

# Lønnsavkastning av utdanning og ferdigheter i offentlig og privat sektor i Nederland

BACHELOROPPGAVE I  
SAMFUNNSØKONOMI

Institutt for Samfunnsøkonomi, NTNU

*Kandidatnr: 10066*

## **INNHALDSFORTEGNELSE**

<b>1. INNLEDNING</b>	<b>2</b>
<b>MOTIVASJON FOR OPPGAVEN</b>	<b>2</b>
<b>PROBLEMSTILLING</b>	<b>2</b>
TILLEGGSPROBLEMSTILLING	2
<b>HVORFOR ER PROBLEMSTILLINGEN INTERESSANT?</b>	<b>2</b>
<b>HVORDAN UNDERSØKE PROBLEMSTILLINGEN</b>	<b>3</b>
<b>2. TEORETISK RAMMEVERK OG TIDLIGERE LITTERATUR</b>	<b>3</b>
<b>TEORETISK RAMMEVERK</b>	<b>3</b>
<b>TIDLIGERE LITTERATUR OG RESULTATER</b>	<b>4</b>
<b>3. DATAMATERIALET</b>	<b>5</b>
<b>BESKRIVELSE OG PRESENTASJON AV DATASET</b>	<b>5</b>
DEFINISJONER OG FORKLARINGER AV VARIABLER	5
DESKRIPTIV STATISTIKK	8
FORDELER VED DATASETET	10
BEGRENSNINGER VED DATASETET	10
<b>4. REGRESJONSANALYSE OG ØKONOMETRISK MODELL</b>	<b>11</b>
<b>VALG AV FUNKSJONSFORM</b>	<b>11</b>
<b>EMPIRISK STRATEGI</b>	<b>13</b>
<b>EMPIRISK STRATEGI FOR TILLEGGSPROBLEMSTILLING</b>	<b>15</b>
<b>5. EMPIRISKE RESULTATER</b>	<b>15</b>
<b>EMPIRISKE HOVEDRESULTATER</b>	<b>15</b>
<b>ANALYSE AV TILLEGGSPROBLEMSTILLING</b>	<b>18</b>
EMPIRISKE RESULTATER TILLEGGSPROBLEMSTILLING	20
<b>6. KONKLUSJON</b>	<b>21</b>
<b>OPPSUMMERING AV HOVEDFUNN</b>	<b>21</b>
<b>BEGRENSNINGER VED ANALYSEN</b>	<b>21</b>
<b>VIDERE FORSKNING</b>	<b>22</b>
<b>7. LITTERATURLISTE</b>	<b>23</b>
<b>KILDER TIL DATAMATERIALET</b>	<b>23</b>
<b>ANDRE KILDER</b>	<b>23</b>

# 1. Innledning

## Motivasjon for oppgaven

Det er regnet som essensielt å opprettholde både en sterk offentlig og privat sektor. De to sektorene utøver ulike funksjoner og tar seg av ulike behov i samfunnet. Offentlig sektor bør ta ansvar for samfunnsoppgaver privat sektor er uegnet til å produsere og tilby, som rettsvesen, forsvar, politi, et sosialt sikkerhetsnett om en husholdnings inntekt faller bort, og gjerne også tjenester som utdanning og helsevesen. Privat sektor vil på sin side ofte i større grad enn offentlig sektor kunne sørge for produktivitetsvekst og dermed økt velstand i samfunnet, nettopp fordi sektoren er konkurranseutsatt. Private bedrifter må hele tiden søke etter å effektivisere produksjonen, og vil grunnet konkurransen ha større incentiver for dette enn aktører innenfor offentlig sektor, som ikke møter samme konkurranse. Privat sektor bidrar også til å finansiere og muliggjøre eksistensen til den offentlige sektoren via skatteinntekter til staten, mens offentlig sektor på sin side muliggjør produksjonen i privat sektor gjennom f.eks. utdanningssystemet og god infrastruktur.

Det er med andre ord essensielt for en moderne blandingsøkonomi å opprettholde både offentlig og privat sektor, og at det rekrutteres gode hoder inn i begge sektorer. Dersom det eksisterer betydelige lønnsforskjeller mellom offentlig og privat sektor, eller lønnsavkastningen av ferdigheter og utdanning er ulik, vil dette kunne ha en effekt på rekrutteringen av høyt utdannede arbeidstakere med gode testresultater på ferdighetstester inn i de to sektorene. Det vil derfor være av myndighetenes og samfunnets interesse at det forblir moderate lønnsforskjeller mellom de to sektorene, og at effekten av utdanning og ferdigheter på lønnen ikke er for ulik mellom de to. Motivasjonen for denne oppgaven ligger derfor i behovet for å undersøke i hvor stor grad utdanning og ferdigheter i form av regneferdigheter har en effekt på lønnsnivået i privat og offentlig sektor innad i et land, men også i å undersøke om det her eksisterer forskjeller mellom land.

## Problemstilling

Er det forskjeller i avkastningen av humankapital i form av utdanning og ferdigheter på lønnen til ansatte i offentlig og privat sektor i Nederland?

## Tilleggsproblemstilling

Er det signifikante forskjeller mellom land på ulike stadier i økonomisk utvikling, her Nederland og Estland, i hvor stor lønnseffekten av økt humankapital er i offentlig og privat sektor?

## Hvorfor er problemstillingen interessant?

Med denne bacheloroppgaven ønsker jeg å utforske i hvor stor grad utdannings- og ferdighetsnivå påvirker ansattes lønn i henholdsvis offentlig og privat sektor i Nederland. Nederland er et land på et høyt utviklet økonomisk nivå, med høy BNP per innbygger innad i OECD<sup>1</sup>. En analyse av data fra Nederland vil kunne være relevant for andre land på tilsvarende nivå i økonomisk utvikling. Den vil også kunne være relevant for land som på

---

<sup>1</sup> Kilde: OECD (2019), Gross domestic product (GDP) (indicator). doi: 10.1787/dc2f7aec-en (Accessed on 05 April 2019)

nåværende tidspunkt har et lavere nivå på BNP per innbygger, men som er i vekst, og som vil kunne nå opp på et tilsvarende nivå i årene som kommer.

Som nevnt er det viktig å sikre en god rekruttering av høyt utdannede arbeidere med høyt utviklede ferdigheter inn i både i offentlig og privat sektor. Å rette oppmerksomhet mot eventuelle store ulikheter i avkastningen av humankapital på lønnen i de to sektorene vil kunne medføre at tiltak settes i gang for å utjevne disse forskjellene, og slik unngå at høyt utdannede med gode testresultater i overvekt søker seg til én sektor, og ikke den andre.

Det er også interessant å studere problemstillingen i to land på ulike stadier i økonomisk vekst, for å se om dette kan være av betydning for avkastningen av humankapital på lønnen. For å studere denne problemstillingen sammenlikner jeg Nederland med Estland, hvor BNP per innbygger er langt lavere.<sup>2</sup> På den ene siden kan landene ses på som to europeiske langt på vei utviklede økonomier. På den andre siden er dette to land av forskjellig størrelse, med svært ulikt innbyggertall, og med en svært forskjellig historie, økonomisk og politisk, som alle er faktorer som kan ha innvirkning på dagens lønssystem. Dette vil ikke denne analysen kunne ta høyde for, men er noe å ha i bakhodet når man sammenlikner de to økonomiene.

### Hvordan undersøke problemstillingen

For å undersøke problemstillingen vil jeg analysere et datasett hentet fra første del av den internasjonale undersøkelsen PIAAC. PIAAC, *Programme for the International Assessment of Adult Competencies*, ble utført av OECD i over 20 land mellom 2011 og 2012, og kartlegger ferdigheter blant voksenbefolkningen innenfor områdene lesing, regning og problemløsning i IKT.<sup>3</sup>

Dataprogrammet Stata/SE benyttes i denne oppgaven til å utføre regresjonsanalyser basert på økonometriske modeller utformet ved hjelp av datasettet. Problemstillingene i oppgaven testes ved hjelp av hypotesetester. Analysen tar utgangspunkt i en enkel økonometrisk modell for effekten av utdanning og ferdigheter på lønnen i Nederland, som så utvides til å inkludere flere kontrollvariabler og interaksjonsledd. Forskjeller i lønnseffekten mellom Nederland og Estland studeres senere i tilleggsproblemstillingen, for slik å kunne gjøre en internasjonal sammenlikning. Som nevnt benyttes BNP per innbygger i respondentåret for undersøkelsen som et mål på nivå på økonomisk utvikling.

## **2. Teoretisk rammeverk og tidligere litteratur**

### Teoretisk rammeverk

Sammenhengen mellom høyere utdanning og lønn er blitt grundig undersøkt i mange studier. Grunnleggende humankapitalteori utarbeidet av Schultz (1961) og Becker (1962) legger til grunn at økt utdanning gir økt inntekt i form av lønn fordi utdanning leder til høyere produktivitet. Utdanning kan dermed betraktes som en investering.

