



Påvirker helsetilstand avkastning av utdanning på lønn?

Empirisk prosjekt med PIAAC-datasett
Norges teknisk-naturvitenskapelig universitet
Institutt for samfunnsøkonomi

FORORD

Denne bacheloroppgaven inngår i det 3-årige bachelorprogrammet for samfunnsøkonomi ved NTNU i Trondheim. Bacheloroppgaven er skrevet i avsluttende semester for bachelorprogrammet og muliggjør at studentene kan anvende tilegnet kunnskap fra foregående emner. Det er gjennomført et empirisk prosjekt der en selvvalgt samfunnsøkonomisk problemstilling er undersøkt og testet. Oppgaven er utarbeidet i et samarbeid mellom fire studenter ved Institutt for samfunnsøkonomi.

Vi ønsker å takke Professor Bjarne Strøm for god oppfølging rundt valg av problemstilling, samt Stipendiat Haakon Andreas Trønnes og Vitenskapelig assistent Turid Alexandra Barkald for god hjelp underveis i prosjektet.

Norges teknisk-naturvitenskapelige universitet, Trondheim

Våren 2019

Innhold

Forord.....	1
1 Innledning	3
2 Tidligere studier, teori og modell.....	4
3 Presentasjon av data.....	5
3.1 Ulemper ved datasettet.....	8
3.2 Fordeler ved datasettet	8
4 Regresjonsanalyse.....	9
4.1 Generell modell.....	9
4.2 Valg av variable	12
4.3 Økonometriske utfordringer.....	13
4.4 Hypotesetesting.....	14
4.5 Empiriske resultater	15
4.5.1 Er lønnsnivået avhengig av utdannelsesnivået?	15
4.5.2 Påvirker helsetilstanden (isolert sett) lønnsnivået?	15
4.5.3 Påvirker helsetilstand avkastningen av utdanning på lønn?	16
4.5.4 Er lønnsgapet forskjellig for kvinner og menn med god og dårlig helse?	16
5 Oppsummering.....	17
Litteraturliste.....	18

1 INNLEDNING

Formålet med denne oppgaven er å gjennomføre empiriske tester med utgangspunkt i datasett som kan knyttes opp mot samfunnsøkonomiske problemstillinger. Disse problemstillingene skal undersøkes ved hypotesetesting.

I denne oppgaven vil det bli presentert en lønnsanalyse som ser på sammenhenger mellom helsetilstand, utdanningsnivå og lønnsnivå. Vi ønsker å undersøke hvordan avkastningen av utdanningsnivå, målt i vedkommendes lønnsnivå, kan variere ut ifra helsetilstanden til individet. Høyere utdanning kan anses som en investering i humankapital som gjør at man som individ får økt kompetanse (Nelson og Phelps, 1966). Utdanning vil gi tapt inntekt i studieårene, men vil på sikt, med stor sannsynlighet, generere høyere inntekt sammenlignet med et individ som ikke har tatt utdanning (Hægeland, 2003). Beslutningen om å ta utdanning kan også tenkes å være påvirket av individets helsetilstand. Derfor lyder vår hovedproblemstilling som følger:

Påvirker helsetilstand avkastningen av utdanningen på lønn?

Ser først på hvordan utdanning påvirker lønnsnivå, og dernest hvordan helsetilstand (isolert sett) påvirker timelønnen. Problemstillingen vil videre bli belyst med å se på hvordan helsetilstand kan være en faktor av betydning for lønnsvariasjon mellom kjønn. Tall fra SSB¹ viser at på overordnet nivå tjener menn mer enn kvinner (SSB, 2019). Følgende spørsmål stilles for å gi et bedre perspektiv på problemstillingen:

Er lønnsgapet mellom kvinner og menn forskjellig for personer med god og dårlig helse?

For å besvare problemstillingen bruker vi innsamlet data fra PIAAC²-undersøkelsen. Dette er en internasjonal undersøkelse iverksatt av OECD³ (Bjørkeng, 2013). Målet med undersøkelsen er å studere ferdighetsnivå på tvers av land. Vi begrenser vår problemstilling til å gjelde kun for norske respondenter. For å løse oppgaven bruker vi dataprogrammet Stata, som er et analyseverktøy for datasett.

