

Anders Martinius Udland

Deltagelse i barnehage

Bacheloroppgave i Bachelor

Veileder: Bjarne Strøm

Mai 2020

Anders Martinius Udland

Deltagelse i barnehage

Bacheloroppgave i Bachelor
Veileder: Bjarne Strøm
Mai 2020

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

CONTENTS

| | |
|--|----|
| 1. Innledning | 2 |
| 1.2 Presentasjons av problemstilling..... | 3 |
| 2. Tidligere Studier | 4 |
| 2.1 «Nye sjanser – bedre læring»..... | 4 |
| 2.2 Coleman- rapporten | 6 |
| 2.4 Oppsummering..... | 7 |
| 3. Metode | 8 |
| 3.1 Regresjonsanalyse..... | 8 |
| 3.2 Determinasjonskoeffisienten R² | 9 |
| 3.3 Funksjonsform | 10 |
| 3.3 Minste kvadraters metode | 10 |
| 3.4 Hypotesetest..... | 13 |
| 3.5 Oppsummering..... | 14 |
| 4. Datamateriale | 14 |
| 4.1 PIRLS..... | 14 |
| 4.2 Variabler | 15 |
| 4.3 Omkoding av variabler | 16 |
| 4.4 Deskriptiv statistikk | 17 |
| 4.4.1 Elevprestasjoner | 17 |
| 4.4.2 Barnehagedeltagelse | 18 |
| 4.4.3 Kjønn..... | 18 |
| 4.4.4 kontrollvariabler..... | 19 |
| 4.4.5 Korrelasjonsmatrise for datamaterialet | 21 |
| 4.6 Styrker og svakheter | 22 |
| 4.7 Oppsummering..... | 23 |
| 5. Regresjonsanalyse..... | 23 |
| 5.1 Barnehagedeltagelse | 24 |
| 5.2 Jenter og gutter..... | 30 |
| 5.3 Funn og utvidelser..... | 34 |

| | |
|----------------------|----|
| 6. Oppsummering..... | 36 |
| Referanser | 37 |

1. INNLEDNING

Jeg har valgt følgende tema for Bacheloroppgaven min:

- *Analyse av betydningen av å ha gått i barnehage (kindergarden) på elevprestasjoner. Et eller flere land.*

Den offentlige debatten i Norge, bærer lite preg av spørsmål om hvorvidt barns deltagelse i barnehager er positivt eller ikke. Det er bred enighet i det politiske spekteret at full barnehagedekning er fordelaktig politikk. Med barnehageforliket i 2003, vedtok stortinget omfattende utbygging av barnehage tilbudet samt en reduksjon av maksprisen for en barnehageplass. Forliket begrunnes blant annet i frigjørelsen av arbeidskraft som tidligere var bundet opp i hjemmeværende mødre. (Korsvold, 2020). Dermed preges ordskiftet i større grad av hvilke konsekvenser barn som ikke deltar i den tidlige innsatsen påføres. Der kontrasten i den politiske uenigheten sterkest belyses i barnehagedebatten, er synet på kontantstøtteordningen. Ordningen sees ofte som en substitutt bar løsning for familier som ønsker å holde barnet lengre i hjemmet. Argumenter på begge sider av det politiske spekteret, fremhever i den sammenheng at familier med innvandringsbakgrunn ofte er brorparten av mottaker av støtten, og konsekvensen vil være en svakere integrering av innvandrere. Integrerings spørsmålet vil være av interesse for videre forskning, men denne oppgaven vil fokusere i større grad på et bredere spørsmål som omhandler hele befolkningen, kjønnsperspektivet. Ordningen legger i utgangspunktet opp til at barn kan være hjemme til og med fylte 2 år. Om kontantstøtten hindrer utviklingen av kognitive evner sammenlignet med deltagelse i barnehage, har ikke denne oppgaven noe grunnlag for å hevde. Derimot vil barn som ikke begynner i barnehagen ved flyte 1 år, bli definert her som ikke-deltagende i barnehage ordningen.

I 2019 presenterte Stoltenbergutvalget, ledet av Folkehelseinstituttets direktør Camilla Stoltenberg, en rapport som omhandler kjønnsforskjeller i skolen. Rapporten fikk stor nasjonal

oppmerksomhet i forbindelse med konklusjonen at jenter gjør det sterkere enn gutter på skolen. Konklusjonen er muligens noe nedslående for det mannlige kjønn, er det slik at gutter gjør det svakere, utelukkende fordi de er gutter? Faktumet at sterke skoleprestasjoner gir økt muligheter i arbeidslivet, kan dermed hindre gutters muligheter for valgfrihet senere i livet.

Stoltenbergutvalget presenterer samtidig teorier om virkningen av sosioøkonomisk bakgrunn og utfordringen lavere samfunnslag har sammenlignet med høyere samfunnslag. Utvalget kommer med en rekke tiltak på tvers av kjønn og økonomisk bakgrunn, som skal gjøre utgangspunktet mer rettferdig og bidra til at alle har like muligheter.

I likhet med Stoltenbergutvalget, skal analysen belyse den statistiske sammenhengen mellom kjønn og sosioøkonomisk bakgrunn og leseferdigheter hos barn, noe som gjør resultatet fra rapporten interessant.

Denne oppgaven vil ikke gjøre et forsøk på å besvare spørsmålet om hvilken politisk løsning myndighetene burde innføre, men heller belyse den statistiske ulikheten mellom mulighetene. Flere hensyn denne oppgaven ikke tar for seg vil være viktige komponenter i en slik debatt, derfor burde ikke oppgaven brukes som et enkeltstående argument for eller mot barnehagedeltagelse.

1.2 PRESENTASJONS AV PROBLEMSTILLING

Opgaven vil gi en kvantitativ undersøkelse av leseferdigheter hos norske barn. Nærmere bestemt vil følgende problemstilling formuleres:

Gir barnehagedeltagelse økt leseferdigheter?

- *Ser vi forskjell på leseferdigheter mellom kjønnene?*
- *Påvirker barnehagedeltagelse kjønnene ulikt?*

Med problemstillingen til grunn, vil hypotesen være at barnehagedeltagelse gir en positiv effekt på leseferdigheter blant barn, samt at jenter presterer bedre på leseundersøkelser, sammenlignet med gutter. Barnehagedeltagelse vil i denne undersøkelsen være definert som om individet har deltatt i en barnehage i alderen 1-5 år før vedkommende begynner på grunnskolen.

Her vil det kontrolleres for elevkarakteristika, spesielt familieforhold, samt skolekarakteristika som lærerkvalitet. Senere vil jeg kontrollere for kjønnsforskjeller, da jeg antar at jenter har sterkere leseferdigheter enn gutter.

Oppgaven resulterer i forkastelse av hypotesen, grunnet usikkerheten tilknyttet validiteten til konklusjonen. Nærmere bestemt vil oppgaven konkludere med et **ikke**-signifikant resultat. Forkastelse av hypotesen, gir en overraskende konklusjon da tidligere studier som presenteres senere konstaterer at deltagelse i barnehage gir positiv effekt på kognitive ferdigheter hos barn. Store deler av avviket begrunnes i sosioøkonomiske variabler, som *par_eduUni*, *par_empFulltid* og hvorvidt eleven er jente. Oppgaven vil gi en analyse på resultatene og fremgangsmåten benyttet.

2. TIDLIGERE STUDIER

Tidligere forskning jeg vil se på vil hovedsakelig ta for meg «*Nye sjanser – bedre læring*» av *Stoltenbergutvalget (2019)*.

Her referes det også til kapittel 13, «*Education production functions*» fra *The Economics of Education*, av *Eric A. Hanushek (2020)*, samt artikkelen «*Utforming av utdanningspolitikken – Hva kan økonomene bidra med?*» av *Hans Bonesrønning (2004)*.

2.1 «NYE SJANSER – BEDRE LÆRING»

Stoltenbergutvalget tar for seg studier som undersøker om barnehagedeltagelse påvirker læringsresultater i skolen. I rapporten konkluderer utvalget med at barnehagedeltagelse vil ha positive effekter på kognitive ferdigheter, sammenlignet med det å være hjemme med en av foreldrene eller i uformelle ordninger (Stoltenbergutvalget, 2019, s.125). Uformelle ordninger referer til alle ordninger der barnet ikke deltar i barnehagen. Derimot, i rammeplanen for barnehager, legges det føringer for barnehagene til å fremme læring, uten at det spesifiseres innhold og på hvilken måte innholdet skal læres. Dermed vil forskjellene innad i barnehagesektoren være av betydning for barnas kognitive utvikling. Der hvor noen barnehager

har større fokus på utvikling av akademiske ferdigheter som skriving og lesing, kan andre barnehager være rettet mot sosiale aktiviteter. Utvalget hevder videre at i et kjønnsperspektiv kan det ikke utelukkes at barnehagedeltagelsen vil forsterke kjønnsroller, hvor jenter trekkes mer mot aktiviteter som er språkrelaterte. Det kan tyde på at de læringsbaserte aktivitetene barnehagen arrangerer virker å interessere jenter til fordel for gutter. (Stoltenbergutvalget, 2019, s.125). Det legges frem tall fra 2018, hvor jenter oppnår gjennomsnittlig 4.6 skolepoeng bedre enn gutter.