---

<sup>2</sup> Kilde: OECD (2019), Gross domestic product (GDP) (indicator). doi: 10.1787/dc2f7aec-en (Accessed on 05 April 2019)

<sup>3</sup> Kilde: OECD (2019), *The Survey of Adult Skills (PIAAC)*, <http://www.oecd.org/skills/piaac/about/#d.en.481111> (Nedlastningsdato: 29.4.2019)

I følge Becker vil høyt utdannedes lønn nesten alltid være langt over gjennomsnittslønnen, uansett hvilket land som observeres. Effekten er sterkere i mindre utviklede land.<sup>4</sup> Ettersom denne studien observerer Nederland og Estland, som tross svært ulikt nivå på BNP per innbygger begge er regnet som høyt utviklede, er det usikkert i hvor stor grad denne effekten vil kunne observeres.

Det kan være ulike årsaker til at vi finner ulikheter i lønnsavkastningen av humankapital i offentlig og privat sektor. Tidligere studier av lønnsavkastning av utdanning i Norge har vist at denne har en større samvariasjon i privat enn offentlig sektor (Bratsberg, Raaum og Hægeland, 2006). Det er derfor ikke utenkelig at vi observerer en liknende effekt i Nederland. Det kan også tenkes at vi oppdager liknende sammenhenger når vi studerer kompetanse i form av testscore på ferdighetstester. Noe som kan tyde på det, som trukket frem i tidligere studier, er at privat sektor i større grad enn offentlig sektor har lønnsystemer bygget opp rundt å belønne effektivitet, for slik å stimulere produktivitetsvekst (Bratsberg, Raaum og Hægeland, 2006).

### Tidligere litteratur og resultater

I Norge er resultatene fra PIAAC-undersøkelsen blitt sammenfattet i rapporten *Ferdigheter i voksenbefolkningen* av Birgit Bjørkeng. Der kommenterer hun mange funn, som de gjennomsnittlige ferdighetene i den norske voksenbefolkningen, forskjellen mellom yngre og eldre voksne, forskjellen mellom menn og kvinner, og den norske befolkningens resultater sett i sammenheng med det internasjonale gjennomsnittet. Hun har derimot ikke inkludert en undersøkelse av hvorvidt lønnseffekten av utdanning og ferdigheter er ulik i offentlig og privat sektor (Bjørkeng, 2013).

I rapporten *Lese- og tallforståelse, utdanning og arbeidsmarkedssuksess* fra 2006, skrevet av Bratsberg, Raaum og Hægeland, om den norske delen av den tidligere utførte ALL-undersøkelsen, inkluderes derimot et kort avsnitt viet til temaet. ALL-undersøkelsen, eller *Adult Literacy and Life Skills Survey*, studerte i stor grad det samme som PIAAC-undersøkelsen, men inkluderte kun lese- og regnetester. Disse to undersøkelsene kan, sammen med den enda tidligere utførte IALS, *International Adult Literacy Survey*, ses i sammenheng med hverandre.

Bratsberg, Raaum og Hægeland har undersøkt om lønnseffekten av ALL-score og antall år utdanning er ulik i de to sektorene. Deres hypotese var at samvariasjonen mellom ALL-score og lønn ville være sterkere i privat enn i offentlig sektor. Dette begrunnet de med at privat sektor i større grad enn offentlig sektor fastsetter lønnen ut ifra individuelle lønnsforhandlinger, og at det benyttes systemer med prestasjonslønn for å skape incentiver for å øke arbeidstakernes effektivitet. De viser også til at tidligere studier har funnet en større effekt av utdanning på lønnen blant menn i privat sektor enn i offentlig sektor. Denne effekten finner de selv igjen når de tester med utvalget fra ALL-undersøkelsen. De går i rapporten ikke inn på hvorvidt det samme gjelder for kvinner, noe som kan tyde på at de ikke har kunnet finne statistisk signifikante resultater som tyder på dette.

Tallene fra ALL-undersøkelsen tyder, til forfatterens overraskelse, derimot ikke på at samvariasjonen mellom score på ferdighetstestene og lønn er ulik i de to sektorene.

---

<sup>4</sup> Kilde: Becker, Gary S., *Human Capital*, <https://www.econlib.org/library/Enc/HumanCapital.html> (Nedlastningsdato: 29.4.2019)

I Nederland inkluderer rapporten *Equal wages for equal jobs? Jobs and wages in the public and private sectors* presentert av Statistics Netherlands, Centraal Bureau voor de Statistiek (CBS) en liknende analyse, men der datamaterialet er hentet inn uavhengig av PIAAC-undersøkelsen. Der studerer de spesielt forskjellene mellom lønnsnivå for kvinner og menn innad i offentlig og privat sektor. Analysen korrigerer for ulikt utdannings og ferdighetsnivå, men presiserer at ferdigheter i tillegg til faglig nivå i stor grad regnes som egenskaper tilegnet gjennom arbeidserfaring. Det er med andre ord ikke benyttet seg av ferdighetstester slik som i PIAAC, ALL og IALS. Rapporten finner at flere er ansatt i privat sektor enn i offentlig sektor. Gjennomsnittslønnen er høyere i offentlig enn privat sektor, men variasjonen er større i privat sektor, noe som begrunnes med at det er mange unge og deltidsansatte som mottar en lav lønn i privat sektor samtidig som at de få aller høyest lønnede mottar en timelønn langt over gjennomsnittet (Geerdinck et al. 2012). Disse opplysningene stemmer i stor grad med den deskriptive statistikken for datasettet analysert i denne oppgaven.

### 3. Datamaterialet

#### Beskrivelse og presentasjon av datasett

I denne oppgaven har jeg benyttet meg av to datasett. Hovedmaterialet er hentet fra OECD sin undersøkelse PIAAC. Datamaterialet er hentet mellom august 2011 og april 2012, og er for landene Nederland og Estland. All data for respondenter under 25 år er utelatt fra det analyserte datasettet for å korrigere for at disse i all hovedsak er studenter uten fast jobb. Jeg har i siste delen av analysen, der de to landene sammenliknes, slått sammen datamaterialet for de to landene for å kunne analysere begge samtidig.

Jeg har også benyttet meg av data for BNP per innbygger i Nederland og Estland i 2011. Datamaterialet er hentet fra OECD sine nettsider. Materialet viser at BNP per innbygger i Nederland i 2011 var på 46 599 US dollar, mens det i Estland var på 24 501 US dollar.

#### Kilder til datamaterialet

Datamaterialet fra PIAAC-undersøkelsen er tilgjengeliggjort av instituttet (ISØ ved NTNU), Datafiler til bacheloroppgave, Blackboard:  
[https://ntnu.blackboard.com/webapps/blackboard/content/listContent.jsp?course\\_id=12825\\_1&content\\_id=595468\\_1&mode=reset](https://ntnu.blackboard.com/webapps/blackboard/content/listContent.jsp?course_id=12825_1&content_id=595468_1&mode=reset) (Nedlastningsdato: 5.4.2019)

Gross Domestic Product (GDP), OECD:  
<https://data.oecd.org/gdp/gross-domestic-product-gdp.htm> (Nedlastningsdato: 5.4.2019)

#### Definisjoner og forklaringer av variabler

Definisjon av avhengig variabel, lønnsvariabelen  $lwage$ :

$$lwage = \log(earnhrppp) = \text{logaritmen til timelønn i PPP US\$ for de over 24 år}$$

Definisjon av sentral forklaringsvariabel:

$pub=1$  hvis offentlig sektor, 0 hvis privat sektor. Offentlig sektor omfatter her stat, kommune, fylkeskommune og frivillige/veldedige organisasjoner.

## Definisjoner av kontrollvariabler:

*numscore1 = Kognitive ferdigheter målt ved resultat på test i tallforståelse. Variabelen er standardisert med gjennomsnitt (forventningsverdi) lik 0 og standardavvik lik 1.*

*Nytt navn på numscore1 i oppgaven er «num».*

*yrsqual = Formelt utdanningsnivå, målt i antall år fullført utdanning.*

*Nytt navn på yrsqual i oppgaven er «educ».*

*female=1 hvis respondenten er kvinne, 0 hvis mann. Nytt navn på female i oppgaven er «fem».*

*c\_q09\_c = Antall år arbeidserfaring. Nytt navn på c\_q09\_c i oppgaven er «exper».*

*c\_q09\_c ^2 = Antall år arbeidserfaring kvadrert. Nytt navn på c\_q09\_c i oppgaven er «exper\_sq». Også brukt i modellene er «exper^2».*

*pub\_num = pub\*num = Interaksjonsledd som sier noe om avkastningen på lønnen av ferdigheter i offentlig sektor sammenliknet med privat sektor.*

*pub\_educ = pub\*educ = Interaksjonsledd som sier noe om avkastningen på lønnen av utdanning i offentlig sektor sammenliknet med privat sektor.*

## Definisjoner av kontrollvariabler i tilleggsproblemstilling:

*neth=1 hvis cntryid==528, Nederlands identifikasjonsnummer i datasettet, og lik 0 ellers, altså for Estland.*

*neth\_pub = neth\*pub = Interaksjonsledd som sier noe om effekten på lønnen av å jobbe i offentlig sektor i Nederland når sammenliknet med Estland.*