1 Statistisk sentralbyrå

2 Programme for the International Assessment of Adult Competencies

3 Organisation for European Economic Co-operation

2 TIDLIGERE STUDIER, TEORI OG MODELL

Det er gjort flere studier som viser sammenheng mellom avkastning av utdanning på lønn, samt på sammenhengen mellom utdanningsnivå og helsetilstand. Aktuell forskning som presenteres her gjelder norske undersøkelser, da dette vil være sammenlignbart med vår problemstilling som omfatter norske respondenter.

I artikkelen «Den samfunnsmessige avkastningen av utdanning» av samfunnsøkonomen Erling Barth, vises det til en klar sammenheng mellom at høyt utdannede personer også opplever en god helsetilstand. Denne sammenhengen er basert på at sykefravær blant de som kun har grunnskole er mer enn dobbelt så høy sammenlignet med personer med høyere utdanning (Barth, 2005). Årsakssammenhengen er ikke helt klar på dette feltet. Det kan tenkes at personer med dårlig helse ikke har mulighet for å fullføre høyere utdanning, og derav får man en sortering der høyt utdannede personer i snitt har bedre helse (Barth, 2005). Ønsker å se nærmere på denne årsakssammenhengen i vår økonometriske modell.

I en annen artikkel utgitt fra SSB, «Økonomisk avkastning av utdanning», av Torbjørn Hægeland, betraktes den individuelle avkastning av utdanning målt i lønnsnivå. Ved bruk av Mincer-modellen viser studien at en person med høy utdanning tjener mer sammenlignet med personer uten utdanning (Hægeland, 2003). Artikkelen tar for seg hvordan avkastningen av utdanning kan variere ut ifra faktorer som sektor (privat eller offentlig), yrkesgrupper og arbeidserfaring.

Naper og Dahl drøfter i artikkelen “Sosialhjelpsmottakerens dødelighet: I hvilken grad kan overdødeligheten tilskrives deres sosiale status?” hvordan lavtlønnede personers selvbilde påvirkes av deres eget lønnsnivå. Dødelighet betyr i denne sammenheng forventet levelader, og kan ikke sammenlignes uten forbehold med helsetilstand som er vår variabel. Videre mener de at lav inntekt (her: sosialhjelp) kan være helsenedbrytende på grunnlag av én av følgende mekanismer: 1) på grunnlag av materiell knappheten eller 2) via psykososiale belastninger ved å være plassert nederst på lønshierarkiet (Naper og Dahl, 2007). I samme studiet drøftes de underliggende faktorer for årsakene til helsenivået hos kvinner og menn. For menn er lønnsnivået mer utslagsgivende på helsenivået enn utdannelsesnivået. For kvinner er det omvendt.

3 PRESENTASJON AV DATA

Benytter datasett fra PIAAC-undersøkelsen. Innsamlingen av data er gjort på initiativ fra Organisasjonen for økonomisk samarbeid og utvikling (OECD) og omfatter 24 land (Bjørkeng, 2013). Formålet med undersøkelsen er å kartlegge de voksnes ferdighetsnivå, innenfor leseferdigheter, tallforståelse og problemløsning innen IKT på tvers av land. Datagrunnlaget er samlet inn i perioden fra august 2011 til april 2012. Voksne er i denne undersøkelsen alle personer over 16 år. I forbindelse med undersøkelsen har respondentene gitt informasjon om blant annet utdanningslengde (antall år), helsetilstand (score fra 1-5), kjønn og alder.

Vi har valgt å begrense lønnsanalysen til å gjelde for nordmenn, og videre personer over 25 år slik at vi utelukker mest mulig studenter med deltidsjobb. Eldste person i undersøkelsen er 65 år og vi velger derfor å ikke sette noen øvre grense på alder, da vi antar at personer er yrkesaktive opp til 65 år (Regjeringen, 2017). På denne måten mener vi at hypotesene er testet på mest reelt datagrunnlag.

Tabell 1 lister opp alle aktuelle variable for denne analysen, med deres tilhørende kodenavn som anvendes i Stata.