Ekspertene i Stoltenbergutvalget påstår at det er en sterk sammenheng mellom foreldres utdanningsbakgrunn og barnas skoleprestasjoner. (Stoltenbergutvalget, 2019, s.13)

Samtidig der utvalget fokuserer på kjønnsperspektivet, hevder rapporten at betydningen av den økonomiske bakgrunnen, mulig har større påvirkning for leseferdigheter enn hva kjønn utgjør. Utvalget i sin rapport, anbefaler inndeling av utdanningsnivå hos foreldre inn i to grupper. De med og de uten universitetsutdanning, det er hensiktsmessig å gjøre, da kjønn fortsatt deles inn i to grupper, jenter og gutter. (Stoltenbergutvalget, 2019, s.13)

Rapporten fremlegger at elever med foreldre med universitetsutdanning gjennomsnittlig oppnår 5.9 grunnskolepoeng mer enn elever uten høyere utdanning (Stoltenbergutvalget, 2019, s.13)

Samtidig oppnår gutter med foreldre uten universitetsutdanning 11 grunnskolepoeng svakere enn jenter i samme situasjon.

Sammenlignes dermed resultatene for kjønnsforskjeller og universitetsutdanning, finner Stoltenbergutvalget at jenter mulig er mindre eksponert for sosioøkonomiske faktorer, eller at gutter i større grad «tjener» på høyere sosioøkonomisk bakgrunn. (Stoltenbergutvalget, 2019, s.13)

Denne oppgaven baseres på undersøkelsen PIRLS, den samme type undersøkelse henviser Stoltenbergutvalget til flere steder i sin rapport, dog en nyere utgave, derfor sees det som hensiktsmessig å anvende konklusjoner utvalget har kommet med som sammenligningsgrunnlag til funn gjort i denne oppgaven.

2.2 COLEMAN- RAPPORTEN

I kapittel 13 kommenterer Hanushek en berømt skolerapport skrevet av sosiologen James Coleman. Undersøkelsen til Coleman benyttet en ny innfallsvinkel til hvordan måle effektiviteten en skole produserer. Der Coleman fokuserte på elevens prestasjoner var tidligere forskning basert på utgifter per elev og skolekarakteristika. Hensikten til Coleman, var dermed å se hva elevens resultat (output) ble ved kostnaden, i motsetning til utgift per elev (input). (Hanushek, 2020, s.164). Konklusjonen til Coleman er debattert hyppig og er ofte stilt spørsmålsteget ved. (Bonesrønning, 2004, s.16). I sin rapport konkluderer Coleman, noe deprimerende, at skolen ikke kunne påvirke utfallet til elevene (Bonesrønning, 2004, s.16). Variabler utenfor myndighetenes kontroll (familie og individ) konkluderte Coleman, hadde sterk påvirkning på elevens prestasjoner. Funnene resulterte i det kjente utsagnet «it's all in the family» (Bonesrønning, 2004, s.16). Utsagnet insinuerer et problem hvor skolens satsning på økt læringsutbytte hos elever fra lavere samfunnslag vil være utfordrende, fordi hvilken sosioøkonomisk bakgrunn gir større effekt på leseferdigheter enn skolens karakteristika. Coleman- rapporten presenterte en skoleproduktfunksjon som danner basisen for grunnmodellen som senere vil presenteres under regresjonsanalysen.

Produktfunksjonen til Coleman er som følger, her hentet fra artikkel skrevet av Bonesrønning:

$$O_{it} - O_{it^*} = f\left(F_i^{(t-t^*)}, P_i^{(t-t^*)}, S_i^{(t-t^*)}\right) + e_{it}$$

Hvor $O_{it} - O_{it^*}$ er elev i's prestasjoner på tidspunkt t og t^* . $F_i^{(t-t^*)}$ er en vektor av familiekarakteristika som påvirker eleven i tidsrommet mellom t og t^* . $P_i^{(t-t^*)}$ er karakteristika ved medelever (peer group effekter). $S_i^{(t-t^*)}$ er skoleinnsatsfaktorer som eksempelvis klassestørrelse og lærerens formelle kvalifikasjoner. (Bonesrønning, 2004, s.16).

Hanushek argumenter for at omfattende forskning siden Coleman rapporten, konkludere med at skolekarakteristika ved lærerkvalitet, faktisk påvirker positivt en elevs prestasjoner. (Hanushek, 2020, s.167). I sin kritikk, ser Hanushek til metoden hvor lærerkvaliteter måles basert på egenskapene nevnt lærer innehar. Metoden gir et feilaktig inntrykk av i hvilken grad læreren produserer gode studenter eller ikke, siden elevresultatene ikke nevnes. (Hanushek, 2020, s.167).

Derimot, argumenterer Hanushek for en alternativ kvantifisering ved å observere utviklingen en elev produserer på tvers av ulike lærere. En god lærer defineres som en der gode resultater akkumuleres kontinuerlig over tid. (Hanushek, 2020, s.167)

2.4 OPPSUMMERING

I dette kapittelet er det presentert teori og et rammeverk for hvordan tidligere studier med lignende problemstillinger har konkludert. Hanushek kritiserer Coleman rapporten i hvordan spesielt lærere ble kvantifisert, jeg ser det produktivt å belyse kritikk av skoleproduktfunksjonen til Coleman, da den vil danne basisen for grunnmodellen jeg presenterer senere i regresjonsanalysen.

3. METODE

Denne delen presenteres teoretisk bakgrunn for oppgaven. I regresjonsanalysen av datasettet vil jeg primært henviser til Thomas, R. L. (2005) «*Using statistics in economics*» for bruk av teori.

3.1 REGRESJONSANALYSE

Viktige statistiske metoder som hjelpemidler i analysen av empirien vil i dette kapitlet presenteres. For å gjennomføre analysen av datasettet, vil jeg anvende en regresjonsanalyse som viser sammenhengen mellom ulike variabler lineært i grunnmodellen. Verktøyet som brukes er Minste kvadraters metode (MKM), hvor hvordan barnehagegang (uavhengig variabel), affektore elevprestasjoner (avhengig variabel) lineær observeres. MKM gir dermed forholdet mellom en endogen variabel og eksogene variabler (både en eller flere). En lineær funksjon vil ha formen:

$$Y = \alpha + \beta X$$

der Y er den avhengige variabelen (endogene), α er konstantleddet som viser hvor kurven skjærer y-aksen, X er forklaringsvariablene/ uavhengige (eksogene), β er et parameter som viser prosentvis effekten forklaringsvariablene, X, har på den avhengige variabelen Y.

I min modell tilsvarer Y elevprestasjoner (read) og X tilsvarer barnehagegang (kinderg_att). Jeg velger å estimere modellen ved bruk av MKM, ut i fra det gitte datasettet estimerer vi α og β til følgende utvalsregresjon:

$$\hat{Y} = \alpha + \beta X$$

På generell form:

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

Hvor Y_i fortsatt tilsvarer den avhengige variabelen, α er konstantleddet, X_i tilsvarer forklaringsvariablene og β_i tilsvarer den prosentvise effekten X_i har på Y_i . ($i = 1, 2, 3, 4, \dots, 2647$ (antall observasjoner for svakeste ledd)) Vi ser ε er lagt til som et restledd, hvor effekter ikke fanget opp av forklaringsvariablene eller konstantleddet kan observeres.

Regresjonsanalysen vil være av kvantitative metode, hvor variabler er data hentet fra PRILS undersøkelsen. Metoden vil derfor ikke ta for seg dybdeintervjuer eller andre kvalitative metodeverktøy, da jeg ikke ser det som hensiktsmessig for å finne sammenhenger mellom avhengig og uavhengig variabel.

3.2 DETERMINASJONSKOEFFISIENTEN R^2

Determinasjonskoeffisienten, R^2 , er et parameter som avgjør forklaringskraften til modellen vi bruker for det gitte datasettet. Parameteret ser om vi har en lineær sammenheng mellom de endogene- og eksogene variablene i regresjonsmodellen.

R^2 er en verdi mellom 1 og 0, $R^2 = 1$ tilsier en 100 % lineær sammenheng mellom variabel X og variabel Y. Dermed kan 100% av variasjonene i den avhengige variabelen Y, bli forklart av tilsvarende variasjon i den uavhengige variabelen X. Der $R^2 = 0$ tilsier en 0 % lineær sammenheng mellom variabel X og variabel Y.