*neth\_educ = neth\*educ = Interaksjonsledd som sier noe om avkastningen av utdanning på lønnen i Nederland når sammenliknet med Estland.*

*neth\_num = neth\*num = Interaksjonsledd som sier noe om avkastningen av ferdigheter på lønnen i Nederland når sammenliknet med Estland.*

*neth\_pub\_educ = neth\*pub\*educ = Interaksjonsledd som sier noe om avkastningen på lønnen av utdanning når man jobber i offentlig sektor i Nederland når sammenliknet med Estland.*

*neth\_pub\_num = neth\*pub\*num = Interaksjonsledd som sier noe om avkastningen på lønnen av ferdigheter når man jobber i offentlig sektor i Nederland når sammenliknet med Estland.*

Når man observerer de tre standardiserte variablene for ferdigheter i datasettet, der *num* er regneferdigheter, *litscore1* er leseferdigheter og *pslscore1* er problemløsningsscore i IKT, oppdager man at de tre er kraftig korrelert med hverandre. (Se tabellen under, der korrelasjonen er blitt undersøkt i det relevante datasettet.) Det er derfor ikke uvanlig i denne typer analyser at man velger kun én av de tre variablene til å inngå i den økonometriske

modellen. Sterkt korrelerte uavhengige variabler vil kunne gi dårlig estimerte regresjonskoeffisienter, høyere standardavvik og høyere p-verdier.<sup>5</sup> Jeg har valgt å benytte nivå på *num* som variabel for respondentens testscore i denne oppgaven. Når begrepet «ferdigheter» benyttes i oppgaven innebærer dette resultat på regnetesten.

Korrelasjon mellom *num*, *litscore1* og *plsscore1*:

Tabell 1, hentet fra Stata:

Korrelasjon	num	litscore1	plsscore1
num	1.0000		
litscore1	0.8590	1.0000	
plsscore1	0.7667	0.8293	1.0000

Vi ser i tabellen at korrelasjonen med lesetesten er på 85,90%. Korrelasjonen med problemløsningstesten i IKT er på 76,67%. Vi kan regne med at dette er en tendens som vil gjelde på tvers av land, ikke bare for Nederland.

Setter også opp korrelasjonsmatrise for mange av de andre viktigste variablene for analysen. Som jeg vil komme tilbake til, er det problematisk om to eller flere av variablene i modellen er sterkt korrelert med hverandre, blant annet fordi det gjør det vanskelig å tolke effekten av hver enkelt variabel på den avhengige variabelen i modellen.<sup>6</sup>

Tabell 2, hentet fra Stata:

Korrelasjon	num	educ	exper	exper_sq	pub_educ	pub_num
num	1.0000					
educ	0.4685	1.0000				
exper	-0.1943	-0.2205	1.0000			
exper_sq	-0.2039	-0.2203	0.9707	1.0000		
pub_educ	0.0739	0.2780	0.0211	0.0191	1.0000	
pub_num	0.5660	0.2945	-0.1126	-0.1137	0.2200	1.0000

Ikke overraskende er *exper* og *exper\_sq* sterkt korrelert med hverandre.<sup>7</sup> Det samme gjelder *pub* og *pub\_num*. Det er også en betydelig korrelasjon mellom utdanning og regneferdigheter, på 46,85%. Ut over dette er korrelasjonen mellom variablene i modellen relativt lav.

Som tidligere nevnt har jeg valgt å benytte den standardiserte variabelen for regneferdigheter, ikke de «rå» testscorene. Dette er for å kunne tolke og sammenlikne regresjonskoeffisientene fra ulike økonomiske modeller. Standardiserte testscores er sammenliknbare grunnet at de er benevningsfrie, og med samme forventningsverdi og standardavvik. Koeffisientene kan dermed tolkes på tvers av måleskalaer og studier, i motsetning til om man bruker rå testcores, som ikke er benevningsfrie, og som vil ha ulike forventningsverdier og gjennomsnitt, og dermed gjøre resultater vanskelige å sammenlikne.

<sup>5</sup> For definisjon av p-verdi se *Empirisk strategi* s. 14.

<sup>6</sup> For mer om problemet med korrelasjon se *Valg av funksjonsform* s. 11 og 12.

<sup>7</sup> For sammenheng mellom *exper* og *exper\_sq* se *Forklaringer og øvrige kommentarer* s. 18.



En viktig variabel i analysen er *d\_q03 Current work – Economic sector*, her omdøpt til *sektor*, som kan ha tre ulike verdier, 1, 2 og 3. Se tabellen under. Verdien 1 impliserer privat sektor, 2 offentlig sektor (som stat og kommune) og 3 frivillige og ideelle organisasjoner som driver med veldedighet, religionsutøvelse o.l. Jeg har valgt å slå sammen 2 og 3 og kalle dette offentlig sektor i denne oppgaven fordi frivillige og ideelle organisasjoner i større grad likner offentlig enn privat sektor bl.a. fordi de ikke streber etter å maksimere profitt. Det er også verdt å nevne at andelen ansatte som jobber i denne sektoren er svært lav på kun 5,69% av alle i arbeid. Jeg har derfor konstruert dummyvariabelen *pub* så den er lik 0 for *d\_q03==1* og lik 1 ellers.

Tabell 3, hentet fra Stata:

Sektor	Prosent
1	66.43
2	27.88
3	5.69
Totalt	100.00

Jeg har valgt å benytte variabelen *earnhrppp*, som gir timelønnen i purchasing power parity US\$, altså er lønnen kjøpekraftskorrigert for prisnivå innad i landet. Slik kan man sammenlikne lønninger på tvers av land til tross for ulike valuta og ulikt prisnivå.

Til tilleggsproblemstillingen der jeg sammenlikner de to landene har jeg generert en variabel *neth=1* hvis *cuntryid==528*, altså Nederlands identifikasjonsnummer i datasettet, for slik å kunne se på effekten på lønnen av å jobbe i Nederland i forhold til Estland, og slik sammenlikne de to.

### Deskriptiv statistikk

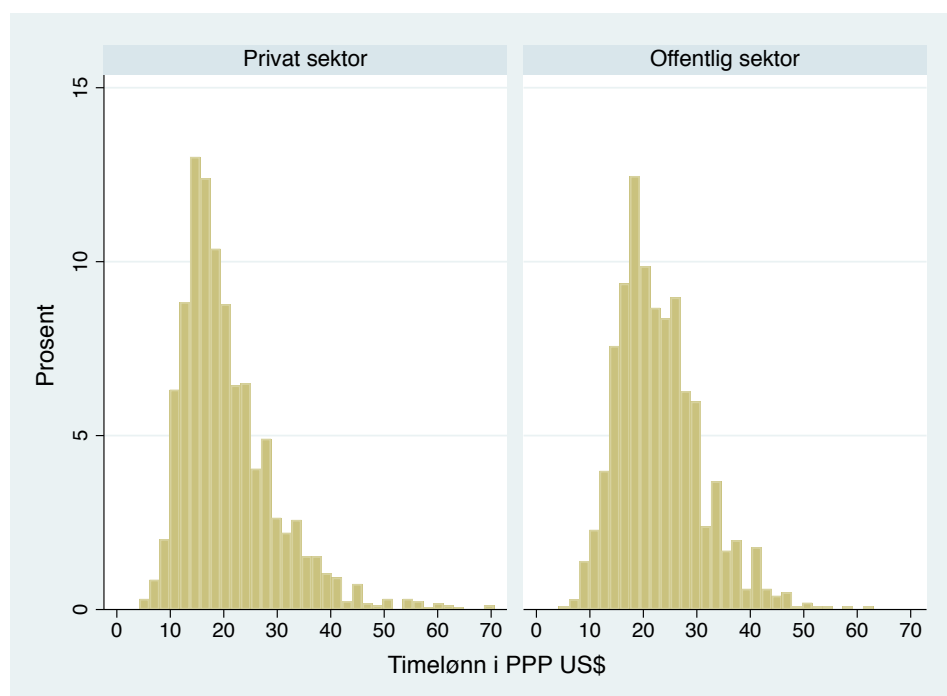
#### Deskriptiv statistikk for avhengig variabel

Tabell 4, hentet fra Stata:

Timelønn	Observasjoner	Gjennomsnitt	Standardavvik
Alle	2,643	21.62	8.70
Offentlig sektor	1,003	23.02	8.02
Privat sektor	1,629	20.78	8.99

Ser at den gjennomsnittlige timelønnen i hele utvalget er på 21,62 PPP \$US. I offentlig sektor er den høyere, på 23,02 PPP \$US, mens den i privat sektor er lavest, på 20,78 PPP \$US. Ser at differansen mellom alle som har oppgitt en inntekt og de som tilsammen jobber i offentlig eller privat sektor er på 11 observasjoner. Det kan være ulike grunner til det, som at respondenten er usikker på hvilken sektor han eller hun jobber innenfor, ikke identifiserer seg med noen av sektorene eller av en annen grunn ikke har ønsket å oppgi denne informasjonen.