Tabell 1: Beskrivelse av relevante variabler

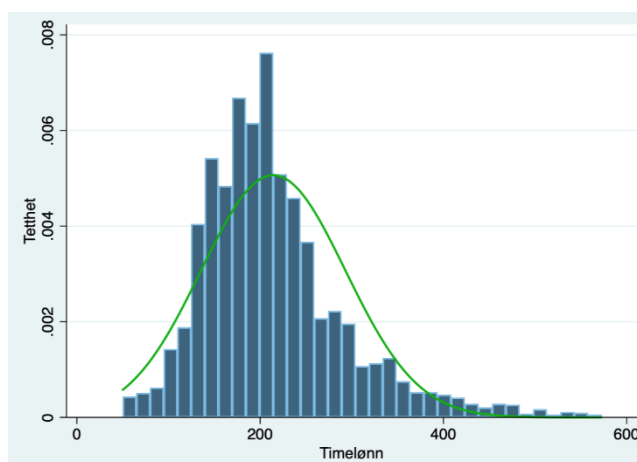
Variabel	Forklaring	Forkortelse
Timelønn	Timelønn i lokal valuta (NOK)	Y
Utdanning	Antall år utdanning	U
Alder	Tar kun med de over 25 år	A
God helse-dummy	Lik 1 hvis god helse (se tabell 3)	GH
Dårlig helse-dummy	Lik 1 hvis dårlig helse (se tabell 3)	DH
Kjønn-dummy	Lik 1 hvis individet er kvinne	K
Andre forklaringsvariabler	Inkluderer flere variabler som antas å påvirke lønn	X_i

X_i representerer forklaringsvariabler som er tilgjengelige fra datasettet som antas å påvirke lønn. Dette er ikke variable som vi undersøker effekten av i egne regresjonsanalyser, men inkluderer i testene som følge av at det gir økt forklaringskraft i modellene (mer om dette i 4.2 *Valg av variable*).

Tabell 2 viser deskriptiv statistikk for de aktuelle variablene i vår oppgave. Gjennomsnittet av Kjønn-variabelen variabelen er 0,48 som vil si at nesten halvparten av de som har besvart undersøkelsen er kvinner. Videre kan man se at den gjennomsnittlige timelønnen er på 227,46 kroner, med et tilhørende standardavvik på 78,78. Den lavest betalte personen som deltok i undersøkelsen fikk kun 50 kroner i timen, men den best betalte fikk nesten 573 kroner i timen.

Tabell 2: Deskriptiv statistikk for relevante variabler

Variabel	Gjennomsnitt	Standardavvik	Max.	Min.
Alder	45,06	11,07	65	26
Utdanning	14,67	2,46	21	7
Lønnsnivå	227,46	76,04	573	50
Helsetilstand	2,53	1,04	1	5
Kjønn (kvinner)	0,48	-	-	-



Figur 1: Lønnsfordeling

Figur 1 viser hvordan lønnsfordelingen ser ut. Tettheten på y-aksen beskriver antall personer av totalen med gitt lønnsnivå. 50% persentilen, altså medianen, er 201,7 kr. Dette er 10% under gjennomsnittsverdien for lønn. Når gjennomsnittsverdien er større enn verdien for medianen sier

man at fordelingen er høyreskjev. Figur 1 illustrerer dette, ved at “halen” på den grønne linja strekker seg mot høyre. Det er altså noen observasjoner som trekker snittet opp.

Variabelen for helsetilstand er tallverdier fra én til fem, der verdiene er definert slik som i tabell 3. Gjennomsnittsverdien til datasettet er 2,31, som betyr at datasettet ligger nærmere god helse enn dårlig helse. Antar at steglengden er den samme mellom alle verdiene. Lager en dummyvariabel for helsetilstand der score 1-3 innføres som «God_helse» og score 4-5 innføres som «Dårlig_helse».

Tabell 3: Kategoriseringen for variabelen helsetilstand

Variabelnavn	Tilstand	Verdi i datasett	Dummyvariabel
i_q08	Excellent	1	
i_q08	Very good	2	God_helse lik 1
i_q08	Good	3	
i_q08	Fair	4	
i_q08	Poor	5	Dårlig_helse lik 1

Tabell 4: Deskriptiv statistikk for timelønn og utdanningslengde ut i fra helsetilstand

Helsetilstand	Gj.snitt timelønn	Gj.snitt ant år utdanning
1 (Excellent)	223,90	14,42
2 (Very Good)	216,55	14,52
3 (Good)	210,00	13,95
4 (Fair)	203,56	13,48
5 (Poor)	184,70	12,93

Tabell 4 viser deskriptiv statistikk for de variabelene helsetilstand og utdanningslengde som vi ønsker å knytte opp mot timelønn. Ser fra tallene at det er en sammenheng mellom at den gjennomsnittlige timelønnen øker med økt nivå* på helsetilstand. Det er ikke en like klar sammenheng mellom utdanningslengde (gj.snitt) og helsetilstand, da respondenter med helsetilstand 2 i snitt har lenger utdanning (antall år) enn personer med helsetilstand 1. Det er likevel en tendens der gjennomsnittlig utdanningslengde øker positivt med en bedret helsetilstand.