Determinasjonskoeffisienten uttrykkes matematisk som følger:

$$R^2 = \frac{SSE}{SST} = \frac{b^2 \sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2}, \Rightarrow R^2 = 1 - \frac{SSR}{SST}$$

Der $SSR = \sum_{i=1}^n e_i^2$

hvor SSR (*Sum of squared residuals*) er summen av de kvadrerte residualene, og SST (*Total sum of squares*) er summen av de kvadrerte avvikene fra gjennomsnittet. SSE (*explained sum of squares*) er avvik forventet fra empiriske verdier og modellen.

3.3 FUNKSJONSFORM

Hvilken funksjonsform som velges, vil ha innvirkning på hvordan kurven til regresjonen vil se ut. Klassisk lineær regresjon er mitt valg av funksjonsform. Dette da vi antar at det er en lineær sammenheng mellom den endogene og de eksogene variablene.

Tabell 1. Funksjonsformer

| Funksjonsform | Formel |
|-------------------------------|------------------------------------|
| Lineær funksjon | $Y = \alpha + \beta X$ |
| Resiprok funksjon | $Y = \alpha + \frac{\beta}{X}$ |
| Ekspontiell funksjon | $Y = \alpha e^{\beta X}$ |
| Dobbel logaritmisk funksjon | $Y = \alpha X^{\beta}$ |
| Semi – logaritmisk funksjon | $Y = \alpha + \beta \ln(X)$ |
| Logaritmisk resiprok funksjon | $Y = (\alpha + \beta \frac{1}{X})$ |

3.3 MINSTE KVADRATERS METODE

Her vil jeg gå nærmere inn på minste kvadraters metode som tidligere var nevnt under regresjonsanalysen 3.1.

Minste kvadraters metode (MKM), er anerkjent som et populært verktøy ved estimering av en modell for et utvalg, med prinsippet om minimering av summen av kvadrerte avvik, gitt forskjellen mellom den faktiske verdien Y , og den predikerte verdien \hat{Y} . Ved MKM, brukes det

estimatorer vi kaller for a og b, disse estimatorene gir oss regresjonslinjen for dataene i utvalget. For at estimatorene skal være holdbare, må det oppfylles forutsetninger knyttet til restleddet ε .

Forutsetninger for restleddet:

- Restleddet ε , er normalfordelt slik at ε_i er $\varepsilon \sim N(0, \sigma^2)$
- Restleddet ε , forventes til å være $\varepsilon = 0$.
- Det forventes en konstant varians til restleddet, og vil være uendelig for samtlige verdier. Det vil si: $Var(\varepsilon_i|X_i) = \sigma^2$
- Restleddet ε , hvis flere, vil være uavhengige av hverandre. Det vil si at:
 $Cov(\varepsilon_i|\varepsilon_j) = 0$
- Restleddet ε_i , har ingen sammenheng med X_i . Det vil si at $E(\varepsilon_i|X_i) = 0$.

Brytes en av forutsetningene for restleddet, vil ikke MKM holde. Hvis derimot forutsetningene oppfylles, vil vi ha følgende egenskaper for estimatorene a og b:

Estimatorene vil være lineære funksjoner av Y:

$$a = \bar{Y} - b\bar{X} = \sum_{i=1}^n \left(\frac{1}{n} - \bar{X} k_i\right) Y_i = \sum_{i=1}^n g_i Y_i$$

$$b = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} = \sum_{i=1}^n \left(\frac{X_i}{\sum_{i=1}^n X_i^2}\right) Y_i = \sum_{i=1}^n k_i Y_i$$

Hvor $k_i = \frac{X_i}{\sum_{i=1}^n X_i^2}$

Forventningsrettet estimatorer for α og β :

$$E(a) = \alpha$$

$$E(b) = \beta$$

Variansminimale blant lineære, forventningsrette estimatorer:

$$V(a) = \sigma_a^2 = \frac{\sigma^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n X_i^2}$$

$$V(b) = \sigma_b^2 = \frac{\sigma^2}{\sum_{i=1}^n X_i^2}$$

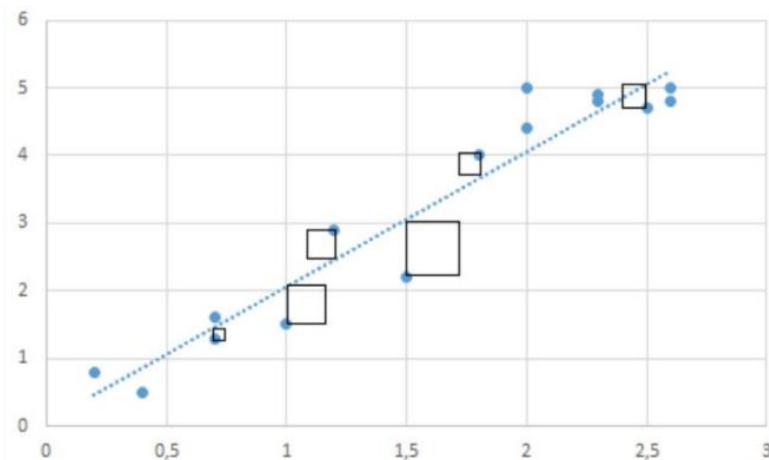
Estimatorene er normalfordelt slik at:

$$a \sim N\left(\alpha, \frac{\sigma^2 \sum_{i=1}^n X_i^2}{n \sum_{i=1}^n X_i^2}\right)$$

$$b \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n X_i^2}\right)$$

Grafisk illustrert, minste kvadraters avvik:

Graf 1. MKM



Residual:

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

$$e_i = Y_i - (a + bX_i)$$

Forklaring:

X_i og Y_i representerer avvik for X og Y. Modellen velges ut ifra hvor bra den passer med dataen i regresjonen. Derfor trengs avvikskvadratet. Avvikskvadratet finnes ved å kvadrere avstanden fra de respektive punktene, til den lineære kurven i *graf 1*, slik at residualen e_i , kan gi en indikasjon på samlet avstand fra punktene til kurven.

3.4 HYPOTESETEST

Jeg vil anvende hypotesetesting som metode for å kunne sikre og teste om regresjonsanalysen som foretas er statistisk signifikant. Senere utføres ulike hypotesetester, der det eksempelvis skal testes for om barnehagedeltagelse har positiv innvirkning på leseferdigheter på skolen. Ved Hypotesetest, fremstilles en nullhypotese H_0 , samt en alternativhypotese H_A . Nullhypotesen vil testes ved et gitt signifikansnivå, α . Normativt testes det for et nivå på $\alpha = 5\%$. Det vil tilsi at ved 5% sannsynlighet vil nullhypotesen feilaktig forkastes, når nullhypotesen er sann (Type I feil).

Alternativt kan $\alpha = 1\%$ anvendes, som tilsier at hypotesen har 1% sannsynlighet for feilaktig forkastelse. Følgelig vil et lavere signifikansnivå føre til lavere sannsynlighet for en feilaktig forkastelse av nullhypotesen. Nullhypotesen vil testes for betakoeffisient lik 0. $\beta_0 = 0$. Det vil tilsi at den endogene hovedvariabelen Y, er total uavhengige av den eksogene interessevariabelen X.

T – test:

I samspill med den kritiske verdien (signifikansnivået), vil testobservatoren (TS) avgjøre til hvilke verdier for nullhypotesen skal forkastes. Testobservatoren ved generell formel ser følgende ut:

$$TS = \frac{b_j}{S_{b_j}} \sim N(n - k)$$

Ved tilfelle der testobservatorens absolutte verdi overstiger kritisk verdi, α , forkastes nullhypotesen da en systematisk statistisk sammenheng ikke kan sies å være signifikant.

F- test:

F- testen anvendes ved multiple hypoteser på samme avhengige variabel. Testen vil konkludere om effekten hypotesene tester for, er signifikante. Verdien av effekten determineres av betakoeffisienten, derfor ser F- testen samlet på alle betakoeffisientene i regresjonen.

Eksempelvis ved hypotesetesting:

$$H_0: [\beta_1, \beta_2, \beta_3] = 0$$

$$H_A: [\beta_1, \beta_2, \beta_3] > 0$$

På generell formel vil testobservatoren se følgende ut:

$$F = \frac{SSR_R}{SSR} * \frac{N - i}{R} \sim F_{R, N-i}$$

Hvor R er antall restriksjoner på parameteret, SSR_R er SSR med restriksjonene. Det følger dermed at når F-testens absolutte verdi er større enn signifikansnivået (kritisk verdi), forkastes nullhypotesen.

3.5 OPPSUMMERING

Dette kapitlet presenterte teori som vil anvendes i regresjonsanalysen. Metodeverktøyet valgt i oppgaven er lineær funksjonsform med minste kvadraters metode (MKM), MKM vil minimere estimeringen av funksjonen, dette vil anvendes ved bruk av dataverktøyet stata.