Histogram 1, hentet fra Stata:



Histogrammet viser lønnsfordelingen i prosent for ulike nivå på timelønn i PPP US\$ i de to sektorene. Ser at disse er relativt like, men at privat sektor har flere som mottar en timelønn over 60 PPP US\$, men også flere arbeidere som mottar en timelønn mellom 10 og 20 PPP US\$ eller lavere enn det offentlig sektor har.

#### Deskriptiv statistikk for sentral forklaringsvariabel

Tabell 5, hentet fra Stata:

pub	Prosent
0	66.43
1	33.57
Totalt	100.00

Av respondentene er det 66,43% av de i arbeid som jobber i privat sektor, 33,57% jobber i enten offentlig sektor eller i en frivillig organisasjon. Dette kunne vi også se av tabell 3.

#### Deskriptiv statistikk for kontrollvariabler

Tabell 6, hentet fra Stata:

	Alle		Offentlig sektor		Privat sektor	
	Gjennomsnitt	Standardavvik	Gj.snitt	Std.avvik	Gj.snitt	Std.avvik
Utdanning	13.45	2.64	14.35	2.47	13.50	2.43
Regneferdigheter	-.02	1.02	.14	.93	.10	.96
Kvinner (prosent)	.51	.50	.61	.49	.41	.492
Arbeidserfaring	22.66	11.57	23.32	10.93	22.36	10.85

Gjennomsnittlig utdanningslengde for respondenter over 25 år er mellom 13 og 14 år, med et standardavvik på mellom 2 og 3 år. Vi ser at gjennomsnittlig utdanningslengde er noe lenger i offentlig enn privat sektor, men at forskjellen ikke er stor. Gjennomsnittlig utdanningslengde er lenger i begge sektorer enn i utvalget som helhet. Dette er sannsynligvis fordi utvalget rommer alle respondenter mellom 25 og 65 år, ikke bare de som er i arbeid, slik som statistikken for offentlig og privat sektor naturligvis gjør.

Det er sannsynligvis av samme grunn at vi ser at regneferdighetene i utvalget som helhet er på et lavere nivå enn i offentlig og privat sektor. Offentlig sektor har arbeidere med i gjennomsnitt litt høyere ferdigheter enn i privat sektor.

61% av arbeidere i offentlig sektor, og 41% av arbeidere i privat sektor, er kvinner. Det er med andre ord en betydelig høyere andel kvinner i offentlig enn privat sektor. Dette sammenfaller med resultater fra tidligere forskning og litteratur. Vi ser at 51% av respondentene totalt er kvinner, det er altså omtrent like mange kvinner og menn som har svart på undersøkelsen.

Gjennomsnittlig arbeidserfaring er også noe lenger i offentlig enn privat sektor, men differansen er på under ett år. Gjennomsnittlig arbeidserfaring i befolkningen som helhet er faktisk høyere enn i privat sektor, men vi ser at standardavviket er noe større for hele befolkningen.

#### Fordeler ved datasettet

Fordeler ved datasettet er at den har mange respondenter fra ulike aldersgrupper, av begge kjønn, fordelt over alle sektorer. Det er med andre ord en bred gruppe som er blitt undersøkt. Dette gjør det mulig å kontrollere for kjønn, arbeidserfaring, sektor, utdanning og ferdigheter når vi analyserer effekter på lønnen. Kun en begrenset mengde av den totale datamengden innhentet i PIAAC-undersøkelsen benyttes i denne oppgaven. Om man derimot benytter seg av datasettet i sin helhet er det mulig å sammenlikne lønns-, ferdighets- og utdanningsnivå, i tillegg til mange andre forhold, på tvers av mange OECD-land. Ettersom data fra de tidligere IALS og ALL-undersøkelsene er blitt bearbeidet for å være sammenliknbare med tall fra PIAAC er det også mulig å studere utviklingen i de landene som har deltatt i flere av undersøkelsene over tid. Innholdet i datasettene gjør det med andre ord mulig å gjennomføre mange ulike typer analyser og undersøkelser av interesse for både myndigheter og samfunn som helhet.

#### Begrensninger ved datasettet

Som nevnt er datasettet stort, men om man analyserer kun en svært spesifikk datamengde, slik som gjøres i denne oppgaven, er mengden data derimot mer begrenset. Antall observasjoner er fremdeles stor nok til å kunne teste hypoteser, og dermed si noe om generelle forhold, men den kunne gjerne vært enda større for å sikre at man analyserer et representativt utvalg for den totale populasjonen.

I denne oppgaven ønsker jeg i tilleggspørsmålet å se på land i ulike stadier av økonomisk utvikling, og analysere effekten av nivå på BNP per innbygger på lønnsavkastningen på utdanning og ferdigheter. Datasettet inneholder derimot kun data for OECD-land, som alle kan ses på som høyt utviklede, til tross for forskjell i nivå på BNP per innbygger. For å enda tydeligere kunne observere hvorvidt nivå på økonomisk utvikling er av betydning kunne det vært forholdsmessig å se på land med en større differanse i levestandard.

## 4. Regresjonsanalyse og økonometrisk modell

### Valg av funksjonsform

I tråd med annen litteratur på feltet velges en log-lineær funksjonsform på den økonometriske modellen. Avhengig variabel er her logaritmen til lønnen, ikke lønnen i seg selv. Å benytte seg av logaritmen til lønnen gjør det mulig å se tilnærmet relativ, eller prosentvis, endring i lønn som følge av en endring i kontrollvariablene. Estimert koeffisient til den standardiserte testscoren, *num*, vil tilnærmet kunne tolkes som den relative lønnsøkningen av et standardavviks økning i regneferdigheter når alle de andre variablene forblir uendret. Tilsvarende vil koeffisienten til *educ* kunne tolkes som tilnærmet relativ lønnsøkning av et år ekstra utdanning. Den estimerte koeffisienten i et ledd måler altså effekten av en marginal endring i den tilhørende kontrollvariablen gitt at alle de andre kontrollvariablene er konstante.

Også i tråd med tidligere litteratur benyttes OLS, minste kvadraters metode, til regresjonsanalysen. Vi har her å gjøre med en multippel regresjon ettersom det er mange inkluderte variabler som påvirker logaritmen til lønnen. Vi antar da at det eksisterer et lineært forhold mellom logaritmen til lønnen og forklaringsvariablene som inngår i modellen. Ettersom vi ikke kjenner til populasjonens regresjonslinje må denne estimeres ved å finne utvalgets regresjonslinje. Som vist over har vi i dette utvalget tilgang på  $n = 2632$  observasjoner for timelønn der respondenten har oppgitt sektor vedkommende jobber i.

Forutsetninger for den klassiske multiple regresjonsmodellen estimert ved OLS:

Hver forklaringsvariabel  $X_j$  ( $j = 2, 3, \dots, k$ ) antas:

- (i) Å være ikke-stokastisk.
- (ii) Å ha verdier som er fastsatt i gjentatte utvalg.
- (iii) Å være slik at når utvalgsstørrelsen  $n \rightarrow \infty$  vil variansen til utvalget bevege seg mot en fast konstant.
- (iv) Vi har ikke perfekt multikollinearitet. Det eksisterer altså ikke et lineært forhold mellom utvalgsverdiene til noen av forklaringsvariablene.

$\varepsilon$  er et stokastisk restledd som fanger opp virkningene av andre variabler som ikke er inkludert i modellen, men som påvirker logaritmen til timelønnen.

Forutsetninger om  $\varepsilon$ :

$i = 1, \dots, n$

- (i)  $E(\varepsilon) = 0$ , forventningsverdien til  $\varepsilon$  er lik 0 og konstant for alle observasjoner. Gjennomsnittet av verdiene på  $\varepsilon$  i utvalget blir med andre ord 0, positive og negative forstyrrelser nuller hverandre ut.
- (ii)  $V(\varepsilon_i) = E(\varepsilon_i - E[\varepsilon_i])^2 = E(\varepsilon_i^2) = \sigma^2$ , variansen til  $\varepsilon$  er konstant for alle observasjoner, vi har homoskedastisitet.
- (iii)  $Cov(\varepsilon_i, \varepsilon_j) = E(\varepsilon_i - E[\varepsilon_i])(\varepsilon_j - E[\varepsilon_j]) = E(\varepsilon_i \varepsilon_j) = 0$ , kovariansen til  $\varepsilon_i$  og  $\varepsilon_j$  er lik 0, det er ingen systematisk samvariasjon mellom de to.  $i \neq j$
- (iv)  $\varepsilon_i$  er normalfordelt for alle observasjoner.

Forutsetning (iv) er nødvendig for at vi skal kunne utføre hypotesetester.

Under disse forutsetningene kan det vises at de estimerte koeffisientene vil være forventningsrette og effisiente, det eksisterer med andre ord ikke estimatorer med lavere varians, og de vil i gjennomsnitt gi rett anslag på parameteren:

$$E(a) = \alpha \text{ og } E(b_j) = \beta_j$$

Der  $b_j$  er forventningsrett estimator for  $\beta_j$  og  $a$  er forventningsrett estimator for  $\alpha$ .

Gitt at  $\varepsilon_i$  er normalfordelt vil også den avhengige variabelen i modellen være normalfordelt. Dette innebærer at også  $a$  og  $b_j$  er normalfordelte.