*Økt nivå på helsetilstand betyr her at respondenten har lavere score, ettersom den laveste scoren (1) indikerer det beste nivået på helsetilstand-variabelen.

3.1 ULEMPER VED DATASETET

På spørsmålet om timelønn er det kun 2160 av totalt 2781 observasjoner. Det vil si at omtrent 22 % ikke har besvart dette spørsmålet. De andre parameterne har heller ikke en svarprosent lik 100.

En annen ulempe ved datasettet er at type utdanning eller bestått grad ikke tas med i beregningen, da type utdanning vil påvirke lønnsnivået (Hægeland, 2003).

Utdanningsvariabelen er kun oppgitt som antall år man har studert. Forklaringskraften til utdanningsvariabelen vi anvender er derfor ikke tilstrekkelig nok for å forklare timelønna alene.

Variabelen helsetilstand er sentral i vår problemstilling. Dette er en variabel som måler en score fra 1-5 der respondenten blir bedt om å svare på hvordan de opplever sin egen helsetilstand på tidspunktet de er med i undersøkelsen. Det er rimelig å tro at for å kartlegge helsetilstand til et individ på en god og omfattende måte, bør det gjøres en grundigere undersøkelse enn å besvare ett spørsmål på hvordan man føler seg. Verdien 1 representerer «Excellent», altså det beste nivået på helsetilstand. Hvordan en respondent oppfatter hva som er «Excellent» vil være individuelt.

Antyder med dette at ved bruk av de variablene vi har tilgjengelig er det sannsynlig at forklaringskraften til modellen kan bli lav.

3.2 FORDELER VED DATASETET

En fordel med datasettet er at det er et representativt utvalg og et stort antall observasjoner.

Ifølge SSB utgjør andelen kvinner omtrent 49% av befolkningen i Norge. Kvinnene i undersøkelsen utgjør 48%. Dette tallet er nokså representativt for kjønnsfordeling i Norge.

Gjennomsnittlig månedslønn for 3. kvartal i 2011 var 38 100 kroner (SSB, 2019). Antar man 150 timer per måned som er normal arbeidsmengde resulterer dette i en timelønn på 254 kroner, som er omtrent 11,9% høyere enn modellens lønnsnitt.

4 REGRESJONSANALYSE

4.1 GENERELL MODELL

For å besvare oppgaven bruker vi minste kvadraters metode (MKM). Denne estimeringsmetoden går ut på å minimere variansen til det estimerte resultatet. Variansen er gitt ved kvadratet av differansen mellom den estimerte verdien og den faktiske måleverdien. Ved å bruke denne metoden estimerer vi en modell som viser sammenhengen mellom timelønna og om våre relevante variabler har effekt, eventuelt hvilken, på timelønnen.

For å kunne bruke denne metoden må følgende antagelser ligge til grunn:

- Utvalget er tilfeldig og tilstrekkelig stort
- Ingen perfekt korrelasjon mellom forklaringsvariablene
- Restleddets faktorer må være uavhengige av hverandre
- Forventningen til restleddet er null gitt ulike variabelverdier. $E(\varepsilon|X_i) = 0$
- Ingen av faktorene i restleddet kan være korrelert til de andre variablene. $Cov(\varepsilon_i, X_i) = 0$
- Restleddet må være normalfordelt. $\varepsilon \sim N(0, \sigma^2)$
- Restleddet må ha konstant varians. $E(\varepsilon_i^2) = \sigma^2, Var(\varepsilon|X_i) = \sigma^2$

Ved brudd på disse antagelsene vil modellen kunne svekkes (se 4.3 *Økonometriske utfordringer*).

Den generelle funksjonsformen for MKM:

$$Y_i = f(X_{1i}, X_{2i}, \dots, X_{ki}) + \varepsilon_i \quad (\text{A})$$

Y_i er den endogene verdien (timelønn) som antas å være avhengig av de eksogene variablene X_{1i}, \dots, X_{ki} . ε er restleddet som fanger opp de verdier resten av modellen ikke klarer å modellere. Restleddet fanger opp de ikke observerbare effektene på timelønn, og representerer et forbedringspotensial i modellen vår.