I hypotesetestene anvendes et signifikantnivå på 5%, hvor validitet bestemmes ved t- testing.

4. DATAMATERIALE

Datamaterialet presentert i denne oppgaven er hentet fra *Progress in International Reading Literacy Study*, herunder PIRLS. Innledningsvis vil kilden for datasettet gjøres rede for, videre vil variablene presenteres, samt forklare deres innhold. Dette gjøres ved å presentere deskriptiv statistikk der vi finner observasjoner, gjennomsnittsverdier, maksimumverdier, minimumsverdier og standardavvik. Det avsluttes med å se på styrker og svakheter i datasettet.

4.1 PIRLS

Opgaven vil analysere et datasett basert på en leseundersøkelse om leseferdigheter blant elever i fjerde klasse i 35 ulike land i 2001. Undersøkelsen heter « *Progress in International Reading Literacy Study* » (PIRLS). Hensikten med PIRLS er å kartlegge forskjellen mellom

leseferdigheter til barn ved 4. klasse i ulike land. Deltagere og deres foreldre, vil i tillegg til testen for leseferdigheter, besvare en undersøkelse med fokus på forhold i hjemme.

I Norge er det en prosjektgruppe ved Senter for leseforskning, Høgskolen i Stavanger, som har hatt ansvaret for gjennomføringen av undersøkelsen. Det formelle ansvaret ligger hos Læringscenteret. (Norsk innledning)

4.2 VARIABLER

Tabell 2. Variabler

| Variabler | Beskrivelse | Type |
|-------------|--|------------------|
| Read | Karakter på leseferdighetstest | Kontinuerlig |
| Girl | 1 hvis jente, 0 hvis gutt | Dummy |
| Kinderg_att | 1 hvis Gått i barnehage, 0 hvis ikke | Dummy |
| Par_edu | Høyeste oppnådde utdanning til forelder | Kategorivariabel |
| Books_home | Antall bøker i husholdningen | Kategorivariabel |
| Income | Husholdningens inntekt målt i amerikanske dollar | Kategorivariabel |
| Clsize | Antall studenter i klassen | Kontinuerlig |
| Teacher_exp | Lærers ansiennitet | Kontinuerlig |
| Schoolsize4 | Antall elever i 4.klasse på skolen | Kontinuerlig |

Denne oppgaven vil analysere i hvilken grad barnehagegang påvirker leseferdigheter i fjerdeklasse på barneskolen, derfor vil interessevariabelen være *kinderg_att*. Senere utvides modellen til å se på kjønn samt hvorvidt eleven er født i Norge.

4.3 OMKODING AV VARIABLER

Variablene *books_home*, *par_emp* og *par_edu* (familiekarakteristika), er kategorivariabler. Det vil tilsa en inndeling av nivåer der hvert nivå innehar en gitt «mengde» av variabelen.

Books_home innehar 5 nivåer fra 1 – 5, med følgende informasjon: Nivå 1 tilsier at husholdningen eier 0 -10 bøker, nivå 2 11-25 bøker, nivå 3 26- 100 bøker, nivå 4 101- 200 bøker, nivå 5 over 200 bøker. Denne variabelen vil omkodes til en dummy variabel, *books_home100*, der husholdningen innehar fler enn 100 bøker, vil variabelen tilegnes verdien 1. Motsatt vil variabelen tilegnes verdien 0 hvis husholdningen innehar færre enn 100 bøker.

Par_emp innehar 4 nivåer fra 1 – 4, med følgende informasjon: Nivå 1 tilsier at begge foreldre jobber fulltid, nivå 2 tilsier at en av foreldrene jobber fulltid, nivå 3 tilsier at begge jobber mindre enn fulltid, nivå 4 tilsier at begge er arbeidsledige. Denne variabelen vil omkodes til dummyvariabelen, *par_empFulltid*, hvor skillet legges ved at minst en av foreldrene er i full jobb. Skillet legges ved minst en foreldre i full jobb, grunnet antagelsen om stabiliteten minst en foreldre bringer til en familie.

Par_edu innehar 5 nivåer fra 1 – 5, med følgende informasjon: nivå 1 tilsier universitetsutdannelse, nivå 2 tilsier utdanning etter videregående skole som ikke kategoriseres som universitetsutdannelse, nivå 3 tilsier videregående skole, nivå 4 tilsier ungdomsskole og nivå 5 tilsier ikke gjennomført ungdomsskole. Denne variabelen omkodes til dummyvariabelen *par_eduUni*, hvor skillet legges ved nivå 1, universitetsutdannelse. Dersom foreldrene har universitetsutdannelse vil variabelen tilegnes verdien 1, og lik 0 dersom lavere utdannelse.

4.4 DESKRIPTIV STATISTIKK

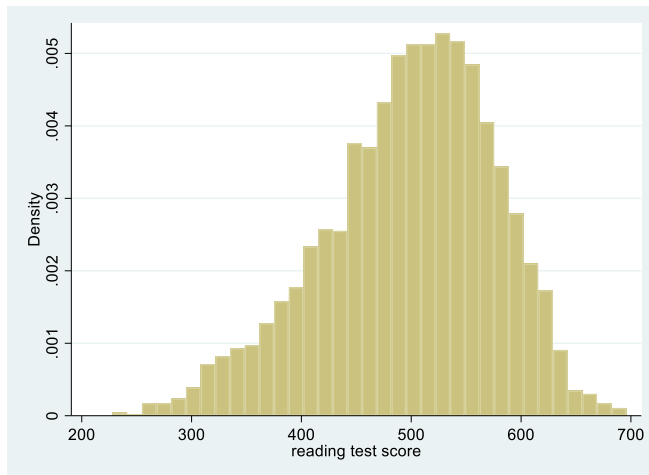
Jeg vil i denne delen legge frem deskriptiv statistikk av variablene som senere vil bruke i regresjonsanalysen. Det vil legges frem eksplisitte tabeller av *read*, *kinderg_att* og *girl* med tilhørende histogram til *read*.

4.4.1 ELEVPRESTAJONER

Tabell 3. *read*

| Elevprestasjoner | Verdier |
|--------------------------|----------|
| Gjennomsnittlig resultat | 498.2563 |
| Standardavvik | 78.36616 |
| Minimum testresultat | 228.0606 |
| Maksimum testresultat | 695.8717 |
| Antall observasjoner | 3 459 |

Tabell 4. *read* histogram



Fra tabell 2 kan det observeres at elevundersøkelsen for Norge består av 3459 observasjoner. Ekstremverdiene, det vil si maksimum og minimum resultat på undersøkelsen, er følgelig 695.8717 (max) og 228.0606 (min). Resultatene fra ekstremverdiene vil kunne gi et skjevt bilde av gjennomsnittet da de kan trekke sterkt i en retning. Gjennomsnittet (mean) er på 498.2563, med et tilhørende standardavvik (std.dev) på 7.36616.

4.4.2 BARNEHAGDELTADELSE

Tabell 5. *kinderg_att*

| Barnehagedeltagelse | Verdier |
|-------------------------------|-----------|
| Gjennomsnittlig resultat | 0.8603762 |
| Standardavvik | 0.3466516 |
| Minimum testresultat (Dummy) | 0 |
| Maksimum testresultat (Dummy) | 1 |
| Antall observasjoner | 3 137 |

Tabell 6. *kinderg_att* Deskriptiv statistikk

. tab *kinderg_att*

| <i>kinderg_att</i> | Freq. | Percent | Cum. |
|--------------------|-------|---------|--------|
| 0 | 438 | 13.96 | 13.96 |
| 1 | 2,699 | 86.04 | 100.00 |
| Total | 3,137 | 100.00 | |

Av tabell 6 vises det at 86,04 % av deltakerne i undersøkelsen har deltatt i barnehage med et tilhørende standardavvik på 34%. Ser at utvalget består av 3137 observasjoner, som tilsier at 322 observasjoner fra *read* ikke har svart om vedkommende deltok i barnehage.

4.4.3 KJØNN

Tabell 7. *girl*

| Kjønn | Verdier |
|-------------------------------|-----------|
| Gjennomsnittlig resultat | 0.481035 |
| Standardavvik | 0.4997137 |
| Minimum testresultat (Dummy) | 0 |
| Maksimum testresultat (Dummy) | 1 |
| Antall observasjoner | 3 401 |

Tabell 8. girl Deskriptiv statistikk

. tab girl

| girl | Freq. | Percent | Cum. |
|-------|-------|---------|--------|
| 0 | 1,765 | 51.90 | 51.90 |
| 1 | 1,636 | 48.10 | 100.00 |
| Total | 3,401 | 100.00 | |

Av tabell 8 vises det 3401 observasjoner. Av totale observasjoner består 1636 personer av jenter, noe som tilsvarer 48,10 %.