Det kan vises at de har variansen:

$$\text{Var}(a) = \sigma_a^2 = \frac{\sigma^2 \sum x_i^2}{\sum (x_i - \bar{x})^2} \text{ og } \text{Var}(b_j) = \sigma_{b_j}^2 = \frac{\sigma^2}{\sum (x_{ji} - \bar{x})^2 (1 - r_{jk}^2)}$$

Der  $r_{jk}$  er korrelasjonen mellom variablene  $X_j$  og  $X_k$ .  $\sigma_a$  og  $\sigma_{b_j}$  er tilhørende standardavvik.

Variansen avhenger av estimert restleddvariens, variasjonen i de respektive variablene og av korrelasjonen mellom variablene. Høy korrelasjon mellom variablene målt ved  $r_{jk}$  vil når alt annet er likt gi høy estimert varians til OLS-estimatorene, og som konsekvens upresise estimater på de sanne populasjonsparameterne. Dette vil kunne gi store konfidensintervaller rundt parameterne, høyere p-verdier, og dermed en tendens til å ikke kunne forkaste nullhypoteser.

Fordi variansene som regel er ukjente må disse estimeres. Det kan vises at den forventningsrette estimatoren for variansen  $\sigma^2$ ,  $s^2$ , kan finnes ved hjelp av utvalgsstørrelsen ( $n$ ), antall parametre i modellen ( $k$ ) og summen av de kvadrerte residualene ( $\sum e_i^2$ ).

Forventningsrett estimator for  $\sigma^2$ :

$$s^2 = \frac{\sum e_i^2}{n - k}$$

Ved OLS minimeres summen av de kvadrerte residualene, altså differansen mellom det observerte, faktiske nivået på  $Y$  for hver observasjon, og det predikerte nivået på  $Y$  for samme observasjon, estimert ved hjelp av utvalget.

Residual for hver observasjon:

$$e_i = Y_i - \hat{Y}_i$$

Der  $Y_i$  = observert nivå på  $Y$  og  $\hat{Y}_i$  = predikert nivå på  $Y$ .

Vi finner estimater som minimerer summen av de kvadrerte residuale ved å løse optimeringsproblem ved hjelp av førsteordensbetingelser.

Hvor godt modellen egentlig forklarer variasjon i den avhengige variabelen, logaritmen til lønnen, avhenger av modellens føyningsmål målt ved determinasjonskoeffisienten  $R^2$ . Nivået på  $R^2$  sier noe om modellens forklaringskraft, altså hvor stor andel av den totale variasjonen i den avhengige variabelen som kan forklares ved variasjon i variablene på høyresiden i modellen.

$$R^2 = \frac{SSE}{SST} = \frac{SST - SSR}{SST} = 1 - \frac{SSR}{SST} \Rightarrow 0 \leq R^2 \leq 1$$

Der SST = Total variasjon, SSE = Forklart variasjon, SSR = Residualvariasjon. Følgelig innebærer dette at  $R^2 = 0$  må bety at modellen ikke forklarer noe av variasjonen i den avhengige variabelen, mens  $R^2 = 1$  betyr at modellen forklarer all variasjon i den avhengige variabelen, og at dermed alle observasjonene befinner seg på regresjonslinja. Begge deler er ekstremtilfeller det er liten grunn til å tro man vil observere i denne modellen, men det er verdt å merke seg nivået på føyning.

Nivået på føyning er langt lavere i modeller der avhengig variabel er logaritmisk og ikke lineær, fordi predikerte verdier på variabelen da alltid vil være mye lavere.  $R^2$  kan med andre ord ikke sammenliknes på tvers av modeller med ulik funksjonsform. I denne oppgaven benyttes derimot utelukkende log-lineære funksjoner, og modellenes forklaringskraft kan dermed sammenliknes. Men dersom to  $R^2$  ligger svært nært hverandre vil vi heller ikke kunne benytte oss av deres relative størrelse for å sammenlikne modellenes forklaringskraft (Thomas, 2005).

### Empirisk strategi

For å teste hypoteser om utvalget benyttes t-test eller F-test.

#### Students t-test

Vi benytter t-test når vi kun tester hypoteser om en enkelt parameter i modellen. Under de gitte forutsetningene for OLS har vi testobservatoren:

$$TS = \frac{b_j - \beta_j}{s_{b_j}} \sim t_{\alpha(n-k)} \text{ under } H_0$$

Som er t-fordelt med n-k frihetsgrader under  $H_0$ .

$\beta_j$  = Den sanne populasjonsparameteren.

$b_j$  = Estimat for  $\beta_j$  gitt av utvalget.

$s_{b_j}$  = Estimat for  $\sigma_{b_j}$  gitt av utvalget

n = utvalgsstørrelse, k = antall variabler i modellen

Beslutningskriteriet er å forkaste  $H_0$  dersom testobservatoren (TS) overskrider det kritiske nivået for t-testen til gitt signifikansnivå,  $\alpha$ .

$$TS > t_{\alpha}(n - k)$$

Eventuelt benyttes testens p-verdi, som forklares under F-test.

#### F-test

Vi benytter oss av F-test når vi tester hypoteser om flere parametere i modellen samtidig. F-testens intuisjon ligger i å teste endringen i residualkvadratsummen når en restriksjon pålegges. Residualkvadratsummen, SSR, er som tidligere forklart andelen av total variasjon i

den avhengige variabelen som ikke kan forklares ved variasjon i variablene på høyresiden i modellen. En stor økning i SSR vil tyde på en «ugyldig» restriksjon pålagt i  $H_0$ .

$$TS = \frac{(SSR_R - SSR_U)/h}{(SSR_U)/(n - k)} \sim F_{\alpha(h, n-k)} \text{ under } H_0$$

Som er F-fordelt med  $h$  frihetsgrader i telleren og  $n-k$  frihetsgrader i nevneren under  $H_0$ .

$SSR_R$  = Residualkvadratsummen for modellen pålagt restriksjonen under  $H_0$ .

$SSR_U$  = Residualkvadratsummen for modellen uten pålagte restriksjoner.

$h$  = antall variabler vi tester for,  $n$  = utvalgsstørrelse,  $k$  = antall variabler i modellen

Husker at  $R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$ . Vi dividerer TS med SST i teller og nevner, og får da:

$$TS = \frac{(R_U^2 - R_R^2)/h}{(1 - R_U^2)/(n - k)} \sim F_{\alpha(h, n-k)} \text{ under } H_0$$

Som er F-fordelt med  $h$  frihetsgrader i telleren og  $n-k$  frihetsgrader i nevneren under  $H_0$ .

$R_U^2$  = Multiplert determinasjonskoeffisient for generell modell uten restriksjoner.

$R_R^2$  = Multiplert determinasjonskoeffisient for restriktiv modell.

Vi kan benytte oss av  $R^2$  ettersom avhengig variabel er logaritmen til timelønnen i begge modellene.

Beslutningskriteriet er å forkaste  $H_0$  dersom testobservatoren (TS) overskrider det kritiske nivået for F-testen til gitt signifikansnivå,  $\alpha$ .

$$TS > F_{\alpha}(h, n - k)$$

Eventuelt kan også testens p-verdi benyttes til å si hvorvidt vi forkaster nullhypotesen eller ikke. Stata oppgir p-verdier når man utfører både t- og f-tester. Det er derfor denne metoden jeg vil benytte når jeg tester hypoteser i denne analysen. Testens p-verdi sier noe om det laveste mulige signifikansnivået vi kan forkaste  $H_0$  ved. Om p-verdien  $< 0,05$  kan vi derfor forkaste på et 5% signifikansnivå. Om p-verdien  $< 0,01\%$  kan vi også forkaste på et 1% signifikansnivå. Jo lavere p-verdi, jo lettere kan vi forkaste en nullhypotese.

Estimerer flere økonometriske modeller, utvider underveis:

$$(i) \quad lwage = \alpha + \beta_2 educ + \beta_3 num + \varepsilon$$

Ser på generell effekt på lønnen av utdanning og ferdigheter.

$$(ii) \quad lwage = \alpha + \beta_2 educ + \beta_3 num + \delta_1 pub + \varepsilon$$

Ser på generell effekt på lønnen av ferdigheter, men korrigerer nå for sektor.

$$(iii) \quad lwage = \alpha + \beta_2 educ + \beta_3 num + \delta_1 pub + \rho_1 pub * educ + \rho_2 pub * num + \varepsilon$$

Ser om effekten på lønnen av utdanning og ferdigheter er lik i offentlig og privat sektor.

$$(iv) \quad lwage = \alpha + \beta_2 educ + \beta_3 num + \delta_1 pub + \delta_2 fem + \delta_3 exper + \delta_4 exper^2 + \rho_1 pub * educ + \rho_2 pub * num + \varepsilon$$

Ser om effekten på lønnen av utdanning og ferdigheter er lik i offentlig og privat sektor, som i modell (ii), men kontrollerer nå for kjønn og arbeidserfaring.

### Empirisk strategi for tilleggsproblemstilling

Hypotesetesting utføres på samme måte som i hovedproblemstillingen.