Vil benytte en semi-logaritmisk transformasjon av funksjonen der man tar den naturlige logaritmen til den avhengige variabelen (timelønn), mens forklaringsvariabelene holdes lineære. Det generelle uttrykket for dette vil bli slik:

$$\ln(Y_i) = f(X_{1i}, X_{2i}, \dots, X_{ki}) + \varepsilon_i \quad (\text{B})$$

Ved å gjøre en transformasjon gjør det oss i stand til å se på relative endring (prosentvis) for timelønn når forklaringsvariabelene har en absolutt marginalendring. Kan nå presentere regresjonsuttrykkene som er brukt i vår hypotesetesting.

Sammenhengen mellom utdanning og lønn vil dermed bli

$$\ln(Y_1) = \beta_0 + \beta_1 U + \varepsilon_1 \quad (1)$$

Koeffisientene med symbol β uttrykker hvor stor effekt den enkelte forklaringsvariabel har på timelønn. Første ledd, β_0 , er konstantleddet i alle modellene. Denne kan tolkes som en representasjon på timelønn gitt at de andre forklaringsvariabelene er konstante (lik 0). Fra tidligere empiri er det rimelig å anta at β_1 er større enn null, som vil si at det gir en positiv effekt på timelønn ved å ta et år ekstra utdanning (Hægeland, 2003). Vi ønsker å se om helsetilstand kan ha en effekt på avkastningen av utdanning målt i timelønn, samt isolert sett.

$$\ln(Y_2) = \beta_0 + \beta_1 U + \beta_2 GH + \beta_3 U * GH + \varepsilon_1 \quad (2)$$

Ved å legge til interaksjonsleddet $\beta_3 * K * GH$ vil denne effekten kunne testes. Antar at både β_2 og β_3 er positive, ut ifra tabell 5, som viser tendensen om at de med god helse tjener bedre enn de med dårlig helse. Helsetilstand er en innført dummyvariabel hvilket vil si at den tar verdien 1 eller 0. GH tar inn respondentene med god helse, som tidligere definert er for score 1-3 på helsetilstand-scoren. Fra forskning og tall presentert vil kjønn kunne være en variabel av betydning for lønn. Legger til denne ved dummyvariabelen K, som tar verdien 1 for kvinner og 0 for menn. Har her interaksjonsleddet $K * GH$ som skal fange opp samspillseffekten de to dummyvariablene (kjønn og helsetilstand) kan tenkes å ha på hverandre.

$$\ln(Y_3) = \beta_0 + \beta_1 U + \beta_2 GH + \beta_3 U * GH + \beta_4 K + \beta_5 K * GH + \varepsilon_2 \quad (3)$$

$$\ln(Y_4) = \beta_0 + \beta_1 U + \beta_2 GH + \beta_3 U * GH + \beta_4 K + \beta_5 K * GH + \beta_i X_i + \varepsilon_2 \quad (4)$$

X_i er andre forklaringsvariabler som tas med i regresjonsanalysen for at vi skal få økt forklaringskraft i modellen vår, og isolere effekten av den enkelte variabel vi studerer.

Restleddet, ε , vil bli redusert ved å ta inn flere forklaringsvariabler som timelønnen antas å være avhengig av.

Gir en fremstilling av de ulike regresjonsuttrykkene 1-4, med verdier for koeffisientene og forklaringskraften i tabell 5 nedenfor.

Tabell 5: Oversikt over regresjonsuttrykkene med tilhørende verdier

	[1]	[2]	[3]	[4]
Log timelønn				
Utdanning	0,48	0,06	0,05	0,04
God helse		-0,08	-0,02	-0,03
God helse*utdanning		0,01	0,00	0,01
Female			-0,13	-0,06
Female*god helse			-0,03	-0,06
Tallforståelse				0,04
Leseferdigheter				-0,01
Problemløsningsferdigheter				0,39
Alder				0,00
Innvandrerbakgrunn				-0,07
Stillingsbrøk				-0,00
Sektor				-0,07
Lavt studienivå forelder				-0,11
Middels studienivå forelder				0,01
Arbeidserfaring				0,01
Konstant	4,66	4,70	4,73	4,67
R²	0,05	0,13	0,19	0,28
N	2958	2958	2958	2649

4.2 VALG AV VARIABLE

Den regresjonsmodellen (ligning 4) som er valgt for å teste de ulike hypotesene inkluderer variabler som antas å ha påvirkning på timelønnen. I oppgaven har vi kun valgt å fokusere på enkelte av disse variablene, men likevel valgt å ta med flere variabler som kan påvirke timelønn for å få økt forklaringskraft i modellen, samt et bedre grunnlag for å gjøre test av de andre variablene.