4.4.4 KONTROLLVARIABLER

Tabell 9. Kontrollvariabler

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---------------|-------|----------|-----------|-----|-----|
| income | 2,994 | 4.064128 | 1.550755 | 1 | 6 |
| par_eduUni | 3,459 | .4862677 | .4998837 | 0 | 1 |
| par_empFull~d | 3,459 | .819601 | .384575 | 0 | 1 |
| books_ho~100 | 3,459 | .0942469 | .2922141 | 0 | 1 |
| clsiz | 3,416 | 20.96165 | 4.82684 | 4 | 32 |
| teacher_exp | 3,354 | 16.51014 | 11.08948 | 1 | 42 |
| schoolsize4 | 3,243 | 41.18409 | 21.41166 | 3 | 192 |

Tabell 9, gir deskriptiv statistikk for variabler som skal kontrollere sammenhengen mellom *read* og *kinderg_att*. Kontrollvariablene er verktøy i regresjonsanalysen som inkluderes stegvis slik at endringer i forklaringskraften til *kinderg_att*, visualiseres.

Beskrivelse av kontrollvariabler:

Kontrollvariablene *par_eduUni*, *par_empFulltid* og *books_home100* er kodet om til dummyvariabler. For disse variablene ser vi bort i fra standardavviket, dette grunnet at verdien på variabelen enten kan være 1 eller 0, med en perfekt fordeling vil det innebære et standardavvik lik 0.5.

Variablene *clsiz*, *teacher_exp* og *schoolsize4*, er kontinuerlige variabler hvor minimum og maksimum verdi representerer ekstremverdier på en skala hvor observasjonene er fordelt. For

clsze (klassestørrelse), vil skalaen strekke seg fra en størrelse på 1 elev per klasse til 42 elever per klasse. Det vil derimot ikke tilsi at alle heltall fra 1 – 42 er representert som størrelse på en klasse. Gjennomsnittlig vil en klasse bestå av 20.96 elever med et tilhørende standardavvik på 4.82. Tilsvarende logikk vil følge for *teacher_exp* og *schoolsize4*.

Income representerer som eneste variabel i regresjonsanalysen, en kategorivariabel.

Kategorivariabelen representerer en skala fra 1 til 6, der følgende inntektsnivå er representert ved skalaen:

Nivå 1: < 20 000 \$

Nivå 2: 20 000 \$ - 29 999\$

Nivå 3: 30 000 \$ - 39 999\$

Nivå 4: 40 000\$ - 49 999\$

Nivå 5: 50 000\$ - 59 999\$

Nivå 6 > 60 000\$

Gjennomsnittsverdien til *income*, leses til 4.06. Hvilket innebærer nivå 4: 40 000 \$ - 49 999\$.

Samtidig er standardavviket tilhørende noe høy, som tyder på spredning i verdiene. Som observert ved tabell 10:

Tabell 10. *income*

| <i>income</i> | Freq. | Percent | Cum. |
|---------------|-------|---------|--------|
| 1 | 193 | 6.45 | 6.45 |
| 2 | 370 | 12.36 | 18.80 |
| 3 | 512 | 17.10 | 35.91 |
| 4 | 625 | 20.88 | 56.78 |
| 5 | 565 | 18.87 | 75.65 |
| 6 | 729 | 24.35 | 100.00 |
| Total | 2,994 | 100.00 | |

4.4.5 KORRELASJONSMATRISER FOR DATAMATERIALET

Tabell 11. Korrelasjonsmatrise

```
. corr read kinderg_att girl income par_eduUni par_empFulltid books_home100 clsiz teacher_exp schoolsize4
(obs=2,647)
```

| | read | kinderg_att | girl | income | par_eduUni | par_empFulltid | books_home100 | clsiz | teacher_exp | schoolsize4 |
|----------------|---------|-------------|---------|---------|------------|----------------|---------------|---------|-------------|-------------|
| read | 1.0000 | | | | | | | | | |
| kinderg_att | 0.0747 | 1.0000 | | | | | | | | |
| girl | 0.1287 | -0.0411 | 1.0000 | | | | | | | |
| income | 0.2258 | 0.1099 | 0.0189 | 1.0000 | | | | | | |
| par_eduUni | 0.3012 | 0.1454 | 0.0063 | 0.3965 | 1.0000 | | | | | |
| par_empFulltid | 0.1263 | 0.0459 | -0.0313 | 0.2997 | 0.1560 | 1.0000 | | | | |
| books_home100 | -0.0048 | -0.0109 | 0.0251 | -0.0177 | -0.0427 | -0.0237 | 1.0000 | | | |
| clsiz | 0.0556 | -0.0316 | 0.0201 | 0.1404 | 0.1128 | -0.0103 | 0.0032 | 1.0000 | | |
| teacher_exp | -0.0485 | -0.0647 | 0.0047 | -0.0323 | -0.0248 | -0.0183 | -0.0119 | -0.0568 | 1.0000 | |
| schoolsize4 | 0.0559 | 0.0057 | 0.0174 | 0.1628 | 0.1333 | 0.0226 | 0.0018 | 0.4573 | 0.0266 | 1.0000 |

Tabell 11 gir oss en korrelasjonsmatrise hvor det kan leses til hvilken grad hver enkel variabel korrollerer med andre variabler. Matrisen er likevel av størst interesse for den avhengige variabel *read* samt interessevariabelen *kinderg_att*. Matrisen er relevant i den grad vi ser motsetninger i hvilke variabler som påvirker positivt og negativ den avhengige variabelen *read*.

Kontrollvariablene for elevkarakteristika virker å trekke i samme retning på *read*. Følgelig virker spesielt *par_eduUni* og *income* å korrelere sterkt med leseferdigheter. Elever med universitetsutdannede foreldre virker dermed å gjøre det bedre på undersøkelsen.

Variabler med høy korrelasjon til *read*, er viktig å inkludere i regresjonen. Ved utelatelse av høyt korrelerte uavhengige variabler vil sannsynligheten for å konkludere på feil grunnlag øke. En feil som best beskrives på engelsk som «Omitted variable bias (OVB)». Konsekvensen av OVB resulterer at eksogene variabler tilegnes forklaringskraft det ikke er grunnlag for. (Bias)

Korrelasjonsmatrisen viser derimot ingen kausal sammenheng. Derfor anvendes ikke matrisen som konklusjonsgrunnlag, men som et verktøy for å finne hvilke variabler det vil være hensiktsmessig å inkludere.

4.6 STYRKER OG SVAKHETER

Styrker:

Datasettet tar for seg rundt 3000 observasjoner. Ved et slikt utvalg vil det kunne antas at observasjonene tar for seg et representativt utvalg sett i lys av at undersøkelsen er gjennomført av et svært anerkjent organ, IAE, hvor et norsk offentlig utvalg (Stoltenbergutvalget) referer til undersøkelsen.

Variablene PRILS anvender er relevante for analysen gjort i denne oppgaven. Oppgaven baserer seg på i hvilken grad barnehagedeltagelse påvirker skoleprestasjoner ved å kontrollere for person/ familie karakteristika og skolekarakteristika. Datasettet består av passende variabler som gjør at vi kan skille mellom personlig forhold og instrumenter myndighetene har tilgjengelig. Samtidig som variablene sammenfaller godt med oppgaven, er datakilden åpen, slik at undersøkelser basert på dataen kan etterprøves og dermed kontrolleres for.

Svakheter:

I den grad variablene stemmer overens med hvordan oppgaven kan besvares, består datasettet likevel av relativt få variabler. Det er vanskelig å se for seg at 100% av leseferdigheter kan beskrives med variabelutvalget vi har tilgjengelig. Det vil være av interesse å få innsyn i psykologiske faktorer som eksempelvis IQ samt nevrotiske trekk.

Datasettet har noen utfordringer med lukkede variabler. *Kinderg_att* er en variabel som tar alle barnehager over en kam, og tar gjennomsnittet av alle barnehager og gir et resultat. Videre forskning her, kan ta for seg ulikhetene en barnehage tilbyr og analysert hvilke kognitive aktiviteter en god barnehage gjennomfører sett mot en med svakere resultater hos elever deltatt i deres barnehage.

PIRLS undersøkelsen som anvendes i denne oppgaven, stammer fra 2001. Hvilket gjør rådataene noe foreldet. Ved å analysere ferskere data hentet fra Stoltenbergutvalget som anvendte data fra 2019 undersøkelsen til PRILS, observeres det en markant endring i resultatet. Spesielt finner jeg ulikheter i spredningen.