#### Økonometrisk modell:

$$(i) \quad lwage = \alpha + \beta_2 educ + \beta_3 num + \delta_1 neth + \delta_2 pub + \rho_1 neth * pub + \rho_2 neth * educ + \rho_3 neth * num + \rho_4 neth * pub * educ + \rho_5 neth * pub * num + \varepsilon$$

Korrigerer nå for Estland. Ser om lønnsavkastningen av ferdigheter i offentlig sektor i forhold til privat sektor er lik i Nederland og Estland.

## 5. Empiriske resultater

### Empiriske hovedresultater

Tabell 7, hentet fra Stata:

VARIABLES	(i) m1 lwage	(ii) m2 lwage	(iii) m3 lwage	(iv) m4 lwage
educ	0.0521*** (0.00311)	0.0491*** (0.00316)	0.0478*** (0.00402)	0.0618*** (0.00381)
num	0.0791*** (0.00808)	0.0819*** (0.00809)	0.0962*** (0.0101)	0.0930*** (0.00951)
pub		0.0696*** (0.0140)	0.0267 (0.0909)	0.112 (0.0845)
fem				-0.0751*** (0.0132)
exper				0.0239*** (0.00242)
exper_sq				-0.000279*** (5.03e-05)
pub_educ			0.00350 (0.00649)	-0.00377 (0.00605)
pub_num			-0.0410** (0.0169)	-0.0202 (0.0157)
Constant	2.272*** (0.0428)	2.287*** (0.0428)	2.303*** (0.0540)	1.793*** (0.0597)



Observations	2,642	2,631	2,631	2,630
R-squared	0.207	0.215	0.217	0.325

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

Modell (i), m1:

Tester ved hjelp av t-test om utdanning har en betydning for lønnen. Nullhypotesen er at utdanningslengde ikke er av betydning for lønnen:

$$H_0: \beta_2 = 0, HA: \beta_2 \neq 0$$

Tester så ved hjelp av t-test om ferdigheter har betydning for lønnen. Nullhypotesen er at ferdigheter ikke er av betydning for lønnen:

$$H_0: \beta_3 = 0, HA: \beta_3 \neq 0$$

Ser av begge testenes p-verdier, som er lik 0, at vi kan forkaste  $H_0$  på både 1% og 5% signifikansnivå. Både utdanning og regneferdigheter er med andre ord av betydning for lønnen. Tolker av koeffisientene at ett ekstra år utdanning gir en gjennomsnittlig økning i lønnen på 5,21%, mens en marginal økning i resultatet på ferdighetstesten vil gi en økning i lønnen på 7,91%.

Modell (ii) m2:

Tester ved hjelp av t-test om lønnslikningen er lik for ansatte i offentlig og privat sektor. Nullhypotesen er med andre ord at koeffisienten til pub er lik 0, at det ikke er forskjell mellom de to lønnslikningene:

$$H_0: \delta_1 = 0, HA: \delta_1 \neq 0$$

Ser av testens p-verdi, som er lik 0, at vi kan forkaste nullhypotesen på både 5% og 1% signifikansnivå. Det er altså forskjell mellom lønnsfastsettelsen til ansatte i offentlig og privat sektor. Koeffisienten til pub viser at å være ansatt i offentlig sektor i gjennomsnitt fører til 6,96% høyere lønn.

Modell (iii), m3:

Tester ved hjelp av t-test om lønnseffekten av økt utdanning er lik i offentlig og privat sektor. Nullhypotesen er at det ikke er forskjell mellom lønnsavkastningen av utdanning i de to sektorene:

$$H_0: \rho_1 = 0, HA: \rho_1 \neq 0$$

Ser av testens p-verdi at laveste mulige signifikansnivå  $H_0$  kan forkastes på er 5,90%. Vi kan altså ikke forkaste  $H_0$  på et 5% signifikansnivå, men ville forkastet på et 10% signifikansnivå. Vi kan altså ikke si på et 5% signifikansnivå at det er forskjell i lønnsavkastningen av utdanning i offentlig og privat sektor.

Tester ved hjelp av t-test om lønnseffekten av økte regneferdigheter er lik i offentlig og privat sektor. Nullhypotesen er at det ikke er forskjell mellom lønnsavkastningen av ferdigheter i de to sektorene:

$$H_0: \rho_2 = 0, HA: \rho_2 \neq 0$$

Ser av testens p-verdi at laveste mulige signifikansnivå  $H_0$  kan forkastes på er 0,15%. Vi kan altså ikke forkaste på både 5% og 1% signifikansnivå hypotesen om at det ikke er forskjell i lønnsavkastningen av regneferdigheter mellom offentlig og privat sektor.

Vi kunne også testet ved hjelp av F-test om lønnslikningen for ansatte i offentlig og privat sektor nå er like. Nullhypotesen er da igjen at det ikke er forskjell mellom de to sektorene, altså at lønnslikningen for ansatte i offentlig og privat sektor er lik:

$H_0: \delta_1 = \rho_1 = \rho_2 = 0$ ,  $H_A: \text{Minst én av } \delta_1, \rho_1 \text{ eller } \rho_2 \neq 0$

Ser av testens p-verdi, som er lik 0, at vi kan forkaste nullhypotesen på både 5% og 1% signifikansnivå. Det er altså forskjell mellom lønnsfastsettelsen til ansatte i offentlig og privat sektor. Dette ser vi også fordi vi kunne avvise nullhypotesen om at  $\rho_2 = 0$  i t-testen over.

Koeffisientene i den utvidede modellen kan sammenliknes med de fra modell (i) og (ii). Ser at effekten av utdanning har falt noe. Ett år ekstra utdanning gir nå en gjennomsnittlig lønnsøkning på 4,78%. Koeffisienten for ferdigheter har derimot økt til at en marginal økning i resultat på testen nå gir 9,62% høyere lønn. T-testen kunne ikke avvise at koeffisienten  $\rho_1 = 0$ , så vi kan altså ikke si at effekten på lønnen av utdanning er forskjellig i offentlig og privat sektor. Vi har derimot at  $\rho_2 \neq 0$ , altså vitner resultatet på hypotesetesten om at effekten av ferdigheter er ulik i de to sektorene. Koeffisienten til `pub_num` tilsier at effekten på lønnen av økte ferdigheter er 4,10% lavere i offentlig enn privat sektor.

#### Modell (iv) m4:

Tester ved hjelp av t-test om lønnseffekten av økt utdanning fremdeles er lik i offentlig og privat sektor når vi korrigerer for kjønn og arbeidserfaring. Nullhypotesen er at det ikke er forskjell mellom lønnsavkastningen av utdanning i de to sektorene:

$H_0: \rho_1 = 0$ ,  $H_A: \rho_1 \neq 0$

Ser av testens p-verdi at laveste mulige signifikansnivå  $H_0$  kan forkastes på er 5,33%. Vi kan altså ikke forkaste på et 5% signifikansnivå, men ville forkastet på et 10% signifikansnivå. Vi kan altså ikke si på et 5% signifikansnivå at det er forskjell i lønnsavkastningen av utdanning i offentlig og privat sektor, selv etter at vi har korrigert for andre faktorer. Vi ser derimot at vi er nærmere å kunne forkaste  $H_0$  enn vi var da vi testet i modell (iii).

Tester ved hjelp av t-test om lønnseffekten av økte regneferdigheter fremdeles er ulik i offentlig og privat sektor når vi korrigerer for kjønn og arbeidserfaring. Nullhypotesen er at det ikke er forskjell mellom lønnsavkastningen av ferdigheter i de to sektorene:

$H_0: \rho_2 = 0$ ,  $H_A: \rho_2 \neq 0$

Ser av testens p-verdi at laveste mulige signifikansnivå  $H_0$  kan forkastes på er 0,20%. Vi kan altså fremdeles forkaste på både 5% og 1% signifikansnivå hypotesen om at det ikke er forskjell i lønnsavkastningen på regneferdigheter mellom offentlig og privat sektor, også når vi har korrigert for andre faktorer. Vi ser derimot at signifikansnivået man kan forkaste på er noe høyere enn det var i modell (iii).

Vi trenger ikke teste ved F-test om lønnslikningen for ansatte i offentlig og privat sektor nå er like. Vi ser, som i modell (iii) at siden vi også her kunne avvise nullhypotesen om at  $\rho_2 = 0$  i t-testen over vil vi kunne forkaste nullhypotesen om at lønnslikningen er lik også her.

Koeffisienten til `educ` er i modell (iv) enda høyere enn i den første (i), med beregnet økning i lønnen på i gjennomsnitt 6,18% for hvert år ekstra utdanning. Effekten av økt resultat på ferdighetstesten er derimot noe lavere enn i den forrige modellen, modell (iii). En marginal økning i resultat på ferdighetstesten gir i modell (iv) en gjennomsnittlig lønnsøkning på 9,30%. T-testen kunne i likhet med i modell (iii) ikke avvise at koeffisienten  $\rho_1 = 0$ , så vi kan heller ikke nå si at effekten på lønnen av utdanning er forskjellig i offentlig og privat sektor.