De variable som ikke inngår i egen hypotesetest, men som inkluderes er symbolisert med X_i . Her inngår innvandrerbakgrunn, foreldrenes utdanningsnivå, stillingsbrøk (fulltid eller deltid), sektor, arbeidserfaring (antall år i jobb) og ferdigshetsnivå (tallforståelse, leseferdigheter og problemløsning innenfor IKT).

En viktig output for regresjonsanalyse er modellens forklaringskraft, R^2 , som er kvadratet av modellens korrelasjon. Denne outputen gir verdier mellom 0 og 1, og jo nærmere 1 verdien er, jo bedre er modellens forklaringssevne. Dersom $R^2=1$ er det perfekt lineær sammenheng mellom Y og X -verdiene. Alle punktene ligger på regresjonslinjen. Dersom $R^2=0$ er det ingen lineær sammenheng og de endogene variablene har ingen sammenheng med Y -verdien. Ved å legge til flere ledd og vurdere R^2 opp mot leddene vi inkluderer kan vi finne en modell som vi mener best forklarer lønna. Ser at forklaringskraften øker fra 0,05 til 0,28 ved å inkludere flere variabler i regresjonsuttrykket. Velger ligning 4 basert på forklaringskraften.

Vi ser allikevel at timelønna ikke kan forklares utelukkende ved de inkluderte variablene.

4.3 ØKONOMETRISKE UTFORDRINGER

Viser til overnevnte betingelser (se del 4.1 *Generell metode*) som må være tilstede for at modellen kan benyttes på godt grunnlag. Hvis forutsetningene ikke stemmer vil det kunne gi svakheter i modellen.

4.3.1 Heteroskedastisitet

Dersom restleddet ikke har konstant varians har vi heteroskedastisitet: $E(\varepsilon_i^2) \neq \sigma^2 = \text{konstant}$.

For at standardfeil ikke skal oppstå må vi ha homoskedastisitet, altså konstant varians. Da vil observasjonene i restleddet ikke forstyrre og man unngår at resultatene i hypotesetestingen blir feil. Ved å bruke "White heteroscedasticity test" kan man avklare om modellen har homoskedastisitet, og dermed om den kan vurderes å være god nok for videre hypotesetesting (Brooks, 2008). Bruker Stata til å gjennomføre en test for heteroskedastisitet. Formulerer en nullhypotese hvor vi har homoskedastisitet og alternativhypotese hvor vi har heteroskedastisitet. Ser at P-verdien er lik null for alle variable og vi må derfor forkaste nullhypotesen vår, hvilket betyr at noen av forklaringsvariabelene påvirker variansen.