4.7 OPPSUMMERING

Her er det presentert deskriptiv statistikk for den avhengige variabelen *read*, samt interessevariabelene *kinderg_att* og *girl*. Deretter er det begrunnet hvorfor noen variabler har blitt omprogramert til dummyvariabler, og hvorfor *income* ikke blir omprogramert. Videre er det presentert en tabell med kontrollvariablene for datasettet samt en korrelasjonsmatrise. Avslutningsvis er det presentert styrker og svakheter ved datasettet, hvor det kan merkes at datasettet er 19 år gammelt. Den deskriptive statistikken vil bli brukt i regresjonsanalysen hvor jeg vil teste for forklaringsstyrken mot den avhengige variabelen *read*.

5. REGRESJONSANALYSE

I denne oppgaven, skal hypotesen om hvorvidt barnehagedeltagelse vil virke positivt inn på leseferdigheter ved senere skolegang besvares. Derfor ønsker jeg svar på hvor robust korrelasjonen mellom barnehagedeltagelsen vil være ved introduksjon av ulike kontrollvariabler.

Ut fra produktfunksjonen presentert i teorikapitlet fra James Coleman, vil jeg formulere en grunnmodell som tar for seg hovedvariabelen *read* og interessevariabelen *kinderg_att*. Videre vil introduseres kontrollvariablene for elevkarakteristika og fullføre modellen ved introduksjon av skolekarakteristika. Avslutningsvis introduseres kjønnsperspektivet, hvor en sammenligningsvariabel genereres, slik at nullhypotesen der jenter har større utbytte av barnehagedeltagelse enn gutter testes.

I ligningene presentert vil betakoeffisienten til variablene se ulik ut basert på hvilken type variabelen er. Dummyvariabler vil ha betakoeffisient lik β , kategorivariabler lik δ og kontinuerlige variabler lik γ . Dette gjøre slik at hvilken type variabel som testes for, er tydelig kommunisert.

5.1 BARNEHAGEDELTAGELSE

Hypotese:

- *Deltagelse i barnehage vil gi høyere resultat på skoleprestasjoner.*

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 > 0$$

Jeg formulere en meget restriktiv lineær grunnmodell basert på skoleproduktfunksjonen presentert av James Coleman i 1966, hvor den avhengige variabel er *read*, (elevprestasjoner) og interessevariabelen er *kinderg_att*.

Modellen visualiserer den direkte sammenhengen mellom hovedvariabelen og interessevariabelen. Modellen er restriktiv i den grad ingen kontrollvariabler inngår i regresjonene, hvilket tilsier at vi med relativt høy sannsynlighet går i «omitted variables bias» (kapittel, 4.4.5 s.20) fellen. Med teorien i bunn, forventes det at vi har en signifikant positiv sammenheng av barnehagedeltagelse på leseferdigheter.

Ved valgt signifikansnivå på 0.05 ved ensidig t – test, gir det en kritisk verdi på 1.64 (Thomas, 2005, s. 587). Gjennomgående i hele denne oppgaven vil det velges signifikansnivå på 0.05 og ensidig t – test.

Lineær grunnmodell med kontrollvariabler:

$$1) \text{read}_i = \alpha_i + \beta_1 \text{kinderg_att}_i + \varepsilon_i$$

Jeg estimerer modellen i Stata, som gir følgende data.

Tabell 12. Grunnmodell

| VARIABLES | (1) I read |
|--------------|---------------------|
| kinderg_att | 17.14*** (4.002) |
| Constant | 486.9*** (3.712) |
| Observations | 3,137 |
| R-squared | 0.006 |

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

Grunnmodellen viser at barnehagedeltagelse gir et signifikant positivt resultat på leseferdigheter. Avkastningen ved deltagelse i barnehagen vil tilsvare en økning på 17,14 i verdi på lesetesten. Det indikerer at individene i snitt vil gjøre det 17,14 poeng bedre på undersøkelsen ved å delta i barnehage i Norge.

Stata modellerer sammenhengen til å være signifikant, jeg velger likevel å vise utregningen for TS:

$$TS = \frac{17.14}{4.002} \Rightarrow 4.2829$$

$$|4.2829| > 1.64$$

Testobservatoren har høyere absolutt verdi, hvilket tilsier et signifikant resultat hvor H_0 kan forkastes til fordel for H_A .

Derimot er R^2 svært lav. Fra kapittel (3.2 s.8), leser vi at forklaringskraften til grunnmodellen er svak, noe som tyder på at den avhengige variabelen påvirkes betydelig av andre faktorer, hvilket kan bety at forventningen om «omitted variables bias» feilen, er oppfylt.

Utvider modellen for elevkarakteristika i henhold til Coleman:

Videre utvides grunnmodellen med elevkarakteristika. Resultatene fra modell 1 ser utelukkende på deltagelse i barnehage som endogen variable. Intuitivt er det flere variabler som spiller inn på et barns leseferdigheter, blant dem, familieforhold. Colemans berømte utsagn «It's all in the family», gjør at vi utvider modellen for testing av familie og individuelle forhold. I likhet med grunnmodellen vil det anvendes lineær regresjon med dummyvariabler samt kategorivariabelen, *income*.

Det forventes at interessevariabelen vil reduseres noe, da introduksjonen av kontrollvariabler vil tilegne seg forklaringskraften som interessevariabelen tok til seg i grunnmodellen. I henhold til teorien presentert tidligere, forventes det at ved høyere sosioøkonomisk bakgrunn, vil eleven gjøre det bedre på undersøkelsen.

Det vil videre i regresjonsanalysen, henvises til tabellene produsert av STATA ved tolkning av hvorvidt variablene er signifikante eller ikke. Derimot vil *kinderg_att* analyseres matematisk.

$$2) \text{ read}_i = \alpha_i + \beta_1 \text{kinderg_att} + \beta_2 \text{girl} + \beta_3 \text{Income} + \beta_4 \text{par_eduUni} + \beta_5 \text{par_empFulltid} + \beta_6 \text{books_home100} + \varepsilon_i$$

Tabell 13. Utvidet grunnmodell

| VARIABLES | (1) I read | (2) II read |
|----------------|---------------------|---------------------|
| kinderg_att | 17.14*** (4.002) | 6.238 (3.974) |
| girl | | 19.71*** (2.695) |
| income | | 5.518*** (0.980) |
| par_eduUni | | 37.43*** (2.967) |
| par_empFulltid | | 15.65*** (4.901) |
| books_home100 | | 7.751 (23.13) |
| Constant | 486.9*** (3.712) | 430.3*** (6.036) |
| Observations | 3,137 | 2,938 |
| R-squared | 0.006 | 0.122 |

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

Ved utvidelsen av grunnmodellen til inkludering av elevkarakteristika, finner jeg en markant reduksjon i forklaringskraften til interessevariabelen min. Med elevkarakteristika følger det at t-testen tilhørende *kinderg_att*, ikke er signifikant. Bevist:

$$TS = \frac{6.238}{3.974} = 1.5924$$

$$|1.5924| < 1.64$$

Det tyder på at vi må beholde nullhypotesen som tilsier at barnehagedeltagelse vil være uten effekt på leseferdigheter. Dette gjøres da jeg ikke med sikkerhet kan hevde at resultatet er korrekt.

I henhold til teori tilknyttet Coleman, kan vi lese at familiekarakteristika, her representert ved dummy variablene *par_eduUni*, *par_empFulltid*, *books_home100*, har relativ sterk påvirkningskraft på leseferdigheter.

Utdannelsen til foreldre, her universitetsutdannelse, har et signifikant positiv resultat på gjennomsnittlige resultat hos elevene. Gjennomsnittlig vil en elev prestere 37,43 poeng bedre, enn elever med foreldre uten universitetsutdannelse.

Sysselsetningsgraden til foreldre, her kontrollert for at begge foreldre har full jobb, gir signifikant positiv effekt på gjennomsnittlig resultat hos elevene. Gjennomsnittlig vil eleven prestere 15.65 poeng bedre, enn elever hvor foreldrene har mindre sysselsetningsgrad.

Antall bøker i hjemmet *books_home100*, blir funnet til å korrelere positivt med leseferdigheter, men til ikke å være signifikant.

Den utvidede grunnmodellen har en betydelig økt R^2 sammenlignet med grunnmodellen. Med en verdi på 0.122, tyder det på at 12,2 % av endringen i vår endogene variabel kan forklares med en tilsvarende endring i de eksogene variablene.

Utvider modellen fullstendig for skolekarakteristika i henhold til Coleman:

Modellen utvides videre slik at skolens karakteristika, herunder klassestørrelse, antall år lærer har undervist samt antall elever på klassetrinnet testes for. Dette gjøres for å teste i hvilken grad hypotesen er robust overfor inkluderingen av skolens eksogene variabler.