Vi har også her at  $\rho_2 \neq 0$ . Men her tilsier  $\rho_2$  at forskjellen i effekten på lønnen av økte ferdigheter er mindre enn i modell (iii), i gjennomsnitt kun 2,02% lavere i offentlig enn privat sektor.

### Forklaringer og øvrige kommentarer

Vi kan ikke tolke *exper* og *exper\_sq* separat. *Exper\_sq* er inkludert fordi vi antar at til tross for at økt erfaring har en positiv effekt på lønnen til ansatte er denne effekten avtakende over tid. Koeffisienten til *exper\_sq* er derfor negativ. Avkastningen på økt erfaring er høyere for en mindre erfaren arbeider enn en som alt har jobbet i mange år.

$$\frac{d \text{lwage}}{d \text{exper}} = \delta_3 + 2 * \delta_4 * \text{exper} = 0.0239 + 2 * (-0.000279) * \text{exper}$$

Ettersom alle modellene her er på samme funksjonsform og har identisk avhengig variabel kan vi sammenlikne nivå på føyning dem imellom.  $R^2$  er tilnærmet lik i de tre første modellene, på henholdsvis 20,7%, 21,5% og 21,7%. Det er derfor vanskelig å sammenlikne modellenes forklaringskraft. Modell (iv), som inkluderer langt flere kontrollvariabler, forklarer derimot 32,5% av variasjonen i den avhengige variabelen. Her er forskjellen større, og vi kan derfor sammenlikne denne modellens forklaringskraft med de tre første. Vi kan konstatere at ved å inkludere flere variabler som kunne tenkes å påvirke respondentenes timelønn oppnår vi at modellen i større grad kan forklare variasjonen i timelønnen. Men at variasjon i variablene på høyresiden i modellen fremdeles kun kan forklare 32,5% av variasjonen i den avhengige variabelen tyder på at det er mange andre faktorer som spiller inn i lønnsfastsettelsen til en ansatt enn de som er studert i denne oppgaven.

### Analyse av tilleggsproblemstilling

#### Deskriptiv statistikk for tilleggsproblemstilling

Nå som datasettet inneholder observasjoner fra både Estland og Nederland er det interessant å se hvor mange observasjoner datasettet fra hvert land inneholder.

Tabell 8, hentet fra Stata:

Land	Antall obs.	Prosent
Estland	6,277	59.48
Nederland	4,276	40.52
Totalt	10,553	100.00

Etter å ha ekskludert all data for respondenter under 25 år ser vi at det er 6277 observasjoner fra Estland, og 4276 fra Nederland.

Deskriptiv statistikk for sysselsatte inndelt etter sektor i Estland:

Tabell 9, hentet fra Stata:

pub	Prosent
0	70.91
1	29.09
Total	100.00

Ser at en litt lavere andel av sysselsatte arbeider i offentlig sektor enn privat sektor sammenliknet med i Nederland. 70,91% jobber i privat sektor, mens 29,09% jobber i offentlig sektor eller frivillige organisasjoner.

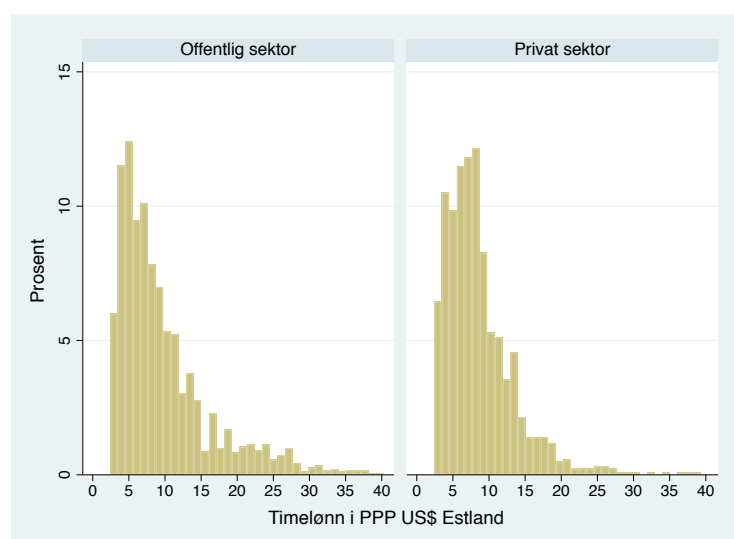
Deskriptiv statistikk for timelønn i Estland:

Tabell 10, hentet fra Stata:

Timelønn Estland	Observasjoner	Gjennomsnitt	Standardavvik
Alle	3,526	9.416299	6.013214
Offentlig sektor	1,209	8.701026	4.826731
Privat sektor	2,302	9.804269	6.520986

Gjennomsnittlig timelønn for alle respondenter i Estland er på 9,41 PPP US\$. I offentlig sektor er denne noe lavere, på 8,70 PPP US\$, mens den i privat sektor er høyest, på 9,80 PPP US\$. Ser derimot at standardavviket er større i privat enn offentlig sektor, slik som i Nederland. Legger merke til at gjennomsnittlig timelønn i Estland er under halvparten av gjennomsnittlig timelønn i Nederland. Som i datasettet for Nederland mangler det noen observasjoner når man legger sammen de som jobber i offentlig og privat sektor og sammenlikner med de som oppgir å motta lønn, her 15 observasjoner.

Histogram 2, hentet fra Stata:



Ser av histogrammet til Estland, som har en mer tydelig høyrehale enn Nederlands fordeling, at en større andel av befolkningen har en lavere timelønn.

Empiriske resultater tilleggsproblemstilling

Tabell 11, hentet fra Stata:

VARIABLES	(i) lwage
educ	0.0503*** (0.00330)
num	0.137*** (0.00870)
neth	0.821*** (0.0824)
pub	-0.159*** (0.0167)
neth_pub	0.185 (0.121)
neth_educ	-0.00258 (0.00627)
neth_num	-0.0408** (0.0159)
neth_pub_educ	0.00350 (0.00859)
neth_pub_num	-0.0410* (0.0224)
Constant	1.483*** (0.0408)
Observations	6,141
R-squared	0.552

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

Modell (i):

Tester ved hjelp av t-test hvorvidt effekten på lønnen av å jobbe i offentlig sektor er statistisk signifikant nå som vi har inkludert data fra Estland. Nullhypotesen er derfor at koeffisienten til pub er lik 0, og at lønnslikningen er identisk for ansatte i offentlig og privat sektor.

$H_0: \delta_2 = 0$ ,  $H_A: \delta_2 \neq 0$

Ser av testens p-verdi, som er lik 0, at vi kan forkaste nullhypotesen på både 5% og 1% signifikansnivå. Det er altså forskjell mellom lønnsfastsettelsen til ansatte i offentlig og privat sektor også når vi inkluderer data fra Estland. Koeffisienten til pub viser derimot at å være ansatt i offentlig sektor i gjennomsnitt fører til 15,85% lavere lønn. Dette er en forandring fra da vi testet for bare Nederland, der å være ansatt i offentlig sektor i gjennomsnitt ga høyere lønn. Dette stemmer med at offentlig ansatte i gjennomsnitt har lavere timelønn enn privat ansatte i Estland.

Tester ved hjelp av t-test om forskjellen i effekten av økt utdanning på lønnen i offentlig og privat sektor er ulik for ansatte i Nederland og Estland. Nullhypotesen er at det ikke er forskjell:

$H_0: \rho_4 = 0$ ,  $H_A: \rho_4 \neq 0$

Ser av testens p-verdi at laveste mulige signifikansnivå  $H_0$  kan forkastes på er 6,84%. Vi kan altså ikke forkaste på et 5% signifikansnivå, men ville forkastet på et 10% signifikansnivå. Vi kan altså ikke si på et 5% signifikansnivå at forskjellen i effekten av økt utdanning på lønnen mellom de to sektorene er ulik i Nederland og Estland.

Tester ved hjelp av t-test om forskjellen i lønnseffekten av ferdigheter i offentlig og privat sektor er ulik for ansatte i Nederland og Estland. Nullhypotesen er at det ikke er forskjell:  
 $H_0: \rho_5 = 0$ ,  $H_A: \rho_5 \neq 0$

Ser av testens p-verdi at laveste mulige signifikansnivå  $H_0$  kan forkastes på er 0,67%. Vi kan altså forkaste på både 5% og 1% signifikansnivå hypotesen om at forskjellen i lønnsavkastningen på regneferdigheter i offentlig og privat sektor er ulik for ansatte i Nederland og Estland. Ser at koeffisienten til `neth_pub_num` er negativ. Lønnsavkastningen av ferdigheter i offentlig sektor i forhold til privat sektor er med andre ord lavere i Nederland enn i Estland.

## 6. Konklusjon

### Oppsummering av hovedfunn

I tråd med humankapitalteori og tidligere forskning viser hypotesetestene om datasettet at økte ferdigheter og økt utdanning medfører økt lønn for ansatte, uavhengig av sektor. Men basert på tidligere undersøkelser av tall fra Norge, som viste at lønnsavkastningen på utdanning var større i privat enn offentlig sektor, mens lønnsavkastningen av ferdigheter var lik i begge sektorer, er funnene noe overraskende.