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of ln_earnhr

chi2(1)      =    16.66
Prob > chi2  =    0.0000
```

Figur 1: Utsnitt fra Stata som viser test for homoskedastisitet

Ved å korrigere for heteroskedastisitet vil standardavvikene endre seg noe. Forskjellen mellom de robuste standardfeilene og de vanlige standardfeilene er så små at de ikke vil gi betydelig utfall på testene. Dette er ofte tilfellet ved mange observasjoner som gjelder vårt datasett. Det er med andre ord heteroskedastisitet i våre tester men dette tas ikke høyde for da effekten anses som neglisjerbar.

4.3.2 Multikollinearitet

Et annet problem som kan oppstå er hvis man har multikollinearitet, som gis ved en verdi i intervallet [-1, 1], der 1 indikerer perfekt positiv korrelasjon mellom to variable. Dersom to variable har sterk korrelasjon vil det bli vanskelig å se hvilken variabel som påvirker timelønnen

hver for seg. Korrelasjon mellom variablene kan man enkelt undersøke i Stata, og for vår modell har vi ingen multikollinearitet. Korrelasjonen mellom f.eks alder og utdanning er 0,23. Det er en viss positiv korrelasjon som kan begrunnes med at det tar tid å ta utdanning.

4.4 HYPOTESETESTING

For å konkludere i våre problemstillinger gjennomføres det hypotesetester av regresjonsanalysene. Ved hypotesetesting tester man om det er tilstrekkelig bevis for de observasjonene man ønsker å trekke en konklusjon om. Vi etablerer én nullhypotese og én alternativ hypotese. Vi tester om nullhypotesen er sterk nok til å beholdes, basert på dens underliggende deskriptive statistiske resultater. Dersom nullhypotesen ikke er signifikant må nullhypotesen forkastes. Dermed må alternativhypotesen aksepteres.

Ved en tosidig hypotesetest vil hypotesene se slik ut:

$$H_0: \beta_i = 0 \text{ mot } H_A: \beta_i \neq 0$$

Ved en t-test er testobservatoren gitt ved differansen mellom estimert og faktisk koeffisient delt på koeffisientens standardavvik:

$$TS = \frac{b_i - \beta_i}{s_{bi}}$$

For å verifisere nullhypotesen må tilhørende signifikansnivå til testen avgjøres, uttrykt ved et valgt konfidensintervall. Dette kan også uttrykkes med p-verdi som er det laveste signifikansnivået som gjør at den observerte verdien for testobservatoren er signifikant. Nullhypotesen forkastes derfor ved lave p-verdier og beholdes ved høye.

Ved en tosidig test vil signifikansnivået og dets kritiske verdi være $\pm 1,960$ for 5%-signifikansnivå, og $\pm 2,575$ for 1%.

Ved en tosidig test forkaster man nullhypotesen dersom verdien av testobservatorens absoluttverdi er større enn absoluttverdien av den kritiske verdien. I våre tester velger vi konfidensintervall lik 5%.

4.5 EMPIRISKE RESULTATER

4.5.1 Er lønnsnivået avhengig av utdannelsesnivået?

Her ser vi om leddet for antall år studert vil ha en innvirkning på lønna. Intuisjon, og henvisning til tidligere studier, tilsier at det er positiv korrelasjon mellom disse faktorene: høyere grad av utdanning vil også lede til høyere lønnsnivå. Dette ser ut til å stemme basert på vår regresjonsanalyse, der den derivert av lønn med hensyn på utdanning er et positivt tall.

Modellen vår tilsier at ett års ekstra utdanning skal gi 4% høyere lønn, siden $\frac{\partial(\text{lønn})}{\partial(\text{utdanning})} =$

0,04. Vi tester følgende nullhypotese: Lønnsnivået er uavhengig av utdannelsesnivået. P-verdien for denne hypotesen er lik 0, hvilket vil si at vi kan forkaste nullhypotesen for alle konfidensintervaller. Vi kan derfor konkludere med at utdannelsesnivået har innvirkning på lønnsnivået. Dette stemmer også overens med Hægeland's artikkel om økonomisk avkastning av utdanning.

4.5.2 Påvirker helsetilstanden (isolert sett) lønnsnivået?

Basert på regresjonsanalysen ser vi at det er negativ sammenheng mellom personers helsetilstand og lønnsnivå. I dette tilfellet ser vi på helsefaktoren som en kontinuerlig variabel, fordi vi ønsker å se på kun helsens innvirkning direkte på lønnsnivået. Vi skiller derfor ikke mellom god og dårlig helse, men bruker variabelen slik den er fra datasettet. Siden skalaen for helsenivået er rangert fra 1 til 5, der 1 er den beste helseverdien, gir dette mening, da en avtakning av helsefaktoren betyr bedre helse. Vi tester dette, med nullhypotesen om at lønnsnivået er uavhengig av helsetilstanden. Denne testens p-verdi er 0,003 hvilket betyr at vi kan forkaste nullhypotesen for vårt konfidensintervall. Vi konkluderer med at helsetilstanden isolert sett påvirker lønnsnivået. Årsakssammenhengen er ikke klar på dette feltet. Fra tallene i vårt datasett ser vi en sammenheng der økt nivå på helsetilstand har et økt gjennomsnittlig nivå på gjennomsnittlig timelønn.