Hanushek kritiserer metoden Coleman anvendte når han konkluderte med «It's all in the family». Denne oppgaven baseres på produktfunksjonen presentert av Coleman, derfor antas det at resultatet fra den utvidete grunnmodellen vil sammenfalle med konklusjonen til Coleman rapporten.

$$\begin{aligned} read_i = & \alpha_i + \beta_1 kinderg_att + \beta_2 girl + \delta_3 income + \beta_4 par_eduUni \\ & + \beta_5 par_empFulltid + \beta_6 books_home100 + \gamma_7 Csize + \gamma_8 Teacher_exp \\ & + \gamma_9 schoolsizes4 + \varepsilon_i \end{aligned}$$

Tabell 14. Komplette grunnmodell

| VARIABLES | (1) I read | (2) II read | (3) III read |
|----------------|---------------------|---------------------|----------------------|
| kinderg_att | 17.14*** (4.002) | 6.238 (3.974) | 6.363 (4.148) |
| girl | | 19.71*** (2.695) | 19.83*** (2.826) |
| income | | 5.518*** (0.980) | 5.126*** (1.035) |
| par_eduUni | | 37.43*** (2.967) | 37.99*** (3.120) |
| par_empFulltid | | 15.65*** (4.901) | 16.12*** (5.183) |
| books_home100 | | 7.751 (23.13) | 6.975 (23.01) |
| clsiz | | | 0.172 (0.322) |
| teacher_exp | | | -0.252** (0.128) |
| schoolsize4 | | | -0.00488 (0.0742) |
| Constant | 486.9*** (3.712) | 430.3*** (6.036) | 432.2*** (9.115) |
| Observations | 3,137 | 2,938 | 2,647 |
| R-squared | 0.006 | 0.122 | 0.125 |

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

Av tabell 14, kan det leses at variablene *clsiz*e og *schoolsize4* ikke har en signifikant verdi. Skolens innsats, (basert på nevnte variabler) vil ikke med sikkerhet kunne påvirke leseferdigheter ved endring av klassestørrelse (*clsiz*e) og antall elever på trinnet (*schoolsize4*) Derimot er variabelen *teacher_exp*, signifikant, med en negativ korrelasjon på 0,252.

Totalt sett vil resultatet av innføringen av skoles karakteristika, ikke endre interessevariabelen *kinderg_att*, nevneverdig. Teorien James Coleman fremsatte «it's all in the family», vil kunne i lys av regresjonsanalysen ovenfor, antas holdbar vedrørende barnehagens evne til å forbedre individet til grunnskolen. Bevist:

$$TS = \frac{6..363}{4.148} = 1.5340$$

$$|1.5340| < 1.64$$

5.2 JENTER OG GUTTER

Hypotese:

- *Jenter har høyere utbytte av barnehagedeltagelse enn gutter*

$$H_0: \beta_{10} = 0$$

$$H_A: \beta_{10} > 0$$

Videre legges interaksjonsvariabelen *intraksjon* til. *Intraksjon* består av *kinderg_att* og *girl* Variabelen gir oss et sammenligningsgrunnlag for videre analyse om avkastningen jenter og gutter har ved å delta i barnehage.

Stoltenbergutvalget observerer store forskjeller mellom kjønnene i resultatene hos elevene. Oppgaven vil videre kartlegge hvilke forskjeller som observeres mellom jenter og gutter som har deltatt i barnehage og de som ikke har deltatt.

Det forventes her i henhold til litteraturen at jenter vil ha en markant foredel ved tidlig læring i barnehagen basert på konklusjonen til Stoltenbergutvalget der barnehager tendenserer mot undervisning tilpasset forventet kjønnsroller.

$$\begin{aligned}
 read_i = & \alpha_i + \beta_1 kinderg_att + \beta_2 girl + \delta_3 income + \beta_4 par_eduUni \\
 & + \beta_5 par_empFulltid + \beta_6 books_home100 + \gamma_7 Clsize + \gamma_8 Teacher_exp \\
 & + \gamma_9 schoolsizes4 + \beta_{10} intraksjon + \varepsilon_i
 \end{aligned}$$

Tabell 15. Komplet grunnmodell med interaksjonsvariabel

```
. reg read kinderg_att girl income par_eduUni par_empFulltid books_home100 clsize teacher_exp schoolsizes4 intraksjon
```

| Source | SS | df | MS | Number of obs | = | 2,647 |
|----------|------------|-------|------------|---------------|---|--------|
| Model | 1990875.4 | 10 | 199087.54 | F(10, 2636) | = | 37.85 |
| Residual | 13864310.8 | 2,636 | 5259.60198 | Prob > F | = | 0.0000 |
| | | | | R-squared | = | 0.1256 |
| | | | | Adj R-squared | = | 0.1222 |
| Total | 15855186.2 | 2,646 | 5992.13387 | Root MSE | = | 72.523 |

| read | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] | |
|----------------|-----------|-----------|-------|-------|----------------------|----------|
| kinderg_att | 2.553667 | 6.045636 | 0.42 | 0.673 | -9.301006 | 14.40834 |
| girl | 13.69372 | 7.626978 | 1.80 | 0.073 | -1.261747 | 28.64919 |
| income | 5.146837 | 1.034836 | 4.97 | 0.000 | 3.117663 | 7.17601 |
| par_eduUni | 37.85624 | 3.123911 | 12.12 | 0.000 | 31.73068 | 43.9818 |
| par_empFulltid | 16.06592 | 5.18393 | 3.10 | 0.002 | 5.900934 | 26.2309 |
| books_home100 | 6.671992 | 23.01453 | 0.29 | 0.772 | -38.45638 | 51.80036 |
| clsize | .1847256 | .32238 | 0.57 | 0.567 | -.4474179 | .816869 |
| teacher_exp | -.2498484 | .1279454 | -1.95 | 0.051 | -.5007319 | .0010351 |
| schoolsizes4 | -.0049553 | .0742143 | -0.07 | 0.947 | -.1504795 | .1405689 |
| intraksjon | 7.11313 | 8.211631 | 0.87 | 0.386 | -8.988764 | 23.21502 |
| _cons | 435.2603 | 9.78825 | 44.47 | 0.000 | 416.0669 | 454.4537 |

| VARIABLES | (1) I read | (2) II read | (3) III read | (4) IV read |
|----------------|---------------------|---------------------|----------------------|----------------------|
| kinderg_att | 17.14*** (4.002) | 6.238 (3.974) | 6.363 (4.148) | 2.554 (6.046) |
| girl | | 19.71*** (2.695) | 19.83*** (2.826) | 13.69* (7.627) |
| income | | 5.518*** (0.980) | 5.126*** (1.035) | 5.147*** (1.035) |
| par_eduUni | | 37.43*** (2.967) | 37.99*** (3.120) | 37.86*** (3.124) |
| par_empFulltid | | 15.65*** (4.901) | 16.12*** (5.183) | 16.07*** (5.184) |
| books_home100 | | 7.751 (23.13) | 6.975 (23.01) | 6.672 (23.01) |
| clsize | | | 0.172 (0.322) | 0.185 (0.322) |
| teacher_exp | | | -0.252** (0.128) | -0.250* (0.128) |
| schoolsize4 | | | -0.00488 (0.0742) | -0.00496 (0.0742) |
| intraksjon | | | | 7.113 (8.212) |
| Constant | 486.9*** (3.712) | 430.3*** (6.036) | 432.2*** (9.115) | 435.3*** (9.788) |
| Observations | 3,137 | 2,938 | 2,647 | 2,647 |
| R-squared | 0.006 | 0.122 | 0.125 | 0.126 |

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

Regresjonen med en ferdig estimert modell, viser en ikke signifikant positiv korrelasjon mellom deltagelse i barnehage og leseferdigheter gitt som elevprestasjon. Betakoeffisienten for *kinderg_att*, er på 2,554. T- testen viser: $P > |t| = 0.673$.

$$TS = \frac{2.554}{6.046} = 0.4224$$

$$|0.4224| < 1.64$$

Resultatet innebærer at nullhypotesen består, mens alternativhypotesen med $\beta_1 > 0$ forkastes.

Videre vil det analyseres til hvilken grad kjønnsperspektivet spiller inn.

Her vil det analyseres ulike kombinasjoner av barnehagedeltagelse på tvers av kjønnene. Dette gjøres for å belyse problemstillingen om jenters fordel av å delta i barnehagen.

Gjør en F-test, for variabler for jenter, dette gjøres for å teste om det er en signifikant forskjell mellom jenter og gutter i leseferdigheter.

Test 1. Jenter

```
. test kinderg_att girl intraksjon
```

```
( 1) kinderg_att = 0
( 2)  girl = 0
( 3) intraksjon = 0
```

```
F( 3, 2636) = 17.18
Prob > F = 0.0000
```

Ser at P- verdien er lik 0, som tilsier at jenter faktisk gjør det sterke på undersøkelsen, og vi kan beholde nullhypotesen med sikkerhet.

Først sees jenter som deltar i barnehage, mot gutter som deltar i barnehage.

