultat ved leseferdigheter som de vi fant i de tidligere modellene. Vi ser også samme tendenser til at estimeringskoeffisienten til

yrsqual_norway minker når ytterligere variabler legges til. I dette tilfellet vil avkastningen på utdanning kun være 3% lavere i Norge sammenliknet med USA. Lønnsavkastningen på utdanning i Norge blir her $(-0.0302+0.0695) = 0.0393$, altså 3.9% mot USAs lønnsavkastning på utdanning på 6.9%. Igjen ser vi en gradvis reduksjon i avviket mellom landenes avkastning på utdanning. Ønsker nå å teste om funnene er statistisk signifikante gjennom en f-test. Benytter samme fremgangsmåte som i likning (5.2.8). Vi tester følgende:

$$H_0: \beta_3 = \beta_7 = 0, \quad H_A: \beta_3 \neq 0 \text{ eller } \beta_7 \neq 0 \quad (5.3.1)$$

Kjører vi i testen i Stata finner vi at testobservatoren ligger på 157.87. Mot den kritiske verdien på 2.9957 kan vi med det forkaste nullhypotesen om at utdanning i Norge og leseferdigheter i Norge ikke har noen betydning for timelønna.

Videre ser vi på effekten av tallforståelse på timelønna i Norge, gitt ved *numscore1_norway*, i **modell G**. Resultatene her er svært like som i modell F, der lønnsavkastningen av tallforståelse i Norge, gitt ved β_8 , er 5.4% lavere sammenliknet med USA.

Lønnsavkastningen på tallforståelse i Norge finner vi ved addere koeffisienten for β_5 og β_8 , hvilket gir følgende $(-0.0545+0.137) = 0.0825$, altså 8.2% mot USAs avkastning på 13.7%.

Det er tydelig at tallforståelse har en stor innvirkning på lønna, gitt ved både de høye estimeringskoeffisientene for begge landene i modell G, og sett ut i fra forklaringskoeffisienten som er høyest i modell D og G der variablene for tallforståelse er inkludert. Estimeringskoeffisienten β_3 , som i denne modellen tilsvarer på -2.9%, har samlet sett nådd sitt laveste nivå, hvilket igjen underbygger mitt poeng om at tallforståelse har stor betydning for timelønna. Lønnsavkastningen på utdanning i Norge er her $(-0.0297+0.0643) = 0.0346$, altså 3.5% mot USAs lønnsavkastning på utdanning som tilsvarer 6.4%. Hanushek m.fl. (2015, p. 114) finner liknende resultater, der avkastningen på tallforståelse er omtrent dobbelt så høy som avkastningen på utdanning. Igjen tester vi en hypotese ved bruk av en f-test. Vi tester følgende:

$$H_0: \beta_3 = \beta_8 = 0, \quad H_A: \beta_3 \neq 0 \text{ eller } \beta_8 \neq 0 \quad (5.3.2)$$

Finner at $TS = 222.74$, mot kritisk verdi på 2.9957. Vi forkaster nok en gang nullhypotesen om at utdanning i Norge og tallforståelse i Norge ikke har betydning for timelønna.

Til slutt har vi i **modell H** inkludert et interaksjonsledd som ser på effekten av å være førstegenerasjons innvandrere i Norge sammenliknet med å være førstegenerasjonsinnvandrere i USA. Ettersom vi tidligere oppdaget noen ualmennelige estimatorer for koeffisienten til leseferdigheter, og det faktum at tallforståelse har vist å ha stor betydning for lønna, har jeg i denne modellen valgt å inkludere sistnevnt og ekskludere førstnevnte. Vi ser i første omgang at estimeringskoeffisienten til β_9 , ligger på -0.0791. Dette forteller oss at lønnsgapet mellom førstegenerasjonsinnvandrere og innfødte i Norge er 7.9% større en lønnsgapet mellom innvandrere og innfødte i USA. Sett ut i fra timelønna, er det med andre ord bedre å være innvandrere i USA enn i Norge. Dette er statistisk signifikant på 5% signifikansnivå. Adderer vi β_6 og β_9 finner vi at effekten av å være innvandrere i Norge på lønna er -8.8%, mot USAs