Hovedfunnene i oppgaven tilsier at mens lønnseffekten av utdanning i Nederland er lik i begge sektorer, er derimot effekten av ferdigheter større i privat enn offentlig sektor. Dette er med andre ord ulikt funnene fra Norge. En mulig forklaring på dette kan være at personlige ferdigheter og kompetanse vektlegges i større grad i privat sektor i Nederland enn i privat sektor i Norge, mens utdanning fra formelle utdanningsinstitusjoner veier tyngre i Norge enn i Nederland. Analysen har ikke gått inn på forskjeller mellom Norge og Nederland, men dette er verdt å kommentere da tidligere forskning vist til her i all hovedsak er produsert i Norge og ikke i Nederland. Det er også verdt å nevne at forskjellen i avkastning er marginal. I modell (iv), som inkluderte flest kontrollvariabler, var forskjellen i effekten på lønnen av økte ferdigheter på kun 2,02% mellom offentlig og privat sektor.

Derimot er det ikke overraskende at det er i nettopp privat sektor vi ser størst innvirkning på lønnen av økt humankapital. Dette stemmer overens med antakelsene om at privat sektor i større grad fastsetter lønnen via individuelle lønnsforhandlinger og systemer for prestasjonslønn enn det offentlig sektor gjør.

### Begrensninger ved analysen

Analysen har flere begrensninger. Ingen av de estimerte modellene tar hensyn til hvorvidt respondentene jobber fulltid eller deltid, en faktor Statistics Netherlands' rapport finner å ha en stor innvirkning på lønnen, og hvor offentlig og privat sektor skiller seg fra hverandre ved at flertallet av deltidsansatte er ansatt i privat sektor. En mulig løsning kunne vært å korrigere for deltid ved å opprette en kontrollvariabel. En annen løsning er å ekskludere data for deltidsansatte og kun studere lønnsstatistikken for heltidsansatte. Dette hadde derimot trolig

ekskludert mange observasjoner fra datasettet, og gjort det enda mer begrenset. Det er også et poeng å få med forskjellen i lønnseffekten av humankapital for heltids- og deltidsansatte, og undersøke om det her er en forskjell mellom offentlig og privat sektor.

Selve analysen inkluderer også bare data for de som har oppgitt en timelønn, altså de som er i arbeid. Respondenter som tidligere har vært i arbeid, men som var arbeidsledige da undersøkelsen ble utført, er derfor ikke inkludert.

En annen begrensning ved analysen er at vi ikke vet noe om kvaliteten på respondentenes utdanning, kun lengden. Her vil det kunne være forskjeller både innad i et land, men kanskje spesielt store forskjeller mellom land. Dette vil kunne ha en innvirkning på resultatene uten at det kan forklares ved hjelp av datasettet.

Som nevnt vil naturligvis flere faktorer enn de inkludert i modellen spille inn og påvirke ansattes lønn. En begrensning ved analysen er at ikke flere kontrollvariabler er inkludert. Eksempler på slike kunne vært respondentenes helse, etniske bakgrunn, innvandrerstatus, hvilken bransje de arbeider innenfor, eller som nevnt over, om de er heltids- eller deltidsansatte.

### Videre forskning

Som nevnt er det mange begrensninger ved analysen. Videre forskning ville kunne ta utgangspunkt i noen av disse for å se hvilken effekt de vil ha på resultatene. Det kunne i denne sammenhengen vært interessant å studere hvorvidt humankapital gir ulik avkastning på lønnen innad i hver sektor avhengig av bransje. Eller om det er forskjeller innad i privat sektor, mellom henholdsvis ansatte i bedrifter og selvstendig næringsdrivende. En forskjell her vil kanskje kunne påvirke den enkelte arbeiderens beslutning om å bli i bedriften eller starte opp selv, og dermed ha en generell effekt på gründervirksomhet og antall nydannelser av bedrifter innad i et land. Det er naturlig å tenke at en større avkastning på lønnen av økt utdanning og økte ferdigheter som selvstendig næringsdrivende vil skape incentiver for dannelsen av flere oppstartsselskaper. En betydelig større avkastning på lønnen som ansatt i en bedrift vil motsatt trolig fungere som et hinder for gründervirksomhet. En siste mulighet er at avkastning på lønnen av økt humankapital ikke påvirker beslutningen om å starte eget selskap, dette kunne vært interessant å studere i den sammenheng.

Man kunne også ha utført en liknende studie som den som ble utført av CBS i Nederland, og analysere hvorvidt ulike kjønn og aldersgrupper opplever ulik avkastning av humankapital på lønnen i henholdsvis offentlig og privat sektor. Det kunne også vært spennende å trekke inn forskjeller mellom flere land, for å se om det gir liknende resultater som de funnet i denne studien. Som nevnt i avsnittet om begrensninger ved datasettet er PIAAC-undersøkelsen kun gjennomført i OECD-land, mens det kanskje kunne vært desto mer spennende å utføre en studie av land på et lavere nivå i økonomisk vekst. Det kunne i denne sammenhengen vært interessant å sammenlikne land som er langt mer ulike enn de som er studert og sammenliknet i denne analysen. Det kan tenkes at man på denne måten vil kunne si mer om hvordan politisk retning og styresett påvirker lønnsavkastningen av humankapital både generelt og i de ulike sektorene, eller om religion eller politisk bakhistorie kan ha en innvirkning. Det kunne i denne sammenhengen også vært et poeng å studere land i ulike verdensdeler. Det er alt i alt mange retninger å ta i forbindelse med videre forskning på temaet.

## 7. Litteraturliste

### Kilder til datamaterialet

Datafiler til bacheloroppgave, Blackboard:

[https://ntnu.blackboard.com/webapps/blackboard/content/listContent.jsp?course\\_id=128251&content\\_id=595468\\_1&mode=reset](https://ntnu.blackboard.com/webapps/blackboard/content/listContent.jsp?course_id=128251&content_id=595468_1&mode=reset) (Nedlastningsdato: 5.4.2019)

Gross Domestic Product (GDP), OECD:

<https://data.oecd.org/gdp/gross-domestic-product-gdp.htm> (Nedlastningsdato: 5.4.2019)

### Andre kilder

OECD (2019), *The Survey of Adult Skills (PIAAC)*,

<http://www.oecd.org/skills/piaac/about/#d.en.481111> (Nedlastningsdato: 29.4.2019)

Becker, Gary S., *Human Capital*, <https://www.econlib.org/library/Enc/HumanCapital.html> (Nedlastningsdato: 29.4.2019)

Bjørkeng, Birgit, *Ferdigheter i voksenbefolkningen, Resultater fra den internasjonale undersøkelsen om lese- og tallforståelse (PIAAC)*, Statistisk sentralbyrå (2013),

[https://www.ssb.no/utdanning/artikler-og-publikasjoner/\\_attachment/141211?ts=1416e80e8e0](https://www.ssb.no/utdanning/artikler-og-publikasjoner/_attachment/141211?ts=1416e80e8e0) (Nedlastningsdato: 24.3.2019)

Bratsberg, Hægeland og Raaum, *Tøffere krav? – Ferdigheter og deltagelse i arbeidslivet*, Idunn (2011), <https://www.idunn.no/spa/2011/04/art06> (Nedlastningsdato: 24.4.2019)

Bratsberg, Raaum og Hægeland, *Lese- og tallforståelse, utdanning og arbeidsmarkedssuksess*, Universitetet i Stavanger, Lesesenteret (2006),

<https://www.nb.no/nbsok/nb/c8614a44623b875276fa6854241d1f66?lang=no> (Nedlastningsdato: 24.4.2019)

Geier og Grini, *Brattere trapp til lønnstoppen, Månedslønn og ulikhet gjennom 20 år*,

Statistisk sentralbyrå (2018), <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/brattere-trapp-til-lonnstoppen> (Nedlastningsdato: 22.4.2019)

McCracken, McIvor, Treacy, Wall, *Human capital theory: assessing the evidence for the value and importance of people to organisational success*, CIPD (2017),

[https://www.cipd.co.uk/Images/human-capital-theory-assessing-the-evidence\\_tcm18-22292.pdf](https://www.cipd.co.uk/Images/human-capital-theory-assessing-the-evidence_tcm18-22292.pdf) (Nedlastningsdato: 7.5.2019)

Moonen og Pleijers, *On average, higher educated earn twice as much as lower educated*,

Statistics Netherlands (CBS) (2011), <https://www.cbs.nl/en-gb/news/2011/12/on-average-higher-educated-earn-twice-as-much-as-lower-educated> (Nedlastningsdato: 22.4.2019)

Thomas, R. L., *Using Statistics in Economics*, McGraw-Hill Education (2005)

Geerdinck, Geijtenbeek, Graham, Sluiter, Wagner, *Equal wages for equal jobs? Jobs and wages in the public and private sectors*, Statistics Netherlands (CBS) (2012),

<https://www.cbs.nl/en-gb/background/2012/44/equal-wages-for-equal-jobs-jobs-and-wages-in-the-public-and-private-sectors-2010--dutch-only--> (Nedlastningsdato: 22.4.2019)