4.5.3 Påvirker helsetilstand avkastningen av utdanning på lønn?

Ut ifra uttrykket ser vi at lønnseffekten og effekten av utdanning på lønn avhenger positivt av helseni vået. Vår nullhypotese er at helsetilstanden ikke har innvirkning på avkastning av utdanningen på lønn. Resultatet gir oss en p-verdi lik 0,026 og testen er derfor signifikant med vårt konfidensintervall. Vi kan derfor forkaste vår nullhypotese. Vi kan med dette konkludere med at helsetilstanden påvirker avkastning av utdanning på lønn.

4.5.4 Er lønns gapet forskjellig for kvinner og menn med god og dårlig helse?

I denne testen sjekker vi om helseni vået gir utslag på lønnen, avhengig av om du er mann eller kvinne. Vi tester derfor om variabelen $K*GH$ er signifikant eller ikke. Nullhypotesen vår er at $K*GH$ er null. Dermed antar vi at lønnslikningen er lik for både menn og kvinner knyttet til helse. Restultatet gir oss en p-verdi lik 0,056, hvilket betyr at vi ikke kan forkaste nullhypotesen med vårt konfidensintervall lik 5%. Vi ser allikevel at sannsynligheten for nullhypotesen er relativt stor da vi faktisk kan forkaste nullhypotesen ved signifikansnivå lik 6%.

5 OPPSUMMERING

Har i denne oppgaven analysert samfunnsøkonomiske empiriske sammenhenger med bruk av datasettet fra PIAAC-undersøkelsen. Ved bruk av minste kvadraters metode som regresjonsmodell er sammenhengen mellom lønnsnivå, avkastningen av utdanning gitt ulik helsetilstand studert og drøftet. Forskning presentert viser sammenhenger mellom utdanning og helse, samt utdanning og lønnsnivå.

Gjennom vårt empiriske prosjekt har det også blitt vist at det er en sammenheng mellom helsetilstand og avkastningen av utdanning på lønnsnivå. Som antatt hadde helsetilstand og utdanningslengde positiv virkning på lønnsnivået. Ettersom det er rimelig å tro at flere faktorer kan påvirke timelønnen bør disse inkluderes i analysen slik at den betydningen av den enkelte variabel vi studerer blir mer korrekt. Ser i presentasjonen av variabelene at den gradvise utvidelsen av regresjonsuttrykkene ga økt forklaringskraft. Det er valgt å bruke et regresjonsuttrykk med relativt høy forklaringskraft, til tross for at ikke alle variablene er testet.

En svakhet ved oppgaven er at ikke alle variable som antas å ha innvirkning på timelønn testes. Vi vet ikke effekten av variablene som inngår med symbolet X_i da dette ikke er vektlagt i våre analyse. Videre har analysen vår en forklaringskraft på 28%, som betyr at det er flere forhold enn det som er presentert som påvirker timelønn.

Prosjektet etterlater seg flere åpne spørsmål som ikke er studert, men kan tenkes å ha innvirkning på vår konklusjon;

«Hvor stor del av karrierevalg hos et individ kan forklares ut ifra helsetilstanden til vedkommende?», «Har lavtlønnede jobber negativ innvirkning på helsetilstand sammenlignet med høytlønnede jobber?», «Er helse en faktor som påvirker muligheten et individ har til å få en høytlønnet jobb?».

Det kan konkluderes med at helsetilstand påvirker avkastning av utdanning på lønn basert på våre resultater. Det empiriske prosjektet som er gjennomført begrenser seg til denne konklusjonen og kan ikke brukes til å si noe om årsakssammenhengen mellom variablene.

LITTERATURLISTE

- Barth, E. (2005). *Den samfunnsmessige avkastningen av utdanning*. Tilgjengelig fra: <https://www.ssb.no/a/publikasjoner/pdf/sa74/kap-8.pdf> (Hentet 23. mars 2019).
- Bjørken, B. (2013). *Ferdigheter i voksenbefolkningen. Resultater fra den internasjonale undersøkelsen om lese- og tallforståelse (PIAAC)*. Rapporter 42/2013. Oslo-Kongsvinger: Statistisk Sentralbyrå.
- Brooks, C. (2008). *Introductory Econometrics for Finance*. (Second Edition). New York: Cambridge University Press.
- Hægeland, T. (2003). *Økonomisk avkastning av utdanning*. Tilgjengelig fra: <https://www.ssb.no/a/publikasjoner/pdf/sa60/kap-10.pdf> (Hentet: 23. mars 2019).
- Naper, S. O. & Dahl, E. (2007). *Sosialhjelpsmottakeres dødelighet: I hvilken grad kan overdødeligheten tilskrives deres sosiale status?* (1). Oslo: Høgskolen i Oslo
- Nelson, R. R. & Phelps, E. S. (1966). *Investment in Humans, Technological Diffusion, and Economic Growth*. (No 1/2). American Economic Association.
- Regjeringen. (2017, 25. april). Alderspensjon fra folketrygden. Hentet fra: <https://www.regjeringen.no/no/tema/pensjon-trygd-og-sosiale-tjenester/innsikt/pensjoner/alderspensjon/id449656/>
- Thomas, R.L (2005) *Using statistics in economics*. New York: McGrawHill