effekt som ligger på -0.85%. Vi ser imidlertid at estimeringskoeffisienten til β_6 ikke er statistisk signifikant på 5% signifikansnivå, hvilket betyr at innvandringsstatus ikke har noen betydning for lønna i USA. Kjører vi en t-test i Stata, finner vi at testobservatoren til *migrant_first* ligger på 0.12, som er langt under kritisk verdi på 1.960. Videre ser vi at β_3 har gått tilbake til nesten tilsvarende initial verdi som i de først modellene. Lønnsavkastningen på utdanning i Norge er i denne modellen estimert til å tilsvare $(-0.0402+0.0715) = 0.0313$, som vil si at ett års ekstra utdanning gir en økning i lønna på 3.1%, mot USAs lønnsavkastning på 7.2%. Tester nullhypotesen om lønnsavkastningen av utdanning og innvandringsstatus ikke er forskjellig mellom Norge og USA:

$$H_0: \beta_3 = \beta_9 = 0, \quad H_A: \beta_3 \neq 0 \text{ eller } \beta_9 \neq 0 \quad (5.3.3)$$

Vi finner at $TS = 214.95$. Mot kritisk verdi på 2.9957 kan vi forkaste nullhypotesen om at utdanning i Norge og innvandringsstatus i Norge ikke har betydning for timelønna.

5.4 Diskusjon

Det store spørsmålet etter denne analysen er *hvorfor USA har større lønnsavkastning på utdanning og ferdigheter sammenliknet med Norge*. Som nevnt tidligere representerer Norges sammenpressede lønnsstruktur og USAs desentraliserte forhandlingsmodell to vidt forskjellige inntektspolitiske tilnærminger. Hanushek m.fl. (2015) finner solide indikatorer på at høyere grad av fagorganisering, strengere arbeidslovgivning og større offentlige sektorer har en negativ innvirkning på lønnsavkastningen av ferdigheter. Det er verdt å nevne at studien kontrollerer for en rekke andre faktorer, derav blant annet industritilhørighet. Industri- og sektortilhørighet er særlig interessant, ikke bare fordi vi vet at lønnsavkastningen er høyere i privat sektor i forhold til offentlig sektor, men fordi det kan gi en viss forklaring på hvorfor den generelle fagforeningsoppslutningen gradvis går ned. I Norge har nærings sammensetningen endret seg ved at flere jobber innen service og færre innen vareproduksjon, der sistnevnte ofte er karakterisert ved høy fagorganisering og motsatt ved førstnevnte (Nergaard, Barth, & Dale-Olsen, 2015). Vi ser de samme tendensene i USA, samt i store deler av den vestlige verden. En interessant problemstilling for videre forskning hadde vært å undersøke hvordan lønnsavkastningen varierer mellom ulike industrier, og om den endrede nærings sammensetningen gjennom årene har ført til en lavere lønnsavkastning på utdanning som følge av lavere fagforeningsoppslutning. Til tross for at variablene for fagorganisering ikke er inkludert i regresjonsanalysen, finner vi store forskjeller i lønnsavkastningen på både utdanning og ferdigheter mellom Norge og USA. Funnene er stort sett statistisk signifikante, også når vi kontrollerer for øvrige forhold.

Et annet interessant spørsmål er *hvorfor tallforståelse har en større effekt på timelønna enn leseferdigheter*. En mulig forklaring er at bedrifter og næringer som etterspør realfagskompetanse ofte er karakterisert ved høyere gjennomsnittslønn enn andre næringer. Prestisjetunge yrker som lege, sivilingeniør og siviløkonom er alle kjennetegnet ved høy gjennomsnittslønn og høy grad av kompetanse innen realfag og matte.

Til slutt kan man stille spørsmålet om *hvorfor førstegenerasjons innvandrere tjener mindre enn innfødte, samt hvorfor vi finner et tilsynelatende avvik mellom innvandrere og innfødte i Norge sammenliknet med innvandrere og innfødte i USA*. Mulige mekanismer og forklaringer kan være manglende språkkunnskaper, eller lavt utdanningsnivå, hvilket reduserer sysselsettingsmulighetene betraktelig (NOU, 2011, p. 127). Videre trekkes det frem i NOU rapporten (2011) at dette forsterkes i særlig grad i Norge ettersom få stillinger utlyses uten krav til utdanning. Andre mulige forklaringer kan være diskriminering mot innvandrere på arbeidsmarkedet.