Ved å addere betakoeffisienten til interaksjonsvariabelen, *intraksjon* med interessevariabelen *kinderg_att* finner jeg forskjellen i poengsum mellom jenter og gutter ved deltagelse i barnehage. $7.11313 + 13.69 = 20.80$. En jente som deltar i barnehagen vil dermed oppnå 20.80 poeng bedre i snitt enn gutter ved deltagelse i barnehage.

Videre sees jenter som deltar i barnehage mot gutter som ikke deltar i barnehage.

Jeg adderer betakoeffisienten til kontrollvariabelen *girl* med verdien funnet for jenter som deltar i barnehagen. $13.69 + 20.80 = 34.49$. En jente som deltar i barnehage, vil oppnå 34.49 poeng bedre i gjennomsnitt enn gutter som ikke deltar i barnehagen.

Ser deretter på jenter som ikke deltar i barnehage mot gutter som deltar i barnehage.

Denne verdien er representert i interaksjonsvariabelen for jenter i barnehage. Dette grunnet interessevariabelen, innebærer gutter som deltar i barnehagen. Jenter som ikke deltar i barnehage gjør det derfor 7.113 poeng bedre i gjennomsnitt enn gutter som deltar i barnehage.

Avslutningsvis analyseres jenter som ikke deltar i barnehage mot gutter som ikke deltar i barnehage.

Denne verdien er representert ved kontrollvariabelen *girl*. Variabelen representerer verdien jenter gjør det ulikt fra gutter. Her er den positiv lik 13.69. En jente vil dermed ved ikke å delta i barnehage, gjøre det 13.69 poeng bedre enn gutter som heller ikke deltar i barnehage.

5.3 FUNN OG UTVIDELSER

Kinderg_att testes systematisk mot inkludering av karakteristika for personlige forhold samt skolens innvirkning. Gjennomgående i analysen finner denne oppgaven ingen grunnlag for å kunne forkaste nullhypotesen med lineær sammenheng mellom leseferdigheter og deltagelse i barnehage.

Det er i tilfellene hvor gutter og jenter er i samme situasjon, det er mest interesse å videre utbre. Det kommer klart frem at jenter jevnt over presterer sterkere enn gutter ved leseferdighetsundersøkelsen. Uten tidlig læring i barnehagen, observeres det at jenter gjør det 13.69 poeng bedre enn guttene. Det tilsier at uten påvirkning fra myndighetene sees det en forskjell i leseferdigheter, som mulig kan forklares i jenters robusthet mot sosioøkonomiske faktorer. Jenters evne til å prestere uten systematiske forhold, ser kan være mulig forklaring på observasjonen.

Der begge kjønn deltar i barnehage, observeres det en økning i forskjellen fra 13.69 til 20.80 poeng. Det tilsier at deltagelsen i barnehagen vil øke forskjellene mellom kjønnene, sammenlignet med situasjonen hvor ingen deltar i barnehagen. Som argumentert i teorien, kan

tilfelle være at barnehager tilpasser sin undervisningssituasjon til å utfylle forventede kjønnsroller, og derfor tilpasser undervisningen til fordel for jenter.

Resultatet gir derimot **ikke** grunnlag for en systematisk statistisk sammenheng, derfor forkastes alternativhypotesen der hvor betakoeffisienten til interaksjonsleddet er større enn null. $\beta_{10} > 0$.

Det bemerkes at determinasjonskoeffisienten, R^2 , er lik 0.126. Det tilsier en marginal endring i forklaringskraften til modellen etter utvidelsen av interaksjonsvariablen, dermed observeres det tilnærmet lik lineær sammenheng mellom den avhengige og de uavhengige variablene som før introduksjonen av samspillsleddet. Determinasjonskoeffisienten, er ikke spesielt høy. Grunnen til dette kan være at sterke forklaringsvariabler som påvirker leseferdigheter er utelatt fra modellen jeg presenterer. Derfor har modellen noen restriksjoner som gjør validiteten begrenset. Mulige variabler som intuitivt påvirker leseferdigheter, er om vedkommende behersker norsk som språk. Sammen med variabler som fanger opp om vedkommende er født i Norge, samt foreldres nasjonalitet vil tenkes påvirke leseferdighetene til eleven. Denne oppgaven vil kunne utvides for testing av innvandringsstatus som fanger opp nettopp dette.

Samtidig kan andre karakteristika ved skolen være mulige utvidelser, med oppgave å kontrollere robustheten barnehagedeltagelse har ovenfor verktøy myndigheter har ved disposisjon.

Eksempelvis kan være et dypere dykk i lærers kvalifikasjoner som utdanning, samt skolens evne til å tilrettelegge for svakere elever.

Her vil det i tillegg være aktuelt å sammenligne resultater konkludert med fra Norge, opp mot utenlandske elever. Her anbefales det både å teste norske resultater opp mot nærliggende land som Sverige og Danmark, men også analysere opp mot land med helt ulik tilnærminger til skoleløpet, eksempelvis Russland.

6. OPPSUMMERING

Det er bred enighet i Norge om viktigheten av full barnehagedekning. Gjennom utvalg som Stoltenbergutvalget bekreftes hvorfor myndighetene bruker store summer over statsbudsjettet på satsing mot tidlig innsats. Den offentlige debatten preges i større grad av graden av kontantstøtteordningen, samt private mot offentlige løsninger. Oppgaven har ikke tatt stilling til politiske problemstillinger, rettere om beslutninger allerede gjennomført har grobunn i forskning. I denne oppgaven er det belyst at det finnes utfordringer tilknyttet tidlig innsats. I motsetning til Stoltenbergutvalget, kan ikke analysen her resultere i samme konklusjon, derimot vises det ingen systematisk statistisk sammenheng mellom deltagelse i barnehage og leseferdigheter i fjerdeklasse på grunnskolen. Endelig beholdes nullhypotesen om ingen lineær sammenheng grunnet usikkerheten tilknyttet resultatet.

Der tidlig innsats ikke er signifikant, finner oppgaven en klar sammenheng mellom leseferdigheter og kjønn. Gjennomgående i funnene, finner oppgaven det konsekvent bedre resultater for jenter ved testing av leseferdigheter. Som tidligere studier har konkludert med, kan dette ha sammenheng med jenters robusthet overfor faktorer som sosioøkonomiske variabler samt tidlig innsats fokus mot jenter. Konklusjonen er derimot at nullhypotesen beholdes begrunnet med ingen signifikante verdier.

Antageligvis belyser ikke denne oppgaven andre faktorer som gjør at tidlig innsats fungerer. Ved unnlattelse av spørsmål tilknyttet innvandring samt sammenligning av forskjellige barnehager, kan resultatet virke noe unyansert.

REFERANSER

1. Bonesrønning, H (2004) Utforming av utdanningspolitikken – Hva kan økonomene bidra med?
<https://samfunnsokonomene.no/wp-content/uploads/2019/05/Trykkutgave-4-2012.pdf>
2. IEA (17.04.2020) PIRLS
<https://www.iea.nl/studies/iea/pirls>
3. Lung Fei, Lee (1982 02.10) Specification error in multinomial logit models: Analysis of the omitted variable bias
<https://www.sciencedirect.com/science/article/abs/pii/0304407682900197>
4. Regjeringen (2018 20.08) Rett til barnehageplass
<https://www.regjeringen.no/no/tema/familie-og-barn/barnehager/innsikt/Rett-til-barnehageplass/id2344761/>
5. Solheim, R,G. & Tønnessen, E, F (2003) En norsk kortversjon av den internasjonale rapporten om 10- åringers lesekunnskaper (Bak innloggingsmur)
https://learn-eu-central-1-prod-fleet01-xythos.s3-eu-central-1.amazonaws.com/5def77a38a2f7/3638714?response-content-disposition=inline%3B%20filename%2A%3DUTF-8%27%27Solheim_T%25C3%25B8nnesen_En%2520norsk%2520kortversjon%2520av%2520den%2520internasjonale%2520rapporten%2520om%252010-%25C3%25A5ringers%2520lesekunnskaper_2003.pdf&response-content-type=application%2Fpdf&X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20200511T082522Z&X-Amz-SignedHeaders=host&X-Amz-Expires=21600&X-Amz-Credential=AKIAZH6WM4PLYI3L4QWN%2F20200511%2Feu-central-1%2Fs3%2Faws4_request&X-Amz-Signature=7e7c480cb2f82b66c9641c5390afb5767c10b668c6df00b53960936310e0af62
6. Stoltenbergutvalget (2019 04.02) Nye sjanser – bedre læring
<https://nettsteder.regjeringen.no/stoltenbergutvalget/files/2019/02/nou201920190003000dddpdfs.pdf>
7. Thomas, R. L. (2005): Using statistics in economics. McGraw-Hill.