Svakheter ved oppgaven og av relevante variabler som har blitt utelatt er det verdt å trekke frem sektor- og industritilhørighet, arbeidserfaring og organisasjonsgrad. Det er en fin balanse mellom å inkludere relevante variabler, for å unngå utelatt variabelskjevhet, men samtidig ikke trekke inn for mange variabler som potensielt kan korrelere og dermed lede til multikollinearitet. Vi kan imidlertid se ut i fra forklaringskoeffisienten til de ulike modellene, at de inkluderte variablene kun forklarer omkring 23-29% av variasjonen i timelønna. Et annet interessant aspekt er hvorvidt sammenhengen mellom utdanning og lønn best forklares ved å kvadrere variabelen for utdanning. Ved å gjøre dette ser vi på lønn som en positiv, men avtakende funksjon av utdanning. Dette kan også gjøres ved å se på forskjellen i avkastningen på utdanning ved å generere ulike dummyvariabler for ulike nivåer på utdanningsløpet. På den måten kan vi undersøke forskjellen i lønnsavkastningen på grunn-, videregående-, og høyere utdanning i de to landene.

6. Oppsummering

I denne oppgaven har jeg undersøkt om hvorvidt vi observerer forskjeller i lønnsavkastningen på utdanning og ferdigheter mellom Norge og USA. Landene ble valgt på grunnlag av to vidt forskjellige inntektspolitiske tilnærminger, der Norges tilnærming reflekteres i høy organisasjonsgrad, mens USA hviler på en mer desentralisert forhandlingsmodell. Hovedspørsmålet er altså *om* det fremkommer ulikheter i avkastningen på utdanning mellom landene, og *hvorfor* vi observerer en eventuell forskjell. Selv om avkastningen på utdanning har vært av primærinteresse, har jeg i tillegg valgt å undersøke forskjeller i avkastningen på ferdigheter mellom landene. Mer spesifikt har jeg derfor undersøkt hvorvidt vi observerer forskjeller i lønnsavkastningen på tallforståelse og leseferdigheter mellom Norge og USA. Til slutt har jeg kontrollert for etnisitet eller opprinnelse, ved bruk av en dummyvariabel for innvandrere og innfødte. For å kunne besvare disse spørsmålene har jeg valgt å presentere et teoretisk rammeverk, bestående av tidligere studier om temaet og litt om relevant økonomisk teori. Jeg har benyttet meg av et utvalg av datamateriale fra PIAAC-undersøkelsen, og ved bruk av dataprogrammet Stata, kjørt en trinnvis regresjonsanalyse basert på åtte egenkomponerte lønnslikninger. Estimeringsmetoden som har blitt benyttet er minste kvadraters metode.

6.1 Hovedfunn i oppgaven

- Resultatet fra de estimerte lønnsmodellene viser at Norge har en lønnsavkastning på utdanning som er klart lavere enn lønnsavkastningen på utdanning i USA. Avviket er statistisk signifikant. Lønnsavkastningen på utdanning i Norge varierer mellom å være 2.9 til 4.2% lavere enn lønnsavkastningen på utdanning i USA når vi kontrollerer for ulike variabler.
- Alle hypotesene presentert i delkapittel 4.4 ble testet ved bruk av enten t-test eller f-test. Alle funnene var statistisk signifikante på 5% signifikansnivå.
- Norges lønnsavkastning på utdanning er lavest i modell D, når vi kontrollerer for tallforståelse, hvilket gir indikasjoner på at tallforståelse har stor effekt på lønna.
- I det generelle tilfellet gir tallforståelse størst effekt på lønna, tilsvarende 11.1%, mens leseferdigheter gir en effekt på lønna tilsvarende 8.4%.
- Generelt tjener innvandrere 5.8% mindre i timen enn innfødte.
- Også når vi sammenlikner lønnsavkastningen på ferdigheter, finner vi at Norge har et lavere nivå enn USA. Lønnsavkastningen på leseferdigheter i Norge er 5.6% lavere, mens avkastningen på tallforståelse er 5.4% lavere.
- I det spesifiserte tilfellet vil lønnsavkastningen på leseferdigheter i Norge være 5.4%, mot USAs avkastning på 11%. Videre er lønnsavkastningen på tallforståelse lik 8.2% i Norge og 13.7% i USA.
- Lønnsgapet mellom innvandrere og innfødte i Norge er 7.9% større enn lønnsgapet mellom innvandrere og innfødte i USA. Innvandrersstatus synes dog å ha liten betydning for lønnsavkastningen på utdanning i Norge, og ingen betydning for lønna i USA.

7. Litteraturliste

- Barth, E., Jon Marius, V. I., Schöne, P., von Simson, K., & Strøm, B. (2016). Sammenhengen mellom ferdigheter og arbeidsmarkedsituasjon for utsatte grupper. In *Skills and labour market attachment for vulnerable groups*: Institutt for samfunnsforskning.
- Becker, G. S. (1964). Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education. *University of Chicago Press*.
- Bjørkeng, B. (2013). *Ferdigheter i voksenbefolkningen : resultater fra den internasjonale undersøkelsen om lese- og tallforståelse (PIAAC)*(Vol. 42/2013).
- Hanushek, E. A., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, 73(C), 103-130. doi:10.1016/j.euroecorev.2014.10.006
- Hanushek, E. A., & Woessmann, L. (2008). The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, 46(3), 607-668. doi:10.1257/jel.46.3.607
- Nergaard, K., Barth, E., & Dale-Olsen, H. (2015). Lavere organisasjonsgrad, et spørsmål om nykommere? *Søkelys på arbeidslivet*, 32(01-02), 91-110. Retrieved from http://www.idunn.no/spa/2015/01-02/lavere_organisasjonsgrad_et_spoersmaal_om_nykkommere_.
- NOU. (2011). *Bedre integrering – Mål, strategier, tiltak*. Retrieved from Oslo: <https://www.regjeringen.no/contentassets/f5f792d1e2d54081b181655f3ca2ee79/no/pdfs/nou201120110014000dddpdfs.pdf>
- NOU. (2013). *Lønnsdannelsen og utfordringer for norsk økonomi*. Retrieved from Oslo: <https://www.regjeringen.no/contentassets/7682d9ecd19b4794bb687c160a020f07/no/pdfs/nou201320130013000dddpdfs.pdf>
- Thomas, R. L. (2005). *Using statistics in economics*. London: McGraw Hill.
- Tønnessen, M. (2018). Innvandrere. *Store norske leksikon*. Retrieved from 6. mai 2019 <https://snl.no/innvandrere>.
- World Bank. (1995). *Priorities and strategies for education: a World Bank review* (1-280-01718-X,9786610017188,0-585-16217-4). Retrieved from Washington, D.C.: <http://documents.worldbank.org/curated/en/117381468331890337/pdf/multi-page.pdf>

Appendiks

Modell	A	B	C	D	E	F	G	H
Variabler	<i>learnhrpp</i>	<i>learnhrppp</i>	<i>learnhrpp</i>	<i>learnhrpp</i>	<i>learnhrpp</i>	<i>learnhrpp</i>	<i>learnhrppp</i>	<i>learnhrpp</i>
	<i>p</i>		<i>p</i>	<i>p</i>	<i>p</i>	<i>p</i>		<i>p</i>
<i>yrsqual</i>	0.0716*** (0.00215)	0.0921*** (0.00290)	0.0755*** (0.00312)	0.0702*** (0.00309)	0.0719*** (0.00311)	0.0695*** (0.00349)	0.0643*** (0.00345)	0.0715*** (0.00311)
<i>norway</i>	0.185*** (0.0114)	0.833*** (0.0635)	0.803*** (0.0625)	0.791*** (0.0617)	0.782*** (0.0617)	0.646*** (0.0706)	0.644*** (0.0699)	0.797*** (0.0621)
<i>yrsqual_norway</i>		-0.0442*** (0.00426)	-	-	-	-	-0.0297*** (0.00479)	-
<i>litscore1</i>			0.0413*** (0.00420)	0.0404*** (0.00415)	0.0399*** (0.00414)	0.0302*** (0.00483)		0.0402*** (0.00415)
			0.0835*** (0.00634)		-	0.110*** (0.0102)		
<i>numscore1</i>				0.111*** (0.00634)	0.138*** (0.0121)		0.137*** (0.0101)	0.103*** (0.00662)
<i>Migrant_first</i>					-	-	-0.0573*** (0.0168)	-0.00860 (0.0246)
<i>litscore1_norway</i>						-	0.0561*** (0.0129)	
<i>numscore1_norway</i>							-0.0545*** (0.0130)	
<i>migrant_first_norway</i>								-0.0791** (0.0327)
Konstant	1.937*** (0.0319)	1.644*** (0.0424)	1.869*** (0.0451)	1.939*** (0.0446)	1.923*** (0.0447)	1.960*** (0.0497)	2.026*** (0.0491)	1.922*** (0.0450)
Observasjoner	5,087	5,087	5,087	5,087	5,087	5,087	5,087	5,087
R ²	0.231	0.247	0.271	0.289	0.292	0.276	0.293	0.291

Standardavvik i parentes

*** p<0,01, ** p<0,05, * p<0,1

Tabell A1: Estimert forskjell i lønnsavkastning av utdanning mellom Norge og USA basert på modellene A-H. Estimeringsmetode: OLS.

